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*

Abstract

Many online platforms 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 to 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 improve a user's reputation. After finding a new job, users contribute 23.7% less in reputation-generating activity. By contrast, they reduce their non-reputation-generating activity by only 7.4% 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 prospects. We provide direct evidence against alternative explanations such as integer constraints, skills mismatch, and dynamic selection effects.

JEL Classification Numbers: H41, D82, D83, J24, J22, M51, L86

1. Introduction

One fascinating and economically important consequence of the rise of the Internet is the growing prevalence of private contributions to collective projects such as Wikipedia, bulletin boards, or open source software. As Lerner and Tirole (2002) put it, to an economist the behavior of individual contributors appears somewhat puzzling: is it a case of altruism, or are there ulterior motives behind private contributions to a public good?

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. [1]

Affiliated with 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-world individuals. Thus we construct complete histories of each individual's online trajectory. This includes their contributions to SO as

[1] In the classic theory of career concerns in Holmström (1982/99), the performance of the current job serves as a signal of one's ability to future employers. Thus a job seeker makes efforts to improve the current performance, in order to signal a higher ability, thus earning a higher salary from the new job. In this paper, "career incentive" and "career concern" will be used interchangeably to denote any career-related incentives, such as higher salary, more job offers, etc. In Section 5, we provide in-depth discussion on the information being signaled through online activity.

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 provides more and better Answers. However, right after the job change, there is a significant drop in Answers both in terms of quality and quantity.

However, the causal link inherent to the career-concerns hypothesis cannot be established based on this piece of evidence alone. A job seeker's behavior can be explained by multiple confounding factors. Most important, users with new jobs might have a busier work schedule, which prevents them from contributing to SO. Accordingly, we adopt a modified version of the difference-in-differences (DD) approach. Instead of comparing the behavior of different individuals, we focus on the same group of individuals and compare their behavior across different types of activities before and after a job change.

In order to show the rationale behind this novel approach, we first build a theoretical model of user contributions. We assume that agents derive utility from different activities and are subject to a time constraint. Specifically, there are three different activities: online contributions that improve an agent's reputation; online contributions that have no effect on an agent's reputation; and work (revenue generating) activities. Finally, we assume that the probability of finding a new job (a better revenue generating activity) is increasing in an agent's reputation.

The model's equilibrium implies that, upon obtaining a new job, the relative time spent on reputation-increasing online activity (relative to no-reputation increasing online activity) decreases. This theoretical result forms the basis of our empirical identification strategy.

In particular, the DD approach compares 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 23.7% right after a job change, of which 12.4–16.3% are due to (the removal of) career concerns. Apart from examining both short and long-term activity changes over time, we also consider the heterogeneous responses to job changes for users with different characteristics, such as education levels, types of degree, work experience, and existing online reputation. All of the results are consistent with the career-concerns hypothesis. As with any other DD specification, the validity of our identification hinges on the parallel-trend assumption. We address several major alternative explanations that can potentially invalidate this assumption, including integer constraints, skills mismatch, and dynamic selection effects.[2]

Our results contribute to understanding the motivations behind online voluntary contributions by private individuals. We demonstrate clear evidence of a widely-held hypothesis: career concerns matter. 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 public goods 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 use of online activity for employer hiring decisions.

Our results also provide important policy implications for platform companies. 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 thorough understanding of the motivations behind user participation is therefore crucial for the success of a platform, especially a platform that relies on voluntary contributions of user-generated content.[3] Our results imply that career concerns can be a way through which platforms encourage active user engagement, by aligning the private interest of contributors to that of the platform.

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. At a conceptual level, Lerner and

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

[3]This type of contribution is also called *crowdsourcing*.

Tirole (2001, 2002, 2005), Blatter and Niedermayer (2008) and Mehra, Dewan, and Freimer (2011) show how contributors to OSS projects improve their career prospects. 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). 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 financial returns from participation in OSS projects. von Krogh and von Hippel (2006); von Krogh et al. (2012) present an excellent survey of this literature.

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 how social effects (measured as group size) affect activities on Wikipedia using a natural experiment. Algan, Benkler, and Morell (2013) subsequently confirms that the long-term contribution on Wikipedia is more sustained by social effects, rather than altruism, through experiments. 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 Holmström (1982/99) and Gibbons and Murphy (1992). However, empirical work that estimates the causal effect of career incentives on people's behavior 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 effects of career concerns using various approaches and discusses related results. Section 6 examines the main assumption needed to identify the career-concern effects. It also addresses various challenges to this assumption and conducts robustness checks. Section 7 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 an 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 tasks. 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 t, they choose w_t, a_t, e_t so as to maximize value $V_t(s)$, where $s = 0, 1$. The value functions are determined recursively as follows:

$$V_t(s) = \max_{w_t, a_t, e_t} g_s(w_t) + f(a_t, e_t) + \delta\, E\, V_{t+1}(s')$$
$$\text{subject to: } w_t + e_t + a_t = T$$

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

(1) $$a_t|_{s=1} < a_t|_{s=0}$$

Moreover,

(2) $$\left.\frac{a_t}{e_t}\right|_{s=1} < \left.\frac{a_t}{e_t}\right|_{s=0} \quad \textit{iff} \quad p'(\cdot) > 0$$

Proof: See Appendix.

Proposition 1 establishes two effects of a job change: a decline in the time spent on providing Answers; and a decline in the relative time spent on Answers vis-à-vis Edits. The first effect (decline of Answers activity) 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 due to diminished career incentives. 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 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 disproportionally to changes in time availability. In

[4] In addition to Answers, Questions is 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.

Section 6, several tests are provided to test the validity of this assumption.

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.[5]

[FIGURE 1 about here.]

[FIGURE 2 about here.]

[FIGURE 3 about here.]

[FIGURE 4 about here.]

[FIGURE 5 about here.]

SOC is a related job matching website that hosts programming-related job listings as well as resumes of job candidates. 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.[6] 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 their 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.[7]

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.[8] 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 a Question or Answer: Edits. Edits help make the questions and Answers more readable and understandable to future viewers.[9]

[5] Founded 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.

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

[7] As 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.

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

[9] Most 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.

Votes Finally, registered users can give up-votes or down-votes to Questions and Answers *but not to Edits*. Votes reward reputation points of the owner of a post: each up-vote on a Question gives the asker five points, whereas each up-vote on an Answer worths 10 points.[10]

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.[11]
- 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. [12]
- 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).[13]
- Multiple job switches: Some users experienced more than one switch. More stable employment is associated to a sharper change in career incentives than temporary or transitional employment. So we exclude such switches if they are less than 8 months apart.[14]
- Profiles with links to SO: We limit our dataset to users with links to their SO profiles, because the ability to track users' online activities requires this link.

Applying this series of criteria results in a sample of 1301 users with 1520 job switches.[15] 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 activities on SO.

[TABLE 1 about here.]

Table 1 provides some descriptive statistics of SO activities from the sample of 1301 users. A typical SO user is not active in writing Questions or Answers. The activity distributions are fairly right-skewed, suggesting that a few users are disproportionally 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 since 2008).

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 to incentivize user activity, then we expect a drop in such activity once the goal (a better job) is attained. Since no one expects to remain

[10] Older 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 an Answer as given.

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

[12] Results from changes in employment status, i.e. from unemployed to employed, can also be interesting. However, from the CV data, we are unable to distinguish unemployment from other activities such as vacations.

[13] We also test the robustness of the result by altering the time periods in this selection.

[14] Other criteria are also tested.

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

in the same job for the rest of their lives, career concerns might not entirely disappear; but at least they are diminished following a change in jobs.

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

To account for these effects, we 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.

[FIGURE 6 about here.]

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; 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 about here.]

Figure 7 provides preliminary evidence regarding our hypothesis. It plots the monthly average of the logarithm of user activities 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 1,301 users who were subject to 1,520 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. Empirical Specification

As illustrated in Figure 6 and 7, our identification strategy is based on a standard difference-in-differences approach. However, instead of comparing the behavior of different individuals, we focus on the same set of individuals and compare their behavior across different activities before and after a job change:

$$y_{ijt} = \alpha_{ij} + \beta S_{it} + \gamma J_j S_{it} + \lambda_{jt} + D_{ijt}\theta + \varepsilon_{ijt} \quad (3)$$

In regression 3, the dependent variable y_{ijt} includes two different types of activities, including one votes-generating activity (VGA) and one non-VGA, which are indicated by subscript j. VGAs can be Answers ($j = a$), Votes gained from Answers ($j = v$), Questions ($j = q$); non-VGA is Edits ($j = e$). All the activities are measured in logarithms. One advantage of this approach is that the coefficients can be readily interpreted as percent variations. S_{it} is the state dummy variable: $S_{it} = 0$ corresponds to the periods before a job change takes place for user i, whereas $S_{it} = 1$ corresponds to the periods after a job change takes place. J_j is a dummy variable that takes on value 1 if the activity is a VGA ($j = a, v, q$) and 0 otherwise ($j = e$).

α_{ij} are individual fixed effects for each type of activity, which can control for many individual characteristics that could influence online contribution levels, such as ability, personal preference, gender, age, etc. The fixed effects are added at both individual and activity level due to contributors' preference of one task over another. For example, some contributors ask many Questions but rarely give any Answers.

β measures changes in Edits activity before and after a job switch. The main parameter of interest is the DD coefficient γ. γ measures the additional change in a VGA (Answers or Questions) over the changes to a non-VGA (Edits) after a job change.

The two parts of Proposition 1 can be expressed by the regressions coefficients β and γ. Specifically, we expect the level of SO activity to drop subsequently to a job shift, that is, we expect β and $\beta + \gamma$ to be negative. Moreover, we expect the drop in Answers to be greater than that of Edits, so that $\gamma < 0$, in addition to $\beta < 0$.

Seasonality and Duration Effects. In order to obtain a more accurate estimate of the effect of career concerns, we include additional activity data from a large sample of 96k active SO users, which can control for variations due to seasonality and duration effects.[16]

Online contribution might be more active in certain months than others; job changes can also occur more often in certain months of the year. We include additional year and month dummies to control for such potential effects, denoted by λ_{jt} in regression 3. Duration effects include the initial excitement of discovering SO, which can change over time and have heterogeneous effects on Answers and Edits activity. We measure duration as the count of the number of months since the first activity on SO for each user and include dummies for all distinct values of duration, denoted by X_{ijt} in regression 3. A separate set of seasonality and duration dummies is added for each type of activity, in order to control for the heterogeneous effects of seasonality and duration on different activities.

5.2. Main Effects of Career Concerns

[TABLE 2 about here.]

Table 2 presents our core results. The results are organized into two panels, using the number of Answers and Votes gained from Answers as measures of Answers activity. For each panel, the first regressions (columns 1 and 3) shows our base results without controlling for seasonality and duration effects. We thus have 18,192 observations (1,516 job switches (from 1,301 contributors) times 6 months: three prior to the job switch, three subsequent to the job switch, and times 2 activities: Answers and Edits). The second regressions in each panel (columns 2 and 4) shows the results while controlling for seasonality and duration effects, using activities from a large sample of SO users.

[16] Although we do not have their CV information and job status, we do observe their online activity over time. The additional data 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.

Column 1 shows that after switching to a new job, Edits activity experiences a significant drop of 7.38%. Moreover, the DD coefficient shows an additional drop of 16.27% in Answers activity, which we attribute to career concerns. The total changes in Answers activity can be calculated by $-7.38\% - 16.27\% = -23.65\%$. The results confirm the predictions from Proposition 1 that both coefficients are negative. Column 2 adds a set of dummies that control for seasonality and duration effects. This reduces our estimate of the treatment effect only slightly, to a statistically significant 12.36% decline.

Columns 3 and 4 report the same set of estimates using Votes instead of Answers to measure the vote-generating activity. Votes is a measure that includes both quantity and quality of Answers, and it can be a better measure of the amount of effort in contribution activities. The results using Votes give similar but slightly larger estimates than results using Answers. In Section 5.4, we investigate in depth the impact of increased effort on quantity and quality of Answers.

5.3. Month-to-Month Comparison

Table 2 summarizes DD estimates by comparing the differential changes of Answers and Edits activity in the 3-month period before and after a job change. We also explore the effects of career concerns over a longer period of time. Using period -2 as the baseline period, we compare the activity of all other periods to period -2.[17] We also control for seasonality and duration effects using the same 96k SO users mentioned before. We do so by estimating two following specifications:

$$y_{it} = \alpha_i + \sum_{\tau=-20}^{20} \beta_\tau \mathbb{1}(P_{it} = \tau) + \lambda_t + X_{it}\,\theta + \varepsilon_{it} \tag{4}$$

$$y_{ijt} = \alpha_{ij} + \sum_{\tau=-20}^{20} \big(\beta_\tau \mathbb{1}(P_{it} = \tau) + \gamma_\tau J_j \mathbb{1}(P_{it} = \tau)\big) + \lambda_{jt} + X_{ijt}\,\theta + \varepsilon_{ijt} \tag{5}$$

Regression 4 measures how each activity vary over time relative to baseline period -2, which is denoted by β_τ. Regression 5 estimates the differential changes between a VGA (i.e. Answers) and non-VGA (i.e. Edits) between the baseline period -2 and all other periods, and the DD coefficient is denoted by γ_τ.

λ_{jt} and X_{ijt} 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.

[FIGURE 8 about here.]

[TABLE 3 about here.]

[TABLE 4 about here.]

The estimates of β_τ from regression 4 and γ_τ from regression 5 are summarized in Table 3 and 4. They are also plotted in Figure 8.

Figure 8A plots the demeaned values of the estimates of β_τ for Answers and Edits activity. It is essentially Figure 7 but with seasonality and duration effects removed. Answers and Edits activity remain relatively stable from 20 to 5 months before the event of a job change. During the 5-month period before changing to a new job, both Answers and Edits activity experience a rapid increase, with Answers growing more than Edits. Then there is a rapid drop in both Answers and Edits activity starting from one month

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

before the new job, with Answers dropping significantly more than Edits and both keep decreasing over time.

Figure 8B shows the differential changes in Answers and Edits activity over time by plotting the DD estimates γ_τ for $\tau \in [-20, 20]$, as well as the 95% confidence interval. Before switching to a new job, all the DD estimates are negative but not significantly different from zero. Following the job change event, all the DD estimates are significantly negative, i.e. Answers drops significantly more than Edits.

One phenomenon illustrated in both figures is that the 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 4 and 5 underestimate the effect of career concerns. Second, job seekers may form the habit of contributing to SO as they improve their online reputation, which can have long term effects on contribution activities. Without taking into account habit formation, our DD estimates provides a lower bound of the true effect of career concerns.

5.4. 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 who are GitHub users and ask how GitHub activities can help with the job search process.[18] Employers consider merely having a GitHub profile as a good signal. They also evaluate the activities of job applicants on GitHub, specifically the popularities of projects, coding styles, etc. Job seekers, in turn, make efforts to show their passion and expertise to employers.

[FIGURE 9 about here.]

"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 9A 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.

To test this hypothesis, we pick the best Answer (measured by Votes) given by a contributor for each month, and Figure 9B 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

[18]GitHub 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.

random distribution of matching quality, with which the largest order statistic (Max Votes) goes up with a larger sample (number of Answers).

5.5. Questions Activity to Build Reputation

As mentioned earlier, SO users can also receive reputation points through asking Questions. We investigate how Questions activity changes surrounding the event of a job change. In order to plot the average logged activity while controlling for seasonality and duration effects, we plot the demeaned value of β_τ from regression 4 in Figure 10.

[FIGURE 10 about here.]

Figure 10 shows the rate of Questions asked around the time of a job shift. Unlike Answers and Edits, we observe little changes 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.

5.6. Career Concerns By Reputation Levels

The hypothesis of career concerns states that job seekers improve online reputation on SO through reputation-generating activities, which is used by employers to screen job candidates. Since reputation points on SO are cumulative, the signals are carried through both existing and new Answers activities. When pursuing new employment opportunities, job seekers with different levels of reputation might have heterogeneous responses to career incentives. For a job seeker who already enjoys an outstanding reputation on SO, the marginal benefit of extra effort to improve that signal should be relatively small. By contrast, a job seeker with lower SO reputation has very strong incentive to improve the signal; however, she might prefer not to reveal her SO profile to employers.

To examine the heterogeneous responses to career incentives based on different reputation levels, we associate each job switch with the reputation points at the time of the switch, and conduct separate analysis by splitting the sample into four equal groups of job switches.

[TABLE 5 about here.]

Table 5 summarizes the DD estimates from the four groups of job switches based on reputation levels. Job seekers in the second and third quartiles (columns 2 and 3) respond most to career incentives at 22.8% and 24.2%. Those with highest reputations (column 4) show a smaller effect at 13.2%. The most striking result comes from job seekers with lowest SO reputations. The estimate has an opposite sign compared to results from other groups. As mentioned before, one possible explanation is that low reputation users probably choose not to reveal their identity on SO when applying for jobs.

5.7. Career Concerns by Education Levels

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 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), Four-Year or Community College (College), Masters, and Ph.D. degrees. Then we conduct separate analyses for each group.

[TABLE 6 about here.]

Table 6 summarizes the DD estimates by education levels. The number of contributors in high school and Ph.D. groups are very small, and it is possible that the results might suffer from serious selection bias. Those with a high school degree as their highest education respond to a job change the most (-20.5% in column 1); on the other hand, those with a Ph.D. degree shows the least response to a job switch (-8.8% in column 4). Panel B using Votes produces similar results. The magnitude of the DD estimates is roughly consistent to the hypothesis of career concerns. Moreover, a comparison of the DD estimates from Panel A and B implies that the Answers contributed by Ph.D. degree holders receive more votes than those with a high school degree.

6. Testing Identification Assumptions

Our identification relies fundamentally on the 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: Plottings of Online Activities

Figure 7 plots the average logged activity over time. It provides some evidence on the parallel changes of Answers and Edits over time. Figure 8A further plots the same activities, while removing the potential confounding effects from seasonality and duration. In the periods further away from the event of a job change, the level of career incentives should be relatively stable. Figure 8A shows that before periods -5 and after period 10, although there are still variations of activities, the relatively parallel changes in Answers and Edits activity exhibit strong support for the parallel trend assumption.

6.2. Evidence of Parallel Trends: Within-Job Activity

The parallel trends assumption implies that if there were no changes in career incentives, then variations in time availability should have similar effects on Answers and Edits activity. 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 differences between a_t and e_t to remain constant.

[FIGURE 11 about here.]

Figures 11 shows the values of Answers and Edits for months 5 to 42 after an agent's job shift. Consistent with our underlying assumption, the differences between the two are fairly constant.

6.3. Integer Constraints

Job changes cause not only changes in the level of career incentives, but also changes in time availability. The validity of the career-concerns hypothesis relies on the assumption that changes in time availability due to a busier work schedule affect Answers and Edits activity in a similar fashion. However, an alternative interpretation that can explain our previous result that a_t/e_t drops subsequent to a job change is that users are faced with an "Integer Constraint:" [19] Answers activity requires a bigger set-up cost than Edits; and when an agent switches jobs, thus becoming busier, there might be fewer time windows to allocate to Answers rather than Edits. In other words, Edits typically require less time and can thus be fitted into a busy schedule more easily.

The rich dataset of user activities on SO, together with the whole employment history for SOC users, allows us to test whether the integer-constraint problem exists in our study, and if yes, then to what extent, i.e. whether it can reject the hypothesis of career concerns.

1. Weekdays vs. Weekend Activities. A new job with a busier work schedule should affect time availability mostly on 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 7 about here.]

Table 7 shows the results of the DD regressions split into weekdays and weekends. Panel A and B uses Answers and Votes as measures of reputation-generating activities, respectively. If there were no integer constraints and all users contribute to Answers and Edits activity both on weekdays and weekends, then there would be no differences between the DD estimates from weekday and weekend activities.[20]

Without controlling for seasonality and duration effects, Answers experience an additional 16.3% drop during weekdays (column 1) and 12.6% during weekends (column 3), relative to changes in Edits. Both estimates drop to 12.2% and 10.3% when controls are added (columns 2 and 4). Broadly speaking, the coefficient estimates are similar to those in the base model, and the difference between the estimates using weekday and weekend activities is relatively small (16.3% - 12.6% = 3.7% without controls; 12.2% - 10.3% = 1.9% with controls). Regression results using Votes (Panel B) give very similar results.

The difference between the DD estimates from weekday and weekend activities implies that although the integer-constraints problem might exist for certain users, but it cannot explain the entire additional drops in Answers relative to Edits after a job change, i.e. it fails to reject the career-concerns hypothesis.

[TABLE 8 about here.]

2. Internal Promotion. Internal promotion is an important case in two ways: First, a promotion often assumes more managerial duties that lead to more significant changes in work schedule than lateral moves. Thus it is most likely to satisfy the integer-constraint hypothesis. Second, the hypothesis of career concerns says that a job seeker signals to potential employers through online activity due to employers' inability to

[19] Integer Constraint originally comes from dynamic programming when some of the choice variables are restricted to be integers. Thus the agent enjoys less freedom to allocate resources compared to the case when everything is divisible.

[20] The majority of contribution activities take place on weekdays rather than on weekends. The selection requirement of having at least one Answers and Edits activity leads to a smaller sample of users for the analysis of weekend activities.

infer the true quality of a job candidate based on limited information. That is very unlikely in the case of an internal promotion, since the past internal performance is transparent to the current employer. [21]

Our analysis includes both weekday and weekend activities. If integer-constraint problem exists for job seekers who received internal promotions, then for weekday activities, we expect to observe an additional reduction in Answers relative to Edits; however, an internal promotion should affect weekday work schedule only, leaving the job seeker similar levels of freedom to organize her schedule on the weekends, thus we expect not to observe a differential effect of an internal promotion on Answers and Edits activities on the weekend.

Table 8 summarizes DD estimates for both internal moves and internal promotions, using weekday and weekend activities separately. Panel A uses all contributors for comparison purposes, which has the same results as Panel A (column 2 and 4) of Table 7. Panel B focuses on internal moves, i.e. job changes within the same company. The estimates in Panel B are not hugely different from 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 moves are lateral moves to a different department, which does not necessarily require managerial tasks. Panel C focuses on internal promotions using a stricter measure based on job title information. [22] Column 5 shows that Answers experience an additional drop of 15.6% compared to Edits on the weekdays. Although the estimate is insignificant (due to a small sample size), it does support the hypothesis of integer constraints for internally promoted workers. Column 6 shows a negligible DD estimate using weekend activities, which further lends credibility to integer-constraints hypothesis that only weekday activities are affected by integer constraints.

To summarize, Panel C of Table 8 shows the likely presence of integer-constraint problem in our DD analysis. Though it cannot reject the career-concerns hypothesis (as argued in the part 1 of this subsection), it does provide certain explanations with regard to the different DD estimates from weekday and weekend activities in Table 7.

3. Switch to a Similar Job. The integer-constraint hypothesis argues that a new job might have a work schedule that has small blocks of free time enough for making Edits, but not large blocks of free time for giving Answers. Although without data on the actual work schedule, we can infer work schedule changes based on certain information on a CV. In particular, we can focus on jobs that have exactly the same job titles.[23] This type of job changes helps to minimize the influence of changes in the nature of the work performed. The new jobs with the same job title as the old ones are associated with similar responsibilities, thus similar time flexibility, so these job changes are least likely to be affected by integer-constraint problem. If the hypothesis of career concerns is true, then we expect to observe a significant DD estimate using this sample, both from weekday and weekend activities.

[TABLE 9 about here.]

Table 9 summarizes the DD estimates. Out of 1301 contributors, 155 (or 12.5%) of them changes to new jobs that have the exact same job titles as the old ones. Due to the small sample size, most of the results become statistically insignificant but the point estimate tells a story that is consistent career-concerns hypothesis.

[21] It is conceivable that a job seeker might pursue an outside opportunity in order to bargain with the current employer, in which case public signals become valuable for internal promotions. Unfortunately, we do not have the data on how a promotion comes to fruition. However, we do not think most promotions are done in such fashion. Otherwise, firms might establish internal policies that forbid employees from building high-quality public signals.

[22] We define an internal job change as a promotion if one of the following (case insensitive) keywords exist in the new job title but not in the old one: 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.

[23] Some of the most common job titles are: Software Engineer, Software Developer, and Web Developer.

Table 9 shows that, for this sample of job changers who are least likely to subject to the integer-constraint problem, the DD estimates are consistent to those with the full sample (18.1% without controls and 13.8% with controls in columns 2 and 3). At the same time, the changes of Edits activity after a job switch are almost negligible (row 1 of columns 2 and 3), indicating a similar time availability before and after the job change. The DD analysis using weekday activity gives a similar result (16.1% in column 4); however, the regression using weekend activity gave a large significant estimate of 24.4% (column 5), implying a large effect of career incentives for this group of contributors.

6.4. Skills Mismatch

Another alternative interpretation for the drop in Answers following a job shift is that the skills required in the new occupation are different from those in the previous job. For example, a C++ programmer may switch to a job that requires skills in Java; such SO user would then be spending more time learning Java instead of 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. To test whether our estimates are driven by skills mismatch, we can focus on users who switch to new jobs with similar sets of technologies based on the tags information. First, we define a measure of skill-similarity between jobs.[24] Then we re-estimate the DD regressions separately based on the skill-similarity measures.

[TABLE 10 about here.]

Table 10 summarizes the DD estimates using different thresholds of tags 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 -16.27% from our baseline model. Moving to the right, column 2 to 5 gradually lower the thresholds of job similarity and include more job changers in the regression. Column 5 includes all the users, which is identical to our baseline model.

The results in Table 10 show that job changers who switch to positions with similar skill requirements also experience a similar drop in Answers over Edit activity. The magnitudes 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 hypothesis of skills mismatch.

6.5. Dynamic Selection Effects (Ashenfelter's Dip)

Another competing hypothesis that can explain the variations in Figure 7 is called dynamic selection effects, which says that the sample of job switchers are selected due to a special event prior to the job change that only affects the treatment group but not the control group. This hypothesis is commonly referred to in the labor economics literature as Ashenfelter's Dip (AD).[25]

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 changes in user behavior in response to incentives.

[24] Let the set of tags associated with the new job be $S1$, and those with the old job be $S0$. We define JobSimilarity$\equiv \frac{Size(S0 \cap S1)}{(Size(S0)+Size(S1))/2}$.

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

The problem of AD can be shown through simulating job change events using random activities together with a certain likelihood function of job changes. Figure 12 plots the mean logged levels of simulated activities over time. It shows a clear increase in Answers activity before a job change, followed by a reduction afterwards, which resembles Figure 7. In this case, even though a DD specification can generate a significant estimate, it is clearly not due to career concerns. The increase in Answers activity before the job change is purely due to a selection of periods with large Answers shocks; the reduction after the job change is purely due to a recovery from the shock.

[FIGURE 12 about here.]

Undoubtedly, this alternative explanation poses a valid concern to our attempts to identify career concerns from the change in activities 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 in Ashenfelter (1978), the dip is assumed to be due to random shocks. Therefore, the problem can be solved by using periods further away from the time of treatment or by matching treatment group with a properly selected control group who also experience a similar shock. However, in our analysis, the "bump" (or the "reversed" dip) in Figure 7, is central to the career-concerns hypothesis. 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 significant? 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 activities following a certain distribution and then simulate job change status given a likelihood function of job changes. The simulation is then repeated R times and the DD estimates are calculated and plotted. The comparison between simulated estimates and the actual DD estimate can help us to examine whether career-concerns hypothesis can be rejected in favor of AD hypothesis.

The simulation requires two main inputs: 1. Random draws of Answers and Edits activities 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 activities from the actual monthly activities. In order to keep the data as clean as possible —i.e., without potential effects of career incentives — we use the activities 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 these 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 which are fitted to the 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 the lagged Answers activity.[26] 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

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

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.

[FIGURE 13 about here.]

The simulation results are plotted in Figure 13. Panel A plots simulated DD estimates using Answers and Edits drawn directly from the actual activity, both in pairs (blue line) and separately (green line). Both simulations have a mean slightly larger than zero at 2-3%, but neither is significantly different from zero. The red line plots the distribution of the actual DD estimate from column 1 of Table 2, with a mean of 0.1627 and a standard deviation of 0.033. In Panel B, instead of drawing random activity directly from the actual activity, we first fit two negative binomial distributions for Answers and Edits activities using MLE. Then we conduct the same set of simulations 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 .1627 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 SO.[27] 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 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 in our conclusion.

Parameter Values Required to Reject Career Concerns. Figure 13 shows a significant gap between the simulated results and the actual DD estimates, which favors the career-concerns hypothesis. However, the simulated results crucially depend 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:

$$Pr(JC) = Logit(\alpha + \beta * A) = \frac{exp(\alpha + \beta * A)}{1 + exp(\alpha + \beta * A)} \tag{6}$$

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 values in Figure 14.

[FIGURE 14 about here.]

[27]The actual correlation between Answers and Edits is 0.564.

Figure 14 shows that the value of $\beta = 22.76\%$ produces a simulated DD estimate that is closest to our actual DD estimate of 16.27%. $\beta = 22.76\%$ means that one log unit increase of number of Answers increases the likelihood of a job change by 22.76%. For our sample of 1,031 SOC users, an average use contributes 4.05 Answers per month. Thus $\beta = 22.76\%$ means that if a job seeker contributes 6.88 Answers, then the chance of changing to a new job increases by 22.76%.[28] Given the relatively low cost of providing Answers on SO, the benefit of the additional online activities is enormous, which is not the case in reality simply from observation. Therefore, we conclude that the magnitude of the causal effect, which is represented by β, is over-estimated. That is to say, in order to reject the career-concerns hypothesis, the parameter value needed in the job change function is too large to be reasonable.

6.6. External Validity

One challenge to the external validity of our result comes from the representativeness of our sample. The main concern is that the data selection process makes our sample not representative of the whole population of users on SO, so the result that the contributors in our analysis responding to career incentives can not be generalized to other contributors not in our sample.

This is definitely a valid concern that since the majority of SO users do not contribute anything, our sample is mostly drawn from the right-end of the distribution. The two panels of figure 5 contrasts the average monthly contributions between all SO users and our sample. The real question is that with a unrepresentative sample, what can be said with regard to the implications?

We believe that our results still provide valuable information to platform managers. First, on SO, 10% of the users make 90% of all contribution activities. These active contributors are the core users of SO that it cares most. Our result unravels one of the motivations that drives user activity. Second, we believe that other users not in our sample can also be motivated by career incentives. Information asymmetry is always a major problem in any job search activity. Employers often hunger for more information of the job applicants. On the other hand, job seekers try to signal their unobserved ability through numerous channels. Any information that reflects one's ability can be used to reduce the information asymmetry. The reputation system on SO proves to be such a channel. Given the fact that many employers value this information and job seeking activity is common to everyone, we believe that our results can also be extended to those less-active SO users. Admittedly, the way and magnitude that less-active users respond to career incentives can be very different. Our analysis focused on the intensive rather than the extensive margin. Career incentives can potentially help to improve the extensive margins among the inactive contributors.

6.7. Robustness and Sensitivity Checks

In the regression analysis, the data in the dependent variable is transformed by logarithm of one plus the activity count, and then the regression specification is estimated using OLS with fixed effects. However, one might also argue that count data is better analyzed using Poisson regression or Negative Binomial regression. We conducted the same set of specifications with Poisson and Negative Binomial regressions, and the estimates from these two methods are very similar to those from OLS with fixed effects. Our main analysis in Section 5.2 compares three-month before and after a job change. In particular, periods $-3, -2, -1$ and $+2, +3, +4$ are chosen for the analysis in order to avoid the noise stemming from the process of job change. We also tried other choices and they all give similar estimates.

[28] One log unit increase is roughly an increase of 170%. So one log unit increase for a typical user is approximately $4.05 \times 170\% = 6.88$.

7. Discussion and Concluding Remarks

The Internet has created many opportunities for online collaboration and networking. Some platforms have been enormously successful, some less so. Examples of the former include Wikipedia, YouTube, Stack Overflow and Amazon Mechanical Turk; examples of the latter include Yahoo! Answers and Digg. What distinguishes a winner from a loser platform? What makes a platform attract user contributions? 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 concerns provide a strong incentive for users to contribute, namely to answer questions posted on the various SO boards. Our strategy for identifying career-concern-based incentives is to estimate the effect of a job change. Our regressions estimates suggest that achieving the goal of switching to 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 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.

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 of active contributors, 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 around job change periods. Future studies should explore whether our results can be transferred to a broader set of industries, and also investigate the potential long-term crowd-out effect implied by extrinsic motivation. Future research can also focus on practical ways to integrate career incentives into platforms. Stack Overflow implemented this idea through building a careers website, Stack Overflow Careers, and it constantly reminds users of the career benefits of contribution by means of website banners. Some other platforms such as Kaggle combines both immediate monetary compensation from firms who want their problems solved and future employment benefits through a reputation system with an employment website. However, for many other platforms such as Wikipedia, the most effective way to integrate career incentives is still unclear. Future research can shed some light on the implementability and effectiveness of various ways to achieve this goal.

Appendix

Proof of Proposition 1. The value function for the two states ($s = 0, 1$) can be written as the following:

$$V_t(0) = \max_{w_t, a_t, e_t} g_0(w_t) + f(a_t, e_t) + \delta\, p(r_t)\, V_{t+1}(1) + \delta\, (1 - p(r_t))\, V_{t+1}(0)$$
$$V_t(1) = \max_{w_t, a_t, e_t} g_1(w_t) + f(a_t, e_t) + \delta\, V_{t+1}(1)$$

This model has infinite time periods and it represents one job switching process (i.e. $s = 0$ to $s = 1$). So the value functions above can be simplified by removing the time subscript and focusing on the two states as following:

$$V_0 = \max_{w_0, a_0, e_0} g_0(w_0) + f(a_0, e_0) + \delta\, p(r_t)\, V_1 + \delta\, (1 - p(r_t))\, V_0 \tag{7}$$
$$V_1 = \max_{w_1, a_1, e_1} g_1(w_1) + f(a_1, e_1) + \delta\, V_1 \tag{8}$$

First, we can show that $V_1 > V_0$. 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 the value functions 7 and 8 can be written as:

$$V_0 \leq \frac{1}{1-\delta}[g_0(w_0^*) + f(a_0^*, e_0^*)]$$
$$V_1 = \frac{1}{1-\delta}[g_1(w_1^*) + f(a_1^*, e_1^*)]$$

When $s = 1$, the optimal time allocation x_0^* is still feasible. So by choosing $x_1 = x_0^*$ when $s = 1$ results a strictly higher value of V_1, which contradicts with $V_1 \leq V_0$. Thus it must be

$$V_1 > V_0 \tag{9}$$

The homotheticity of $f(a, e)$ means that it can be transformed as $e\, f(k, 1)$ where $k = \frac{a}{e}$. The second argument in $f(k, 1)$ will be kept as a constant one, which means that we can simplify the notation even further as $e\, f(k)$. The value functions become:

$$V_0 = \max_{w_0, a_0, e_0} g_0(w_0) + e_0\, f(k_0) + \delta\, p(r_t)\, V_1 + \delta\, (1 - p(r_t))\, V_0 \tag{10}$$
$$V_1 = \max_{w_1, a_1, e_1} g_1(w_1) + e_1\, f(k_1) + \delta\, V_1 \tag{11}$$

At state s, the agent maximizes V_s subject to $w + e + a = T$. So the first-order conditions of the Lagrangian with respect to w, a, and e at $s = 1$ are given by:

$$\lambda_1 = g_1'(w_1) \tag{12}$$
$$\lambda_1 = f'(k_1) \tag{13}$$
$$\lambda_1 = f(k_1) - k_1\, f'(k_1) \tag{14}$$

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

$$\lambda_0 = g_0'(w_0) \tag{15}$$

$$\lambda_0 = f'(k_0) + \delta\,(V_1 - V_0)\,p'(r_t) \tag{16}$$

$$\lambda_0 = f(k_0) - k_0\,f'(k_0) \tag{17}$$

Equations 13 and 14 imply that $(1+k_1)\,f'(k_1) - f(k_1) = 0$, and equations 16 and 17 imply that $(1+k_0)\,f'(k_0) - f(k_0) = -\delta\,(V_1 - V_0)\,p'(r_t)$, which is less than 0 if and only if $p'(r_t) < 0$. Then based on these two results, we have the following inequality:

$$(1+k_1)\,f'(k_1) - f(k_1) > (1+k_0)\,f'(k_0) - f(k_0)$$

Since $f' > 0$ and $f'' < 0$, it's straight-forward to show that $h(x) = (1+x)\,f'(x) - f(x)$ decreases in x. Thus we have

$$k_1 < k_0 \quad \text{or} \quad \frac{a_1}{e_1} < \frac{a_0}{e_0} \tag{18}$$

Regarding the first part of the Proposition, it helps to first compare equations 14 and 17. The convexity property of f function imply that $f(x) - x\,f'(x)$ increases in x. Together with the second part of the proposition $k_1 < k_0$, we get $\lambda_1 < \lambda_0$. With equations 12 and 15, we get the following inequality:

$$g_1'(w_1) < g_0'(w_1) \tag{19}$$

$g_1' > g_0'$ implies that $g_1'(w_1) > g_0'(w_1)$. Together with 19 we get $g_0'(w_1) < g_1'(w_1) < g_0'(w_0)$, which means that $w_1 > w_0$. Since $a_0 + e_0 = T - w_0$ and $a_1 + e_1 = T - w_1$ and $\frac{a_1}{e_1} < \frac{a_0}{e_0}$, we get $a_1 < a_0$. ■

References

Algan, Y, Y Benkler, and M F Morell. 2013. "Cooperation in a Peer Production Economy Experimental Evidence from Wikipedia." *Working Paper* .

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.

TABLE 1
DESCRIPTIVE STATISTICS

	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 various online activities of the 1301 contributors used in the final DD analysis.

TABLE 2

EFFECTS OF CAREER CONCERNS ON ANSWERS AND EDITS ACTIVITY

	Panel A: $y \in \{$Answers, Edits$\}$		Panel B: $y \in \{$Votes, Edits$\}$	
	(1)	(2)	(3)	(4)
$NewJob\,(S)$	-0.0738***	-0.0742***	-0.0738***	-0.0742***
	(0.019)	(0.019)	(0.019)	(0.019)
$NewJob\,(S) \times Answer/Vote\,(J)$	-0.1627***	-0.1236***	-0.1943***	-0.1536***
	(0.033)	(0.033)	(0.037)	(0.037)
Seasonality dummy		x		x
Duration dummy		x		x
No. of contributors	1301	97723	1301	97723
No. of observations	18192	9105862	18192	9105862
R^2	0.014	0.033	0.014	0.027

Notes: This table summarizes the DD estimates from regression 3. The DD coefficient measures the extent to which Answers activity changes relative to Edits activity in a three-month period before and after switchingt to a new job. The dependent variables include both Answers and Edits activity. Panel A and B use distinct measures of Answers activity: Panel A uses the number of Answers, and Panel B uses the number of Votes received from Answers. All measures of activities are transformed by logarithm of one plus the activity count. Independent Variables: S_{it} indicates whether the current state is a new or old job: $S_{it} = 0$ prior to job switch, $S_{it} = 1$ after job change. J_j indicates the different types of activities: $J_j = 1$ if $k = a, v$ (Answers/Votes), $J_j = 0$ if $k = e$(Edits).

First row of the table presents the estimates of β, which measures the changes in Edits activity after switching to a new job. The second row presents the estimates of DD coefficient γ. The first columns in each Panel estimates the regression without extra controls; the second columns adds seasonality (year and month) and duration (length of time since first activity on SO) dummies. To control these effects, we use the activity data of 96k SO users, which is shown in the "No. of contributors". The estimates from column 1 show that after switching to a new job, Edits activity drops by 7.38%, and Answers activity experiences an additional drop of 16.27%, which we attribute to the removal of career concerns. The magnitude of the estimates slightly drops after adding more controls. The estimates using Votes give similar results.

Number of contributors for DD analysis: 1301; Number of job switches: 1520; Number of contributors used to control for seasonality and duration effects: 96422. Robust standard errors in parentheses, clustered at the level of the individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 3
MONTH-TO-MONTH COMPARISON: FIRST DIFFERENCES

	(1) $y = \text{Answers}$		(2) $y = \text{Edits}$		(3) $y = \text{Votes}$	
Period -20	-0.222***	(0.03)	-0.142***	(0.02)	-0.229***	(0.03)
Period -19	-0.237***	(0.03)	-0.162***	(0.02)	-0.264***	(0.03)
Period -18	-0.234***	(0.03)	-0.149***	(0.02)	-0.269***	(0.03)
Period -17	-0.190***	(0.03)	-0.158***	(0.02)	-0.223***	(0.03)
Period -16	-0.204***	(0.03)	-0.134***	(0.02)	-0.223***	(0.03)
Period -15	-0.186***	(0.03)	-0.120***	(0.02)	-0.210***	(0.03)
Period -14	-0.170***	(0.03)	-0.105***	(0.02)	-0.196***	(0.03)
Period -13	-0.183***	(0.03)	-0.113***	(0.02)	-0.196***	(0.03)
Period -12	-0.220***	(0.03)	-0.120***	(0.02)	-0.241***	(0.03)
Period -11	-0.209***	(0.03)	-0.114***	(0.02)	-0.233***	(0.03)
Period -10	-0.196***	(0.03)	-0.140***	(0.02)	-0.220***	(0.03)
Period -9	-0.212***	(0.03)	-0.155***	(0.02)	-0.224***	(0.03)
Period -8	-0.203***	(0.03)	-0.139***	(0.02)	-0.257***	(0.03)
Period -7	-0.175***	(0.03)	-0.111***	(0.02)	-0.179***	(0.03)
Period -6	-0.182***	(0.03)	-0.099***	(0.02)	-0.184***	(0.03)
Period -5	-0.140***	(0.03)	-0.089***	(0.02)	-0.158***	(0.03)
Period -4	-0.105***	(0.03)	-0.091***	(0.02)	-0.111***	(0.03)
Period -3	-0.011	(0.03)	0.000	(0.02)	-0.023	(0.03)
Period -2 (baseline)	0	(-)	0	(-)	0	(-)
Period -1	-0.020	(0.03)	-0.001	(0.02)	-0.013	(0.03)
Period 0	-0.154***	(0.03)	-0.086***	(0.02)	-0.145***	(0.03)
Period 1	-0.308***	(0.03)	-0.181***	(0.02)	-0.354***	(0.03)
Period 2	-0.224***	(0.03)	-0.082***	(0.02)	-0.268***	(0.03)
Period 3	-0.198***	(0.03)	-0.070***	(0.02)	-0.227***	(0.03)
Period 4	-0.203***	(0.03)	-0.072***	(0.02)	-0.226***	(0.03)
Period 5	-0.265***	(0.03)	-0.145***	(0.02)	-0.288***	(0.03)
Period 6	-0.292***	(0.03)	-0.129***	(0.02)	-0.340***	(0.03)
Period 7	-0.294***	(0.03)	-0.126***	(0.02)	-0.326***	(0.03)
Period 8	-0.330***	(0.03)	-0.153***	(0.02)	-0.382***	(0.03)
Period 9	-0.358***	(0.03)	-0.158***	(0.02)	-0.399***	(0.03)
Period 10	-0.374***	(0.03)	-0.167***	(0.02)	-0.425***	(0.03)
Period 11	-0.404***	(0.03)	-0.197***	(0.02)	-0.441***	(0.03)
Period 12	-0.441***	(0.03)	-0.217***	(0.02)	-0.483***	(0.03)
Period 13	-0.461***	(0.03)	-0.207***	(0.02)	-0.523***	(0.03)
Period 14	-0.463***	(0.03)	-0.232***	(0.02)	-0.522***	(0.03)
Period 15	-0.488***	(0.03)	-0.230***	(0.02)	-0.533***	(0.03)
Period 16	-0.475***	(0.03)	-0.261***	(0.02)	-0.528***	(0.03)
Period 17	-0.483***	(0.03)	-0.233***	(0.02)	-0.537***	(0.03)
Period 18	-0.489***	(0.03)	-0.248***	(0.02)	-0.551***	(0.03)
Period 19	-0.479***	(0.03)	-0.246***	(0.02)	-0.532***	(0.03)
Period 20	-0.480***	(0.03)	-0.247***	(0.02)	-0.554***	(0.03)
Seasonality dummy	x		x		x	
Duration dummy	x		x		x	
No. of observations	4646575		4646575		4646575	
R^2	0.047		0.004		0.038	

Notes: This table summarizes the estimates of β_τ in regression 4 by using Answers, Edits, and Votes as the dependent variable. β_τ measures the differences in logged activities between period τ and -2, while controlling for seasonality and duration effects. Period 1 is the first month when a new job starts. Period -2 is used as the (omitted) baseline period since it has the highest activity level. The estimates show that both Answers and Edits activities rise up gradually until three months before a job change, and then both start to drop. The values of β_τ are plotted in Figure 8A. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 4
MONTH-TO-MONTH COMPARISON: DIFFERENCE-IN-DIFFERENCES

	(1)		(2)	
	$y \in$ {Answers, Edits}		$y \in$ {Votes, Edits}	
Period -20	-0.079	(0.05)	-0.087	(0.06)
Period -19	-0.075	(0.05)	-0.103*	(0.06)
Period -18	-0.084*	(0.05)	-0.120**	(0.06)
Period -17	-0.032	(0.05)	-0.065	(0.05)
Period -16	-0.070	(0.05)	-0.089*	(0.05)
Period -15	-0.066	(0.05)	-0.091*	(0.05)
Period -14	-0.065	(0.05)	-0.091*	(0.05)
Period -13	-0.070	(0.05)	-0.083	(0.05)
Period -12	-0.100**	(0.05)	-0.121**	(0.05)
Period -11	-0.095**	(0.05)	-0.119**	(0.05)
Period -10	-0.055	(0.04)	-0.080	(0.05)
Period -9	-0.057	(0.04)	-0.069	(0.05)
Period -8	-0.064	(0.04)	-0.118**	(0.05)
Period -7	-0.063	(0.04)	-0.068	(0.05)
Period -6	-0.083**	(0.04)	-0.085*	(0.04)
Period -5	-0.050	(0.04)	-0.069	(0.05)
Period -4	-0.015	(0.04)	-0.020	(0.04)
Period -3	-0.011	(0.03)	-0.023	(0.04)
Period -2 (baseline)	0	(-)	0	(-)
Period -1	-0.020	(0.03)	-0.012	(0.04)
Period 0	-0.068*	(0.04)	-0.059	(0.04)
Period 1	-0.127***	(0.04)	-0.173***	(0.05)
Period 2	-0.142***	(0.04)	-0.186***	(0.05)
Period 3	-0.128***	(0.04)	-0.157***	(0.05)
Period 4	-0.131***	(0.04)	-0.154***	(0.05)
Period 5	-0.120***	(0.04)	-0.143***	(0.05)
Period 6	-0.163***	(0.04)	-0.210***	(0.05)
Period 7	-0.169***	(0.04)	-0.201***	(0.05)
Period 8	-0.178***	(0.04)	-0.229***	(0.05)
Period 9	-0.200***	(0.04)	-0.241***	(0.05)
Period 10	-0.207***	(0.04)	-0.258***	(0.05)
Period 11	-0.208***	(0.04)	-0.245***	(0.05)
Period 12	-0.224***	(0.04)	-0.265***	(0.05)
Period 13	-0.254***	(0.04)	-0.316***	(0.05)
Period 14	-0.231***	(0.04)	-0.290***	(0.05)
Period 15	-0.258***	(0.04)	-0.303***	(0.05)
Period 16	-0.214***	(0.04)	-0.267***	(0.05)
Period 17	-0.250***	(0.04)	-0.303***	(0.05)
Period 18	-0.241***	(0.04)	-0.302***	(0.05)
Period 19	-0.233***	(0.04)	-0.286***	(0.05)
Period 20	-0.233***	(0.04)	-0.308***	(0.05)
Seasonality dummy	x		x	
Duration dummy	x		x	
No. of observations	9293150		9293150	
R^2	0.034		0.029	

Notes: This table lists the estimates of γ_τ in regression 5 by using Answers, Votes, Questions, together with Edits, as the dependent variables. Edits activity is used as the control group. γ_τ captures the differences in changes of vote-generating activities relative to changes in Edits between period τ and -2. The demeaned values of γ_τ are plotted in Figure 8B. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 5

EFFECTS OF CAREER CONCERNS BY REPUTATION LEVELS

	Panel A: $y \in \{$Answers, Edits$\}$				Panel B: $y \in \{$Votes, Edits$\}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NewJob\,(S)$	0.157***	-0.067*	-0.166***	-0.251***	0.157***	-0.067*	-0.166***	-0.251***
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
$NewJob\,(S) \times Answer/Vote\,(J)$	0.102	-0.228***	-0.242***	-0.132*	0.133*	-0.213***	-0.325***	-0.211***
	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Seasonality dummy	x	x	x	x	x	x	x	x
Duration dummy	x	x	x	x	x	x	x	x
Reputation	0-25%	25-50%	50-75%	75-100%	0-25%	25-50%	50-75%	75-100%
No. of contributors (DD)	356	350	340	311	356	350	340	311
No. of contributors (control)	96422	96422	96422	96422	96422	96422	96422	96422
N	9092012	9092002	9092012	9092000	9092012	9092002	9092012	9092000
R^2	0.033	0.033	0.033	0.033	0.027	0.027	0.027	0.027

Notes: This table illustrates the heterogeneous responses to career incentives by job seekers with different levels of reputation. It divides the job switches into four groups based on the reputation points at the time of a job change. Panel A and B presents results using number of Answers and number of Votes from Answers as Answers activity. All regressions include controls for seasonality and duration effects.

Job seekers with medium reputation levels respond to career concerns the most (-22.8% and -24.2% in columns 2 and 3). Those who already have an excellent reputation on SO respond less to career concerns at -13.2% (column 4). The estimate from job seekers with low SO reputations shows an insignificant positive value. One potential explanation is that users with low reputations on SO probably choose not to reveal their SO profile when applying for jobs.

Reputation Points: Min: 0; First Quartile: 770; Median: 2,124; Third Quartile: 5,265; Max: 132,067.

Robust standard errors in parentheses, clustered at the level of the individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 6

EFFECTS OF CAREER CONCERNS BY EDUCATION LEVELS

	Panel A: $y \in \{$Answers, Edits$\}$				Panel B: $y \in \{$Votes, Edits$\}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NewJob\,(S)$	0.102	-0.084***	-0.045	-0.036	0.102	-0.084***	-0.045	-0.036
	(0.16)	(0.03)	(0.05)	(0.10)	(0.16)	(0.03)	(0.05)	(0.10)
$NewJob\,(S) \times Answer/Vote\,(J)$	-0.205	-0.121***	-0.164**	-0.088	-0.143	-0.145***	-0.189**	-0.146
	(0.25)	(0.04)	(0.08)	(0.17)	(0.31)	(0.05)	(0.08)	(0.20)
Seasonality dummy	x	x	x	x	x	x	x	x
Duration dummy	x	x	x	x	x	x	x	x
Highest Education Level	HS	College	Masters	PhD	HS	College	Masters	PhD
No. of contributors (DD)	12	778	230	51	12	778	230	51
No. of contributors (control)	96422	96422	96422	96422	96422	96422	96422	96422
No. of observations	8944282	8955016	8947346	8944870	8944282	8955016	8947346	8944870
R^2	0.033	0.033	0.033	0.033	0.027	0.027	0.027	0.027

Notes: This table illustrates the heterogeneous response to career incentives for job seekers with different education background. From the CV, we extract and analyze the education section and conduct separate analysis based on the highest degree obtained: High School (HS), Four-Year or Community College (College), Masters, and Ph.D. degrees. Panel A and B presents results using number of Answers and number of Votes from Answers as Answers activity. All regressions include controls for seasonality and duration effects.

High School degree holders respond to a job change the most, at -20.5% (column 1). College and Masters degree experience similar level of changes at -12.1% (column 2) and -16.4% (column 3), respectively. Those who already have a Ph.D. shows an estimate with the smallest magnitude (column 4). The number of job seekers in the first and last group are quite small, so the results might suffer from serious selection bias. The magnitude of these estimates are consistent with the hypothesis of career concerns, which assumes signaling one's ability through various signals. Those only have a high school degree probably have to rely on other signals when applying for jobs; on the other hand, a Ph.D. degree is always a very strong signal, and one does not need to signal the unobserved ability through other signals.

Robust standard errors in parentheses, clustered at the level of the individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 7

EFFECTS OF JOB CHANGES ON WEEKDAY VS. WEEKEND ACTIVITIES

	Panel A: $y \in$ {Answers, Edits}				Panel B: $y \in$ {Votes, Edits}			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NewJob\,(S)$	-0.070***	-0.072***	-0.061**	-0.060**	-0.070***	-0.072***	-0.061**	-0.060**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$NewJob\,(S) \times Answer/Vote\,(J)$	-0.163***	-0.122***	-0.126***	-0.103**	-0.178***	-0.135***	-0.150***	-0.124***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Seasonality dummy		x		x		x		x
Duration dummy		x		x		x		x
Days	Weekday	Weekday	Weekend	Weekend	Weekday	Weekday	Weekend	Weekend
No. of contributors (DD)	1159	1159	374	374	1159	1159	374	374
No. of contributors (control)	-	96422	-	51296	-	96422	-	51296
No. of observations	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 activities separately. Panel A uses the number of Answers and Panel B uses the number of Votes from Answers as measures of Answers activity. A new job affects work schedule mostly during weekdays, not weekends. Therefore, a significant DD estimate using weekend activities shows that the estimate is not likely to be caused by integer constraints. The magnitude of estimates using weekend activities are slightly smaller than those with using weekday activities (-12.2% in column 2 vs. -10.3% in column 4).

Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 8

EFFECTS OF JOB CHANGES FOR INTERNAL PROMOTIONS

	Panel A: All		Panel B: Same Company		Panel C: Promotion	
	(1)	(2)	(3)	(4)	(5)	(6)
$NewJob\,(S)$	-0.072***	-0.060**	-0.070	-0.039	-0.151**	-0.157
	(0.02)	(0.02)	(0.05)	(0.08)	(0.06)	(0.10)
$NewJob\,(S) \times Answer\,(J)$	-0.122***	-0.103**	-0.090	-0.119	-0.156	-0.008
	(0.04)	(0.04)	(0.09)	(0.13)	(0.11)	(0.19)
Seasonality dummy	x	x	x	x	x	x
Duration dummy	x	x	x	x	x	x
Days	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
No. of contributors (DD)	1159	374	142	41	80	21
No. of contributors (control)	96422	51296	96422	51296	96422	51296
No. of observations	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 B: 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. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 9

EFFECTS OF JOB CHANGES FOR SWITCHING TO JOBS WITH SAME JOB TITLES

	Baseline	Same Job Title			
	(1)	(2)	(3)	(4)	(5)
$NewJob\,(S)$	-0.074***	-0.028	-0.028	-0.008	-0.028
	(0.02)	(0.06)	(0.06)	(0.07)	(0.08)
$NewJob\,(S) \times Answer\,(J)$	-0.124***	-0.181*	-0.138	-0.161	-0.244*
	(0.03)	(0.11)	(0.11)	(0.11)	(0.14)
Seasonality dummy	x		x	x	x
Duration dummy	x		x	x	x
Days	-	-	-	Weekday	Weekend
No. of contributors (DD)	1301	155	155	144	45
No. of contributors (control)	96422	-	96422	96422	51296
No. of observations	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 the same job titles. New jobs with the same job titles should have similar work schedule, thus they are not subject to any integer constraints effect. Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

TABLE 10

EFFECTS OF JOB CHANGES FOR NEW JOBS WITH SIMILAR TECHNOLOGY

	Panel A: $y \in \{$Answers, Edits$\}$			Panel B: $y \in \{$Votes, Edits$\}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$NewJob\,(S)$	-0.013	-0.074**	-0.082***	-0.013	-0.074**	-0.082***
	(0.06)	(0.04)	(0.02)	(0.06)	(0.04)	(0.02)
$NewJob\,(S) \times Answer/Vote\,(J)$	-0.173*	-0.122*	-0.125***	-0.231**	-0.150**	-0.154***
	(0.10)	(0.06)	(0.03)	(0.11)	(0.07)	(0.04)
Seasonality dummy	x	x	x	x	x	x
Duration dummy	x	x	x	x	x	x
Job similarity (by tags)	$\geq$100%	$\geq$50%	$\geq$0%	$\geq$100%	$\geq$50%	$\geq$0%
Contributors	96582	96823	97658	96582	96823	97658
No. of observations	9089694	9092762	9105016	9089694	9092762	9105016
R^2	0.033	0.033	0.033	0.027	0.027	0.027

Notes: This table summarizes the DD estimates based on the similarities of technologies used in the old and the new jobs. We introduce a measure of job similarity and conduct separate analysis based on that measure. A job similarity of 100% means that the new job has exactly the same set of tags as the old one; 0% means that the old and new jobs have no common tags. The goal is to test whether our DD estimate can be explained by the hypothesis of skills mismatch. The result shows that even for job seekers who switch to new jobs with exactly the same set of technology, the DD estimate is still significant at -17.3% (column 1).

Robust standard errors in parentheses, clustered at individual-activity type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

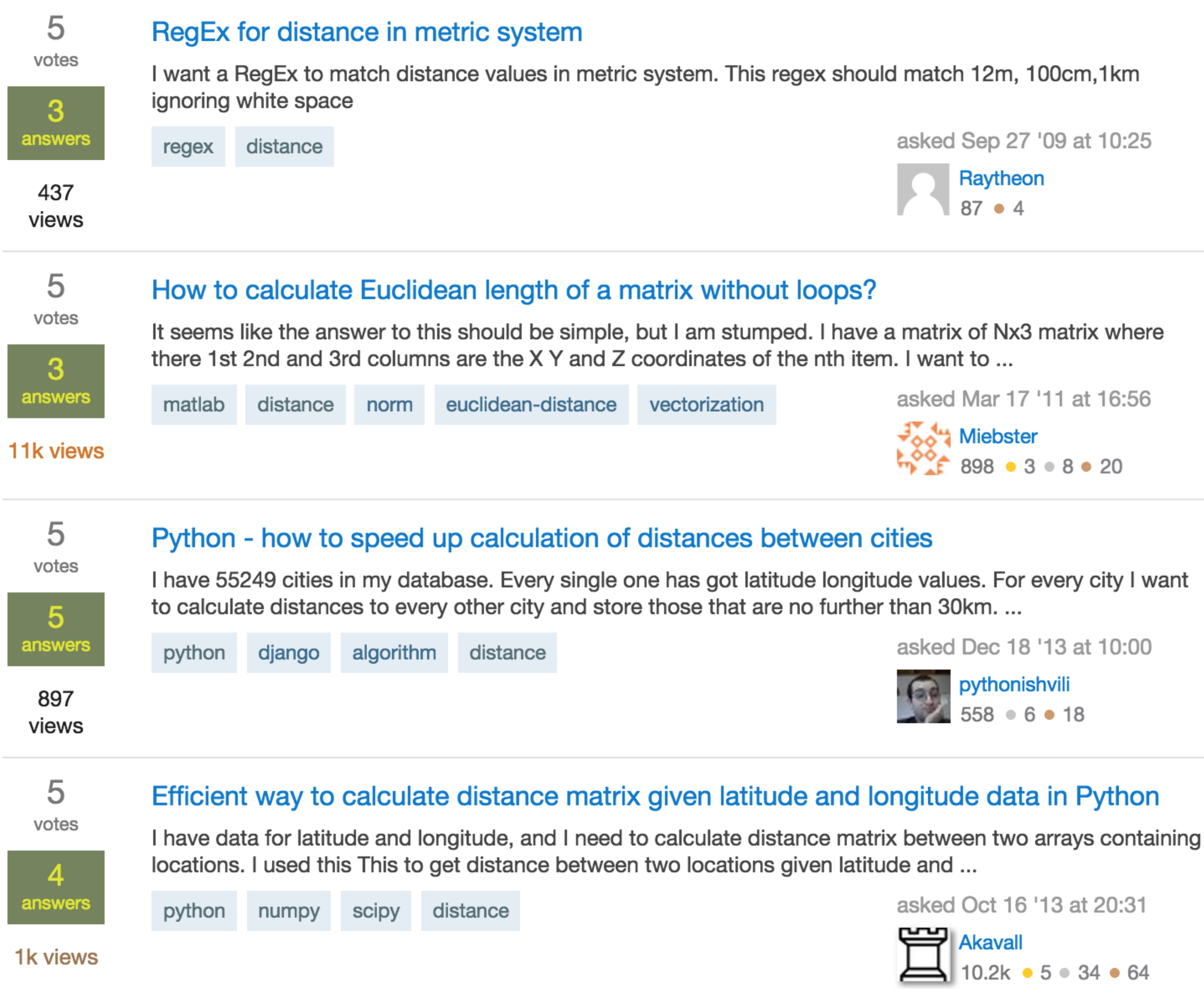


FIGURE 1
Sample List of Questions on Stack Overflow

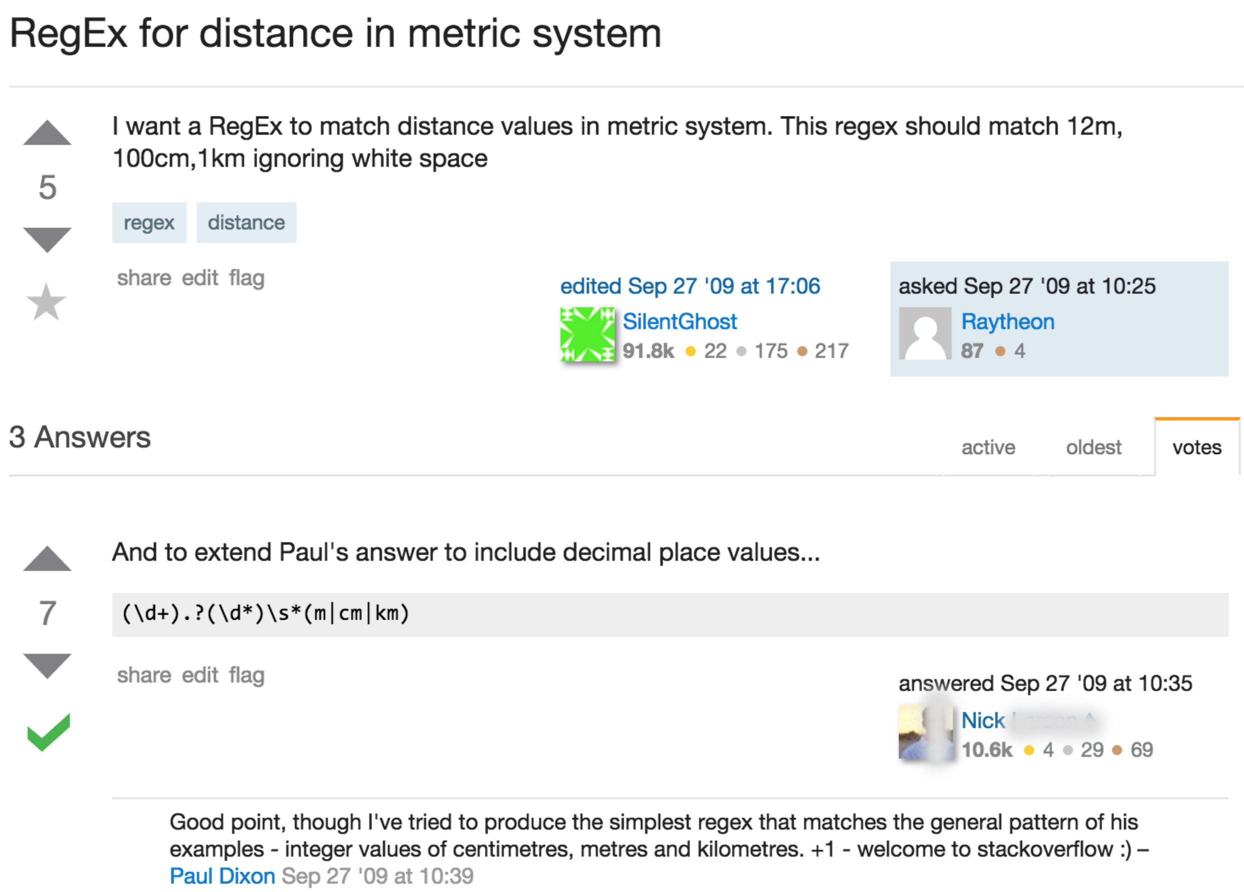


FIGURE 2

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 10 reputation points to the contributor.

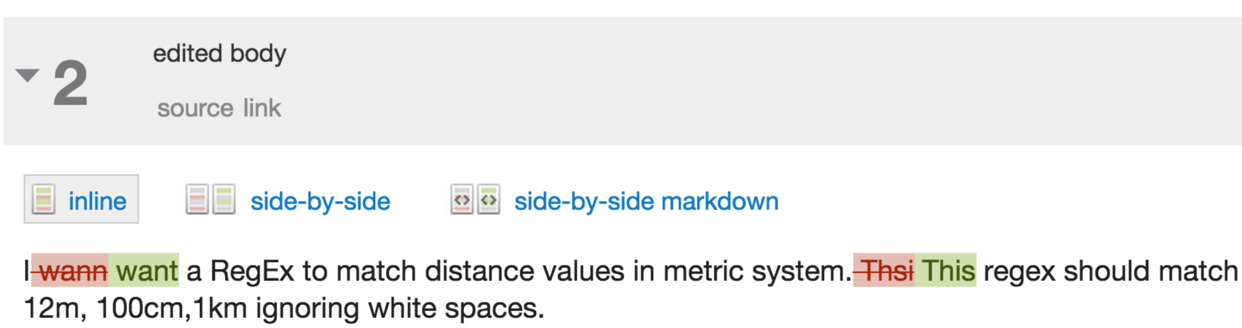


FIGURE 3

Sample of Edits on Stack Overflow

Notes: The majority of Edits on SO are simple corrections to spelling or grammar mistakes. Some also include more significant changes. Users with reputation under 2000 can suggest edits, which rewards them two points if the suggestion is accepted. Users with over 2000 reputation do not get any rewards.

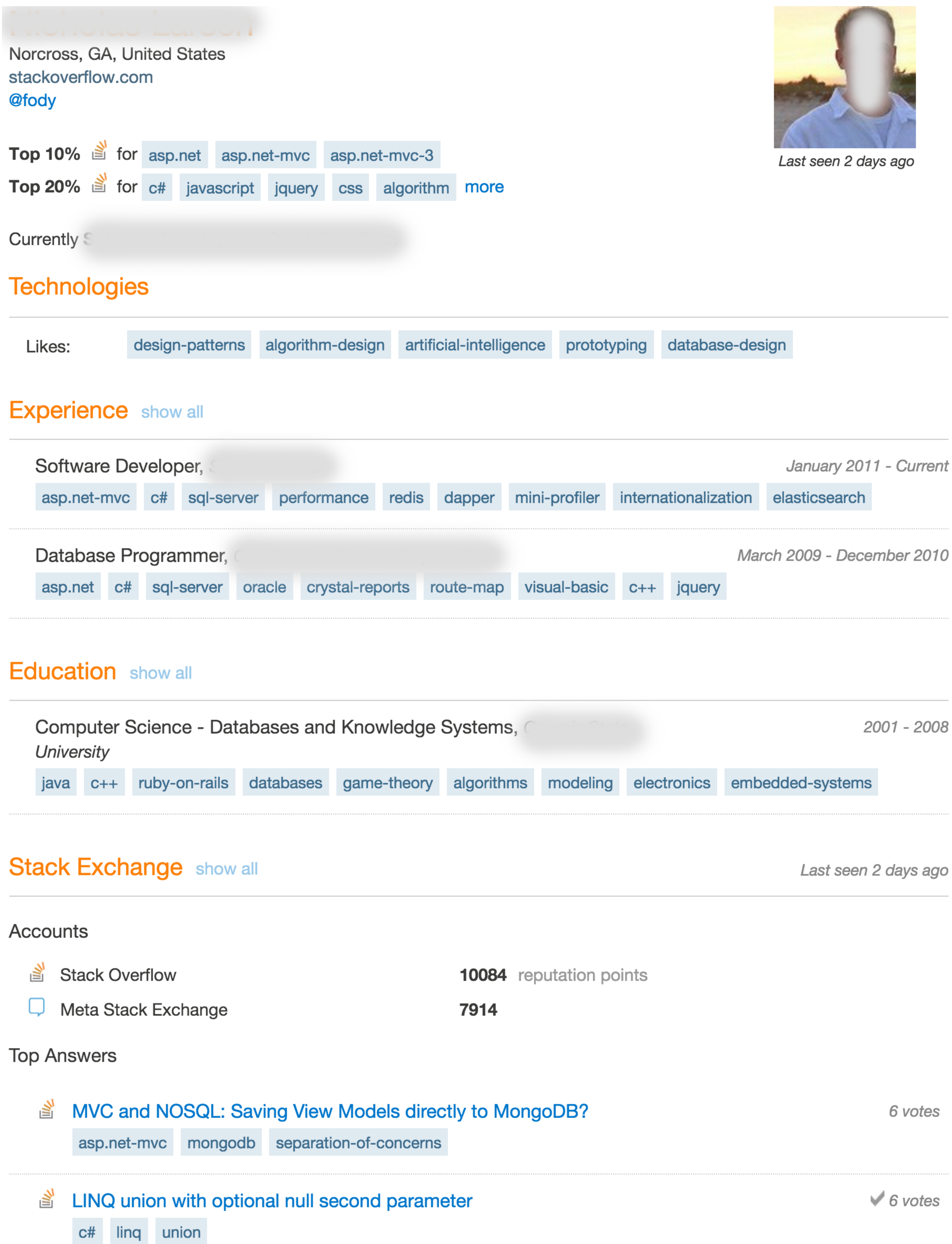


FIGURE 4
Sample of User Profile on Stack Overflow Careers

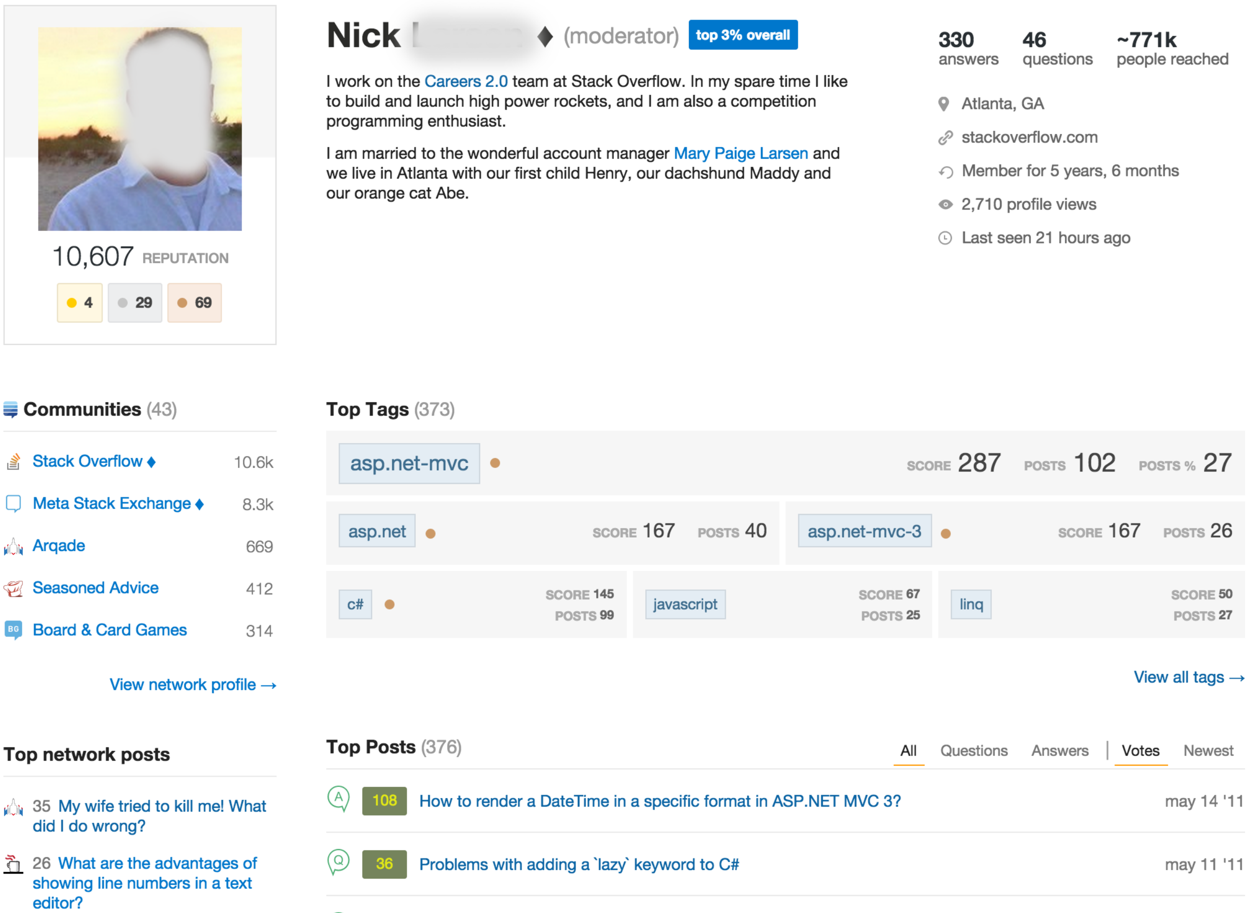


FIGURE 5
Sample of User Profile on Stack Overflow

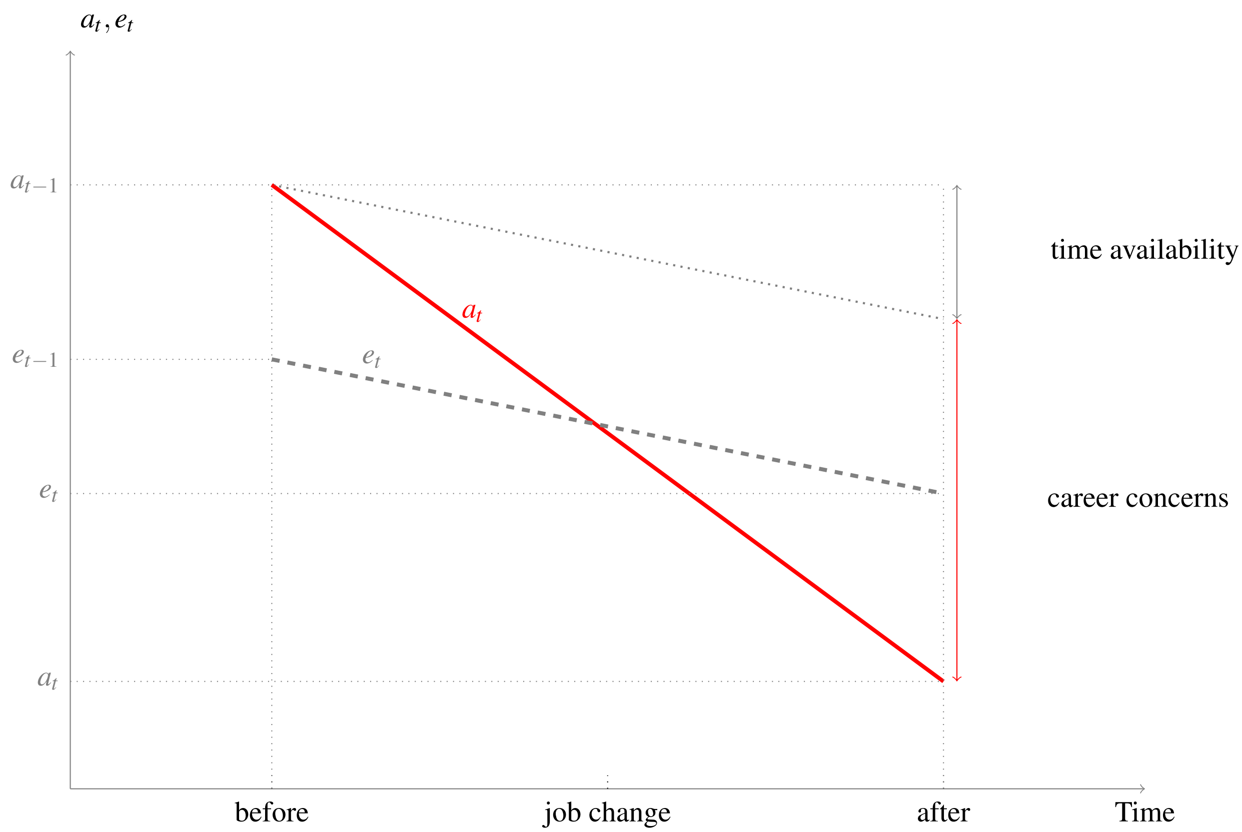


FIGURE 6
Graphical Illustration of Identification Strategy: Difference-in-Differences

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.

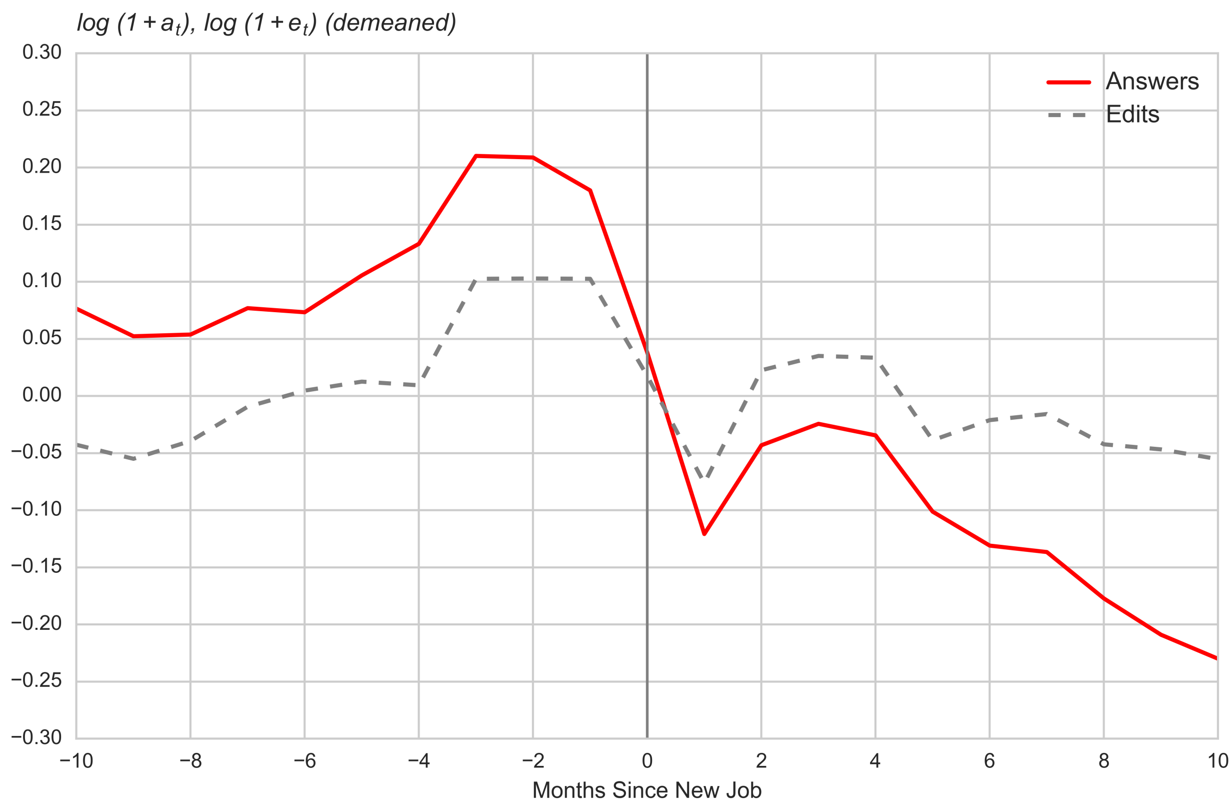
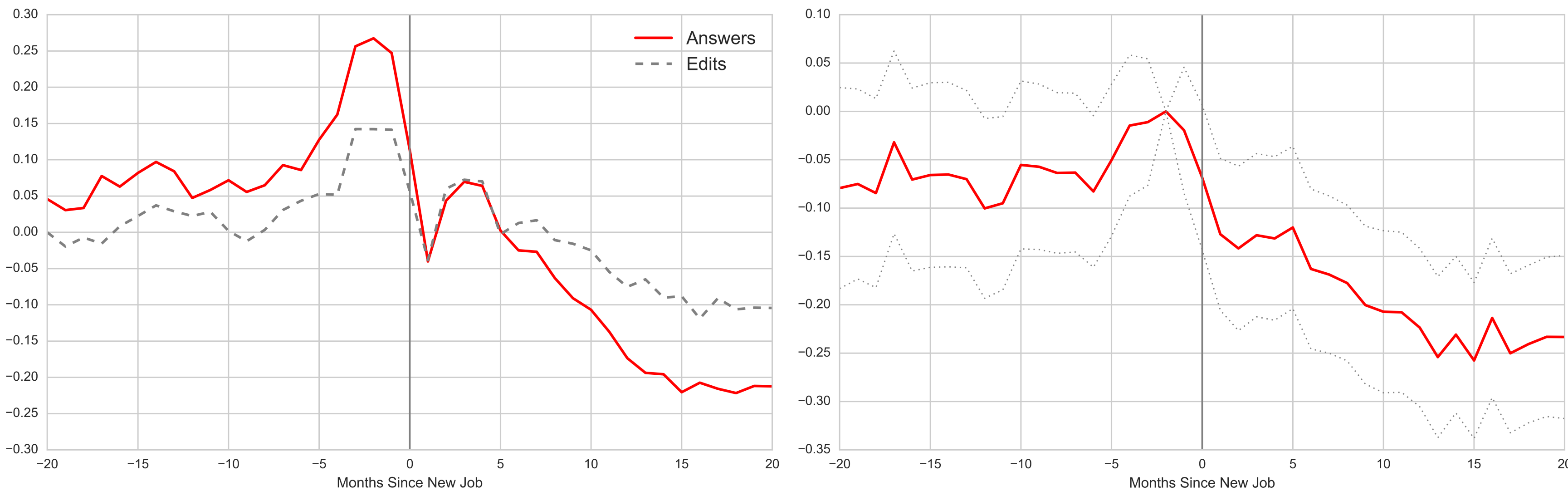


FIGURE 7

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 logarithm of one plus the 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.



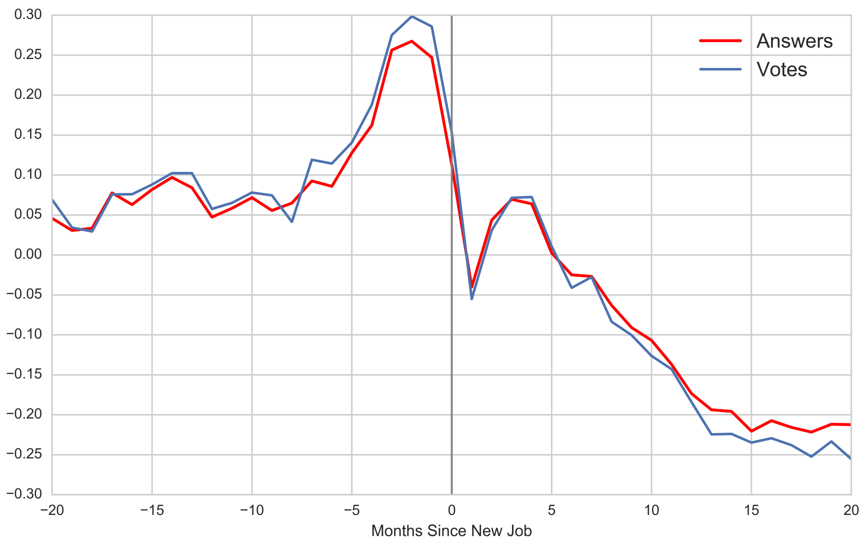
(A) Answers and Edits Activity (With Seasonality and Duration Effects Removed)

(B) Differences-In-Differences Between Answers and Edits Activity With 95-Percent Confidence Interval

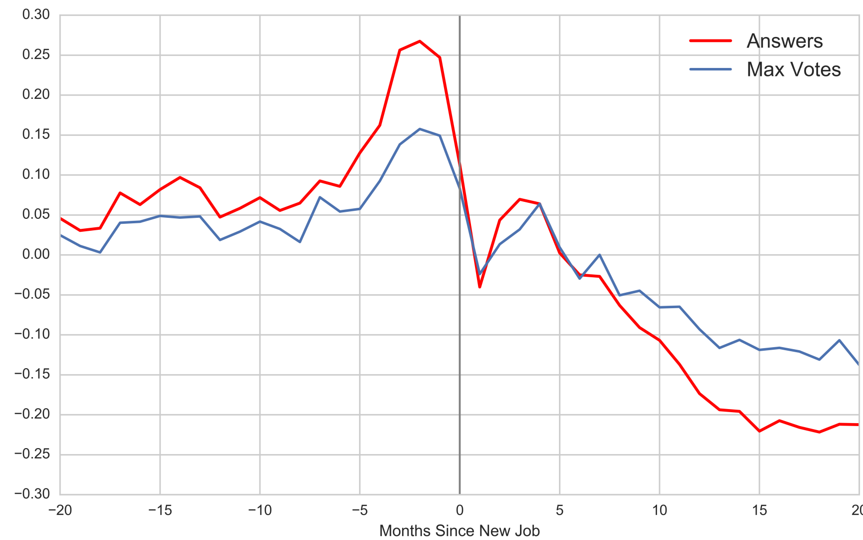
FIGURE 8

Effects of Career Concerns on Answers Activity

Notes: Figure 8A: This figure plots the demeaned values of β_τ in regression 4, using Answers and Edits as dependent variables (or coefficient estimates listed in Table 3). 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.
Figure 8: This figure plots values of γ_τ in regression 5. It uses period -2 as the baseline period, and estimates DD coefficients by comparing the differential changes in Answers and Edits activity between period τ and the baseline period -2. It controls for seasonality and duration effects for Answers and Edits separately.



(A) Number of Answers vs. Votes received from Answers



(B) Number of Answers vs. Votes from a best Answer

FIGURE 9
Quantity and Quality of Answers Activity

Notes: Figure 9A: This figure provides alternative measurements of Answers activity: number of Answers and Votes from Answers up to 30 days after an Answer is given. Votes take into consideration both quality and quantity of Answers activity. This figure plots demeaned values of β_τ in regression 4, using Answers and Votes as dependent variables. It is essentially a plot of changes in logged monthly average activities over time, after controlling for seasonality and duration effects. Both measures give almost identical graph, which is not a surprising result given that number of Votes is closely related to the number of Answers. One plausible conclusion is that the quality of Answers given by job seekers does not change much over time.

Figure 9B: This figure reveals how the average quality of the best Answers changes over time. Many users contribute multiple Answers in a given month, and each Answer receives differnet amount of Votes from other users. For each month, we choose the best Answer given by each user, and plot the average logged Votes of those Answers. Figure 9A implies that the average quality of Answers might not change over time. In contrast, Figure 9B shows that the quality of the best Answers indeed goes up before a job change, which suggests increased efforts to give better Answers by job seekers.

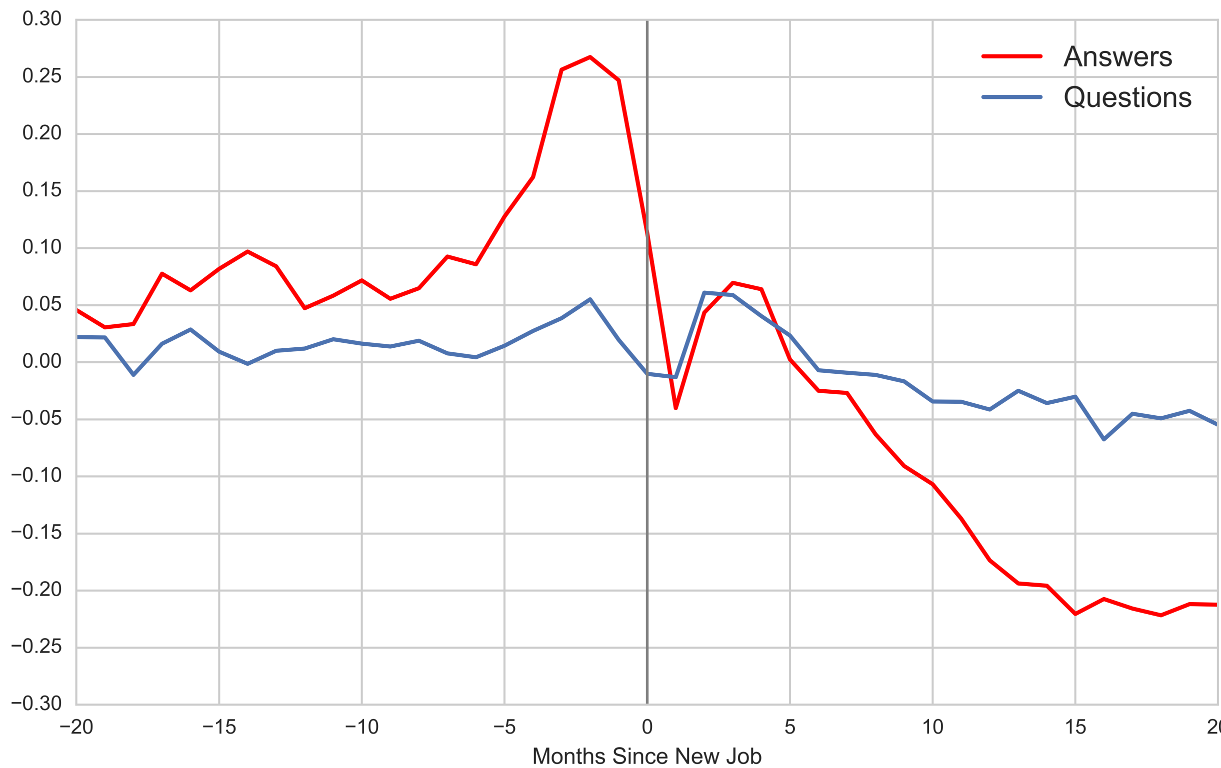


FIGURE 10

Answers and Questions Activity on Stack Overflow

Notes: This figure contrasts how Answers and Questions activity change surrounding the event of a job change. More specifically, it plots the demeaned values of β_τ in regression 4, using Answers and Questions as dependent variables, in order to control for seasonality and duration effects.

A contributor can earn reputation points through both Answers and Questions activities through Votes casted by others. One Vote to an Answer rewards 10 points, and one Vote to a Question rewards 5 points. This figure shows that job seekers respond to career incentives through increasing Answers activity, but not through Questions activity. Several explanations can explain this phenomenon: 1. Answers activity is a better way to improve one's reputation online. 2. Employers might consider a job seeker who ask too many Questions as a negative signal.

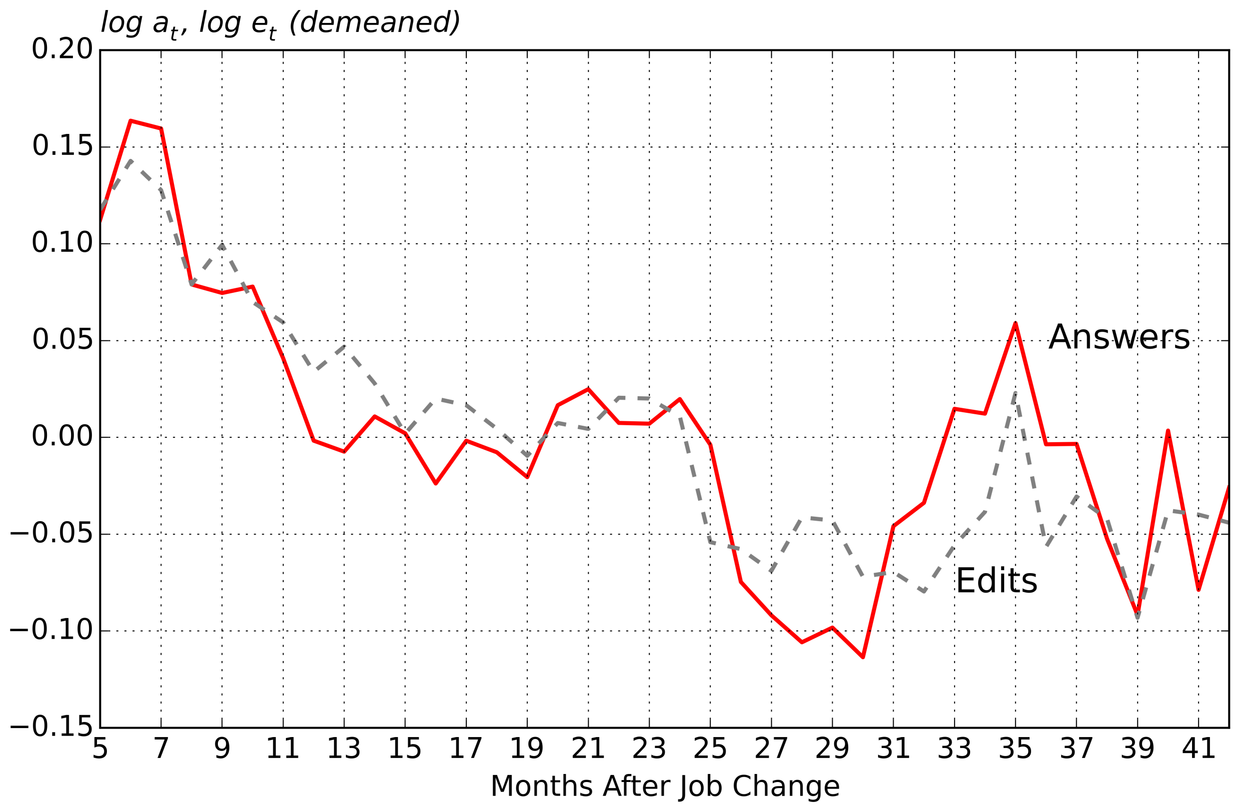


FIGURE 11
Within-Job Variations of Answers and Edits Activity

Notes: This figure plots how Answers and Edits activity vary over time within a job. From the CV data, we extract all jobs with both the starting and ending dates after a user joins SO. We select all the periods that are five months away from the beginning and ending dates. Then the average logged activity levels are plotted in this figure. x-axis shows the number of months after a job starts. Jobs have different lengths, so not all the data points are calculated from the same jobs. Assuming career incentives within a job don't vary much, this figure shows how changes in time availability affects Answers and Edits activity. Both Answers and Edits activity changes more or less in a parallel fashion, which supports the parallel trends assumption required in the DD analysis.

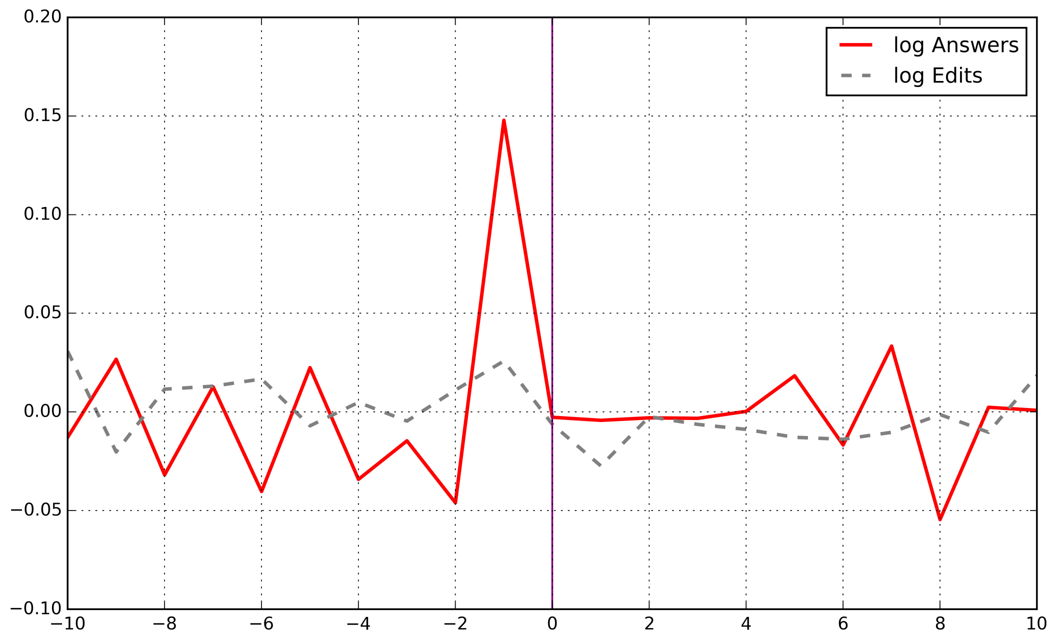


FIGURE 12
Demonstration of Dynamic Selection Effects by Simulation

Notes: This figure demonstrate the potential problem of dynamic selection effects. It is commonly referred to as Ashenfelter's Dip in Labor Economics. 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.

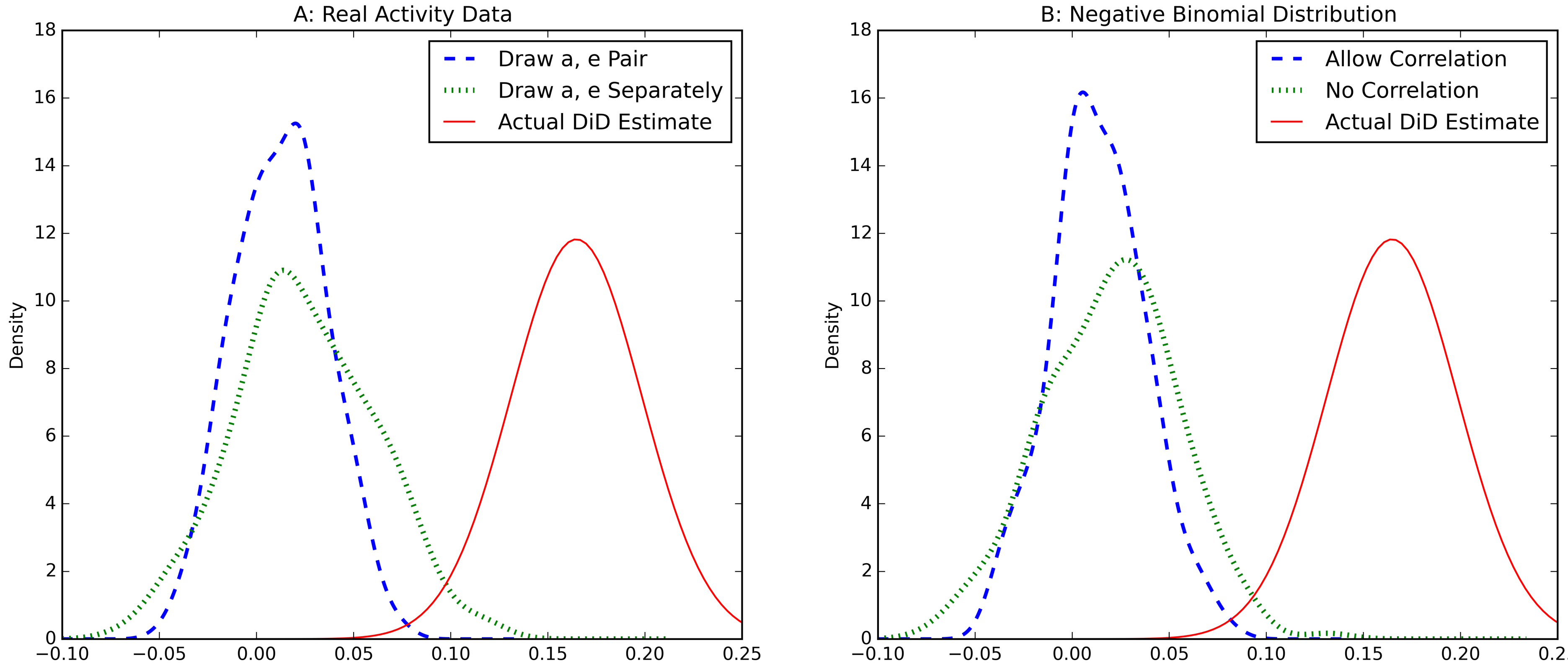


FIGURE 13
Density Plot of Simulated Difference-In-Differences Estimates

Notes: This figure presents to extent to which dynamic selection effects can confound the DD estimates in our main regression using simulated DD. It is done in several ways: Panel A draws data directly from the actual activity; 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 a mean of 0.1627 and standard deviation of 0.033.

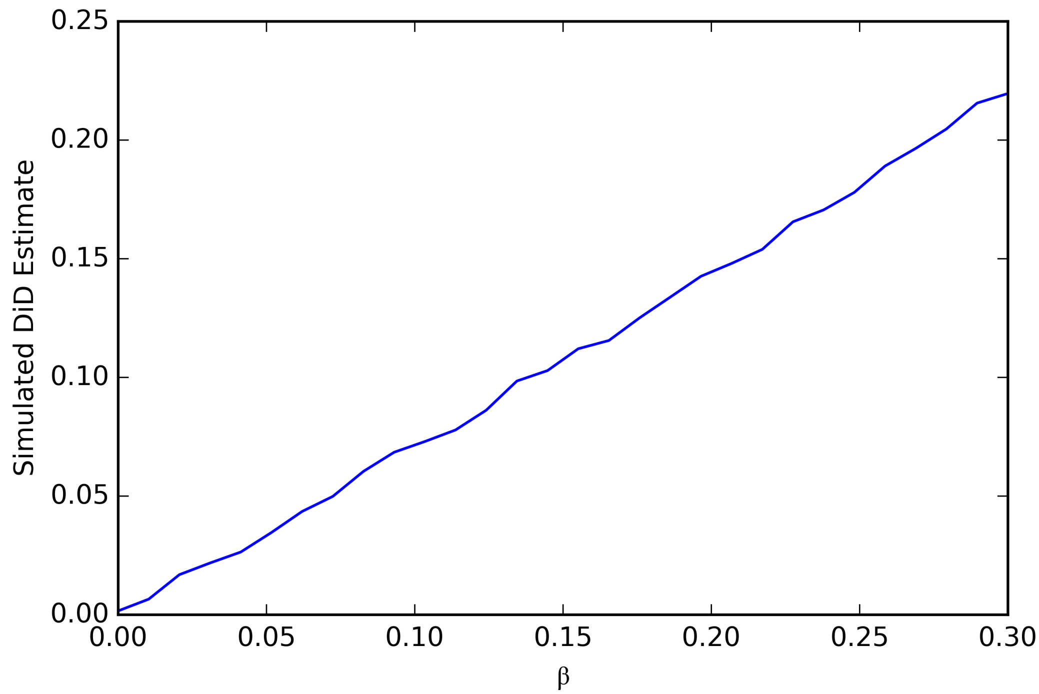


FIGURE 14
Average Simulated DD Estimates Given Values of β

Notes: The figure plots average simulated DD estimates by using different values of β. It shows the value of β needed in order for dynamic selection effects to produce a simulated DD estimate of 16.27%. β is the coefficient of Answers activity in the probability function of encoutering a job change event.