

What Makes Geeks Tick? A Study of Stack Overflow Careers

Lei Xu *

McGill University

Tingting Nian

University of California, Irvine

Luís Cabral

New York University and CEPR

Job Market Paper October, 2015

Latest version at <http://leixu.org/JMP>

Abstract

Many online platforms such as Yahoo! Answers and GitHub rely on users to voluntarily provide content. What motivates users to contribute content for free however is not well understood. In this paper, we use a revealed preference approach to show that career concerns play an important role in user contributions on Stack Overflow, the largest online Q&A community. We investigate how activities that can enhance a user's reputation vary before and after the user finds a new job. We contrast this with activities that do not help in enhancing a user's reputation. After finding a new job, users contribute 25% less in reputation-generating activities on Stack Overflow. By contrast, they reduce their non-reputation-generating activity by only 8% after finding a new job. These findings suggest that users contribute to Stack Overflow in part because they perceive this as a way to improve future employment. We provide direct evidence against alternative explanations such as integer constraints, skills mismatch, and dynamic selection effects. The results suggest that, beyond altruism, career concerns play an important role in explaining voluntary contributions on Stack Overflow.

*I am extremely indebted to Fabian Lange, Luís Cabral, Licun Xue, and Mitchell Hoffman for their advice, guidance, and support. I also would like to thank Avi Goldfarb, Heski Bar-Isaac, Chris Forman, Allan Collard-Wexler, Bryan Bollinger, Kei Kawai, Ariel Pakes, Jean Tirole, Jidong Zhou, Ignatius Horstmann, Philip Oreopoulos, Andrew Ching, Victor Aguirregabiria, Aloysius Siow, Yan Song, and Jie Ma for valuable suggestions; and seminar participants at NYU Stern, McGill, UToronto, as well as WISE 2014, IIOC 2015, SOLE 2015, Platform Strategy Research Symposium, NBER Summer Institute 2015 for helpful comments. I acknowledge support from the McGill University, NYU Stern, and Rotman. All errors are my own.

Xu: Ph.D. Candidate, Department of Economics, McGill University; lei.xu2@mail.mcgill.ca. Nian: Assistant Professor of Information Systems, The Paul Merage School of Business at UC Irvine; tnian@uci.edu. Cabral: Paganelli-Bull Professor of Economics and International Business, Stern School of Business, New York University; luis.cabral@nyu.edu.

1 Introduction

The prevalence of online platforms has attracted many firms to adopt a platform business model. Many tried but failed to launch a successful platform, mostly due to insufficient user participation from one or multiple sides. Due to network effects, a user won't participate without others' participation. A solid and thorough understanding of the motivations behind user participation is crucial for the success of a platform, especially for platforms that rely on voluntary contributions through user-generated content.¹

At the same time, to an economist the behavior of individual contributors also seems a bit startling. As Lerner and Tirole (2002) put it, is it a case of altruism, or are there ulterior motives behind private contributions to a public good?

Our paper addresses this research question using data from Stack Overflow (SO), the largest online Q&A platform 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.

Affiliated to 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.

The data from SO and SOC enables us to link online activity to real individuals. Thus we construct complete histories of each individual's online trajectory. This includes their contributions to SO as well as individual characteristics and employment histories. 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. However, right after the job change, there is a significant drop in Answers both in terms of quality and quantity. We use a difference-in-differences (DD) approach by comparing reputation-generating to non-reputation-generating activities from the same sample of job changers before and after a job switch. We conclude that contribution levels decrease by 24.5% right after a job change, of which 12.5–16.5% are due to (the removal of) career concerns. We provide direct evidence against alternative explanations such as integer

¹This type of contribution is also commonly referred to as *crowdsourcing* in the business world.

constraints, skills mismatch, and dynamic selection effects.²

To the best of our knowledge, ours is the first paper that empirically identifies and estimates the causal relation between changes in career status and voluntary 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 platform settings (Yahoo! Answers, Wikipedia, Stack Overflow, GitHub, YouTube, etc).

Related literature. Most of the theoretical and empirical literature on voluntary contributions and career concerns focus on Open Source Software (OSS). In many ways, the OSS phenomenon is very similar to contributions to sites such as SO, so a review of this literature is warranted.³ At a conceptual level, Lerner and Tirole (2001, 2002, 2005), Blatter and Niedermayer (2008) and Mehra, Dewan, and Freimer (2011) show how contributors to OSS projects improve their 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.

Career concerns is by no means the only motivation behind voluntary contributions. Conceptually, 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). Empirically, Zhang and Zhu (2011) examine social effects on Wikipedia; Lakhani and von Hippel (2003) study learning motivation on Apache field support system; Dobrescu, Luca, and Motta (2013) characterize social connections in book reviews; Luca and Zervas (2015) investigate economic incentives to commit review fraud on Yelp.

There is also a large body of literature on the theory of career concerns, starting with

²The dynamic selection effect is a hypothesis that is also commonly referred to in the labor economics literature as Ashenfelter's Dip.

³von Krogh and von Hippel (2006); von Krogh et al. (2012) present an excellent survey of this literature. 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).

Holmström (1982/99) and Gibbons and Murphy (1992). However, empirical work that estimates the causal effect of career incentives has been limited. Chevalier and Ellison (1999) examine the role of career concerns in investment strategies adopted by mutual fund managers. Kolstad (2013) isolates the role of intrinsic and extrinsic incentives in surgeon responses by examining the effects of the exogenous introduction of physician report cards.

Roadmap. The article is structured as follows: Section 2 introduces a simple dynamic model of user contributions. It develops the idea of career concerns we have in mind and clarifies the assumptions needed for identification. Section 3 describes the data for our analysis. Section 4 discusses how we use various activities to identify the effect of career concerns. Section 5 estimates the effect of career concerns using various approaches. Section 6 examines the main assumption needed to identify the career-concern effects. It also addresses various challenges to this assumption. Section 7 discusses other results related to the career-concerns hypothesis related to SO. Section 8 concludes.

2 Theoretical Model of User Contribution

We propose a simple dynamic model of user contributions. 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:

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

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.⁴

We assume each agent's utility each period is additively separable between work task and SO-related tasks:

$$u_t = g_s(w_t) + f(a_t, e_t)$$

where $f(\cdot, \cdot)$ is a homothetic function and both $f(\cdot, \cdot)$ and $g(\cdot)$ are twice differentiable functions such that $f', g' > 0$ and $f'', g'' < 0$. The homotheticity of $f(\cdot, \cdot)$ means a constant marginal rate of substitution along rays, which implies that the time elasticity of a and e are the same, namely, $\eta_a = \eta_e$ where $\eta_x = \frac{dx/dT}{x/T}$. Many commonly used utility functions satisfy these assumptions, including constant elasticity of substitution (CES) functions where $f(a, e) = (\alpha a^\rho + (1 - \alpha)e^\rho)^{\frac{1}{\rho}}$.

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:

$$V_0 = g_0(w) + f(a, e) + \delta p(a) V_1 + \delta (1 - p(a)) V_0$$

$$V_1 = g_1(w) + f(a, e) + \delta V_1$$

⁴In 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.

Finally, we assume reputation at time t is given by cumulative number of Answers in $t - 1$:⁵

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

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*

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

Moreover,

$$\frac{a_t}{e_t} \Big|_{s=1} < \frac{a_t}{e_t} \Big|_{s=0} \quad \text{iff} \quad p'(\cdot) > 0 \tag{2}$$

Proof: See Appendix.

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-à-vis Edits. The first effect (decline of Answers) can be decomposed into two effects: an increase in the marginal utility of time spent 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 elasticity with respect to changes in T plays an important role. As individuals work more, the assumption is necessary to prevent Edits to respond disproportionately to changes in time availability. In Section (6), several tests are provided to test the validity of this assumption.

⁵In 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.

<p>5 votes</p> <p>3 answers</p> <p>437 views</p> <hr/> <p>5 votes</p> <p>3 answers</p> <p>11k views</p> <hr/> <p>5 votes</p> <p>5 answers</p> <p>897 views</p> <hr/> <p>5 votes</p> <p>4 answers</p> <p>1k views</p>	<p>RegEx for distance in metric system</p> <p>I want a RegEx to match distance values in metric system. This regex should match 12m, 100cm,1km ignoring white space</p> <p><code>regex</code> <code>distance</code></p> <p>asked Sep 27 '09 at 10:25  Raytheon 87 ● 4</p> <hr/> <p>How to calculate Euclidean length of a matrix without loops?</p> <p>It seems like the answer to this should be simple, but I am stumped. I have a matrix of Nx3 matrix where there 1st 2nd and 3rd columns are the X Y and Z coordinates of the nth item. I want to ...</p> <p><code>matlab</code> <code>distance</code> <code>norm</code> <code>euclidean-distance</code> <code>vectorization</code></p> <p>asked Mar 17 '11 at 16:56  Miebster 898 ● 3 ● 8 ● 20</p> <hr/> <p>Python - how to speed up calculation of distances between cities</p> <p>I have 55249 cities in my database. Every single one has got latitude longitude values. For every city I want to calculate distances to every other city and store those that are no further than 30km. ...</p> <p><code>python</code> <code>django</code> <code>algorithm</code> <code>distance</code></p> <p>asked Dec 18 '13 at 10:00  pythonishvili 558 ● 6 ● 18</p> <hr/> <p>Efficient way to calculate distance matrix given latitude and longitude data in Python</p> <p>I have data for latitude and longitude, and I need to calculate distance matrix between two arrays containing locations. I used this This to get distance between two locations given latitude and ...</p> <p><code>python</code> <code>numpy</code> <code>scipy</code> <code>distance</code></p> <p>asked Oct 16 '13 at 20:31  Akavall 10.2k ● 5 ● 34 ● 64</p>
--	---

Figure 1: Sample List of Questions on Stack Overflow

3 Data

Our dataset is derived from the Stack Overflow (SO) and Stack Overflow Careers (SOC) sites. SO is the largest online Q&A site where programmers ask and answer programming-related questions (figures 1 and 2). It provides for Wikipedia-style editing (figure 3); and it includes a system of votes, badges and user reputation that ensures high-quality, peer-reviewed answers. SO is widely used by programmers.⁶

SOC is a related job matching website that hosts programming-related job listings as

⁶Founded in 2008, it currently comprises 4.8 million users. Some summary statistics regarding the site's activity: 7.7 million visits/day; 7.9 thousand questions/day; 10 million cumulative questions, 17 million cumulative Answers.

RegEx for distance in metric system

 I want a RegEx to match distance values in metric system. This regex should match 12m, 100cm,1km ignoring white space
5 

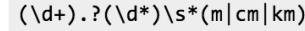
 edited Sep 27 '09 at 17:06
SilentGhost
91.8k ● 22 ● 175 ● 217

 asked Sep 27 '09 at 10:25
Raytheon
87 ● 4

3 Answers

active oldest 

 And to extend Paul's answer to include decimal place values...
7 

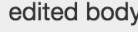

  

 answered Sep 27 '09 at 10:35
Nick Larsen ♦
10.6k ● 4 ● 29 ● 69

Good point, though I've tried to produce the simplest regex that matches the general pattern of his examples - integer values of centimetres, metres and kilometres. +1 - welcome to stackoverflow :) –
Paul Dixon Sep 27 '09 at 10:39

Figure 2: Sample of a Question with its Answers on Stack Overflow

Notes: One question can receive multiple answers which are ranked by Votes by default. Asker of the question can select one answer as the “correct” answer. Users can also comment on questions or answers. Both Questions and Answers receive up-Votes or down-Votes. One up-vote reward to a question rewards the asker 5 points; one up-vote to an answer rewards the answerer 10 points.

 2 


 inline  side-by-side  side-by-side markdown

I ~~wann~~ want a RegEx to match distance values in metric system. ~~This~~ This regex should match 12m, 100cm,1km ignoring white spaces.

Figure 3: Sample of Edits on Stack Overflow

Notes: Majority of Edits on SO are simple corrections to spelling or grammar mistakes. Some also involve more significant changes. Users with reputation under 2000 can suggest edits, which rewards them 2 points if accepted. Users with over 2000 reputation do not get the 2-point reward.

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 4: Sample of User Profile on Stack Overflow Careers



Nick Larsen ♦ (moderator) top 3% overall

I work on the [Careers 2.0](#) team at Stack Overflow. In my spare time I like to build and launch high power rockets, and I am also a competition programming enthusiast.

I am married to the wonderful account manager [Mary Paige Larsen](#) and we live in Atlanta with our first child Henry, our dachshund Maddy and our orange cat Abe.

10,607 REPUTATION

4 29 69

Communities (43)

- Stack Overflow ♦ 10.6k
- Meta Stack Exchange ♦ 8.3k
- Arqade 669
- Seasoned Advice 412
- Board & Card Games 314

[View network profile →](#)

Top Tags (373)

Tag	Score	Posts	Posts %
asp.net-mvc	287	102	27
asp.net	167	40	167
asp.net-mvc-3	167	26	167
c#	145	99	50
javascript	67	25	50
linq	50	27	50

[View all tags →](#)

Top network posts (376)

Post Type	Score	Question Title	Answered
A	108	How to render a DateTime in a specific format in ASP.NET MVC 3?	may 14 '11
Q	36	Problems with adding a 'lazy' keyword to C#	may 11 '11

Figure 5: Sample of User Profile on Stack Overflow

well as candidate resumes. For contributors, creating a resume on the website (figure 4) 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 of expertise.⁷ On the resume, contributors can easily provide a link to their SO profile (figure 5), 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.

Through a paid subscription, SOC helps employers by reducing hiring search costs. 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.⁸

Measures of user activity. There are four major activities by users on SO:

- | | |
|-----------|---|
| 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 existing Questions. ⁹ 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. Edits help make the questions and Answers more readable and understandable to future viewers. ¹⁰ |
| Votes | Finally, registered users can give up-votes or down-votes to Questions and Answers <i>but not to Edits</i> . Votes contribute to the post owner's reputation: each |

⁷The exact criteria is not disclosed by SO. An alternative path to an invitation is to request it on the website.

⁸As of October 24th, 2015, there are 1283 jobs on SOC, with 893 jobs located in the U.S. and Canada. The number is quite small compared to jobs on other popular employment websites such as Monsters.com and Indeed.com, where employers can post jobs free of charge.

⁹A user can also answer his or her own question, but no reputation points are earned to avoid gaming the system.

¹⁰Most Edits correct grammar or spelling mistakes; clarify the meaning of a post; or add related information. Users with reputation under 2000 can suggest edits, which rewards them 2 points if accepted. Users with over 2000 reputation do not get the 2-point reward.

up-vote on a Question give the asker 5 points, whereas each up-vote on an Answer are worth 10 points.¹¹

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.¹²
- Job switchers: the change in the level of career concerns comes from a job switch; we select users who experienced a job change from November 2008 until November 2014, the month when we stopped collecting data. To focus on job switches, we require the gap between two jobs are less or equal to one month. ¹³
- 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 the four-month period before or after the month of a job change. (in other words, we exclude inactive SO users).
- Multiple job switches: Some users experienced more than one switch; we exclude such switches if they are less than 8 months apart.
- 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 1462 job switches.¹⁴ For each user in our sample, we associate their user resumes (which include dates of job changes) to user IDs on SO. With the user IDs at hand, we then collect their activity on SO.

Table 1 provides some descriptive statistics of SO activities from the sample of 1249 users. Looking at the user activity, we see that the typical SO user is not active in writing

¹¹Older Answers have more cumulative Votes. To control for the comparability among Answers given at different time, we measure the total Votes gained on each answer within 30 days after the answering date.

¹²A large fraction of the jobs posted on SOC are located in the United States and Canada.

¹³We are also working on the analysis of contributors who experience a change in employment status, i.e. from unemployed to employed.

¹⁴Obviously, the sample we use is not representative of the whole population, since the majority of users have very few contribution. However, we do think it is a representative sample of active contributors.

Table 1: Descriptive Statistics of Sample of Users on Stack Overflow

	Mean	Median	Std. Dev.	Min.	Max.
User Activity (Monthly)					
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	71.5	2170.283	0	112967
Total UpVotes	334.669	82	800.728	0	15143
Reputation Points	1603.965	150	6204.839	-6	132122
Age	33.889	33	7.433	16	95
Time on SO	4.225	4.337	1.503	0.167	6.507

Notes: This table lists the descriptive statistics of lifetime activities of the 1249 contributors used in our DD analysis.

Questions or Answers. The activity distributions are fairly right-skewed, suggesting that a few users are disproportionately responsible for much of the content created on SO. 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

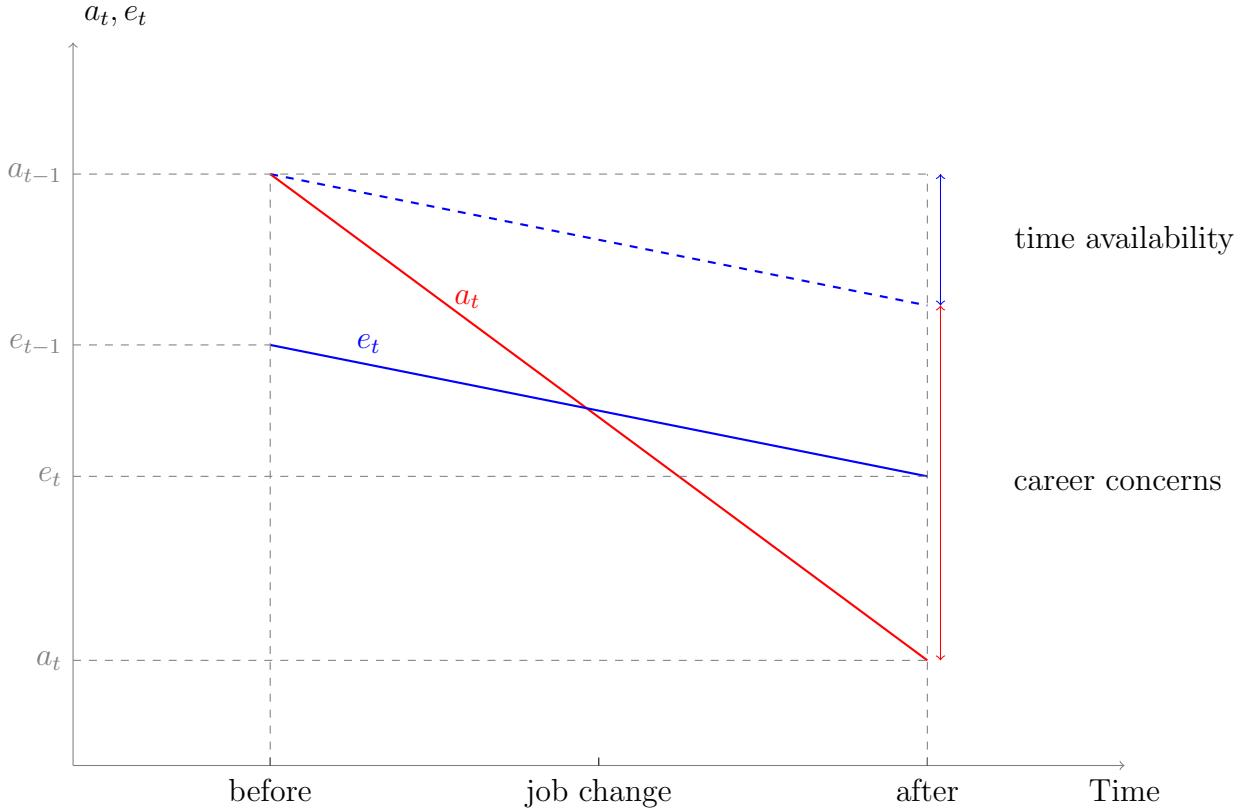


Figure 6: Graphical Illustration of Identification Strategy: Difference-in-Differences

Notes: Treatment group: Answers activity (a); Control group: Edits activity (e). All activity data comes from the same sample of contributors. DD coefficient is calculated as $(a_t - a_{t-1}) - (e_t - e_{t-1})$, which measures the differences of Answers-Edits gap before and after a job change.

(measured by the shift from $g_0(w)$ to $g_1(w)$ in the model).

To account for these effects, we use the differential change in Answers relative to Edits to test the hypothesis that answers are motivated by career concerns. A crucial difference between Edits and Answers is that the latter give rise to Votes, whereas the former do not. Therefore, we expect Answers to decline by more than Edits after individuals switch jobs.

Our DD approach assumes that, aside from changes in job status, Edits and Answers follow a parallel path. Since this is such a crucial assumption, in Section 6 we provide evidence in its support.

Essentially, our DD approach corresponds to the second part of Proposition 1. Figure 6 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;

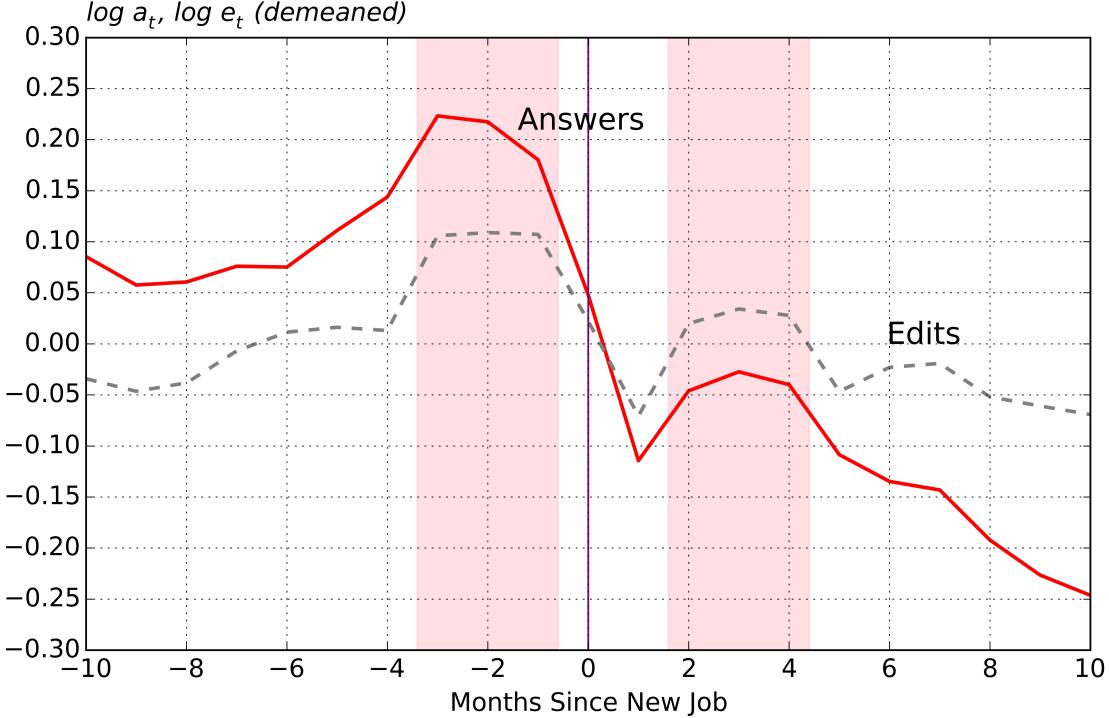


Figure 7: Average Monthly Activity on Stack Overflow (Answers and Edits)

Notes: This figure plots average monthly activity of Answers and Edits. Answers and Edits activity are demeaned $\log(1+\text{activity count})$. x-axis: Number of months since a new job starts. $t = 1$ means the first month of a new job. People with different starting dates are normalized to the same timeline based on number of months since the new job. y-axis: \log differences of activities. The initial set of DD regressions focuses on the 3-month before and after the job change, represented as activities in the pink area.

therefore, the difference between the changes in Answers and in Edits identifies the effect of job change on career-concerns incentives for Answers.

Figure 7 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 larger than the drop in Edits activity.

Naturally, several other alternative hypotheses may explain these dynamics. In Section 6, we present several hypotheses under which the parallel trend assumption could be violated, and evaluate the validity of each hypothesis.

5 Empirical Analysis

We now come to a more formal test of the hypothesis implied by Proposition 1. Our empirical analysis focuses on the sample of 1,249 users who were subject to 1,462 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 period 1 as the month when a job change takes 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 ($-3, -2, -1$); and 3 months subsequent to a job switch ($+2, +3, +4$). We thus exclude months 0 and 1; 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.

5.1 Basic Difference-in-Differences (DD) Results

We use the following regression specifications to estimate the impact of career incentive on voluntary contributions:

$$y_{it} = \alpha_i + \beta S_{it} + \epsilon_{it} \quad (3)$$

$$y_{kit} = \alpha_{ki} + \beta S_{it} + \gamma A_k S_{it} + \epsilon_{kit} \quad (4)$$

Equation 3 calculates the changes of user activities before and after a job switch; equation 4 estimates the additional change in reputation-generating activity over the changes in Edits (that is, equation 4 corresponds to DD specification). In these equations, y represents a generic activity, where subscript k denotes Answers (a), Votes gained from Answers (v), or Edits (e). S_{it} is the state dummy variable: $S_{it} = 0$ corresponds to the periods before a job change takes place by user i , whereas $S_{it} = 1$ corresponds to the period after a job change takes place. A_k is a dummy variable that takes on value 1 if the activity is a vote-generating activity (Answers) and 0 otherwise (Edits). α_{ki} controls for activity type (a, e) interacting with individuals (i) fixed effects. Finally, γ is the DD coefficient.

Equations 3 and 4 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

Table 2: Effect of Job Change on Answers Activity

	(1) $y = a$	(2) $y \in \{a, e\}$	(3) $y = v$	(4) $y \in \{v, e\}$
$NewJob(S)$	-0.245*** (0.03)	-0.080*** (0.02)	-0.259*** (0.03)	-0.080*** (0.02)
$NewJob(S) \times Answer(A)$		-0.165*** (0.03)		-0.179*** (0.04)
Regression	FE	FE	FE	FE
Contributors	1249	1249	1249	1249
N	8772	17544	8772	17544
R^2	0.021	0.015	0.019	0.014

Notes: FE: OLS Regression with Fixed-Effects. Dependent Variable: y : generic activity, a : Answers, e : Edits, v : Votes from Answers. Independent Variable: $S_{it} = 0$ prior to job switch, $S_{it} = 1$ after job change; $A_k = 1$ if $k = a$, $A_k = 0$ if $k = e$. Time Period: November 2008 - November 2014. Number of Contributors: 1249; Number of job switches: 1462. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

in (3) to be negative. Moreover, we expect the drop in Answers to be greater than that of Edits, so that $\gamma < 0$ in (4), in addition to $\beta < 0$.

Table 2 shows the results for our base regressions. There are two pair of regressions, using two measures of reputation-generating activity. For each pair, the first regression is limited to the Answers/Votes activity. We thus have 8,772 observations (1,462 job switches times 6 months: three prior to the job switch, three subsequent to the job switch); the second regression includes both Answers/Votes and Edits, thus doubling the number of observations.

Consider the first pair of regressions (that is, columns 1 and 2), where Answers and Edits are measured in logarithms. One advantage of this approach is that the coefficients can be readily interpreted as percent variations. The first regression has Answers as a dependent variable and S_{it} (job status) as the sole explanatory variable. The coefficient estimate on S_{it} suggests that, upon changing jobs, users decrease their Answers activity by about 24.5%. This is consistent with the first part of Proposition 1. However, as we mentioned earlier, it provides a weak test for our main hypothesis.

Regression 2 has both Answers and Edits as the dependent variable. As independent variables we have S_{it} and S_{it} interacted with A_k , the dummy indicator for Answers 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

16.5% reduction in Answers; this later variation we attribute to career concerns.

Columns 3 and 4 report the same set of estimates using Votes to measure Answers activity. Votes is a measure that includes both quantity and quality of Answers, and it can be a better measure of the effort to gain reputation points through high-quality Answers. The results using Votes give similar but slightly larger estimates than results using number of Answers. In Section 7.1, we investigate in depth the allocation of increased effort on quantity and quality of Answers.

5.2 Difference-in-Differences with More Controls

The regression estimates from table 2 could be potentially confounded by seasonality and duration effects. Both online contribution and job changes can have seasonal variation. Duration effects can include the initial excitement right after joining SO and switching to different tasks over time.

In order to control these effects, we use activity data from a large sample of 85k SO users.¹⁵

$$y_{kit} = \alpha_{ki} + D_{it} * (\beta S_{it} + \gamma A_k S_{it}) + \lambda_{kt} + X_{kit} \theta + \epsilon_{kit} \quad (5)$$

This regression specification builds on the baseline model (regression 4). For seasonality effects, we include yearly and monthly dummies; for duration effects, we include dummies for “SO Duration”, which measures the number of months since one’s first activity on SO. All the dummies are for Answers and Edits activity separately.

D_{it} is a dummy variable that indicates the 1,462 job changes and the 6-month period used in previous analysis; α_{ki} includes fixed effects for all individuals and each activity type; λ_{kt} controls for seasonal effects through year and month dummies; X_{kit} includes dummies for the number of months since first activity on SO.

Table 3 summarizes the DD estimates using a large sample of 85k active SO users as

¹⁵Although we do not have their CV information and job status, we do observe their online activity over time.

Table 3: Effect of Job Change on Answers Activity (With Control Variables)

	Panel A: $y \in \{\text{Answers}, \text{Edits}\}$				Panel B: $y \in \{\text{Votes}, \text{Edits}\}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>NewJob</i> (S)	-0.080*** (0.02)	-0.082*** (0.02)	-0.084*** (0.02)	-0.082*** (0.02)	-0.080*** (0.02)	-0.082*** (0.02)	-0.084*** (0.02)	-0.082*** (0.02)
<i>NewJob</i> (S) \times <i>Answer</i> (A)	-0.165*** (0.02)	-0.120*** (0.03)	-0.126*** (0.03)	-0.124*** (0.03)	-0.179*** (0.03)	-0.132*** (0.04)	-0.138*** (0.04)	-0.136*** (0.04)
Seasonality Dummy	No	Yes	No	Yes	No	Yes	No	Yes
Duration Dummy	No	No	Yes	Yes	No	No	Yes	Yes
# Users for Trend Control	-	85k	85k	85k	-	85k	85k	85k
R^2	0.001	0.023	0.029	0.031	0.001	0.022	0.024	0.026

Notes: All the DD regressions measure the changes in Edits and Answers activity 3-month before and after a new job. First row shows changes in Edits activity; second row shows the additional reduction in Answers activity, which we attribute to the removal of career concerns. Panel A uses number of Answers and Panel B uses number of Votes from Answers as measures of Answers activity. All use activity data of 85k SO users to control for seasonality (yearly and monthly) and duration (# of month on SO) effects. Independent Variable: $S_{it} = 0$ prior to job switch, $S_{it} = 1$ after job change; $A_k = 1$ if $k = a$, $A_k = 0$ if $k = e$. Number of contributors for DD analysis: 1249; Number of job switches: 1462. Robust standard errors in parentheses, clustered at the level of the individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

controls. Panel A uses the number of Answers and Panel B uses the number of Votes from Answers as measures of Answers activity. Coefficients in the first row measure the changes in Edits activity; those in the second row measure the additional reduction in Answers activity, which we attribute to the removal of career concerns. All the DD analysis are conducted by using the same set of 1249 contributors and by comparing activity changes 3-month before and after a job change. In Panel A, the DD estimates change from 16.5% without any control to 12.4% with both seasonality and duration controls. All DD estimates remain statistically significant at similar levels.¹⁶ The results in table 3 imply that after controlling for seasonality duration effects, 12.4% of Answers (column 4) and 13.6% (column 8) of Votes are due to career incentives.

5.3 Month-to-Month Comparison

Previous analysis focused on the 3-month period before and after a job change. In order to show a complete picture of activity changes over time, we use all the available data from the 1249 users. Using period -2 as the baseline period, we compare the activity of all other

¹⁶The additional data from 85k SO users is used to control for seasonality and duration effects only, and it does not contribute directly to the significance of DD coefficients due to a larger dataset.

periods to period -2 .¹⁷ We also control for seasonality and duration effects using the same 85k SO users mentioned before.

$$y_{it} = \alpha_i + D_i * \left(\sum_{\tau=-20}^{20} \beta_\tau \mathbb{1}(P_{it} = \tau) \right) + \lambda_t + X_{it} \theta + \epsilon_{it} \quad (6)$$

$$y_{kit} = \alpha_{ki} + D_i * \left(\sum_{\tau=-20}^{20} (\beta_\tau \mathbb{1}(P_{it} = \tau) + \gamma_\tau A_k \mathbb{1}(P_{it} = \tau)) \right) + \lambda_{kt} + X_{kit} \theta + \epsilon_{kit} \quad (7)$$

Regression 6 estimates month-to-month activity differences using period -2 as the baseline month. Regression 7 estimates DD coefficients for each month. D_i is a dummy variable that indicates the 1,462 job changes. λ_{kt} and X_{kit} control for seasonal and duration effects for each type of activity. P_{it} represents the number of months after a job change, and $\mathbb{1}(P_{it} = \tau)$ is a dummy variable which equals to 1 if the month t for user i corresponds to τ months after a job change. β_τ in regression 6 measures the differences in activity between period τ and baseline period -2 . Moreover, γ_τ in regression 7 is the DD coefficient between period τ and period -2 .

The estimates of β_τ from regression 6 and γ_τ from regression 7 are plotted in figure 8 and 9.

Figure 8 is similar to figure 7 but with seasonality and duration effects removed. It shows a clear rapid increase in Answers activity a few months before a job change.

Figure 9 plots the DD estimates γ_τ for $\tau \in [-20, 20]$. DD estimates before period 0 are all negative, but they are not significantly different from zero. However, after a job change, all the estimates are significantly negative.

One phenomenon shown in both figures is that Answers activity keeps dropping over time after a job change. Several possible theories can explain this phenomenon. First, the first few months are often considered as probationary periods where both employers and employees can freely terminate their contracts. Thus career concerns drop significantly but do not completely disappear as both parties need some time to realize the matching quality. If this is the case, then our results from regressions 6 and 7 underestimate the effect of career

¹⁷Period -2 is used as the baseline period since it has the highest average Answers activity level.

Table 4: Month-to-Month Comparison: First Differences

	(1) Answers	(2) Edits	(3) Votes
Period -20	-0.237*** (0.04)	-0.154*** (0.03)	-0.212*** (0.05)
Period -19	-0.239*** (0.04)	-0.172*** (0.03)	-0.237*** (0.04)
Period -18	-0.228*** (0.04)	-0.150*** (0.03)	-0.223*** (0.04)
Period -17	-0.189*** (0.04)	-0.160*** (0.03)	-0.193*** (0.04)
Period -16	-0.213*** (0.04)	-0.144*** (0.03)	-0.196*** (0.04)
Period -15	-0.206*** (0.04)	-0.137*** (0.03)	-0.195*** (0.04)
Period -14	-0.190*** (0.04)	-0.121*** (0.03)	-0.177*** (0.04)
Period -13	-0.196*** (0.04)	-0.123*** (0.03)	-0.180*** (0.04)
Period -12	-0.236*** (0.04)	-0.133*** (0.03)	-0.232*** (0.04)
Period -11	-0.224*** (0.04)	-0.122*** (0.03)	-0.218*** (0.04)
Period -10	-0.204*** (0.04)	-0.141*** (0.03)	-0.193*** (0.04)
Period -9	-0.221*** (0.04)	-0.155*** (0.03)	-0.206*** (0.04)
Period -8	-0.209*** (0.03)	-0.147*** (0.02)	-0.225*** (0.04)
Period -7	-0.188*** (0.03)	-0.116*** (0.02)	-0.161*** (0.04)
Period -6	-0.190*** (0.03)	-0.096*** (0.02)	-0.167*** (0.04)
Period -5	-0.151*** (0.03)	-0.095*** (0.02)	-0.157*** (0.04)
Period -4	-0.112*** (0.03)	-0.097*** (0.02)	-0.106*** (0.03)
Period -3	-0.012 (0.03)	-0.009 (0.02)	-0.025 (0.03)
Period -2 (Base)	0 (-)	0 (-)	0 (-)
Period -1	-0.030 (0.03)	-0.008 (0.02)	-0.026 (0.03)
Period 0	-0.160*** (0.03)	-0.091*** (0.02)	-0.128*** (0.03)
Period 1	-0.306*** (0.03)	-0.182*** (0.02)	-0.315*** (0.04)
Period 2	-0.224*** (0.04)	-0.087*** (0.03)	-0.252*** (0.04)
Period 3	-0.193*** (0.04)	-0.073*** (0.03)	-0.210*** (0.04)
Period 4	-0.198*** (0.04)	-0.066*** (0.03)	-0.193*** (0.04)
Period 5	-0.260*** (0.03)	-0.148*** (0.03)	-0.251*** (0.04)
Period 6	-0.283*** (0.03)	-0.122*** (0.03)	-0.297*** (0.04)
Period 7	-0.280*** (0.03)	-0.113*** (0.03)	-0.272*** (0.04)
Period 8	-0.333*** (0.03)	-0.147*** (0.03)	-0.341*** (0.04)
Period 9	-0.355*** (0.03)	-0.157*** (0.03)	-0.362*** (0.04)
Period 10	-0.372*** (0.03)	-0.167*** (0.03)	-0.379*** (0.04)
Period 11	-0.409*** (0.03)	-0.202*** (0.03)	-0.399*** (0.04)
Period 12	-0.437*** (0.03)	-0.212*** (0.03)	-0.435*** (0.04)
Period 13	-0.471*** (0.03)	-0.205*** (0.03)	-0.477*** (0.04)
Period 14	-0.466*** (0.03)	-0.222*** (0.03)	-0.472*** (0.04)
Period 15	-0.488*** (0.03)	-0.222*** (0.03)	-0.483*** (0.04)
Period 16	-0.482*** (0.03)	-0.258*** (0.03)	-0.477*** (0.04)
Period 17	-0.474*** (0.03)	-0.224*** (0.03)	-0.477*** (0.04)
Period 18	-0.485*** (0.03)	-0.240*** (0.03)	-0.491*** (0.04)
Period 19	-0.472*** (0.04)	-0.234*** (0.03)	-0.474*** (0.04)
Period 20	-0.474*** (0.04)	-0.233*** (0.03)	-0.500*** (0.04)
Seasonality Dummy	Yes	Yes	Yes
Duration Dummy	Yes	Yes	Yes
<i>N</i>	3774516	3774516	3774516
<i>R</i> ²	0.046	0.004	0.037

Notes: This table lists the estimates of β_τ in regression 6 by using Answers, Edits, and Votes as the dependent variable. β_τ measures the differences in logged activities between period τ and -2. The values of β_τ are plotted in figure 8. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

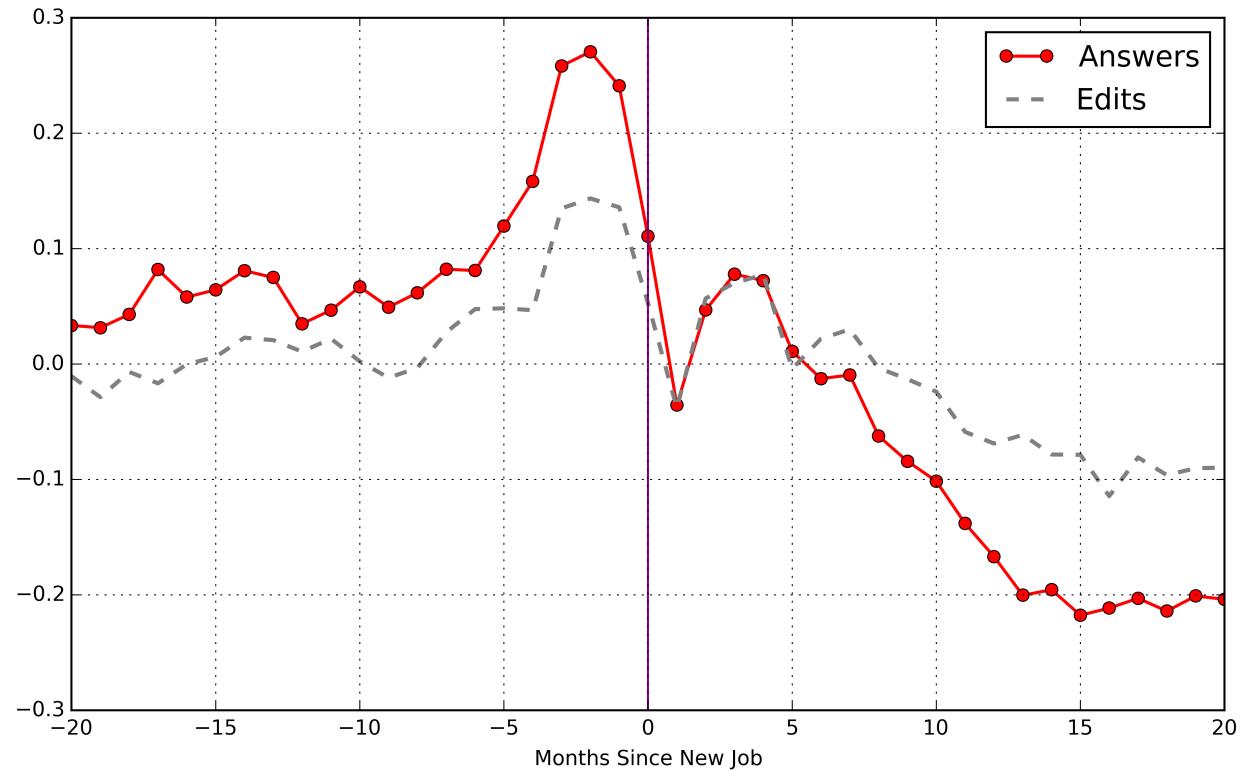


Figure 8: Coefficients Of Periods Relative To Time Of Job Change

Notes: This figure plots values of β_τ in regression 6, using Answers and Edits as dependent variables. It shows how Answers and Edits activity change over time, after removing seasonality and duration effects. Essentially, it is an extended version of figure 7, while controlling for seasonality and duration effects. The detailed value of coefficients are listed in table 5.

Table 5: Month-to-Month Comparison: Difference-in-Differences

	(1) $y \in \{a, e\}$	(2) $y \in \{v, e\}$	(3) $y \in \{q, e\}$
Period -20	-0.083 (0.05)	-0.058 (0.06)	0.110*** (0.04)
Period -19	-0.067 (0.05)	-0.065 (0.05)	0.136*** (0.04)
Period -18	-0.077 (0.05)	-0.073 (0.05)	0.092*** (0.03)
Period -17	-0.029 (0.05)	-0.033 (0.05)	0.129*** (0.03)
Period -16	-0.069 (0.05)	-0.052 (0.05)	0.119*** (0.03)
Period -15	-0.069 (0.05)	-0.058 (0.05)	0.093*** (0.03)
Period -14	-0.069 (0.05)	-0.057 (0.05)	0.068** (0.03)
Period -13	-0.073 (0.05)	-0.058 (0.05)	0.080** (0.03)
Period -12	-0.103** (0.05)	-0.100** (0.05)	0.098*** (0.03)
Period -11	-0.102** (0.05)	-0.097** (0.05)	0.090*** (0.03)
Period -10	-0.063 (0.04)	-0.052 (0.05)	0.106*** (0.03)
Period -9	-0.066 (0.04)	-0.051 (0.05)	0.120*** (0.03)
Period -8	-0.062 (0.04)	-0.078* (0.05)	0.114*** (0.03)
Period -7	-0.073* (0.04)	-0.045 (0.04)	0.076** (0.03)
Period -6	-0.094** (0.04)	-0.071* (0.04)	0.053* (0.03)
Period -5	-0.056 (0.04)	-0.062 (0.04)	0.061** (0.03)
Period -4	-0.015 (0.04)	-0.009 (0.04)	0.072*** (0.03)
Period -3	-0.004 (0.03)	-0.016 (0.04)	-0.005 (0.03)
Period -2 (Base)	0 (-)	0 (-)	0 (-)
Period -1	-0.022 (0.03)	-0.019 (0.04)	-0.026 (0.02)
Period 0	-0.069* (0.04)	-0.037 (0.04)	0.030 (0.03)
Period 1	-0.125*** (0.04)	-0.134*** (0.04)	0.116*** (0.03)
Period 2	-0.137*** (0.04)	-0.165*** (0.05)	0.097*** (0.03)
Period 3	-0.120*** (0.04)	-0.137*** (0.05)	0.083*** (0.03)
Period 4	-0.132*** (0.04)	-0.127*** (0.05)	0.051* (0.03)
Period 5	-0.112** (0.04)	-0.104** (0.05)	0.120*** (0.03)
Period 6	-0.162*** (0.04)	-0.175*** (0.05)	0.064** (0.03)
Period 7	-0.167*** (0.04)	-0.158*** (0.05)	0.055* (0.03)
Period 8	-0.186*** (0.04)	-0.194*** (0.05)	0.087*** (0.03)
Period 9	-0.198*** (0.04)	-0.206*** (0.05)	0.080*** (0.03)
Period 10	-0.205*** (0.04)	-0.212*** (0.05)	0.082*** (0.03)
Period 11	-0.207*** (0.04)	-0.197*** (0.05)	0.111*** (0.03)
Period 12	-0.225*** (0.04)	-0.223*** (0.05)	0.116*** (0.03)
Period 13	-0.266*** (0.04)	-0.272*** (0.05)	0.122*** (0.03)
Period 14	-0.244*** (0.04)	-0.250*** (0.05)	0.132*** (0.03)
Period 15	-0.266*** (0.04)	-0.261*** (0.05)	0.135*** (0.03)
Period 16	-0.224*** (0.04)	-0.219*** (0.05)	0.136*** (0.03)
Period 17	-0.249*** (0.04)	-0.252*** (0.05)	0.134*** (0.03)
Period 18	-0.245*** (0.04)	-0.252*** (0.05)	0.138*** (0.03)
Period 19	-0.238*** (0.04)	-0.240*** (0.05)	0.140*** (0.03)
Period 20	-0.241*** (0.05)	-0.267*** (0.05)	0.120*** (0.03)
Seasonality Dummy	Yes	Yes	Yes
Duration Dummy	Yes	Yes	Yes
N	7549032	7549032	7549032
R ²	0.034	0.029	0.022

Notes: This table lists the estimates of γ_τ in regression 7 by using Answers, Votes, Questions, together with Edits, as the dependent variables. Edits activity is used as the control group. γ_τ compares the activity changes between period τ and -2. The values of γ_τ are plotted in figure 9. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

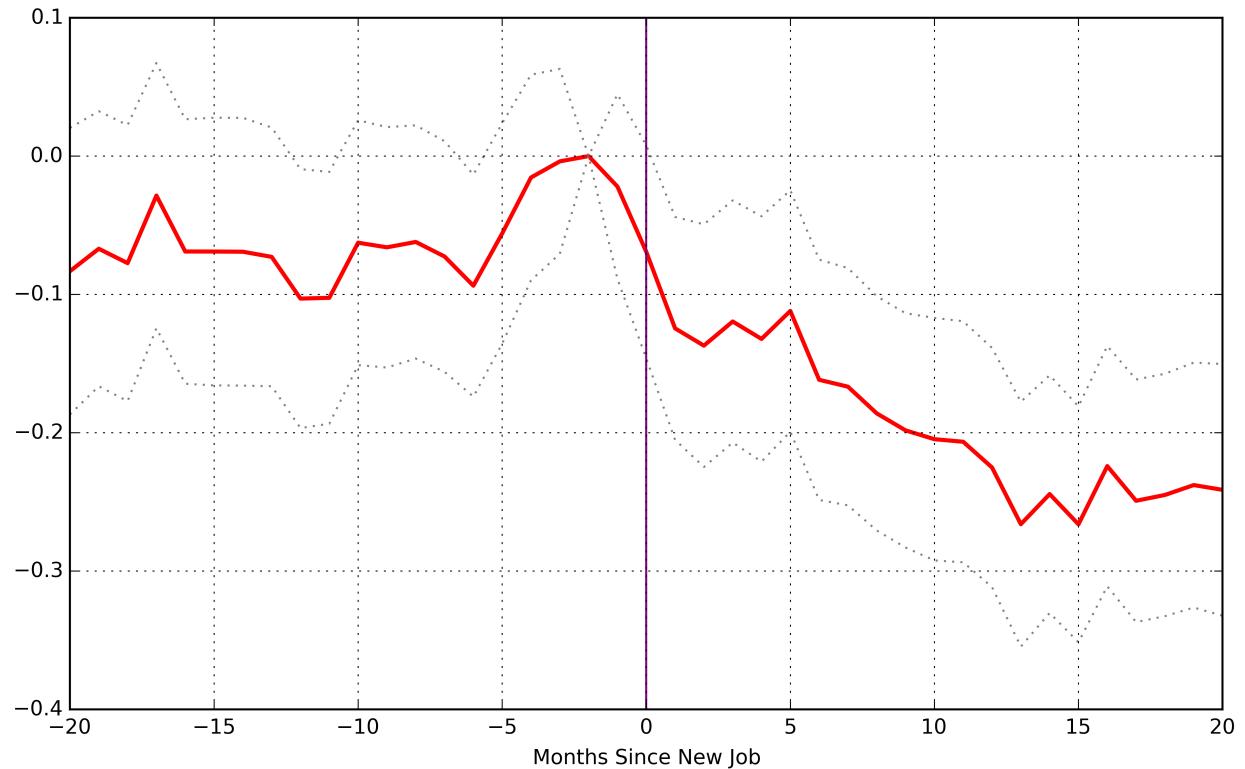


Figure 9: Difference-in-Differences With 95-Percent Confidence Interval

Notes: This figure plots values of γ_τ in regression 7. It uses period -2 as baseline period, and estimates DD coefficients using each of all other periods against the baseline period. It controls for seasonality and duration effects for answers and edits separately.

concerns. Second, contributions to SO can become a habit, which changes over time at a slow pace. Without taking into account habit formation, our DD estimates provides a lower bound of the effect of career concerns.

6 Testing Identification Assumptions

Our identification relies fundamentally on the a parallel trends assumption in Answers and Edits activity. That is to say, if it weren't for a job shift (thus without changes in career incentives), the relative ratio 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. In this section, we first provide some evidence to support the assumption. Then we will discuss and test for several major challenges to the assumption.

6.1 Evidence of Parallel Trends: Within Job Activity

In figure 8, the level of career concerns before period -5 is probably not very strong, and the variations of Answers and Edits activity provide some support to the parallel assumption.

To show some cleaner evidence, we identify a period of time when no job changes take place for each contributor, that is, a period of stable employment. It seems reasonable to assume that, during these periods, though a contributor's time availability fluctuates, 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.

Figures 10 shows the values of Answers and Edits for months 5 to 42 after an agent's job shift. Consistently with our underlying assumption, the ratio between the two is fairly constant.

6.2 Integer Constraints

A job change brings not only a change in career incentives, but also a change in time availability. The validity of the career-concerns hypothesis relies on the assumption that a

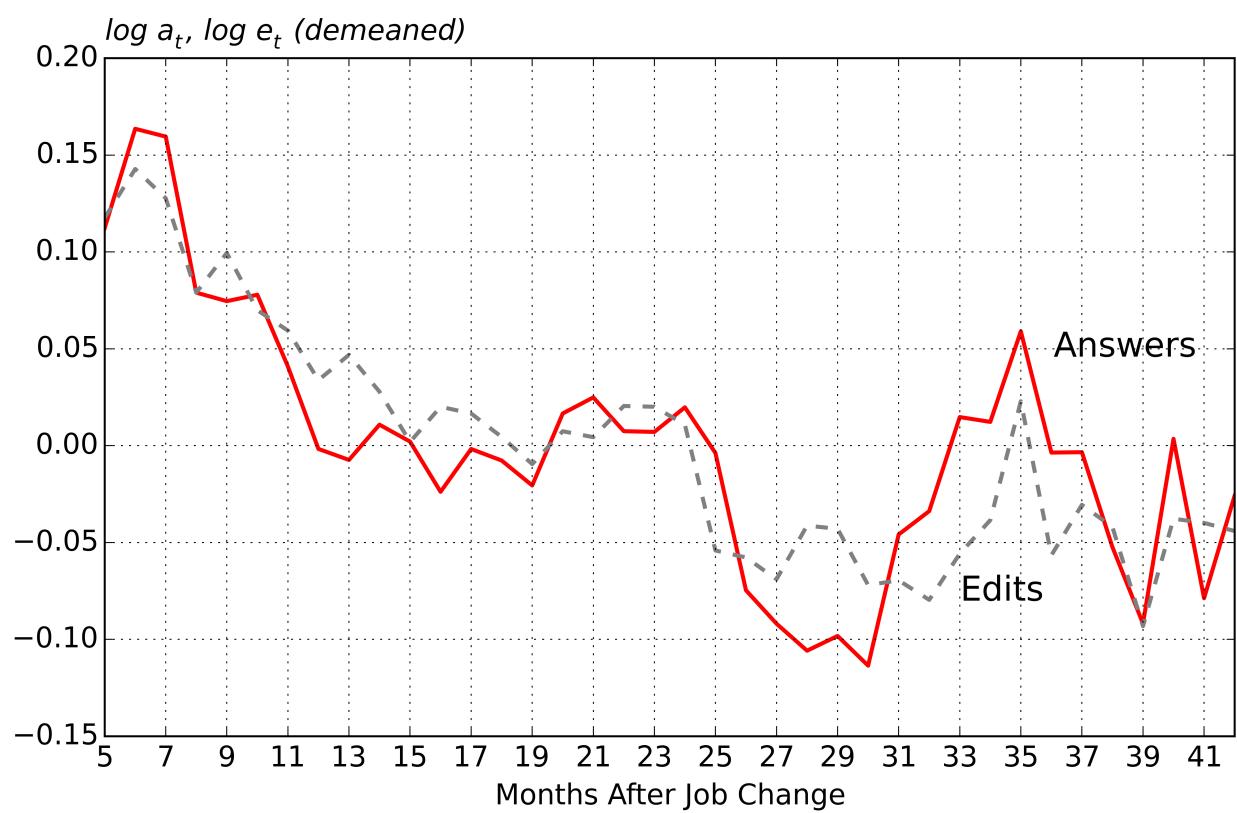


Figure 10: Within-Job Variation of Activities

sudden change in time availability associated to a busier job affect Answers and Edits in a similar fashion. An alternative interpretations for our result that a_t/e_t drops subsequent to a job change is that users are faced with an “integer constraint:” Answers require a bigger set-up cost than Edits; and when an agent switches jobs, thus becoming busier, there are 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.

The ideal way to test the parallel trend assumption is to check the changes in Answers and Edits activity followed by an *exogenous* job change, namely, an un-anticipated job change.¹⁸ With an *exogenous* job change, contributors experience changes in work responsibility thus changes in time availability, but the level of career concerns stay relatively the same. Unfortunately, such information is rarely available from information on a CV. However, with the whole contribution history and employment history for each user, we can focus on time periods with relatively stable level of career concerns, as well as on activities that take place on the weekends.

Weekdays vs. Weekend. A new job with a busier work schedule should affect time availability mostly during weekdays rather than weekends. Accordingly, we split our sample into weekday and weekend activities and conducted separate DD analyses. The idea is that, to the extent that work hours are more highly concentrated on weekdays, the integer-constraint hypothesis should imply a bigger effect on a_t/e_t during weekdays.

Table 6 shows the results of our basic DD regressions split into weekday and weekend days. Panel A and B uses Answers and Votes as measures of reputation-generating activity.

Without controlling for seasonality and duration effects, Answers activity has an additional 16.3% drop during weekdays (column 1) and 12.6% during weekends (column 3). Broadly speaking, the coefficient estimates are similar to those in the base model, and the difference between the estimates using weekday and weekend activity is small (16.3% - 12.6% = 3.7%). Regression results with controls (columns 2 and 4) and using Votes (Panel B) give very similar results.

The small difference between DD estimates using weekday and weekend activity implies

¹⁸Examples include unexpected promotion, or merger and acquisition.

Table 6: Effect of Job Change on Weekday vs. Weekend Activity

	Panel A: $y \in \{\text{Answers, Edits}\}$				Panel B: $y \in \{\text{Votes, Edits}\}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NewJob(S)$	-0.070*** (0.02)	-0.072*** (0.02)	-0.061** (0.02)	-0.060** (0.02)	-0.070*** (0.02)	-0.072*** (0.02)	-0.061** (0.02)	-0.060** (0.02)
$NewJob(S) \times Answer(A)$	-0.163*** (0.04)	-0.122*** (0.04)	-0.126*** (0.04)	-0.103** (0.04)	-0.178*** (0.04)	-0.135*** (0.04)	-0.150*** (0.05)	-0.124*** (0.05)
Seasonality Dummy	No	Yes	No	Yes	No	Yes	No	Yes
Duration Dummy	No	Yes	No	Yes	No	Yes	No	Yes
Days	Weekday	Weekday	Weekend	Weekend	Weekday	Weekday	Weekend	Weekend
# Contributors	1159	1159	374	374	1159	1159	374	374
# Users as Control	-	85k	-	45k	-	85k	-	45k
N	16104	7378770	5004	2895384	16104	7378770	5004	2895384
R^2	0.014	0.031	0.016	0.043	0.013	0.026	0.017	0.031

Notes: This table summarizes DD estimates using weekday and weekend activity separately. Panel A uses number of Answers and Panel B uses number of Votes from Answers as measures of Answers activity. A new job affects work schedule mostly during weekdays, not weekends. Independent Variable: $S_{it} = 0$ prior to job switch, $S_{it} = 1$ after job change; $A_k = 1$ if $k = a$, $A_k = 0$ if $k = e$. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

that the integer-constraints hypothesis cannot explain the additional drops in Answers after a job change, i.e. it fails to reject the career-concerns hypothesis.

However, it is possible that integer-constraints hypothesis does contribute to the small differences between weekday and weekend estimates (1.9% with control variables). In order to further investigate the effect of integer-constraints hypothesis, we focus on the case of internal promotion.

Internal Promotion. Internal promotion is a special case in two ways: First, the hypothesis of integer-constraints is most plausible to a promotion, since it is associated to a more dramatic change in work schedule compared to switching to a similar job. Second, our story of career concerns where a job seeker signals to potential employers through online activity is unlikely to the case of internal promotion. The signals are valuable only when an employer cannot observe the true quality of a job seeker. However, the current employer can perfectly observe the job-seeker quality from internal projects.

Table 7 includes DD estimates from internal move and internal promotion using weekday and weekend activities separately. Panel A uses all contributors, which is the same as Panel A (column 2, 4)) of table 6. It is included here for comparison purposes. Panel B focuses on internal moves, i.e. job switches within the same company. Both estimates in Panel B are

Table 7: Effect of Job Change for Internal Promotions

	Panel A: All		Panel B: Same Company		Panel C: Promotion	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>NewJob</i> (S)	-0.072*** (0.02)	-0.060** (0.02)	-0.070 (0.05)	-0.039 (0.08)	-0.151** (0.06)	-0.157 (0.10)
<i>NewJob</i> (S) \times <i>Answer</i> (A)	-0.122*** (0.04)	-0.103** (0.04)	-0.090 (0.09)	-0.119 (0.13)	-0.156 (0.11)	-0.008 (0.19)
Seasonality Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Duration Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Days	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Contributors	1159	374	142	41	80	21
N	7378770	2895384	7364466	2890896	7363686	2890644
R^2	0.031	0.043	0.031	0.043	0.031	0.043

Notes: This table summarizes DD estimates using internal move (Panel C: job switches within a company) and internal promotion (Panel C: strict measure of promotion to a higher position based on job titles). Integer-constraints should have a large effect for internal promotion, while career-concerns hypothesis is unlikely the case in this case. Independent Variable: $S_{it} = 0$ prior to job switch, $S_{it} = 1$ after job change; $A_k = 1$ if $k = a$, $A_k = 0$ if $k = e$. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

similar to those in Panel A in magnitude but both become insignificant (most likely due to a smaller sample size). One potential concern is that many internal job changes are lateral moves to a different position, which adds unnecessary noise to our analysis. In Panel C, we focus on internal promotions using a stricter measure based on job title information.¹⁹ DD estimate from column 5 of Panel C mean that Answers experience an additional drop of 15.6% compared to Edits on the weekdays. At the same time, DD estimate using weekend data is very close to zero in magnitude, which is consistent to integer-constraint hypothesis. Moreover, Edits activity during both weekdays and weekends experience a large drop of 15%. One possible explanation is that an internal promotion requires more managerial work and less day-to-day programming, thus the time spent on SO also experience a large drop.

Switch to a similar job. The integer-constraint hypothesis argues that the a new job allows for small blocks of free time enough for Edits, but not for Answers. Although without data on the work schedule, we can use the information within job titles to infer work schedule

¹⁹We define an internal job change as a promotion if the keywords in the new job title not contained in the old job title includes one of the following words (case insensitive): Senior, Lead, Manager, Director, Sr., Principal, Specialist, Administrator, Chief, Associate, President, CEO, Vice, Leader, Director, VP, Partner, Management, Head, Advisor, Full, Supervisor, Executive, President, Principle.

Table 8: Effect of Job Change for Switching to Similar Jobs

	Baseline	Same Job Title			
	(1)	(2)	(3)	(4)	(5)
<i>NewJob</i> (S)	-0.082*** (0.02)	-0.028 (0.06)	-0.028 (0.06)	-0.008 (0.07)	-0.028 (0.08)
<i>NewJob</i> (S) \times <i>Answer</i> (A)	-0.124*** (0.03)	-0.181* (0.11)	-0.138 (0.11)	-0.161 (0.11)	-0.244* (0.14)
Seasonality Dummy	Yes	No	Yes	Yes	Yes
Duration Dummy	Yes	No	Yes	Yes	Yes
Days	-	-	-	Weekday	Weekend
# Contributors	1249	155	155	144	45
# Users as Control	85k	-	85k	85k	45k
N	7380646	1992	7365094	7364828	2919924
R^2	0.031	0.010	0.031	0.031	0.043

Notes: This table summarizes DD estimates using new jobs with same job titles. New jobs with same job titles should have similar work schedule, thus they are not subject to any integer-constraints effect. Independent Variable: $S_{it} = 0$ prior to job switch, $S_{it} = 1$ after job change; $A_k = 1$ if $k = a$, $A_k = 0$ if $k = e$. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

changes. By focusing on new jobs with similar job titles, we minimize the influence of changes in the nature of the work performed. Assuming that new jobs with the same job title have similar responsibility thus similar control over one's free time, we conduct separate analyses for new jobs with exactly the same job titles.

Table 8 summarizes the DD estimates. Out of 1249 contributors, 155 (or 12.5%) changes to new jobs with exact same job titles. Due to the small sample size, most of the results become statistically insignificant but the point estimate tells a consistent career-concerns story. Table 8 shows that, for this group of contributors, only Answers activity experience a large drop after a job change (18.1% and 13.8% with and without controls in columns 2 and 3), while Edits stay at a similar level. Regressions using weekday and weekend activity also show similar results (columns 4 and 5).

6.3 Skill Mismatch

An alternative interpretation for the drop in Answers following a job shift is that the new occupation requires skills different from the previous job. For example, a C++ programmer

Table 9: DD Estimates For New Jobs With Similar Technology

	(1) $y \in \{a, e\}$	(2) $y \in \{a, e\}$	(3) $y \in \{a, e\}$	(4) $y \in \{a, e\}$	(5) $y \in \{a, e\}$
<i>NewJob</i> (S)	-0.013 (0.05)	-0.054 (0.05)	-0.071** (0.04)	-0.075*** (0.03)	-0.080*** (0.02)
<i>NewJob</i> (S) \times <i>Answer</i> (A)	-0.209** (0.09)	-0.153* (0.08)	-0.162*** (0.06)	-0.157*** (0.04)	-0.165*** (0.03)
Job Similarity (by Tags)	$\geq 100\%$	$\geq 75\%$	$\geq 50\%$	$\geq 25\%$	$\geq 0\%$
Contributors	162	240	407	778	1249
N	2064	3000	5184	10272	17544
R^2	0.012	0.011	0.014	0.014	0.015

Notes: Dependent Variable: y : generic activity, a : Answers, e : Edits. Independent Variable: $S_{it} = 0$ prior to job switch, $S_{it} = 1$ after job change; $A_k = 1$ if $k = a$, $A_k = 0$ if $k = e$. Time Period: November 2008 - November 2014. Robust standard errors in parentheses, clustered at individual-activity type level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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).

As shown in figure 4, 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 focus on users who switch to new jobs with similar technology based on tags information. First, we define a measure of skill-similarity between jobs.²⁰ Then we re-estimate the basic DD regression using users who have switch to jobs with similar technology.

Table 9 summarizes the DD estimates using different thresholds job similarity. Column 1 focuses 162 job changers whose new jobs have exactly the same tags as the old ones, and DD coefficient gives an estimate of -20.9%, which is a larger magnitude compared to the estimate of -16.5% from our baseline model. Moving to the right, column 2 to 5 gradually lower the thresholds of job similarity and includes more job changers in the regression. Column 5 includes all the users which is identical to our baseline model.

The results in table 9 show that job changers who switch to positions with similar skill requirement also experience a similar drop in Answers over Edit activity. The magnitudes

²⁰Let the set of tags in the new job be S_1 , and the set of tags in the old job be S_0 . We define $JobSimilarity \equiv \frac{Size(S_0 \cap S_1)}{(Size(S_0) + Size(S_1))/2}$.

of the estimates using various thresholds are also comparable to results from the baseline model. Thus we conclude that the DD estimate cannot be explained by the skill mismatch hypothesis.

6.4 Dynamic Selection Effects (Ashenfelter's Dip)

The variations of activity levels plotted in figure 7 can be explained by the dynamic selection effect where the sample of contributors are selected into a job change due to special events before the job change. It is a hypothesis commonly referred to in the labor economics literature as Ashenfelter's Dip (AD).²¹

Suppose that contributors experience random shocks in the number of Answers and Edits in each period. Suppose also that a higher number of Answers helps getting job offers. Then the sample of job shifters tends to include those who experience a large Answers shock in periods immediately preceding a job change. In that case, the “bump” in the number of Answers before job changes (as the one in figure 7) is purely caused by the selection into treatment from random activities, not by a change in user behavior in response to incentives.

Through simulation of random activities and given some job changing function, we can easily simulate the problem of AD. Figure 11 plots the simulated activities over time. It shows an increase in Answers activity before a job change, and a reduction afterwards, which resembles figure 7. However, the activity increase is purely due to selecting periods followed by large Answers shock as time of job change, and the reduction is purely due to recovery from the Answers shock.

Undoubtedly, this alternative explanation poses a valid concern to our attempts to identify career concerns from the change in activity surrounding a job change. It also touches on the issue of reverse causality or selection into treatment in which A_{t-1} causes $NewJob_t$. In the classical AD problem, the dip is assumed to be due to random shocks. Therefore, the correction to the AD concern typically involves removing the random shock by using data from periods away from the treatment period or by matching treatment group with a properly selected control group who also experience a similar shock. However, in our setting

²¹In a more general econometric setting, AD problem can be considered as problems of endogeneity through reverse causality or selection into treatment. Please refer to Ashenfelter (1978) for a detailed discussion.

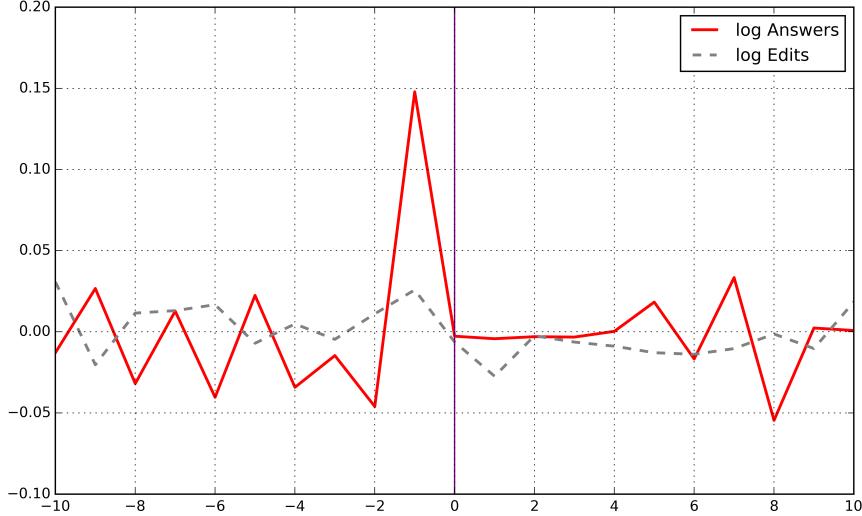


Figure 11: Simulation of AD Problem

Notes: Given random shocks of Answers and Edits activity, and given a job changing function that increases in lagged Answers activity with certain parameter value, one can simulate job changes and plot a graph similar to what we observed using actual SO activity data.

of Stack Overflow, the “bump” in the plot, which is like a reversed “dip”, is key to our identification strategy. In essence, we are estimating the size of the “bump” and interpret it as a behavioral response by contributors due to career concerns, rather than a design problem with selection into treatment from random shocks.

Identification of AD vs. Career Concerns. In the remaining part of this section, we argue, by means of numerical simulations, that AD does not provide compelling evidence against our career concerns story. We attempt to answer the following questions: If we assume everything is random and the level of Answers activity helps getting a new job, then how large would the DD estimates be? Would the estimates be statistically significantly? Would they allow us to reject the career concerns hypothesis? If not, under what conditions would we be able to do so?

First, we draw random Answers and Edits activity following a certain distribution and then simulate job change status given a likelihood function of job change. The simulation is then repeated R times and the DD estimate from each simulation is collected to plot the distribution of the estimates. The comparison between simulated and original DD estimates can show the confidence level with which we can reject the null hypothesis that the “bump”

in the number of Answers is explained purely by selection of random shocks instead of behavioral responses.

The simulation requires two main inputs: 1. Random draws of Answers and Edits activity 2. Probability of a job change given Answers activity from the previous period. The simulation parameters can significantly affect the simulation results.

For the first part of the input, we draw Answers and Edits activity from the actual Answers and Edits activity. In order to keep the data as clean as possible —i.e., without potential effect of career incentive — we use activity during a period that is at least five months away from a job change. We draw Answers and Edits both in pairs and separately. The results using the two approaches help us compare the typical DD approach in settings such as AD to the DD approach developed earlier in the paper. As a robustness check, we also conduct simulations drawn from two separate negative binomial distributions fitted to those of actual activities, for cases both with and without correlation between the two distributions.

For the second part of the input, we use parameters from a logistic regression of job change status on lagged Answers activity.²² There is an endogeneity problem, since job search intensity correlates with both job change status and Answers activity. Unfortunately, we are unable to solve it due to lack of data on both job search intensity and job offers received. However, the estimate from a regression with an endogeneity problem should provide us the upper bound of the true value, which is the worst-case scenario to the career-concerns hypothesis. In other words, by not correcting for endogeneity we are stacking the cards against our preferred hypothesis.

For each simulation method, we simulate the DD results $R = 200$ times. Each simulation includes 1500 job switches which mimics our original DD analysis. Then we plot the estimates using kernel density plot.

The simulation results are plotted in figure 12. Panel A plots simulated DD estimates using Answers and Edits drawn directly from actual activity data, both in pairs (blue line) and separately (green line). Both simulations have a mean slightly larger than zero at 2-3%,

²²We also checked simulation results using alternative specifications, e.g. including or excluding reputation level, lagged Edits, etc. They all produce similar results.

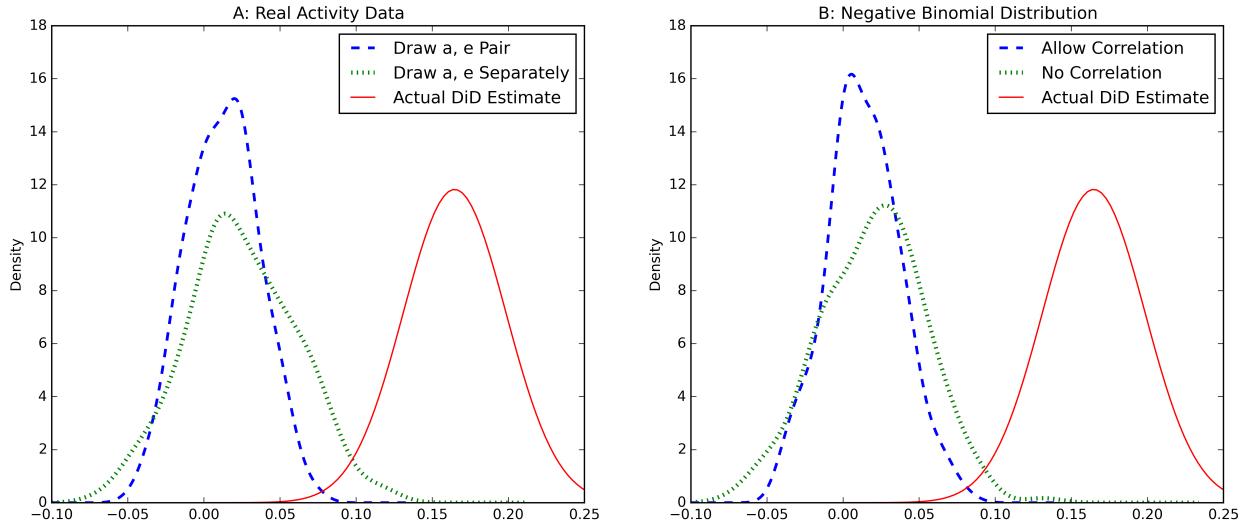


Figure 12: Density Plot of Simulated DD Estimates

Notes: Panel A draws data directly from actual activity data; Panel B draws data from negative binomial distributions fitted from Answers and Edits data. Blue lines allows for correlation between Answers and Edits; Green lines draws Answers and Edits independently. Red lines plot the distribution of the actual DD estimate with mean of 0.165 and standard deviation of 0.034.

but neither is significantly different from zero. The red line plots the distribution of the actual DD estimate from column 2 of table 2, with a mean of 0.165 and a standard deviation of 0.034. In Panel B, instead of drawing random activity directly from actual activity data, we first fit two negative binomial distributions for Answers and Edits activity using MLE. Then we conduct the same set of simulation and plot the distributions of simulated DD estimates. Panel B gives similar plots to Panel A.

The comparison between the simulation and actual DD estimates shows that the DD estimate of .165 cannot be explained by selection into a job change due to random activities, given reasonable ranges of coefficient values.

Simulations using Answers and Edits drawn separately give a wider range of DD estimates than those using data drawn in pairs. In reality, the number of Answers and Edits given in a month by a contributor is always correlated since both are correlated with the time spent on Stack Overflow.²³ If the two activities are perfectly correlated, then simulated DD always gives zero estimates. However, when drawn independently, Answers and Edits are uncorrelated, thus it's more likely to observe high levels of Answers activity with low Edits

²³The actual correlation between Answers and Edits is 0.564.

activity, which presents a graph similar to ours.

Another main reason that the logic of AD problem doesn't invalidate the career-concerns hypothesis is the small effect of Answers activity on new job offers. Though unable to accurately estimate this effect due to the presence of endogenous variables (as we mentioned above), we can obtain an upper bound of the true value. The fact that we cannot reject the career-concerns hypothesis using the upper bound estimate gives us even more confidence of our conclusion.

Parameter Values Required to Reject Career Concerns. Figure 12 shows a significant gap between simulated and actual DD estimates, which helps us to reject the story of AD in favor of career concerns. However, the simulation crucially depends on the two inputs discussed above. In this subsection, we adjust the second input, the probability of job changes given a certain level of Answers activity, and calibrate the parameters in order to mimic the actual DD estimate.

The job change probability is modeled using the following simple logit model:

$$Pr(JC) = Logit(\alpha + \beta * A) = \frac{exp(\alpha + \beta * A)}{1 + exp(\alpha + \beta * A)} \quad (8)$$

First, we calibrate α while holding $A = 0$ by matching the simulated job length to the distribution of actual job lengths from the data. With the calibrated value $\hat{\alpha} = -3.07$, the unconditional rate of job change is $\frac{exp(\hat{\alpha})}{1+exp(\hat{\alpha})} = 4.43\%$. Then, holding $\alpha = -3.07$, we simulate DD estimates for different values of β . For each β , we run the simulation for $R = 100$ times and plot the average value in figure 13.

Figure 13 shows that the value of $\beta = 22.76\%$ produces a simulated DD estimate that is closest to our actual DD estimate of 16.5%. $\beta = 22.76\%$ means that one log unit increase of number of Answers increases the likelihood of a job change by 22.76%. One log unit increase is roughly an increase of 170%. For all our 1,249 SOC users used in DD analysis, the average total number of Answers per month is 4.05, thus an increase of 170% means extra $4.05 \times 170\% = 6.88$ Answers per month. That is to say, for an average person who gives 4 Answers per month, simply by answering 6.88 more Answers can increase the likelihood of a

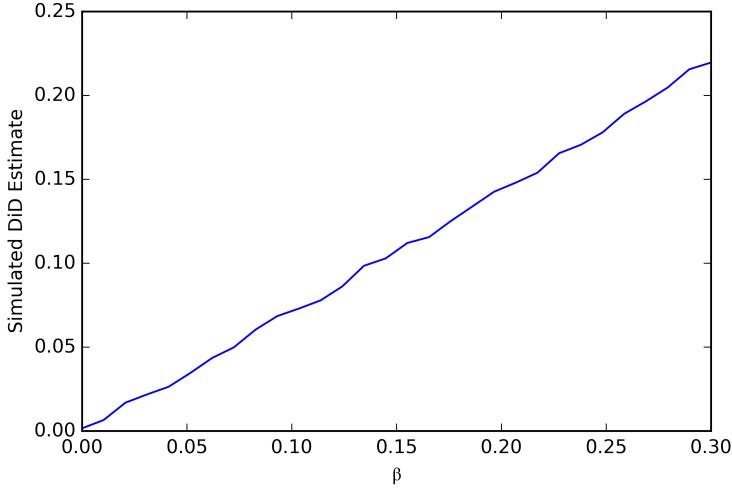


Figure 13: Average Simulated DD Estimates Given Values of β

Notes: The figure plots average simulated DD estimates by using different values of β .

job change in the following period by 22.76%, which is extremely high given the small cost of giving Answers. Therefore, we conclude that the causal effect represented by β is overestimated. That is to say, in order to reject the career-concerns hypothesis, the parameter values needed in the job change function become too large to be reasonable.

7 Extension & Other Results

7.1 Signaling Game: Quality vs Quantity

The classic career-concerns hypothesis in Holmström (1982/99) shows that job seekers exert effort to signal their unobserved ability. Then the natural questions to ask are (a) what are job seekers signaling through SO, and (b) what information do employers get from the online activity of a job seeker? This is an important research question related to but different from the focus of this paper. Marlow and Dabbish (2013) interview several contributors and employers on GitHub regarding job searching activities using activities on GitHub.²⁴ Employers consider the fact of merely having a GitHub profile as a good signal. They also evaluate job applicants' activity, specifically the desire to learn, popularity of a

²⁴GitHub is a online repository hosting service popular among programmers. It offers services including revision control and source code management. As of 2015, GitHub has over 9 million users and over 21.1 million repositories, making it the largest host of source code in the world.

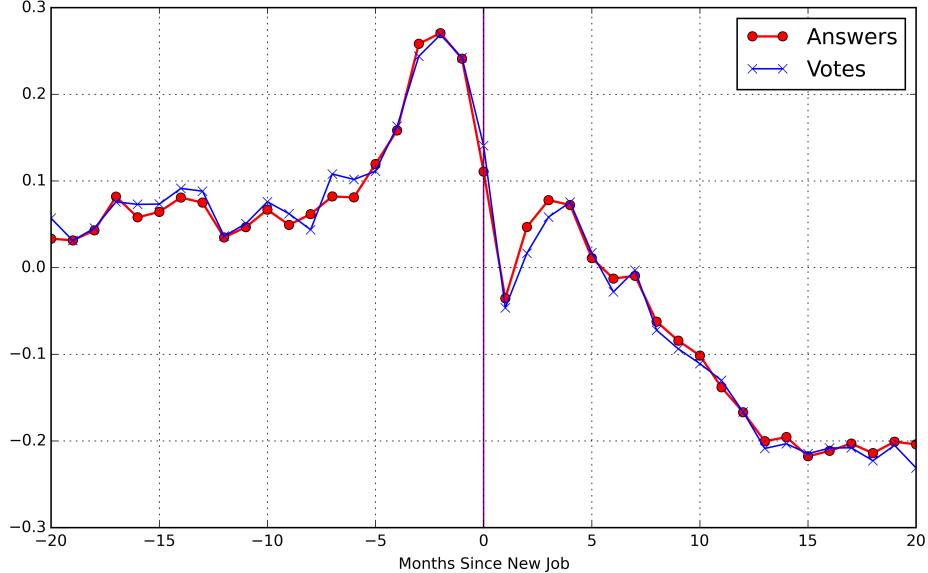


Figure 14: Different Measure of Answers Activity (Number of Answers vs. Votes from Answers)

Notes: This figure presents alternative measurement of Answers activity: Votes from Answers up to 30 days after an answer is given. It takes into consideration of both quality and quantity of Answers. Both measures give almost identical graph shows that the average quality of Answers does not change.

project, and coding style, etc. Job seekers, in turn, expect to show their passion and expertise to employers.

“Popularity” and “expertise” on GitHub are similar to the quality of Answers on Stack Overflow, which can be roughly measured by Votes. In principle, it is possible that the effect of a job shift is also felt in terms of the quality of Answers. Our DD results using number of Answers and Votes in previous sections give very similar estimates. Figure 14 plots the time evolution of Votes and Answers. The correlation between the two measures is remarkably high. The fact that the average quality of Answers remains constant seems to contradict the basic intuition of the career concerns story. However, one cannot conclude that career concerns have no effects on the quality of Answers. Given a fixed supply of questions, the additional efforts to answer questions should lead to *both* better Answers from questions a contributor would answer regardless of career concerns, *and* more Answers from questions a contributor would not answer without career concerns due to low matching quality. Thus one should observe that as the time of a job change approaches, a job seeker gives more Answers, and at the same time, the quality of some Answers are higher but others are lower.

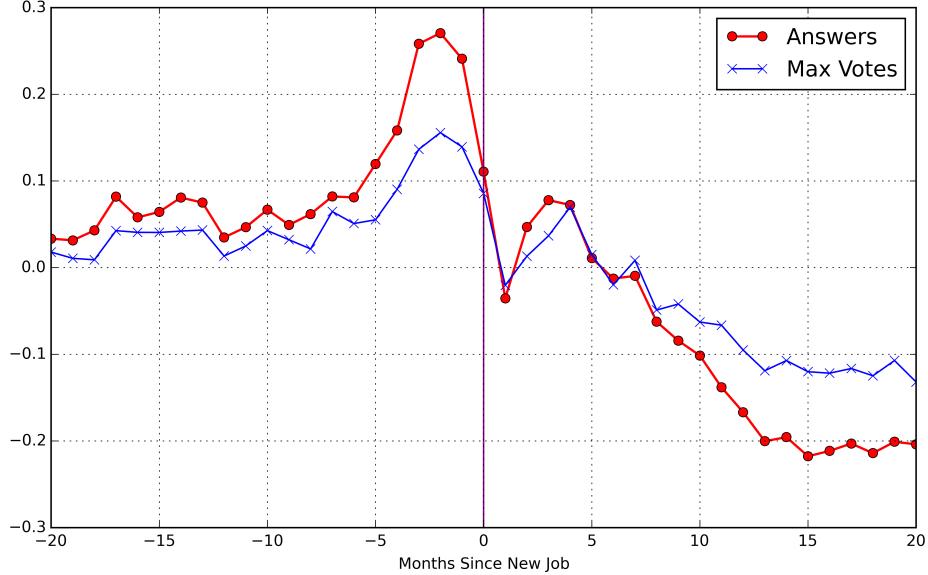


Figure 15: Answer Quality Variation Over Time

Notes: This figure decomposes total Votes from Answers. It shows that even though average quality does not change, the quality of best Answers in a given month indeed goes up before a job change, which indicates increased efforts to give better answers.

To test this hypothesis, we pick the best answer (measured by Votes) given by a contributor for each month, and figure 15 plots the average Votes from the best Answers over time. It shows that the quality of best Answers follows a similar pattern to the number of Answers, which is consistent to the hypothesis that apart from quantity of Answers, contributors also improve the quality of Answers before a job change. However, caution should be taken regarding the causality, since the result can also be explained by the random distribution of matching quality, with which the largest order statistic (Max Votes) goes up with a larger sample (number of Answers).

7.2 Reputation Sizes and Timing of Answers

The career-concerns hypothesis states that job seekers signal their unobserved ability to employers through reputation-generating activity on Stack Overflow, namely Answers. If that is the case, and if employers value the stock of Answers more than the timing of Answers, then career concerns should have heterogeneous effects on job seekers with different levels of reputation. A job seeker with low reputation on Stack Overflow might prefer even not to

Table 10: DD Estimates by Reputation Points

	Panel A: $y \in \{a, e\}$				Panel B: $y \in \{v, e\}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NewJob(S)$	0.161*** (0.04)	-0.065* (0.03)	-0.170*** (0.04)	-0.247*** (0.04)	0.161*** (0.04)	-0.065* (0.03)	-0.170*** (0.04)	-0.247*** (0.04)
$NewJob(S) \times Answer(A)$	0.048 (0.06)	-0.254*** (0.06)	-0.282*** (0.07)	-0.171** (0.07)	0.053 (0.06)	-0.206*** (0.06)	-0.336*** (0.07)	-0.226*** (0.08)
Reputation	0-25%	25-50%	50-75%	75-100%	0-25%	25-50%	50-75%	75-100%
Contributors	358	354	343	314	358	354	343	314
N	4392	4392	4380	4380	4392	4392	4380	4380
R^2	0.018	0.027	0.053	0.048	0.018	0.018	0.055	0.045

Notes: This table summarizes regression estimates by dividing the job changes into four groups by reputation points at the time of job change. Job seekers with medium-reputation respond most to career concerns, whereas low-reputation job seekers probably do not use SO as a signaling channel. Dependent Variable: y : generic online activity, a : Answers, e : Edits, v : Votes from Answers. Reputation Points: Min: 0; First Quartile: 770; Median: 2,124; Third Quartile: 5,265; Max: 132,067 Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

include his or her SO profile on their resume. By contrast, for a job seeker with outstanding reputation the marginal benefit of extra effort to contribute should be relatively small. To examine the heterogeneous responses to career concerns due to existing reputation points, we associate each job change with the reputation level at the time of job change, and split the sample into four equal groups by reputation points for separate analysis.

Table 10 summarizes the DD estimates from four groups of job seekers using number of Answers (panel A) and Votes from Answers (panel B) as measures of reputation-generating activity. Panel A shows that low-reputation users (column 1) do not seem to change their behavior as a result of a job change. By contrast, medium-reputation users (columns 2 and 3) show effects of magnitudes larger than results from our base regressions, and high-reputation users (column 4) very similar to the average. Panel B presents similar result to Panel A.

7.3 Questions Activity to Build Reputation

As mentioned earlier, SO users can also receive reputation points through asking questions. We investigate how Questions activity changes around a job change event using regression 6, which measures month-to-month activity change while controlling for seasonality and duration effects.

Figure 16 shows the rate of questions asked around the time of a job shift. Unlike Answers

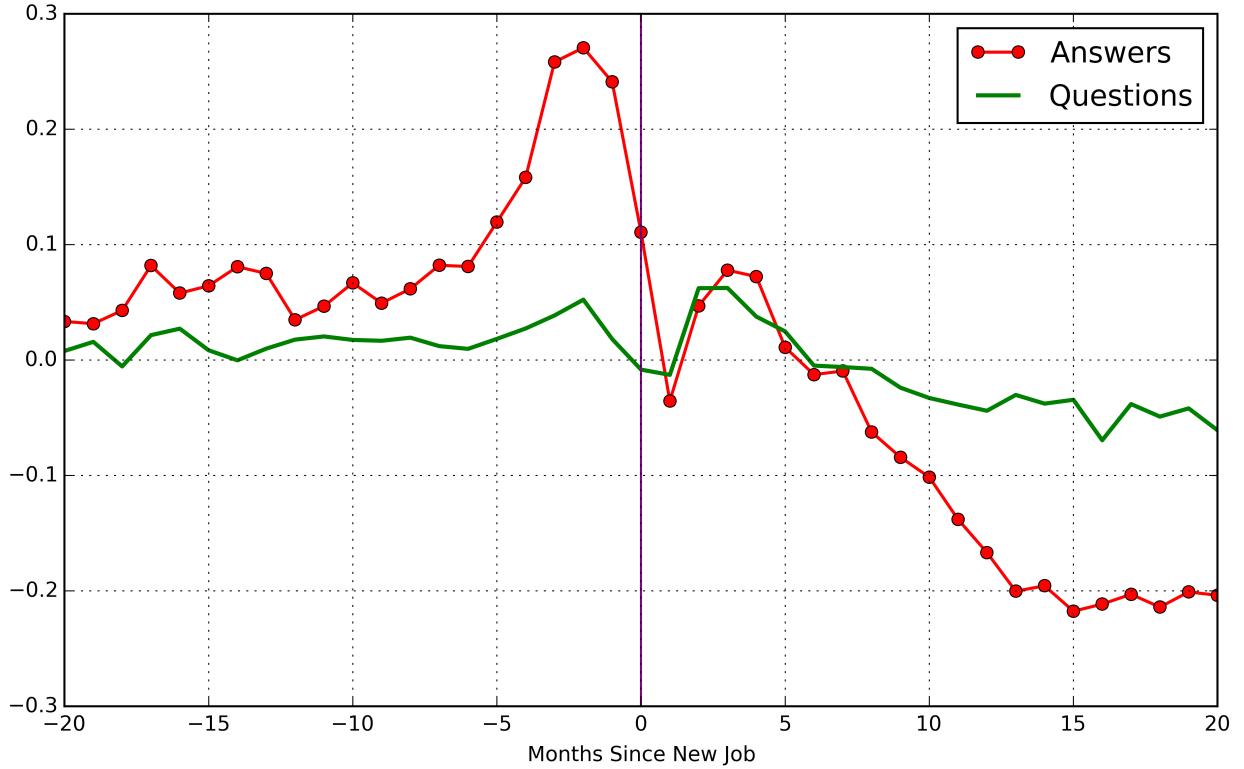


Figure 16: Average Monthly Activity on Stack Overflow (Answers and Questions)

and Edits, we observe little change in the number of questions over time. Questions activity experiences a slight drop at the end of the old job and then rises right after the starting month of a new job. One possible explanation of the rise is that, more than a reputation-increasing activity, questions are used a learning tool; and a shift to a new job creates new learning demands (even for new jobs with the same set of technology), an effect that seems to compensate for the higher opportunity cost of time spent on SO as well as the diminished incentive to build a reputation. Another possible explanation is that asking questions might be perceived as inability to solve problems, and thus job seekers avoid asking questions.

7.4 Career Concerns and Education Level, Technology Degree

The story of career concerns says that job seekers make effort to improve the output which gives information of his or her true quality. In the real world, employers judge the quality of job seekers through multiple aspects: education, work experience, age, online activity, etc. A high school dropout should have more incentive to improve the signal using other

Table 11: Career Concerns and Education Level

	Panel A: Answers				Panel B: Votes			
	(1) $y \in \{a, e\}$	(2) $y \in \{a, e\}$	(3) $y \in \{a, e\}$	(4) $y \in \{a, e\}$	(5) $y \in \{v, e\}$	(6) $y \in \{v, e\}$	(7) $y \in \{v, e\}$	(8) $y \in \{v, e\}$
<i>NewJob</i> (<i>S</i>)	0.101 (0.17)	-0.086*** (0.03)	-0.044 (0.05)	-0.036 (0.09)	0.101 (0.17)	-0.086*** (0.03)	-0.044 (0.05)	-0.036 (0.09)
<i>NewJob</i> (<i>S</i>) \times <i>Answer</i> (<i>A</i>)	-0.276 (0.26)	-0.161*** (0.04)	-0.206*** (0.08)	-0.103 (0.17)	-0.135 (0.32)	-0.172*** (0.04)	-0.221*** (0.08)	-0.148 (0.19)
# Contributors	12	780	232	53	12	780	232	53
Education	HS	College	Masters	PhD	HS	College	Masters	PhD
<i>N</i>	168	10956	3240	780	168	10956	3240	780
<i>R</i> ²	0.011	0.016	0.014	0.004	0.002	0.015	0.014	0.006

Notes: This table summarizes DD coefficients using separate analysis for each education level. Panel A uses number of Answers and Panel B uses number of Votes from Answers as measures of Answers activity. High School degree holders respond most to a job change, whereas PhD holders respond the least. Dependent Variable: *y*: generic online activity, *a*: Answers, *v*: Votes, *e*: Edits. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

activities, including online activity; a PhD holder, on the other hand, probably rely less on online activity to signal his or her ability. We extract the highest degree obtained by a SOC user and divide them into four groups: High School (HS), College, Masters, and PhD. Then we conduct separate analyses for each group.

Table 11 summarizes the DD estimates by education levels. The number of contributors in HS and PhD groups are very small, thus giving an insignificant DD estimates. However, the magnitude of results in each group is consistent to career-concerns hypothesis. HS group respond most to a job change (-27.6% in column 1), whereas PhD holders respond the least (-10.3% in column 4). Panel B using Votes gives similar results. DD estimates from Panel A and B also imply that a much higher quality of Answers given by PhD holders than those by HS group.

The same intuition can also be applied to the type of degree. Assuming all SOC users are seeking technology-related jobs, but not everyone had a technology-related degree (broadly defined as computer science, engineering, etc.). We hypothesize that non-technology-degree holders respond more to career concerns on SO.

Table 12 shows DD estimates for technology-related and non-technology-related degrees. The magnitude of the DD estimates between two groups are very similar. However, the slight difference is consistent to the story of career concerns, where technology-degree holders

Table 12: Career Concerns and Technology Degree

	Panel A: Answers		Panel B: Votes	
	(1)	(2)	(3)	(4)
	$y \in \{a, e\}$	$y \in \{a, e\}$	$y \in \{a, e\}$	$y \in \{a, e\}$
<i>NewJob (S)</i>	-0.078** (0.04)	-0.083*** (0.02)	-0.078** (0.04)	-0.083*** (0.02)
<i>NewJob (S) × Answer (A)</i>	-0.187*** (0.07)	-0.156*** (0.04)	-0.177** (0.07)	-0.176*** (0.04)
# Contributors	310	830	310	830
TechDeg	No	Yes	No	Yes
<i>N</i>	4404	11640	4404	11640
<i>R</i> ²	0.016	0.014	0.013	0.015

Notes: This table summarizes DD coefficients for technology-related and non-technology-related degrees. Panel A uses number of Answers and Panel B uses number of Votes from Answers as measures of Answers activity. The magnitude of estimates are very similar, but the slight difference is consistent to career-concerns hypothesis. Technology degree holders respond less to the event of job change; they also supply higher quality answers. Dependent Variable: y: generic online activity, a: Answers, v: Votes, e: Edits. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

respond less to career incentives compared to non-technology-degree holders. Moreover, combining the results using Votes, the results also indicate that technology-degree holders give higher quality Answers.

8 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 both intrinsic and extrinsic incentives matter. 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 consider the specific case of Stack Overflow and show that career con-

cerns 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 12.5–16.5% can be assigned to a drop in career concerns. This value is both statistically significant and economically significant. We discuss and test the validity of the identifying assumption by showing evidence related to our career-concerns hypothesis as well as alternative explanations. We also

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 possible 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. (Though we are unable to prove that it is a representative sample, we do have some anecdotal evidence from several programmers that it is a common practice to provide links to online profiles such as GitHub and Stack Overflow on the CV when applying for jobs.) Second, 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.

Last, our results suggest several additional areas for future research. Empirically, we think it is important to examine whether signaling through online activity can be generalized to other industries. Our result focuses mostly on programmers during surrounding the periods of a job change. Future studies should explore whether our results can be transferred to a broader set of industries, as well as the potential long-term crowd-out effect by extrinsic motivation. Moreover, future research can also provide practical advice to platforms. Stack Overflow encourages more user contribution through building a careers website, Stack Overflow Careers, and it constantly reminds the career benefits to its users through banners on the website. For platform designers, it's imperative to understand the proper ways to design a platform in order to improve user participation.

Appendix

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

$$V_1 > V_0 \quad (9)$$

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

$$\begin{aligned} \lambda_1 &= g'_1(w_1) \\ \lambda_1 &= f_e(a_1, e_1) \\ \lambda_1 &= f_a(a_1, e_1) \end{aligned} \quad (10)$$

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

$$\begin{aligned} \lambda_0 &= g'_0(w_0) \\ \lambda_0 &= f_e(a_0, e_0) \\ \lambda_0 &= f_a(a_0, e_0) + \delta p'(A + a_0)(V_1 - V_0) \end{aligned} \quad (11)$$

The second and third equations in (10) imply that $f_e(a_1, e_1) = f_a(a_1, e_1)$.²⁵ The second and third equations in (11) imply that $f_e(a_0, e_0) > f_a(a_0, e_0)$. Given that $f_a, f_e > 0$, together with the properties of homothetic functions, it implies that $\frac{f_a(\frac{a_0}{e_0}, 1)}{f_e(\frac{a_0}{e_0}, 1)} < \frac{f_a(\frac{a_1}{e_1}, 1)}{f_e(\frac{a_1}{e_1}, 1)}$. The convexity of $f(\cdot, \cdot)$ with respect to both a and e implies an convex indifference curve, which implies that $MRS_{ae}(x, 1)$ decreases in x . 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 (10) 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 $\delta p'(A + a_0)(V_1 - V_0)$. Given (9) and

²⁵I denote $f_a \equiv \frac{df(a_t, e_t)}{da_t}$ and $f_e \equiv \frac{df(a_t, e_t)}{de_t}$.

to the extent that $p' \geq 0$, again the value of a increases. ■

References

- Ashenfelter, Orley. 1978. “Estimating the Effect of Training Programs on Earnings.” *The Review of Economics and Statistics* 60 (1):47–57.
- 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.” *Quarterly Journal of Economics* 114 (2):389–432.
- Dobrescu, L I, M Luca, and A Motta. 2013. “What makes a critic tick? Connected authors and the determinants of book reviews.” *Journal of Economic Behavior Organization* 96:85–103.
- Gibbons, Robert and Kevin J Murphy. 1992. “Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence.” *Journal of Political Economy* 100 (3):468–505.
- 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.” *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.
- Lakhani, Karim R and Eric von Hippel. 2003. “How open source software works: “free” user-to-user assistance.” *Research Policy* 32 (6):923–943.

- 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.
- . 2005. “The Economics of Technology Sharing: Open Source and Beyond.” *The Journal of Economic Perspectives* 19 (2):99–120.
- Luca, Michael and Georgios Zervas. 2015. “Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud.” *Working Paper* .
- Marlow, Jennifer and Laura Dabbish. 2013. *Activity traces and signals in software developer recruitment and hiring*. 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, Qian, Bin Gu, and Andrew B Whinston. 2012. “Content Contribution for Revenue Sharing and Reputation in Social Media: A Dynamic Structural Model.” *Journal of Management Information* 29 (2):41–76.
- Varian, H.R. 2012. “Public goods and private gifts.” *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).

von Krogh, Georg and Eric von Hippel. 2006. “The Promise of Research on Open Source Software.” *Management Science* 52 (7):975–983.

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.