Career Concerns As Public Good

The Role of Signaling for Open Source Software Development

Lena Abou El-Komboz* Moritz Goldbeck†

January 13, 2024

Latest Version

Abstract

Much of today's software relies on programming code shared openly online. Yet, it is unclear why volunteer developers contribute to open-source software (OSS), a public good. We study OSS contributions of some 22,900 developers worldwide on the largest online code repository platform, *GitHub*, and find evidence in favor of career concerns as a motivating factor to contribute. Our difference-in-differences model leverages time differences in incentives for labor market signaling across users to causally identify OSS activity driven by career concerns. We observe OSS activity of users who move for a job to be elevated by about 16% in the job search period compared to users who relocate for other reasons. This increase is mainly driven by contributions to projects that increase external visibility of existing works, are written in programming languages that are highly valued in the labor market, but have a lower direct use-value for the community. A sizable extensive margin shows signaling incentives motivate first-time OSS contributions. Our findings suggest that signaling incentives on private labor markets have sizable positive externalities through public good creation in open-source communities, but these contributions are targeted less to community needs and more to their signal value.

Keywords: software, knowledge work, digital platforms, signaling, open source, job search

JEL-Codes: L17, L86, H40, J24, J30

*University of Munich and ifo Institute.

†Corresponding author: goldbeck@ifo.de. University of Munich and ifo Institute.

We thank Jean-Victor Alipour, Florian Englmaier, Thomas Fackler, Oliver Falck, Manuel Hoffmann, and Muhammed Yildirim as well as seminar participants at ifo Institute for valuable comments and suggestions. Further, we thank Raunak Mehrotra for excellent research assistance and gratefully acknowledge public funding through DFG grant number 280092119.

1 Introduction

Today's digital economy relies heavily on open-source software (OSS) (Lifshitz-Assaf and Nagle, 2021). While the role of patents in IT decreases (see, e.g., Acikalin et al., 2022), OSS has long become an important mode of software production (Osterloh and Rota, 2007). Numerous modern products and services are built using OSS, including electronic devices, web applications, and AI algorithms. Estimates for 2022 suggest 96% of software codebases contain OSS (Synopsys, 2023). Yet, OSS is often created by a decentralized community of volunteer developers (Nagle, 2022). Because OSS is both non-rival in consumption and non-excludable due to open-source licensing (Lerner and Tirole, 2005b), OSS is a public good. This model of open community-based software development has always been "startling" to economists (Lerner and Tirole, 2002) as the motivation of individual contributors to exert private effort in order to create an openly available public good is hard to rationalize.

One potential rationale behind private contributions to OSS is it allows developers to signal valuable information and communication technology (ICT) skills to potential employers (Lerner and Tirole, 2002) since individual contributions are directly and transparently observable on online OSS platforms. Generally, ICT abilities are highly valued skills in the labor market (Draca et al., 2007; Bresnahan et al., 2002) that yield significant returns (Falck et al., 2021). At the same time, high skill obsolescence (Deming and Noray, 2020) and the inability of formal education to certify job-relevant technical skills (Fuller et al., 2022; Marlow and Dabbish, 2013) lead to information asymmetries that make it difficult for employers to assess individuals' ability. Publicly visible OSS contributions could represent a valuable signal to potential employers (Marlow and Dabbish, 2013; Long, 2009) with respect to the most job-relevant skill in software development: practical programming ability (see, e.g., Wagner and Ruhe, 2018; Surakka, 2007). This implies that, besides private benefits from learning and improved labor market outcomes, signaling activity driven by developers' career concerns might directly generate considerable positive externalities (Leppämäki and Mustonen, 2009) in the form of a public good, open source software.

In this paper, we investigate whether career concerns are indeed a driver of OSS development. To this end, we exploit variation in individual incentives to signal over time. Specifically, because signaling is costly and its value quickly depreciates, individuals economize on the signal and dynamically allocate OSS activity to times of immediate job search in order to signal skill to employers. This allows us to test for the presence of the signaling motive empirically by studying OSS contributions of software developers who move for a job on the largest online code repository platform, *GitHub*. We focus on movers as job changes are often associated with moving (Balgova, 2020; Amior, 2019), especially for the high-skilled (Haussen and Uebelmesser, 2018; von Proff et al., 2017), which might confound our results when not explicitly considered in the empirical model. We, therefore, compare developers relocating for a new job to developers moving to a new location for other reasons in a difference-in-differences design. We argue that while job movers face elevated signaling incentives driven by immediate career concerns in the period prior to moving, the "job search period," these incentives are absent for developers who relocate for other reasons. Consequently, OSS activity attributable to signaling is captured by the difference in OSS contributions between job movers and other movers during relative to outside of the job search period.

Our data comprises all *GitHub* users with changing location information from ten snapshots of the *GHTorrent* database dated between 2015 and 2021. Due to this sample selection approach, we are able to capture typical volunteer developers who occasionally contribute to OSS (Vidoni, 2022). In total, our sample contains some 22,900 movers worldwide, of which around a third simultaneously change their job. Besides location and organizational affiliation, we observe in detail each user's public activity on the platform such as the monthly number of commits in open-source projects, their collaborators, or quality metrics such as stars, followers, and forks. This allows us to investigate not only whether career concerns drive OSS activity, but also if there are systematic shifts in OSS activity when motivated by signaling incentives with respect to the types of projects, usefulness to the community and quality, or user groups.

We find significantly elevated OSS activity by about 16% of job movers in the job search period compared to developers moving for other reasons. Assuming an average job tenure of three years applies to OSS developers and constant (base) activity levels over time, this translates to 6.8% of overall OSS activity being caused by signaling incentives during job transitions. Within the job search period effect size steadily decreases, consistent with stronger incentives during the application preparation phase. Notably, our analysis points to the importance of the extensive margin, inducing first-time contribution to OSS. In general, the effect derives from a broad base of job movers rather than a specific group. But we observe a larger effect for users relocating internationally and for users moving to academia. The signaling effect tends to be smaller for users with new jobs at large firms and especially at big tech companies, where we do not see a signaling effect. Multiple classifications of projects based on programming languages indicate that the effect is mainly driven by contributions to web development and data engineering projects, and to projects using top-paying programming languages. However, signaling projects are starred and forked less by other users, pointing to a lower direct use-value to the OSS community. In general, our results are in line with career concerns motivating significantly increased OSS contributions during the job search period as we observe activity shifts to projects that increase the visibility of existing works or necessitate skills highly valued in the labor market. Additional analyses with respect to model choice and other empirical decisions emphasize the robustness and conservativeness of our preferred specification.

This work makes several contributions. In contrast to most existing studies that follow a stated preferences approach, we deploy a quasi-experimental framework and are therefore able to achieve high internal validity of our results and causally link career concerns to OSS activity under reasonable assumptions. In addition, we improve on external validity by selecting our sample from the near-universe of OSS activity on *GitHub*, the by far largest online code repository platform. Therefore our data includes not only the most active OSS developers but also volunteer developers who only occasionally contribute to OSS, but together make up the vast majority of OSS contributors. We also add to the labor market literature by showing that employees indeed signal ability through OSS activity, which groups are especially likely to signal, and how this motivation impacts the type of projects users engage in. Importantly, we contribute to the literature on private public good provision by pointing out that there are significant positive externalities from private career concerns while, at the same time, the direction of public good creation changes when labor market considerations are prominent.

Our findings have multiple managerial and policy implications. Notably, they highlight an important but neglected channel of public good creation: the positive externalities from individual labor market signaling incentives. We show that these externalities are significant with respect to overall OSS activity and signaling incentives systematically induce first-time contributions of users previously inactive in the OSS community. To increase public good creation and platform growth, both management and policymakers should take these positive externalities of career concerns into account in platform design and public policy. For example, platform design that considers the signaling needs of their users explicitly could further boost growth at the extensive (user) and intensive (activity) margin. At the same time, decision-makers should be aware of the shift in focus towards labor market requirements and away from direct use-value for the OSS community in signaling projects. For labor market and education policy as well as HR professionals, our findings point out the continued shift away from formal (public) skill certification and emphasize greater importance of more fluid and practical skill signals that directly showcase work product. Lastly, innovation policy aiming to foster public good creation in the knowledge economy may consider maximizing positive externalities from signaling incentives, e.g. via adopting open science policies that create synergies between funded and signaling activities.

The remainder of this paper is organized as follows. First, we discuss related literature in Section 2. Section 3 introduces the data. In Section 4, we present the empirical identification strategy. Results are provided in Section 5 and Section 6 concludes with a discussion.

2 Related literature

Economics of Open Source. This project is related to the economics of open source. Literature in this area examines the distinct innovation model of OSS, which is based on volunteer contributions of often decentralized teams and is governed by open licenses (Osterloh and Rota, 2007; Lerner and Tirole, 2005b). As such, open innovation contrasts sharply with traditional ("closed") innovation featuring exclusive intellectual property rights (Lerner and Tirole, 2005a, 2002). These unique properties, combined with the lasting success of OSS and the growing importance of software in general, spurred dedicated research (see, e.g., von Krogh et al., 2003; Lifshitz-Assaf and Nagle, 2021). Compared to volunteer developers, firms are of less significance as in traditional innovation models, but increasingly incorporate OSS in their business models (Butler et al., 2019; Lee and Cole, 2003), for example to increase visibility (Conti et al., 2021) or learn from community feedback (Nagle, 2018). OSS research addresses a wide variety of topics such as productivity effects (Nagle, 2019), team organization (Raveendran et al., 2022; Puranam et al., 2014), geography (Wachs et al., 2022), or innovation and entrepreneurship (Wright et al., 2023; Wen et al., 2016; Colombo et al., 2014; Bitzer and Schröder, 2007).

Naturally, a large literature revolves around the reasons volunteer developers contribute to OSS and broadly distinguishes between internal factors and external rewards (Krishnamurthy, 2006; Hars and Ou, 2002). Von Krogh et al. (2012) cluster motivations into intrinsic (ideology, altruism, kinship, fun), internalized extrinsic (reputation, reciprocity, learning, own use), and extrinsic (career, pay). Empirically, researchers elicit the prevalence of different motivations to contribute predominantly through surveys. These works generally find evidence for mixed motivation, but internal factors tend to be most important (Von Krogh et al., 2012). For example, a survey of *Linux* contributors by Hertel et al. (2003) emphasizes the role of group belonging, identification, and a feeling of indispensability while acknowledging own use-value as another motivator. Likewise, Stewart and Gosain (2006) show that *SourceForge* contributors are more involved because of shared values. Hars and Ou (2002) conduct an e-mail survey among OSS developers, who state that self-determination, learning, and reputation are the main reasons to contribute. Community surveys by Lakhani and Wolf (2003) and Nagle et al. (2020) explicitly stress that external and monetary factors are far less important than intrinsic motivation from creativity and intellectual stimulus. In a survey by Hann et al. (2004), *Apache* developers state own use-value, recreational value, and career impact most often as motivating factors. Gerosa et al. (2021) elicit from survey responses that reputation-building as a motive became more important in recent years, and that learning and career incentives are especially relevant for novice contributors. Shah (2006) finds motivational dynamics, where initial participation is typically driven by own use-value whereas maintainers of OSS are often intrinsically motivated. Roberts et al. (2006) note that motivations interact with each other in complex ways as, e.g., being paid increases status but at the same time is associated with a lower use-value. Indeed, Krishnamurthy et al. (2014) shows that monetary reward can crowd out other motivations. Investigating behavioral changes of developer contribution after being sponsored, both Conti et al. (2023) and Wang et al. (2022) find evidence in favor of a net-positive effect of monetary incentives on activity. Projects with fast feedback and a non-commercial nature are associated with a higher probability of receiving contributions (Smirnova et al., 2022).

Our study adds to this literature in that it broadens the scope in terms of contributors being studied. While existing work mainly focuses on the most active OSS developers, often partly paid for their work, we investigate typical users on the platform, i.e., volunteer developers who sporadically contribute to open-source projects (Vidoni, 2022). The importance of economic benefits and motives for this group of OSS contributors is neglected in the literature, and this study is among the first to study the role of career concerns in a causal identification framework. As such, it sharply contrasts with the prevailing methodological approaches used in existing research on this topic. These works are largely based on surveys, which feature the important caveat of only eliciting stated preferences as opposed to the revealed-preference approach embodied in our causal framework. As a result, we are able to make quantifiable causal claims on the importance of career concerns motive for typical volunteer OSS developers under reasonable assumptions. Our findings suggest a sizable portion of OSS activity is driven by career concerns, and that motivations dynamically change over time, which in turn alters the content of contributions.

Labor market signaling. This article focuses on one specific motivating factor to contribute to OSS, career concerns, and therefore adds to the vast literature on signaling originating from Spence (1973). Subsequent theoretical models explicitly relate career concerns to signaling via observable effort (Holmström, 1999; Chevalier and Ellison, 1999), even when beliefs on ability are precise (Miklós-Thal and Ullrich, 2015). While basic signaling models yield separation of skill types even if signaling has no real effects, Leppämäki and Mustonen (2009) provide a model where signaling activity generates (positive) product market externalities. Empirically, Miklós-Thal and Ullrich (2016) test the career concerns hypothesis in soccer and find confirmatory results for marginal individuals. Pallais (2014) shows detailed public performance records on the online marketplace *oDesk* improved workers' subsequent employment outcomes, especially for the inexperienced. Also on an online platform for contract labor, Agrawal et al. (2016) find standardized and verifiable information important for developing-country candidates' employment probability. For software developers, Xu et al. (2020) find career concerns increase reputation-generating activity in an online community forum. Experimental evidence by Piopiunik et al. (2020) reveals basic IT skills signals in CVs on the broader white-collar labor market significantly increase the probability of receiving a job interview invitation. Apart from this causal evidence, surveys show reputation-building, signaling, and career concerns are important motivations for developers to contribute to OSS (e.g., Gerosa et al., 2021; Marlow and Dabbish, 2013; Hann et al., 2004; Hars and Ou, 2002). Similarly, employers state they regard OSS contribution as a credible and valuable signal. For example, in a survey, Long (2009) finds tech companies value OSS experience of applicants. More specifically, Marlow and Dabbish (2013) surveys recruiting managers who state *GitHub* activity is used in hiring as a signal for technical abilities and motivation, and is regarded as a stronger signal than the applicants' resume with respect to these areas. A survey among developers by Hakim Orman (2008) shows OSS activity and traditional education are seen as complements and not substitutes. However, Bitzer and Geishecker (2010) finds formally educated individuals are underrepresented in the OSS community. For developing-country candidates, Hann et al. (2013) claim that valuable OSS activity is an effective and credible signal as it is associated with significant wage premiums for *Apache* project participants. Huang and Zhang (2016) associate improved outside options from OSS signaling with job-hopping, but also acknowledge retaining effects from learning.

The contribution of this research to this strand of literature is twofold. First, in contrast to most work in this area, we follow a quasi-experimental approach using observational data from the near-universe of OSS developers. This allows us to make causal claims under reasonable assumptions leading to a comparably high degree of internal validity. Furthermore, because we are able to study a large and diverse group of OSS contributors and do not limit our scope to the most active users, the results also feature a higher level of external validity in comparison to the fairly specific and small groups typically studied in existing works thus far. Our second contribution, which received limited attention, is asking to what degree signaling activity is wasteful or productive from a content perspective. Our empirical evidence suggests lower but still positive direct use-value for the community of signaling activity, and therefore adds an empirical perspective to the notion of positive externalities of signaling, which has only been examined theoretically to date (Leppämäki and Mustonen, 2009).

Public good provision. The paper is also connected to the broader literature on private public good provision. In contrast to traditional innovation models that rely on private property, open innovation models like OSS largely depend on voluntary contributions by individual developers and thus can be framed as private public good provision (Lerner and Tirole, 2002). Traditional theory emphasizes group size as the main factor influencing the provision of the good (e.g., Chamberlin, 1974; Bliss and Nalebuff, 1984; Palfrey and Rosenthal, 1984; Bergstrom et al., 1986; Hendricks et al., 1988; Bilodeau and Slivinski, 1996). Explicitly modeling intrinsic motivation, Bitzer et al. (2007) show provision is more likely maintained when OSS programmers value gift benefits and the intellectual challenge, have a long time horizon (i.e., are younger), are patient, face low development cost, and derive a high own use-value. In a model of OSS development, Johnson (2002) shows how own use-value considerations drive the direction of software production. Incorporating own use-value considerations and provision costs, Myatt and Wallace (2002) model a public good provision game and show multiple equilibria can arise. Ignoring intrinsic motives, Bitzer and Schröder (2005) derive joining and exiting dynamics from signaling in a model of repeated contribution. Regarding the licensing regime, Fershtman and Gandal (2004) show that contributions are higher when OSS licensing is less restrictive. Athey and Ellison (2014) model a world where OSS projects can be successful when developers are motivated by reciprocal altruism if customer support is not needed. Zeitlyn (2003) emphasizes the gift economies motivation. Empirically, O'Neil et al. (2022) define contribution territories for firms and individuals in the space of possible innovation to rationalize why certain areas are neglected. Recently, del Rio-Chanona et al. (2023) find public good generation on *StackOverflow* is impacted negatively by large language models, a substitute to online forums.

Our empirical results are important to inform on the applicability of theoretical models depending on their presumptions. Our findings emphasize that external motives are relevant and that considering the dynamic evolution of motivation is important. At the same time, external motives such as career concerns likely do not explain OSS activity entirely. Hence, theoretical models that aim to capture OSS contribution comprehensively should consider modeling multi-dimensional motivations to contribute that include both internal and external motivations and incorporate their dynamic evolution. In general, our study emphasizes the importance of labor market incentives of high-skilled professionals for the private provision of an important public good in the knowledge economy, which likely features considerable positive spillovers both on the private market and in the form of public follow-on innovation in the OSS community.

3 Data

We study software developers on *GitHub*, the by far largest online code repository platform. *GitHub* was founded in 2008, reached 10 million users by 2015, and in 2021 reported 73 million users worldwide (GitHub, 2021; Startlin, 2016). Around a fifth of all code contributions on the platform are made to public repositories, i.e., open-source projects (GitHub, 2021). Repositories are maintained using the integrated version control software *git*. Importantly, the nature of the *git* version control system allows us to track each user's contribution to open-source projects over time as it records and timestamps all activity in public repositories. *GitHub* provides access to public user profiles and repositories via API. Data analyzed in this paper originates from *GHTorrent*, a research project by Gousios (2013) that mirrors the data publicly available via the *GitHub* API and generates a queryable relational database in irregular time intervals.[1] The resulting snapshots contain data from public user profiles and repositories as well as a detailed activity stream capturing all contributions to and events in open-source repositories. This paper relies on ten *GHTorrent* snapshots dated between 09/2015 and 03/2021.[2]

On their *GitHub* profile, users can indicate their location. This self-reported indication is voluntary and is neither verified nor restricted to real-world places by *GitHub*. Goldbeck (2023) finds no systematic bias in the location information provided on the platform, even though only a fraction of users indicates their location. We assign users to cities via exact matching to city names in the *World Cities Database*.[3] Users can also provide an indication of their organizational affiliation, which we use to elicit job changes. Location and organization information is observed only on snapshot dates – i.e., roughly every six months – while user activity is timestamped. We aggregate users' timestamped activity to monthly data to obtain a panel structure. Since the data is highly skewed and most users are inactive (see, e.g., Vidoni, 2022; Luca, 2015), we restrict our sample to users with an observed minimum activity of three months with non-zero commits.

Movers. From the data, we select movers, i.e., users who change their city-level location once in the observation period. Our empirical strategy elicits signaling activity from time-varying incentives around a job change. When people change jobs, they often simultaneously move (Balgova, 2020; Amior, 2019), which is especially the case among high-skilled professionals (Amior, 2015; Machin et al., 2012; Greenwood, 1975, 1973). To attain a meaningful comparison and get rid of any confounding factors associated with moving we, therefore, compare users who move for a job to users who move for other reasons. We infer the reason for moving from changes in the organizational affiliation of users. Whenever there is no affiliation change around the move date we regard a user as moving for other reasons. Conversely, if a new affiliation appears

[1] *GHTorrent* data contains potentially sensitive personal information. Information considered sensitive (e.g., e-mail address or user name) has been de-identified (i.e., recoded as numeric identifiers) by data center staff prior to data analysis by the authors. Data from the *GHTorrent* project is publicly available at `ghtorrent.org`.

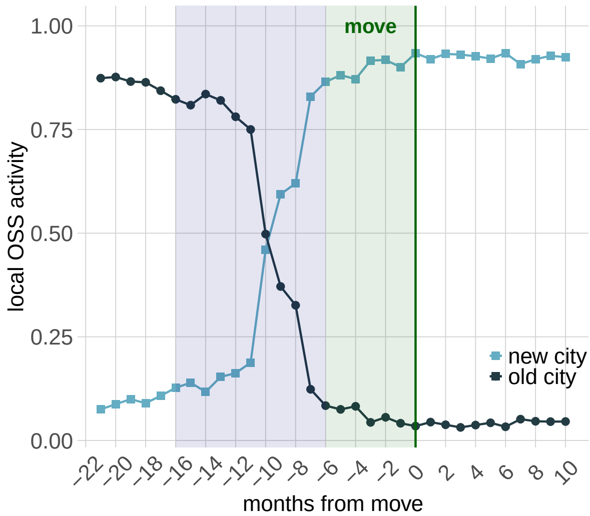
[2] Snapshots are dated 2015/09/25, 2016/01/08, 2016/06/01, 2017/01/19, 2017/06/01, 2018/01/01, 2018/11/01, 2019/06/01, 2020/07/17, and 2021/03/06.

[3] A fraction of 0.25% of users (total: 58) are not matched to a city in the database but rather a state or a country. We do not geocode cities or states with a name that exists multiple times.

around the move date we consider a user as job mover. To implement this, we extract users' move (and job change) dates from the data.

We infer the move date from user-level location information as the month of the first snapshot with a new city indication. There is some uncertainty regarding the actual move date for two main reasons. First, users manually enter (new) location information data on the platform themselves and do this not necessarily exactly at the time of moving. On the one hand, users might be busy during the time period of moving and enter their move late. On the other hand, it might be beneficial to communicate the future location early, maybe even before actually moving, to let peers know about their relevant location as soon as possible. We empirically investigate the plausibility of the move dates attained through the snapshots by looking at team member locations in the projects a user actively contributes to each month. To this end, we assign locations to projects depending on other members' locations. Specifically, we define a user's project as localized in a particular city if the current location of more than 60% of the team members is in that city. This is only possible for a subset of projects as few members share their location and team members can be distributed. Nevertheless, it allows us to get an impression of changes in the spatial collaboration pattern of users in our sample.

Figure 1: User collaboration around relocation date



Notes: Graph shows in-sample users' commits to new- and old-city repositories as a share in users' total commits to repositories with an assigned location. Location is assigned to repositories for which at least 60% of the team members indicate a common city as current location. *Sources:* GHTorrent, own calculations.

Figure 1 plots the share of users' activity in localized collaborative projects by origin and destination city. The dark blue line represents a users' activity share in projects where team members are predominantly located in her origin city while the light blue line represents activity in projects with team members predominantly located in the destination city. The graph shows a clear pattern. Most localized activity is in old city projects up to ten months prior to the estimated move date. This starts to reverse afterward and most localized activity is measured in destination city projects from six months prior to moving until the end of the observation period. It is plausible that users start collaborating with teams in their destination city prior to moving and activity in old-city projects fades out. Importantly, this graph shows user-provided locations systematically and meaningfully relate to collaboration patterns, which validates our measurement of moving. Similarly to the move date, we elicit job changes from users' affiliation indication as the first month the new city location is observed in the data.

Summary statistics. The resulting sample of users comprises 22,896 movers, of which 7,211 (32%) simultaneously change their job.[4] Naturally, since most registered users are inactive, this sample is very different compared to the universe of users in the data and comprises more active users, which is confirmed by the summary statistics in Table A.1. More interestingly, Table 1 provides an overview of our sample and compares job movers and other movers. In general, job movers and movers are comparable in terms of activity, collaboration, and quality metrics. At the same time, there are also some differences between the groups. The median mover has five followers, contributes around 170 commits to open-source projects in the observation period, and has 15 projects with on average 2 to 3 team members. Job movers contribute a bit less to team projects and the average team size is smaller compared to other movers, and their team projects also receive fewer stars and forks. Projects in our sample are very diverse both in terms of programming languages (cf. Table A.7) and topics covered and range from web development to data engineering (cf. Figure A.5).

The differences between job movers and other movers regarding team project behavior is one reason why we look at single projects, i.e., projects in which only the focal user is active. But there is a more important reason derived from theoretical considerations and a practitioner's perspective with respect to labor market signaling through OSS activity. Not all contributions to OSS communities constitute equally valuable signals of ability and thus generate reputation (Xu et al., 2020; Marlow and Dabbish, 2013). In particular, for potential employers, it is difficult and time-consuming to assess individual contributions to collaborative projects even if transparently available (Tubino et al., 2020). In contrast, single-authored projects can be assigned entirely to individual users. At the same time, quality metrics such as stars and forks make assessment effortless and enable non-software developers like HR professionals to perform such assessments. Consequently, using OSS activity in single projects as the main outcome metric ensures a close practical and theoretical relation to actual signaling potential.

Although we look at users moving worldwide, 71% are relocations to another city within the country. About 29% of relocations are international, and 19% of movers or two-thirds of international movers even move inter-continentally. This mirrors the fact that software developers are disproportionally mobile internation-

[4] Figure A.2 reports the moves by data snapshot and shows a similar distribution for job movers and other movers.

Table 1: Summary statistics

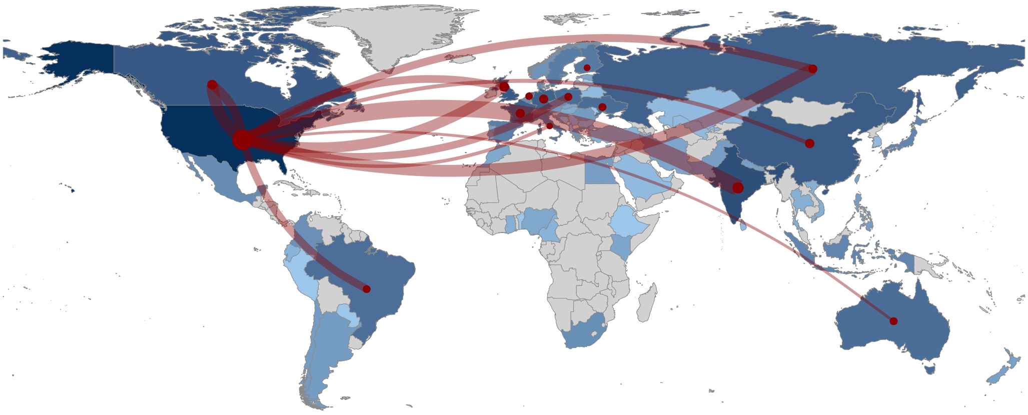
Median	Movers job	Movers other	Δ	%Δ
Activity				
Commits	163	188	–25	13.3%
commits single projects	*72*	*76*	*–4*	*5.3%*
commits team projects	*59*	*80*	*–21*	*26.3%*
Experience	37	42	–5	11.9%
Collaboration				
Projects	14	16	–2	12.5%
single projects	*9*	*9*	*0*	*0.0%*
team projects	*5*	*6*	*–1*	*16.7%*
Project members	2.21	2.82	–0.61	21.6%
Quality				
Followers	5	5	0	0.0%
Stars	1.10	1.88	–0.78	41.5%
stars single projects	*0.09*	*0.12*	*–0.03*	*25.0%*
Forks	0.62	1.11	–0.49	44.1%
forks single projects	*0*	*0*	*0*	*0.0%*

Notes: Experience is measured as tenure on the platform in months since the first commit at the move date. Column Δ reports the absolute difference in median between job movers and other movers. Column %Δ sets this difference in relation to other movers' median. *Sources:* GHTorrent, own calculations.

ally (see, e.g., Adrian et al., 2017; D'Mello and Sahay, 2007; Solimano, 2006). The average relocation distance is 5,324km and there are no significant differences in these statistics between job movers and movers relocating for other reasons (cf. Figure A.1). Figure 2 maps the observed migration flows in our data in more detail. Countries are colored in darker blue the higher the number of domestic relocations and the width of the network edges represents the number of international relocations. The dominance of the USA as the central hub both in terms of domestic moves and as a receiving country is clearly visible even on the logarithmic scale. Domestic moves are observed most frequently in the USA (63.5%), India (7.5%), and the United Kingdom (3.9 %). Table A.4 shows the ten countries with the most domestic moves, which account for over 90% of domestic moves and 65% of all relocations. The most important origin countries are shown in Table A.5. Table A.3 reports the ten largest origin and destination cities, which are predominantly the world's big software industry hubs, e.g., San Francisco and New York. Notably, for international relocations, we observe that users tend to move to richer countries as indicated by per capita GDP increasing on average by USD 9,780 (Figure A.3), with no systematic differences between job movers and other movers.

Users are affiliated with a diverse range of organizations. Most firms in the data are small, but the distribution is highly skewed to the right (Figure A.4). On average, each organization has four affiliated users and 23

Figure 2: Domestic and international user relocations



Notes: Blue country coloring shows the number of domestic movers after logarithmic transformation. There are 73 countries with domestic movers; grey indicates no domestic movers. The size of the red country centroids indicates the number of international moves a country is involved in. 14 countries are associated with international relocations. Red arcs represent edges in the directed country mover network, i.e., the number of international relocations from one country to another, and are scaled logarithmically. For clarity, only edges above 75 are shown. *Sources:* GHTorrent, own calculations.

users are affiliated with the median organization.[5] Table A.2 reports organizational affiliations and job transitions by organization type. As a consequence of the skewness, about 29% of users are affiliated with the 100 largest firms and 7.2% with the big technology firms (i.e., Google, Apple, Meta, Amazon, Microsoft; GAMAMs). Job transitions point out net movements towards larger, and especially big tech, firms and away from academic and small-firm affiliations. This is confirmed by Table A.6 depicting top origin and destination affiliations. While top origin affiliations include mostly students, universities, and freelancers the biggest destination shares almost exclusively are held by large software companies such as the GAMAMs or Red Hat, IBM, and LinkedIn.

4 Empirical strategy

The key idea behind our empirical model setup is to exploit temporary differences in signaling incentives across users. Specifically, we compare the activity of users who move for a job and movers who move for other reasons. The reasoning behind this is that users who move for a job experience increased incentives to signal their ability on the platform to potential employers prior to their move during the job search period, whereas movers who relocate for other reasons do not experience this temporary increase. As already discussed above, we focus on movers since job changers typically simultaneously relocate, which is widely acknowledged in the literature (Balgova, 2020; Amior, 2015) and especially the case for high-skilled pro-

[5] Note that these numbers are not to be confused with the number of employees since not all employees are active OSS contributors on *GitHub* and provide their affiliation.

fessionals (see, e.g., Abreu et al., 2015; Haapanen and Tervo, 2012; Venhorst et al., 2011; von Proff et al., 2017; Kodrzycki, 2001; Ciriaci, 2014; Haussen and Uebelmesser, 2018). Thus, comparing movers leads to improved comparability as it accounts for confounding factors associated with moving.

Figure 3: Adapted difference-in-differences model

From a theoretical perspective, we structure signaling incentive dynamics into three phases, where each phase is governed by a distinct incentive regime. This is illustrated in Figure 3. In the first phase, which we call the pre-period, an eventual mover is still working in her previous arrangement and does not actively prepare to change jobs. In this phase signaling incentives are not entirely absent and are at a normal level as there is no immediate pressure to signal skill in the labor market. In the decisive second phase, the "job search period," the job mover then actively searches for a new employer and prepares to relocate while movers who relocate for other reasons only prepare to relocate. In this phase, job movers face elevated incentives to signal skill to potential employers. Finally, there is a third phase after the move, which we call post-period, in which movers have relocated and the job mover has started to work for her new employer. Movers who relocated for other reasons are still with their old affiliation. In this phase, as job movers just started a new job, signaling incentives vanish and are likely even lower than in the pre-period and compared to other movers because job movers have to settle in to their new job environment, and the especially low signaling incentives.

As a result of these theoretical considerations, we expect elevated OSS activity of users who move for a job compared to users who move for other reasons in the job search period if career concerns are an important factor for OSS contribution. Additionally, we expect to see lower OSS activity of job movers compared to other movers in the post-period. We empirically investigate the dynamics of OSS activity by estimating the following baseline event study model:

$$y_{it} = \beta_1 + \sum_{j=\underline{T}}^{\overline{T}} \left[\beta_j \left(t_j \times \text{JobChanger}_i \right) \right] + \delta_i + \delta_{s(t)} + \delta_{a(i)t} + e_{it}, \tag{1}$$

where y_{it} is the number of commits of user i in relative-to-move month t to single-authored repositories ("signaling projects"). Note that the event study panel is balanced in the job search and pre-period but unbalanced in the post-period as some moves happen during the end of our observation period. The variable JobChanger_i indicates if user i moves for the job, i.e., simultaneously changes her affiliation and location. The core element is the interaction term of JobChanger_i with relative months to the moving month t_j. Coefficients of interest are $\beta_{\mathbf{j}}$ and reveal the difference in the temporal pattern of signaling activity around the move date between users who simultaneously change their job and users who do not. To control for time-constant unobserved user characteristics relevant to their level of OSS activity, we add user fixed effects δ_i. Calendar month fixed effects $\delta_{s(t)}$ account for unobserved factors affecting all users' activity in a given month. We include experience fixed effects $\delta_{a(i)t}$ to account for differences in platform tenure across users that impact OSS activity. Standard errors are clustered at the user level.

Starting from this flexible dynamic model, we adapt the standard difference-in-differences model to estimate the average treatment effect on the treated such that three phases around the move date are considered: a pre-period, a job search period, and a post-period. The reference period is the pre-period, and the temporary treatment of increased incentives to signal using OSS activity is present only during the job search period. In the post-period, signaling incentives for job changers are lower relative to the pre-period because of diminished career concerns and the new job crowding out OSS activity. The resulting model specification is

$$y_{is} = \beta_1 + \beta_2(\text{SearchPeriod}_{s(i)} \times \text{JobChanger}_i) + \beta_3(\text{PostMove}_{s(i)} \times \text{JobChanger}_i) + \delta_i + \delta_s + \delta_{a(i)s} + e_{is}, \quad (2)$$

where $\text{SearchPeriod}_{s(i)}$ is one if calendar month s falls in user i's job search period prior to the move. To account for generally reduced incentives of job switchers to make OSS contributions after the move relative to users who move for other reasons, we interact an indicator for the post-move period, $\text{PostMove}_{s(i)}$, with job changer status. The coefficient of interest β_2 captures the ATT of increased signaling incentives during the job search period, i.e., differences in OSS activity between job movers and other movers in the job search period relative to the period before. Similarly, β_3 represents the average difference in OSS activity between job movers and other movers in the post-move period relative to the pre-period.

Although the inclusion of the post-period is not formally needed for identification, we consider it explicitly in our model for two reasons. First, it adds credibility to the signaling effect estimated from the difference between the pre-period and the job search period if signaling activity declines when taking up a new job, which we assume reduces immediate signaling incentives. Second, validation of parallel trends between job movers and other movers in both the pre- and post-period helps to further assess the validity of our design. And third, although not the main goal of this analysis, estimating the effect of taking a new job on OSS activity is interesting in itself. The three-period specification with the pre-period as reference is superior to alternatives. Taking the post-period as reference neglects the crowding-out of OSS via time constraints of formal work. Combining pre- and post-period as reference attenuates this issue, but leads to potential overestimation due to the same mechanism.

Empirical results from the event study specification guide the selection of appropriate time frames for the three phases in the ATT model. In addition, a priori theoretical and empirical considerations set our expectations. In his classical framework, Blau (1994) divides the job search period into three steps. The first step is the preparation phase, where applicants prepare their application package. Then there is the actual application step in which applicants undergo the formal application process. Finally, the third step is the decision step, in which employers and applicants decide on whether to enter an employment relationship or not. Signaling activity is expected to occur predominantly in the first step, i.e. preparation (Chamberlain, 2015). Recent statistics for the US show hiring time for complex jobs such as software development averages around four months prior to applying (Firaz, 2022), and people start thinking about and preparing for job search likely much earlier. Additionally, there is some fuzziness in our measurement of the move date due to only observing locations about every six months. Therefore, we expect to see most OSS signaling activity in the preparation phase of the job search period somewhere between six and 15 months prior to our estimated move date.

Note that our model specification provides a conservative and incomplete estimation of the role of career concerns for individual OSS activity for multiple reasons. First, signaling incentives are not entirely absent in the pre-period. Career concerns are not binary and we exploit time variation in their strength rather than presence or absence. Second, our estimates are downward biased due to measurement error when some control group movers in fact move for the job, as well, but do not change their affiliation. Third, our focus on movers implies we study a group of users who face significant additional time constraints relative to users who are not relocating and therefore trade-off their time allocation between more activities, potentially leading to less time spent on signaling activity in this group. Finally, the dynamics within the job search period as well as the fact that toward the end of our signaling period, the share of users who already found a job increases biases the ATT downward. Consequently, our estimates should be interpreted as a lower bound to the importance of career concerns for OSS activity.

Our key identifying assumption is that in the absence of signaling incentives for job changers, their activity would have evolved similarly to movers not changing jobs simultaneously, conditional on controls. Although we cannot test this assumption directly we assess it by showing parallel trends in periods when signaling incentives are absent, i.e., both the pre- and post-period. The main remaining threats to our identification strategy are factors unrelated to signaling incentives that affect the user activity of job changers in the job search period prior to the move but not the user activity of movers that do not change jobs or vice versa. One such concern could be due to potentially reduced work ethic of job movers in their old job as it comes to an end and, as a consequence, more time for side projects. However, one could also expect the old job claims more time towards the end as, e.g., projects have to be handed over. Another potential concern is an increased prevalence of learning motives during periods of unemployment between two jobs. This is, however, not only unlikely due to generally short unemployment spells for IT professionals; the median duration of unemployment in the US, for example, is only eight weeks.[6] It is especially unlikely given that our design focuses on movers, and relocating to another city or even country is generally time-consuming and stressful. Nevertheless, in Sections 5.2 and 5.3 we address these concerns and assess related channels

[6]Statistic retrieved from BLS based on the Current Population Survey 2018: https://www.bls.gov/web/empsit/cpseea37.htm. Last accessed on 11/10/2023.

by investigating the kind of OSS activity of job movers and how it differs from other movers to validate if the observed activity can likely be attributed to signaling or not.

5 Results

5.1 Main effect

Figure 4 plots the event study coefficients for user activity around the relocation date resulting from the model in Equation 1. The dynamics are consistent with signaling as a driver of OSS activity and the hypotheses derived from our theoretical considerations. In the pre-period, there are no statistically significant differences in OSS activity between users who move for a job and users who move for other reasons. Similarly, after moving we observe a lower activity level for job movers compared to other movers but the dynamic development is, again, parallel to each other. This absence of differential trends between treatment and control group users is reassuring of the validity of our empirical design as it provides confidence that our key identifying assumption holds. Importantly, during the period prior to moving, OSS activity of job movers is significantly elevated relative to other movers conditional on controlling for time, user, and experience fixed effects. We claim this increase is driven by immediate career concerns in the period of job search which incentivizes signaling activity.

Figure 4: Event study estimates

Notes: Estimates for $t_j \times \text{JobChanger}_i$ based on Equation 1 with user, experience and calendar month fixed effects. The outcome is IHS-transformed commits to single-authored projects. The reference month is $t = -16$. Bars show 95% confidence intervals. Standard errors are clustered at the user level. *Sources:* GHTorrent, own calculations.

The dynamic activity pattern during the job search period is consistent with signaling behavior, too. Signaling activity is strongest at the beginning of the job search period 10 to 14 months before the move month with activity in signaling projects being elevated by up to 24.5% for job movers. The effect then declines steadily to substantially lower levels before the move date around 6-10% before returning to a permanently lower, stable, activity level from the move month onward, with estimates centering around -7 to -10%. Model (3) in Table A.13 provides estimates for each period. This pattern is in line with our theoretical considerations predicting more intense signaling in the preparation step of the job search period as users generally have an incentive to have their signal ready by the time of application which is likely earlier in the job search period. In addition, more and more users finding a job during the job search period or moving earlier than the observed move month, both leading to reduced incentives to signal.

Because of sparsity, we transform the dependent variable using the inverse hyperbolic sine transformation in order to retain zero-valued observations (Bellemare and Wichman, 2020). At the same time, this transformation approximates the natural logarithm and is commonly interpreted in a similar way (Burbidge et al., 1988; MacKinnon and Magee, 1990). As our data typically features right-skewed but low numbers of commits, we do not rescale the dependent variable prior to transformation. Estimates are generally sensitive to scaling and as there is no overarching guideline, scaling choice is described as a data fitting problem in the econometric literature (Aihounton and Henningsen, 2021). As rescaling typically leads to larger estimates our choice with respect to dependent variable scaling is conservative (Chen and Roth, 2023).[7] The effect size of the resulting coefficient estimates thus is not only statistically highly significant but also economically sizable as we estimate between 5 and 25% higher OSS activity of job movers compared to other movers in the job search period, depending on the month relative to move date.

Table 2: Difference-in-differences model

IHS(single commits)	(1)	(2)	(3)
Job mover × job search	0.3621***	0.2962***	0.1646***
	(0.0137)	(0.0144)	(0.0141)
Job mover × post move	-0.2608***	-0.2208***	-0.1036***
	(0.0189)	(0.0203)	(0.0190)
User FE	×	×	×
Month FE		×	×
Experience FE			×
Adjusted R^2	0.289	0.308	0.359
Observations	1,946,413	1,946,413	1,946,413
Users	22,896	22,896	22,896

Notes: Results from estimation of Equation 2. experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * p $>$ 0.01, ** p $>$ 0.05, and *** p $>$ 0.1. *Sources:* GHTorrent, own calculations.

The dynamic event study specification validated by theoretical and empirical evidence from the literature informs our definition of the job search period. We identify the period of distinctly elevated OSS activity

[7] We discuss model specification in more depth in Section 5.3.

in the 15 months prior to the month of moving as job search period. Using this definition of the job search period allows us to estimate the average treatment effect on the treated (ATT) per Equation 2. Table 2 provides the ATT estimates of our adapted three-period difference-in-differences specification. As expected, job movers OSS activity is elevated during the job search period relative to other movers and is lower in the post period. The inclusion of calendar month and experience fixed effects considerably improves model fit as described by adjusted R^2. The coefficient(s) of interest are attenuated as a result. Our preferred specification in Model (3) estimates that job movers contribute about 16.5% more on average in the job search period compared to other movers.

While the ATT effect size as such is suitable in assessing the importance of signaling incentives for individuals' OSS contributions during a job transition, we are further interested in the broader relevance of this motivation for the OSS community. Because our definition of the job search period is broad and includes periods with only moderately elevated signaling incentives, the ATT is best interpreted relative to the length of the job search period by performing a back-of-the-envelope calculation. Recent statistics state average job tenure in the US is around four years and only two years for software developers (Firaz, 2022). Assuming an average job tenure of three years applies to OSS developers, constant (base) activity levels across users and over time, and using our estimates ATT coefficient implies 6.8% of overall OSS activity is caused by signaling incentives during job transitions.[8] This suggests career concerns are a significant motivation for software developers and causes a sizable portion of contributions to OSS.

5.2 Heterogeneity

A natural question that arises from our main finding is whether there are systematic shifts in job movers' OSS activity during the job search period. This not only improves our understanding of how the signaling motive impacts users and activities differently but provides further validation of the signaling as the motive behind increased OSS activity. In particular, we explore two main dimensions of heterogeneity. First, we ask if job movers systematically focus their OSS activity during the job search period on certain types of projects, e.g., projects that are especially valuable as signal in the labor market. Second, we investigate if particular groups of job movers exhibit significant differences in effect size or if the effect size derives from all job movers equally.

We investigate effect heterogeneity with respect to the type of projects users contribute to during the job search period in Table 3. For this purpose, we use information on the main programming languages of projects and classify them into categories to distinguish broad project types. Our classification is documented in Table A.7 in the Appendix. This project-level approach requires using the number of contributions to each project type as outcome variable in user-level regressions. Thus, we run separate regressions of the

[8] Calculated as: $\hat{\beta}_2 * \frac{\#months_{JobSearch}}{\#months_{JobTenure}} = 16.46\% * \frac{15}{36}$.

Table 3: Heterogeneity by project type

IHS(single commits)	(1) low-level	(2) data eng.	(3) app dev.	(4) web dev.	(5) routine	(6) other
Job mover × job search	0.0136**	0.0426***	0.0256***	0.0607***	0.0277***	0.0353***
	(0.0061)	(0.0082)	(0.0051)	(0.0109)	(0.0073)	(0.0072)
Job mover × post move	-0.0047	-0.0177*	-0.0068	-0.0852***	-0.0145	0.0015
	(0.0077)	(0.0107)	(0.0077)	(0.0144)	(0.0098)	(0.0089)
User FE	×	×	×	×	×	×
Month FE	×	×	×	×	×	×
Experience FE	×	×	×	×	×	×
Adjusted R^2	0.26051	0.26955	0.29500	0.28444	0.28765	0.33629
Observations	1,946,413	1,946,413	1,946,413	1,946,413	1,946,413	1,946,413
Users	22,896	22,896	22,896	22,896	22,896	22,896

Notes: Results from estimation of Equation 2 with IHS-transformed number of commits to single-authored projects featuring main programming language of the respective class. Classification of programming languages according to Table A.7. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, own calculations.

model in Equation 2 for each project type. Results show significant differences in the ATT effects.[9] Notably, we obtain the largest effects for web development and data engineering projects. Low-level programming, program routine, and app development projects experience much smaller increases in the job search period. These results are consistent with, first, job movers focusing on web development because such projects are a way to showcase their work product and thus skill in existing works. Secondly, job movers might signal more through data engineering projects as skills related to such projects are especially valuable in the labor market.

To investigate the second channel in more detail, we classify programming languages directly by their valuation in the labor market as stated in the *StackOverflow* list of top-paying technologies.[10] Using the same method as above, we compare the ATT for programming languages listed as top-paying technologies compared to non-listed programming languages. Among top-paying programming languages, we further separate the top 30 best-paying from other listed programming languages. Which languages are in each category is shown by Table A.8 in the Appendix. According to survey evidence by *StackOverflow*, programming languages in the best-paying category are associated with about USD 16,500 higher total annual compensation compared to other listed languages, a 24% premium. Table 4 displays the estimation results. While job movers significantly increase OSS activity during the job search period in all groups, the increase is by far the largest for the best-paying programming languages. Compared to the increase in languages lower on the list, the increase in OSS activity in projects using best-paying programming languages is about

[9]Note that increased sparsity leads to a loss of quantitative comparability to the main results in favor of comparability between project-type regression estimates.

[10]The list is available at https://survey.stackoverflow.co/2023/#technology-top-paying-technologies. Last accessed on 11/03/2023.

twice as large. The effects in the other two categories are not statistically distinguishable. This provides further indication that job movers focus their signaling activity on projects requiring skills especially valuable in the labor market.

Table 4: Heterogeneity by labor market value

IHS(single commits)	listed		
	(1) top 30	(2) other	(3) not listed
Job mover × job search	0.0842***	0.0456***	0.0396***
	(0.0095)	(0.0104)	(0.0076)
Job mover × post move	-0.0181	-0.0703***	-0.0165*
	(0.0126)	(0.0132)	(0.0094)
User FE	×	×	×
Month FE	×	×	×
Experience FE	×	×	×
Adjusted R^2	0.23914	0.24635	0.27395
Observations	1,946,413	1,946,413	1,946,413
Users	22,896	22,896	22,896

Notes: Results from estimation of Equation 2 with IHS-transformed number of commits to single-authored projects featuring main programming language of the respective class. Classification of programming languages according to Table A.8. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * p > 0.01, ** p > 0.05, and *** p > 0.1. *Sources:* GHTorrent, own calculations.

As an alternative method to classify projects, we tap project descriptions and deploy a keyword-based NLP approach (Gentzkow et al., 2019). Only about one fourth of projects in our sample have descriptions and descriptions are typically brief. Therefore, we use a bag-of-words representation of all project descriptions and create a list of keywords associated with five project categories (education, data(base), website, code, and files) from analyzing the most frequently appearing words.[11] We then assign projects to a cluster when their description contains at least one associated keyword.[12] This approach naturally results in a smaller sample due to few project with description and strict requirements from the keyword list. Yet, using appropriate keywords is a targeted approach and increases the confidence in our classification. Estimating our baseline model for commits to the project types generated with this method yields similar results, reported in Table A.12. We obtain the largest effect for coding projects, followed by files and websites. These findings are generally in line with the programming language-based approach. Notably, we find no effect for educational projects, consistent with signaling rather than learning motives.

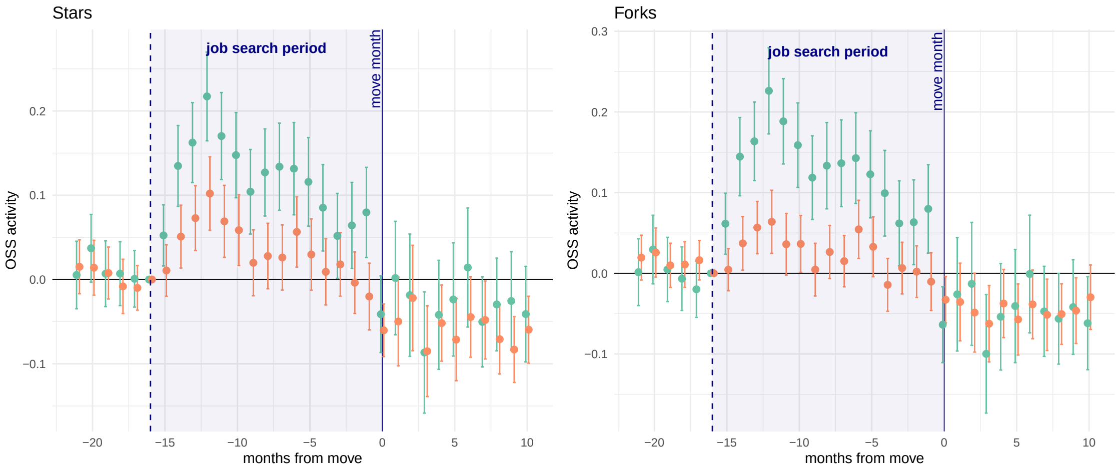
To distinguish whether career concerns induce job movers to start contributing to OSS, we formulate the model as a linear probability model (LPM) with an indicator for contribution rather than the number of con-

[11] The keywords are reported in Table A.9.

[12] As a result, projects may be assigned to multiple clusters.

tributions as recommended in Chen and Roth (2023). Estimation results are shown in Table A.10 and suggest a 7% higher probability of job movers contributing during the job search period relative to other movers. To investigate the extensive margin further, we run our baseline event study model using contributions to new projects, defined as projects initiated (i.e., first commit date) during the month under consideration and compare new single projects to new team projects. Results in Figure A.7 show that job movers especially start working on new single projects during the job search period. Together, these findings suggest the extensive margin plays a significant role, and job movers specifically engage in OSS activity that is unambiguously attributable to themselves, which is advantageous in order to signal personal ability.

Figure 5: Heterogeneity by community use-value



Notes: Estimates for $t_j \times \text{JobChanger}_i$ based on Equation 1 with user and calendar month fixed effects. The outcome is IHS-transformed commits to single-authored projects with (orange) or without (green) stars (left) or forks (right), respectively. The reference month is $t = -16$. Bars show 95% confidence intervals. Standard errors are clustered at the user level. *Sources:* GHTorrent, own calculations.

When thinking about the relevance of OSS contributions spurred by career concerns as a public good, quality is an important factor. On *GitHub*, projects may receive stars and can be forked by other users on the platform. Stars are a way for other users to indicate they find the project useful and to bookmark them for future reference. Forking refers to a process that copies a project into a new repository of the forking user so that she can use and alter the code in her own projects. Forking thus indicates other users' interest. We use both quality indicators and estimate the event study model, differentiating between OSS activity in projects with and without stars or forks, respectively. Figure 5 depicts the results and shows most OSS contributions of job movers during the job search period are in low-quality projects. This implies other users do not find signaling projects immediately useful. However, we found before that many signaling projects are websites that likely do not contain new code but rather showcase existing work more clearly. Such repositories are rarely starred or forked since usage is mostly off-platform. This might explain why the selected quality indicators suggest low quality and does not necessarily mean that projects are perceived as not valuable. Rather, the value could lie in making existing works more visible and accessible to the community. Nevertheless, these findings do suggest a lower direct use-value of signaling projects for the OSS community regarding the usefulness of code in other projects on the platform.

Table 5: International relocations

IHS(single commits)	international		upward moves	
	(1) international	(2) inter-continental	(3) income group	(4) GDP p. c.
Job mover × job search	0.1461***	0.1472***	0.1620***	0.1625***
	(0.0158)	(0.0150)	(0.0146)	(0.0144)
Job mover × job search × indicator	0.0619**	0.0923***	0.0295	0.0450
	(0.0260)	(0.0313)	(0.0393)	(0.0452)
Job mover × post move	-0.1040***	-0.1038***	-0.1038***	-0.1038***
	(0.0190)	(0.0190)	(0.0190)	(0.0190)
User FE	×	×	×	×
Month FE	×	×	×	×
Experience FE	×	×	×	×
Adjusted R^2	0.35948	0.35949	0.35945	0.35945
Observations	1,946,413	1,946,413	1,946,413	1,946,413
Users	22,896	22,896	22,896	22,896

Notes: Results from estimation of Equation 2 adding a triple interaction which features an indicator variable to separate heterogeneous effects of interest. Upward income group moves are defined as moves from developing to developed countries. Upward moves in GDP per capita are based on current 2021 PPP USD. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, World Development Indicators, own calculations.

Labor market signaling via OSS activity might be valuable to a different extent for job movers. We, therefore, investigate whether the effect is broad-based among all users or driven by a group of users with a particularly large increase in OSS activity during the job search period. For this purpose, we first explore heterogeneity with respect to followers comparing quartiles and find no significant differences (cf. Figure A.6). Second, we investigate whether signaling activity differs for users moving internationally by interacting dummy variables for types of moves to our baseline model. The results are reported in Table 5. Model (1) indicates that users moving internationally engage in 42% more labor market signaling via OSS compared to domestic movers. Likewise, inter-continental job movers signal even more and feature a 63% higher effect compared to non-intercontinental movers as shown by Model (2). Models (3) and (4) suggest that the effect differences are especially driven by international movers relocating to higher-income countries, though the coefficients lack statistical significance. These results are in line with existing evidence (e.g., Agrawal et al., 2016; Hann et al., 2013) suggesting that OSS signals could substitute formal certification, which is less transferrable and accepted internationally, particularly for developing countries.

Table 6 shows that there is some heterogeneity in signaling activity depending on users' origin (old) and destination (new) affiliation. Importantly, users who obtain new jobs at big tech firms do not engage in labor

Table 6: Heterogeneity by affiliation

IHS(single commits)	destination			origin	
	(1) median	(2) big tech	(3) academia	(4) median	(5) academia
Job mover × job search	0.1784***	0.1753***	0.1578***	0.1631***	0.1601***
	(0.0198)	(0.0144)	(0.0145)	(0.0142)	(0.0502)
Job mover × job search × indicator	-0.0219	-0.1460***	0.0930**	0.0843	-0.0114
	(0.0234)	(0.0480)	(0.0457)	(0.0999)	(0.0652)
Job mover × post move	-0.1038***	-0.1042***	-0.1032***	-0.1040***	-0.1693***
	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0528)
User FE	×	×	×	×	×
Month FE	×	×	×	×	×
Experience FE	×	×	×	×	×
Adjusted R^2	0.35946	0.35950	0.35947	0.35946	0.36126
Observations	1,946,413	1,946,413	1,946,413	1,946,413	1,406,169
Users	22,896	22,896	22,896	22,896	22,896

Notes: Results from estimation of Equation 2 adding a triple interaction which features an indicator variable to separate heterogeneous effects of interest. Median split refers to median size of affiliation in terms of users in the full *GHTorrent* sample. Big tech refers to Google, Amazon, Meta, Apple and Microsoft. Academia refers to students and university affiliations. Specifically, users stating *university, college, institute, universiteit, universidad, universität* or *student* in their affiliation are assigned to academia. Destination (origin) refers to users' affiliation before (after) the affiliation change. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, own calculations.

market signaling through OSS activity to a significant extent. In contrast, users changing jobs to academic affiliations signal significantly more. There is no statistically significant difference in signaling activity depending on the old affiliation, but an economically significant point estimate for above-median firm size points towards more signaling activity by users coming from larger firms. These results, though weak, are consistent with an arguably generally greater role of open source in academia while large corporations like the big tech firms emphasize proprietary software more, and users qualified for a job at the big tech firms typically do not need (additional) ability signals from OSS activity as they tend to have the highest credentials anyways.

5.3 Robustness

We choose a model that uses the inverse hyperbolic sine (IHS) transformation of the outcome variable as the preferred specification, which has the mentioned advantages of retaining zeros while approximating the logarithmic transformation (see, e.g., Bellemare and Wichman, 2020; MacKinnon and Magee, 1990; Burbidge et al., 1988). A related and widely-used transformation is the logarithmic transformation and specifically $\log(y+1)$ (Bellégo et al., 2022). The challenge with these transformations is that they are scale-dependent, but this problem is more severe for high-valued and sometimes-zero outcomes (Mullahy and Norton, 2022; Chen and Roth, 2023). Aihounton and Henningsen (2021) frame scaling as a data fitting exercise. Since our data is low-valued and sparse, we opt for a conservative quantitative interpretation arising from IHS transformation of the unscaled dependent variable. Another class of alternative models are Poisson models such as the PPML estimator. These models are the established go-to choice in trade (Larch et al., 2019) and other applications with high-valued count data featuring zeros such as investment, profit, or revenue data (Cohn et al., 2022). However, these models perform poorly in practice on low-valued sparse panel data such as ours and there is no standard econometric approach yet. Additionally, our data features sparsity not only across units but also within. For such applications, IHS or logarithmic transformations are the preferred choice in practice, e.g. in Xu et al. (2020) or Bahar et al. (2022).

Apart from being conservative in our preferred model specification, we assess the robustness of our results by estimating several alternative models. Results are reported in Table A.10 in the Appendix. First, we show that the most widely-used alternative way to transform the dependent variable in similar applications (e.g., Xu et al., 2020), a logarithmic transformation, yields similar coefficient estimates. Second, we run two types of frequently used count data models: a negative binomial and a Poisson fixed effects model. Both models are known to frequently exhibit performance issues with fixed effects and convergence issues (Bellégo et al., 2022; Correia et al., 2019). The PPML model results in similar coefficient estimates for the job search period and an increased estimate for the post-period. The negative binomial model estimates are significantly inflated by a factor of three to four compared to our preferred specification. These findings indicate the robustness of our results with respect to model specification and confirm that our estimated effect size is conservative. Furthermore, we follow state-of-the-art best practices (Chen et al., 2022) in that we explicitly consider intensive and extensive margin effects. The formulation of our model as LPM suggests reasonably high importance of the extensive margin (see Model (3) in Table A.10). Note that through our sample selection of active OSS contributors only, extensive margin effects are likely downward biased. At the same time, this implicit conditioning decreases potential bias of the intensive margin in our main specification (Hersche and Moor, 2020).

Measurement error in the move date possibly introduces bias in our estimates due to observing location data only every six months and users entering their new location after relocation. The event study results in Figure 4 partly alleviate this concern as there is a discontinuous drop in OSS activity of job movers at the proxied move date. Nevertheless, it is unclear whether the downward trend during the job search period is due to already-moved job movers still in the treatment group or, e.g., due to decreased signaling incentives of users who already found a job. We address this by varying the job search period definition and separately estimating a coefficient for the period for which we are unsure if the user actually already moved. This adjustment generally increases the estimated effect by up to three percentage points to about 19.5%. Note that although this introduces upward bias in our estimates it simultaneously alters the length of the job search period and, as a result, leads to a mechanic downward adjustment in the interpretation when thinking about overall OSS activity attributable to career concerns.

Our approach exploits the specific timing of elevated career concerns during the job search period. Still, coinciding increases in other motives are a potential concern. Specifically, if people disproportionately learn new skills in between jobs and this activity is conducted in public repositories on *GitHub*, our model would wrongly attribute such activity to career concerns. One of our project types in the keyword-based classification are educational projects. This category captures repositories associated with coursework, assignments, or online education (e.g., *Coursera*). Table A.12 shows no effect on the activity in educational projects, suggesting that activity driven by learning motives does not drive our effect. In addition, we investigate projects not owned by the mover, such as company projects, or projects consisting of initial forks (a copy of existing repositories). We find no evidence for a significant relevance of these channels (see Table A.11).[13]

For completeness, we report estimation results for the event study specification in Table A.13 and, similarly as in Table 2 for the ATT, show the results for the models without experience and calendar month fixed effects, as well. Figure A.8 plots event study coefficients for variations of the baseline model. Further, we establish the robustness of our results to alternative sample definitions with respect to geocoding and job changes in Models (3) and (4) of Table A.10. For user-level heterogeneity analyses using interaction terms, alternative model specifications based on separate regressions with redefined outcome variables similar to the project-derived heterogeneity analyses (Tables A.15, A.16, and A.17) show qualitatively similar results.

6 Conclusion

We show private career concerns of software developers induce significant contributions to open-source software, a public good. By exploiting temporal variation in signaling incentives in a quasi-experimental design, we establish a causal increase of OSS activity of job movers compared to users relocating for other reasons in the job search period by about 16%. These positive externalities of labor market signaling are sizable from both the individual and the community perspective but often neglected in existing works that predominantly emphasize other motives to contribute to OSS development. A broad base of users on the largest online code repository platform, *GitHub*, engages in labor market signaling during the job search period and signaling opportunity even attracts first-time contributors. OSS activity driven by signaling motives is disproportionately directed to projects that increase external visibility of existing works or are written in programming languages highly valued in the labor market. At the same time, signaling projects are starred and forked less by other users on the platform. This suggests OSS activity induced by career concerns is targeted less to the direct use-value of the OSS community and more to their value as a labor market signal.

Our study has limitations. Data does not contain information on users besides activity on the platform, location, and affiliation and cannot be linked to other data on the individual level, which constrains the number

[13] Note that project ownership is prone to measurement error, as it might wrongly capture the same individual as distinct persons, e.g., when committing to projects using two different e-mail addresses as identification or using multiple devices. Thus, it is not surprising that there is a small significant effect for non-own projects in Table A.11.

of possible heterogeneity analyses. Furthermore, location and affiliation changes are only observed at snapshot frequency, i.e., roughly every six months. This leads to blurriness in the proxied move (and affiliation) change months and likely biases our estimates downwards. In general, we opt for a conservative model specification as a quantitative interpretation of our effect size depends on econometric choices regarding model class and outcome scaling and transformation. It should also be noted that although our empirical strategy identifies the causal effect of temporarily elevated signaling incentives under reasonable assumptions, it by no means captures all OSS activity attributable to labor market signaling and therefore should be interpreted as a lower bound estimate. Similarly beyond the scope of this work is to assess the extent to which OSS signals improve individual-level labor market outcomes.

Despite these limitations, our findings have several managerial implications. Importantly, decision-makers aiming to increase OSS activity should take into account career concerns as a significant motivating factor for developers. Platform design addressing the signaling needs of users explicitly might grow the platform at both the intensive (activity) and the extensive (users) margin. Measures that foster public visibility, transparency as well as accessibility for non-experts might contribute to this goal, e.g., through easily understandable activity metrics, skill badges, or lists of spoken programming languages on user profiles. At the same time, platform managers should be aware that signaling motives might steer OSS activity towards projects with lower direct use-value for the community whenever there is a gap between signaling value and community value of projects. For hiring managers, our results emphasize that OSS is a commonplace and potentially valuable signal of skill for developer talent. Consequently, it should receive attention in employee search and assessment.

Finally, our study provides several insights for public policy. In general, the positive externalities of career concerns on public good creation merit attention due to likely significant positive spillovers of OSS on the private sector and innovation. Innovation policy that enables and encourages publicly funded software development to be hosted and shared on online open-source platforms may increase the motivation of the funded developer teams while at the same time generating OSS, a public good that potentially spurs further innovative activity. With respect to labor market and educational policy, our results point to the continued shift away from (public) skill certification in occupations related to software development and emphasize a greater role of more fluid and practical skill signals directly showcasing work product. Educational institutions should acknowledge both the labor market value of OSS activity for their students and the positive societal externalities from such activity and consider encouraging students to engage in OSS development or even explicitly integrate OSS projects into curricula.

References

Abreu, Maria, Alessandra Faggian, and Philip McCann, "Migration and Inter-industry Mobility of UK Graduates," *Journal of Economic Geography*, 2015, *15* (2), 353–385.

Acikalin, Utku U., Tolga Caskurlu, Gerard Hoberg, and Gordon M. Phillips, "Intellectual Property Protection Lost and Competition: An Examination Using Machine Learning," *Working Paper*, 2022.

Adrian, Favell, Feldblum Miriam, and Peter Smith Michael, *The Human Face of Global Mobility: A Research Agenda*, Routledge, 2017.

Agrawal, Ajay, Nicola Lacetera, and Elizabeth Lyons, "Does Standardized Information in Online Markets Disproportionately Benefit Job Applicants from Less Developed Countries?," *Journal of International Economics*, 2016, *103*, 1–12.

Aihounton, Ghislain and Arne Henningsen, "Units of Measurement and the Inverse Hyperbolic Sine Transformation," *The Econometrics Journal*, 2021, *24* (2), 334–351.

Amior, Michael, "Why Are Higher Skilled Workers More Mobile Geographically?: The Role of the Job Surplus," *CEPR Working Paper*, 2015.

_ , "Education and Geographical Mobility: The Role of the Job Surplus," *CEPR Working Paper*, 2019.

Athey, Susan and Glenn Ellison, "Dynamics of Open Source Movements," *Journal of Economics & Management Strategy*, 2014, *23* (2), 294–316.

Bahar, Dany, Prithwiraj Choudhury, Do Yoon Kim, and Wesley W. Koo, "Innovation on Wings: Nonstop Flights and Firm Innovation in the Global Context," *Management Science*, 2022.

Balgova, Maria, "Leaping Into the Unknown? The Role of Job Search in Migration Decisions," *Working Paper*, 2020.

Bellégo, Christophe, David Benatia, and Louis Pape, "Dealing With Logs and Zeros in Regression Models," *Working Paper*, 2022.

Bellemare, Marc F. and Casey J. Wichman, "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 50–61.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian, "On the Private Provision of Public Goods," *Journal of Public Economics*, 1986, *29* (1), 25–49.

Bilodeau, Marc and Al Slivinski, "Toilet Cleaning and Department Chairing: Volunteering a Public Service," *Journal of Public Economics*, 1996, *59* (2), 299–308.

Bitzer, Jürgen and Ingo Geishecker, "Who Contributes Voluntarily to OSS? An Investigation Among German IT Employees," *Research Policy*, 2010, *39* (1), 165–172.

_ **and Philipp J. H. Schröder**, "Open Source Software, Competition and Innovation," *Industry and Innovation*, 2007, *14* (5), 461–476.

— **and Philipp JH Schröder**, "Bug-fixing and Code-writing: The Private Provision of Open Source Software," *Information Economics and Policy*, 2005, *17* (3), 389–406.

— **, Wolfram Schrettl, and Philipp JH Schröder**, "Intrinsic Motivation in Open Source Software Development," *Journal of Comparative Economics*, 2007, *35* (1), 160–169.

Blau, Gary, "Testing a Two-Dimensional Measure of Job Search Behavior," *Organizational Behavior and Human Decision Processes*, 1994, *59* (2), 288–312.

Bliss, Christopher and Barry Nalebuff, "Dragon-slaying and Ballroom Dancing: The Private Supply of a Public Good," *Journal of Public Economics*, 1984, *25* (1-2), 1–12.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-level Evidence," *The Quarterly Journal of Economics*, 2002, *117* (1), 339–376.

Burbidge, John B., Lonnie Magee, and A. Leslie Robb, "Alternative Transformations to Handle Extreme Values of the Dependent Variable," *Journal of the American Statistical Association*, 1988, *83* (401), 123–127.

Butler, Simon, Jonas Gamalielsson, Björn Lundell, Christoffer Brax, Johan Sjöberg, Anders Mattsson, Tomas Gustavsson, Jonas Feist, and Erik Lönroth, "On Company Contributions to Community Open Source Software Projects," *IEEE Transactions on Software Engineering*, 2019, *47* (7), 1381–1401.

Chamberlain, Andrew, "Why is Hiring Taking Longer," *Glassdoor Policy Brief*, 2015.

Chamberlin, John, "Provision of Collective Goods as a Function of Group Size," *American Political Science Review*, 1974, *68* (2), 707–716.

Chen, Chinchih, Carl Benedikt Frey, and Giorgio Presidente, "Disrupting Science," *Working Paper*, 2022.

Chen, Jiafeng and Jonathan Roth, "Logs with Zeros? Some Problems and Solutions," *Working Paper*, 2023.

Chevalier, Judith and Glenn Ellison, "Career Concerns of Mutual Fund Managers," *The Quarterly Journal of Economics*, 1999, *114* (2), 389–432.

Ciriaci, Daria, "Does University Quality Influence the Interregional Mobility of Students and Graduates? The Case of Italy," *Regional Studies*, 2014, *48* (10), 1592–1608.

Cohn, Jonathan B., Zack Liu, and Malcolm I. Wardlaw, "Count (and Count-like) Data in Finance," *Journal of Financial Economics*, 2022, *146* (2), 529–551.

Colombo, Massimo G., Evila Piva, and Cristina Rossi-Lamastra, "Open Innovation and Within-industry Diversification in Small and Medium Enterprises: The Case of Open Source Software Firms," *Research Policy*, 2014, *43* (5), 891–902.

Conti, Annamaria, Christian Peukert, and Maria P. Roche, "Beefing IT Up for Your Investor? Open Sourcing and Startup Funding: Evidence from GitHub," *Working Paper*, 2021.

_ , Vansh Gupta, Jorge Guzman, and Maria P. Roche, "Incentivizing Innovation in Open Source: Evidence from the GitHub Sponsor Program," *Working Paper*, 2023.

Correia, S., P. Guimarães, and T. Zylkin, "ppmlhdfe: Fast Poisson Estimation with High-dimensional Fixed Effects," *Working Paper*, 2019.

del Rio-Chanona, Maria, Nadzeya Laurentsyeva, and Johannes Wachs, "Are Large Language Models a Threat to Digital Public Goods? Evidence from Activity on Stack Overflow," *arXiv preprint*, 2023.

Deming, David J. and Kadeem Noray, "Earnings Dynamics, Changing Job Skills, and STEM Careers," *The Quarterly Journal of Economics*, 2020, *135* (4), 1965–2005.

Draca, Mirko, Raffaella Sadun, and John Van Reenen, "ICT and Productivity: A Review of the Evidence," *Handbook of Information and Communication Technologies*, 2007.

D'Mello, Marisa and Sundeep Sahay, ""I am kind of a nomad where I have to go places and places"… Understanding Mobility, Place and Identity in Global Software Work from India," *Information and Organization*, 2007, *17* (3), 162–192.

Falck, Oliver, Alexandra Heimisch-Roecker, and Simon Wiederhold, "Returns to ICT Skills," *Research Policy*, 2021, *50* (7), 104064.

Fershtman, Chaim and Neil Gandal, "The Determinants of Output per Contributor in Open Source Projects: An Empirical Examination," *CEPR Working Paper*, 2004.

Firaz, Abbasi, "How Long Do Software Engineers Stay at a Job?," *LinkedIn Blog*, 2022.

Fuller, Joseph, Christina Langer, and Matt Sigelman, "Skills-based Hiring is on the Rise," *Harvard Business Review*, 2022, *11.*

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as Data," *Journal of Economic Literature*, 2019, *57* (3), 535–574.

Gerosa, Marco, Igor Wiese, Bianca Trinkenreich, Georg Link, Gregorio Robles, Christoph Treude, Igor Steinmacher, and Anita Sarma, "The Shifting Sands of Motivation: Revisiting What Drives Contributors in Open Source," *IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, 2021, pp. 1046–1058.

GitHub, "The 2021 State of the Octoverse," *Company Report*, 2021.

Goldbeck, Moritz, "Bit by Bit: Colocation and the Death of Distance in Software Developer Networks," *Working Paper*, 2023.

Gousios, Georgios, "The GHTorent Dataset and Tool Suite," in "IEEE 10th Working Conference on Mining Software Repositories (MSR)" 2013, pp. 233–236.

Greenwood, Michael J., "The Geographic Mobility of College Graduates," *The Journal of Human Resources*, 1973, *8* (4), 506–515.

_ , "Research on Internal Migration in the United States: A Survey," *Journal of Economic Literature*, 1975, pp. 397–433.

Haapanen, Mika and Hannu Tervo, "Migration of the Highly Educated: Evidence from Residence Spells of University Graduates," *Journal of Regional Science*, 2012, *52* (4), 587–605.

Hann, Il-Horn, Jeff Roberts, and Sandra Slaughter, "Why Developers Participate in Open Source Software Projects: An Empirical Investigation," *Proceedings of the International Conference on Information Systems*, 2004.

_ , Jeffrey A. Roberts, and Sandra A. Slaughter, "All Are Not Equal: An Examination of the Economic Returns to Different Forms of Participation in Open Source Software Communities," *Information Systems Research*, 2013, *24* (3), 520–538.

Hars, Alexander and Shaosong Ou, "Working for Free? Motivations for Participating in Open-source Projects," *International Journal of Electronic Commerce*, 2002, *6* (3), 25–39.

Haussen, Tina and Silke Uebelmesser, "Job Changes and Interregional Migration of Graduates," *Regional Studies*, 2018, *52* (10), 1346–1359.

Hendricks, Ken, Andrew Weiss, and Charles Wilson, "The War of Attrition in Continuous Time with Complete Information," *International Economic Review*, 1988, pp. 663–680.

Hersche, Markus and Elias Moor, "Identification and Estimation of Intensive Margin Effects by Difference-in-difference Methods," *Journal of Causal Inference*, 2020, *8* (1), 272–285.

Hertel, Guido, Sven Niedner, and Stefanie Herrmann, "Motivation of Software Developers in Open Source Projects: An Internet-based Survey of Contributors to the Linux Kernel," *Research Policy*, 2003, *32* (7), 1159–1177.

Holmström, Bengt, "Managerial Incentive Problems: A Dynamic Perspective," *The Review of Economic Studies*, 1999, *66* (1), 169–182.

Huang, Peng and Zhongju Zhang, "Participation in Open Knowledge Communities and Job-Hopping," *MIS Quarterly*, 2016, *40* (3), 785–806.

Johnson, Justin Pappas, "Open Source Software: Private Provision of a Public Good," *Journal of Economics & Management Strategy*, 2002, *11* (4), 637–662.

Kodrzycki, Yolanda K., "Migration of Recent College Graduates: Evidence from the National Longitudinal Survey of Youth," *New England Economic Review*, 2001, pp. 13–34.

Krishnamurthy, Sandeep, "On the Intrinsic and Extrinsic Motivation of Free/Libre/Open Source (FLOSS) Developers," *Knowledge, Technology & Policy*, 2006, *18* (4), 17–39.

_ , Shaosong Ou, and Arvind K. Tripathi, "Acceptance of Monetary Rewards in Open Source Software Development," *Research Policy*, 2014, *43* (4), 632–644.

Krogh, Georg Von, Stefan Haefliger, Sebastian Spaeth, and Martin W. Wallin, "Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development," *MIS Quarterly*, 2012, pp. 649–676.

Lakhani, Karim R. and Robert G. Wolf, "Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects," *Open Source Software Projects*, 2003.

Larch, Mario, Joschka Wanner, Yoto V. Yotov, and Thomas Zylkin, "Currency Unions and Trade: A PPML Re-assessment with High-dimensional Fixed Effects," *Oxford Bulletin of Economics and Statistics*, 2019, *81* (3), 487–510.

Lee, Gwendolyn K. and Robert E. Cole, "From a Firm-based to a Community-based Model of Knowledge Creation: The Case of the Linux Kernel Development," *Organization Science*, 2003, *14* (6), 633–649.

Leppämäki, Mikko and Mikko Mustonen, "Skill Signalling with Product Market Externality," *The Economic Journal*, 2009, *119* (539), 1130–1142.

Lerner, Josh and Jean Tirole, "Some Simple Economics of Open Source," *The Journal of Industrial Economics*, 2002, *50* (2), 197–234.

_ and _ , "The Economics of Technology Sharing: Open Source and Beyond," *Journal of Economic Perspectives*, 2005, *19* (2), 99–120.

_ and _ , "The Scope of Open Source Licensing," *Journal of Law, Economics, and Organization*, 2005, *21* (1), 20–56.

Lifshitz-Assaf, Hila and Frank Nagle, "The Digital Economy Runs on Open Source. Here's How to Protect It," *Harvard Business Review*, 2021.

Long, Ju, "Open Source Software Development Experiences on the Students' Resumes: Do They Count?," *Journal of Information Technology Education: Research*, 2009, *8* (1), 229–242.

Luca, Michael, "User-Generated Content and Social Media," in "Handbook of Media Economics," Vol. 1, Elsevier, 2015, pp. 563–592.

Machin, Stephen, Kjell G. Salvanes, and Panu Pelkonen, "Education and Mobility," *Journal of the European Economic Association*, 2012, *10* (2), 417–450.

MacKinnon, James G. and Lonnie Magee, "Transforming the Dependent Variable in Regression Models," *International Economic Review*, 1990, pp. 315–339.

Marlow, Jennifer and Laura Dabbish, "Activity Traces and Signals in Software Developer Recruitment and Hiring," *Proceedings of the Conference on Computer Supported Cooperative Work*, 2013, pp. 145–156.

Miklós-Thal, Jeanine and Hannes Ullrich, "Belief Precision and Effort Incentives in Promotion Contests," *The Economic Journal*, 2015, *125* (589), 1952–1963.

_ and _ , "Career Prospects and Effort Incentives: Evidence from Professional Soccer," *Management Science*, 2016, *62* (6), 1645–1667.

Mullahy, John and Edward C. Norton, "Why Transform Y? A Critical Assessment of Dependent-variable Transformations in Regression Models for Skewed and Sometimes-zero Outcomes," *NBER Working Paper*, 2022.

Myatt, David P. and Chris Wallace, "Equilibrium Selection and Public-good Provision: The Development of Open-source Software," *Oxford Review of Economic Policy*, 2002, *18* (4), 446–461.

Nagle, Frank, "Learning by Contributing: Gaining Competitive Advantage Through Contribution to Crowdsourced Public Goods," *Organization Science*, 2018, *29* (4), 569–587.

_ , "Open-Source Software and Firm Productivity," *Management Science*, 2019, *65* (3), 1191–1215.

_ , "Strengthening Digital Infrastructure: A Policy Agenda for Free and Open Source Software," *Brookings Institution Policy Brief*, 2022.

_ , David A. Wheeler, H. Lifshitz-Assaf, H. Ham, and J. Hoffman, "Report on the 2020 FOSS Contributor Survey," *The Linux Foundation Core Infrastructure Initiative*, 2020.

Orman, Wafa Hakim, "Giving it Away for Free? The Nature of Job-market Signaling by Open-source Software Developers," *The BE Journal of Economic Analysis & Policy*, 2008, *8* (1).

Osterloh, Margit and Sandra Rota, "Open Source Software Development: Just Another Case of Collective Invention?," *Research Policy*, 2007, *36* (2), 157–171.

O'Neil, Mathieu, Laure Muselli, Xiaolan Cai, and Stefano Zacchiroli, "Co-producing Industrial Public Goods on GitHub: Selective Firm Cooperation, Volunteer-employee Labour and Participation Inequality," *New Media & Society*, 2022.

Palfrey, Thomas R. and Howard Rosenthal, "Participation and the Provision of Discrete Public Goods: A Strategic Analysis," *Journal of Public Economics*, 1984, *24* (2), 171–193.

Pallais, Amanda, "Inefficient Hiring in Entry-level Labor Markets," *American Economic Review*, 2014, *104* (11), 3565–3599.

Piopiunik, Marc, Guido Schwerdt, Lisa Simon, and Ludger Woessmann, "Skills, Signals, and Employability: An Experimental Investigation," *European Economic Review*, 2020, *123*, 103374.

Puranam, Phanish, Oliver Alexy, and Markus Reitzig, "What's "new" About New Forms of Organizing?," *Academy of Management Review*, 2014, *39* (2), 162–180.

Raveendran, Marlo, Phanish Puranam, and Massimo Warglien, "Division of Labor Through Self-selection," *Organization Science*, 2022, *33* (2), 810–830.

Roberts, Jeffrey A., Il-Horn Hann, and Sandra A. Slaughter, "Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects," *Management Science*, 2006, *52* (7), 984–999.

Shah, Sonali K., "Motivation, Governance, and the Viability of Hybrid Forms in Open Source Software Development," *Management Science*, 2006, *52* (7), 1000–1014.

Smirnova, Inna, Markus Reitzig, and Oliver Alexy, "What Makes the Right OSS Contributor Tick? Treatments to Motivate High-skilled Developers," *Research Policy*, 2022, *51* (1), 104368.

Solimano, Andrés, *The International Mobility of Talent and Its Impact on Global Development: An Overview*, ECLAC, 2006.

Spence, Michael, "Job Market Signaling," *The Quarterly Journal of Economics*, 1973, pp. 355–374.

Startlin, "History of GitHub," 2016.

Stewart, Katherine J. and Sanjay Gosain, "The Impact of Ideology on Effectiveness in Open Source Software Development Teams," *MIS Quarterly*, 2006, pp. 291–314.

Surakka, Sami, "What Subjects and Skills are Important for Software Developers?," *Communications of the ACM*, 2007, *50* (1), 73–78.

Synopsys, "Open Source Security and Risk Analysis Report 2023," *Report*, 2023.

Tubino, Laura, Andrew Cain, Jean-Guy Schneider, Dhananjay Thiruvady, and Niroshinie Fernando, "Authentic Individual Assessment for Team-based Software Engineering Projects," in "Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering Education and Training" 2020, pp. 71–81.

Venhorst, Viktor, Jouke Van Dijk, and Leo Van Wissen, "An Analysis of Trends in Spatial Mobility of Dutch Graduates," *Spatial Economic Analysis*, 2011, *6* (1), 57–82.

Vidoni, Melina, "A Systematic Process for Mining Software Repositories: Results from a Systematic Literature Review," *Information and Software Technology*, 2022, *144*, 106791.

von Krogh, Georg, Sebastian Spaeth, and Karim R. Lakhani, "Community, Joining, and Specialization in Open Source Software Innovation: A Case Study," *Research policy*, 2003, *32* (7), 1217–1241.

von Proff, Sidonia, Matthias Duschl, and Thomas Brenner, "Motives Behind the Mobility of University Graduates: A Study of Three German Universities," *Review of Regional Research*, 2017, *37*, 39–58.

Wachs, Johannes, Mariusz Nitecki, William Schueller, and Axel Polleres, "The Geography of Open-Source Software: Evidence from GitHub," *Technological Forecasting and Social Change*, 2022, *176*, 121478.

Wagner, Stefan and Melanie Ruhe, "A Systematic Review of Productivity Factors in Software Development," *arXiv preprint arXiv:1801.06475*, 2018.

Wang, Yaxin, Liang Wang, Hao Hu, Jing Jiang, Hongyu Kuang, and Xianping Tao, "The Influence of Sponsorship on Open-Source Software Developers' Activities on GitHub," *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2022, pp. 924–933.

Wen, Wen, Marco Ceccagnoli, and Chris Forman, "Opening Up Intellectual Property Strategy: Implications for Open Source Software Entry by Start-up Firms," *Management Science*, 2016, *62* (9), 2668–2691.

Wright, Nataliya Langburd, Frank Nagle, and Shane Greenstein, "Open Source Software and Global Entrepreneurship," *Research Policy*, 2023, *52* (9), 104846.

Xu, Lei, Tingting Nian, and Luis Cabral, "What Makes Geeks Tick? A Study of Stack Overflow Careers," *Management Science*, 2020, *66* (2), 587–604.

Zeitlyn, David, "Gift Economies in the Development of Open Source Software: Anthropological Reflections," *Research Policy*, 2003, *32* (7), 1287–1291.

A Appendix

A.1 Tables

Table A.1: Sample selection

Median	All users	Movers	Δ
Activity			
Commits	6.00	170.00	164.00
commits single projects	*2.00*	*73.00*	*71.00*
commits team projects	*1.00*	*65.00*	*64.00*
Experience	34.00	39.00	5.00
Collaboration			
Projects	2.00	15.00	13.00
single projects	*2.00*	*9.00*	*7.00*
team projects	*2.00*	*5.00*	*3.00*
Quality			
Followers	0.00	5.00	5.00
Stars	0.00	1.30	1.30
stars single projects	*0.00*	*0.10*	*0.10*
Forks	0.00	0.76	0.76
forks single projects	*0.00*	*0.00*	*0.00*

Notes: Experience is measured as tenure on the platform in months since the first commit at the move date. Column Δ reports the absolute difference in median between movers in our sample and all users in the ten *GHTorrent* snapshots we utilize (N = 28,802,543). Column %Δ sets this difference in relation to other movers' median. *Sources:* GHTorrent, own calculations.

Table A.2: Affiliation and job transitions

Affiliation	all movers	job movers	other movers	Δ
Largest 100 firms	28.9 %	28.9 %	27.2 %	+1.7 p.p.
Big tech	*7.2 %*	*7.3 %*	*4.9 %*	*+2.4 p.p.*
Academic	8.9 %	9.0 %	6.3 %	+2.7 p.p.
Other	55.1 %	54.8 %	61.6 %	-6.8 p.p.
Job transitions	anytime	origin	destination	Δ
Largest 100 firms	28.9 %	20.3 %	26.8 %	+6.5 p.p.
Big tech	*7.2 %*	*2.0 %*	*7.1 %*	*+5.1 p.p.*
Academic	8.9 %	9.1 %	7.2 %	-2.0 p.p.
Other	55.1 %	68.6 %	58.9 %	-9.6 p.p.

Notes: Table reports affiliations and job transitions by organization type in shares of the respective sample. Column Δ reports the percentage point difference between job and other movers. *Sources:* GHTorrent, own calculations.

Table A.3: Top origin and destination cities

Origin	**Users**	**Share**	**Destination**	**Users**	**Share**
New York, USA	650	2.84 %	San Francisco, USA	1,307	5.71 %
San Francisco, USA	618	2.70 %	New York, USA	936	4.09 %
London, UK	421	1.84 %	London, UK	763	3.33 %
Bangalore, India	325	1.42 %	Seattle, USA	708	3.09 %
Chicago, USA	311	1.36 %	Bangalore, India	559	2.44 %
Boston, USA	305	1.33 %	Los Angeles, USA	379	1.66 %
Los Angeles, USA	305	1.33 %	Austin, USA	345	1.51 %
Moscow, Russia	305	1.33 %	Toronto, Canada	331	1.45 %
Seattle, USA	273	1.19 %	Chicago, USA	318	1.39 %
Paris, France	247	1.08 %	Boston, USA	315	1.38 %
Cumulative share		15.09 %	Cumulative share		26.05 %

Notes: Table reports the ten largest origin and destination cities in terms of the number of users in our sample. *Sources:* GHTorrent, own calculations.

Table A.4: Domestic moves

Country	Users	Share	
		all	domestic
United States	10,348	45.20 %	63.49 %
India	1,219	5.32 %	7.48 %
United Kingdom	638	2.79 %	3.91 %
Canada	620	2.71 %	3.80 %
China	522	2.28 %	3.20 %
France	436	1.90 %	2.68 %
Germany	417	1.82 %	2.56 %
Russia	375	1.64 %	2.30 %
Poland	195	0.85 %	1.20 %
Australia	194	0.85 %	1.19 %
		65.36 %	91.81 %

Notes: Table reports the ten largest countries in terms of the number of domestic movers in our sample. Shares reported in the third and fourth columns refer to all and to domestic movers, respectively. *Sources:* GHTorrent, own calculations.

Table A.5: Top origin and destination countries

International movers					
Origin	*Users*	*Share*	*Destination*	*Users*	*Share*
United States	1,831	0.28	United States	2,011	0.30
India	817	0.12	United Kingdom	774	0.12
United Kingdom	491	0.07	Canada	506	0.08
Russia	386	0.06	Germany	319	0.05
Canada	384	0.06	Russia	306	0.05
France	267	0.04	Netherlands	290	0.04
Australia	186	0.03	Australia	240	0.04
Italy	165	0.03	Poland	228	0.03
Brazil	163	0.02	France	182	0.03
Germany	151	0.02	Brazil	169	0.03
Inter-continental movers					
Origin	*Users*	*Share*	*Destination*	*Users*	*Share*
United States	1,453	0.34	United States	1,583	0.37
India	793	0.18	United Kingdom	428	0.10
United Kingdom	284	0.07	Russia	287	0.07
Russia	203	0.05	Canada	275	0.06
Australia	180	0.04	Australia	229	0.05
France	144	0.03	Germany	177	0.04
China	130	0.03	Poland	159	0.04
Canada	105	0.02	France	116	0.03
Italy	72	0.02	Netherlands	111	0.03
Poland	72	0.02	Italy	96	0.02

Notes: Table reports the ten largest origin and destination countries in terms of the number of international and inter-continental movers in our sample. *Sources:* GHTorrent, own calculations.

Table A.6: Top origin and destination affiliations

Origin	**Share**	**Destination**	**Share**
Student	0.92 %	Microsoft	2.08 %
Microsoft	0.72 %	Google	2.00 %
University of Washington	0.62 %	Amazon	1.37 %
Freelancer	0.51 %	Facebook	1.00 %
IBM	0.41 %	Red Hat	0.64 %
New York University	0.41 %	Shopify	0.44 %
University of California	0.41 %	IBM	0.37 %
University of Florida	0.41 %	Stanford University	0.31 %
University of Oxford	0.41 %	LinkedIn	0.28 %
Amazon	0.31 %	Apple	0.26 %
	5.13 %		8.75 %

Notes: Table reports the ten most frequently stated affiliations as a percentage of all users with non-empty affiliation information. *Sources:* GHTorrent, own calculations.

Table A.7: Classification of programming languages

Classification	**programming language**	**share**	
		lang.	class.
App development	Ruby	5.68 %	
	Go	4.06 %	
	Swift	1.09 %	
	Objective-C	0.65 %	11.48 %
Data engineering	Python	13.03 %	
	R	1.22 %	
	Jupyter Notebook	1.18 %	
	Scala	0.89 %	16.32 %
Low-level programming	C++	5.37 %	
	C	3.33 %	
	C#	2.30 %	
	Rust	1.40 %	
	Assembly	0.08 %	12.48 %
Program routine	Shell	3.16 %	
	PowerShell	0.22 %	3.38 %
Web development	JavaScript	20.91 %	
	HTML	6.65 %	
	Java	6.19 %	
	PHP	4.36 %	
	CSS	4.28 %	
	TypeScript	3.21 %	42.39 %
Other			10.74 %

Notes: The 27 most-used programming languages in terms of commits in the *GHTorrent* are classified, 21 of which are represented in our sample. Classified programming languages account for 89.26% of commits in our sample. *Sources:* GHTorrent, own calculations.

Table A.8: Top-paying programming languages

Classification	programming language	share		median pay	
		lang.	class. cumul.	lang.	class. avg.
Top 30 top-paying languages	Zig	0.009 %		$103,611	
	Erlang	0.145 %		$99,492	
	F#	0.091 %		$99,311	
	Ruby	5.749 %		$98,522	
	Clojure	0.399 %		$96,381	
	Elixir	0.383 %		$96,381	
	Scala	0.894 %		$96,381	
	Perl	0.491 %		$94,540	
	Go	4.087 %		$92,760	
	OCaml	0.365 %		$91,026	
	Objective-C	0.646 %		$90,000	
	Rust	1.365 %		$87,012	
	Swift	1.041 %		$86,897	
	Groovy	0.202 %		$86,271	
	Shell	3.347 %		$85,672	
	Haskell	0.771 %		$85,672	
	Apex	0.015 %		$81,552	
	PowerShell	0.23 %		$81,311	
	SAS	0.002 %		$81,000	
	Lua	0.312 %		$80,690	
	Nim	0.016 %		$80,000	
	Raku	0.001 %		$79,448	
	Python	12.933 %		$78,331	
	Kotlin	0.438 %		$78,207	
	APL	0 %		$77,500	
	Crystal	0.041 %		$77,104	
	TypeScript	3.074 %		$77,104	
	Assembly	0.078 %		$77,010	
	Fortran	0.132 %		$76,104	
	Cobol	0.001 %		$76,000	
	C#	2.314 %	39.572 %	$74,963	$86,008
Other top-paying languages	C++	5.516 %		$74,963	
	Julia	0.416 %		$74,963	
	R	1.217 %		$74,963	
	SQL	0.12 %		$74,963	
	C	3.438 %		$74,351	
	JavaScript	20.381 %		$74,034	
	Solidity	0.007 %		$72,701	
	Ada	0.013 %		$72,656	
	HTML	6.653 %		$71,500	
	CSS	4.264 %		$70,148	
	Prolog	0.018 %		$70,000	
	Delphi	0 %		$69,608	
	GDScript	0.021 %		$69,608	
	VBA	0.002 %		$65,698	
	Visual Basic	0.096 %		$65,000	
	Matlab	0.215 %		$61,735	
	PHP	4.375 %		$58,899	
	Dart	0.221 %	46.973 %	$55,862	$69,536
Not listed			13.455 %		

Notes: Table reports programming languages on the *StackOverflow* list of top-paying technologies. We further distinguish between the top 30 and other listed programming languages. Classified programming languages account for 86.54% of commits in our sample. *Sources:* GHTorrent, StackOverflow, own calculations.

Table A.9: Keywords

Cluster	keywords	% projects
Code	adventofcode; algorithm; algorithms; android; api; app; application; apps; c; class; framework; functions; game; hacktoberfest; ios; javascript; library; module; nodejs; plugin; python; react; server; software; template; testing; tictactoe; tool; ui	7.06
Website	blog; personal; personalwebsite; portfolio; resume; site; website	2.11
File	collection; docs; document; documentation; dotfiles; file; files; githubslideshow; presentation; presentations; scripts	1.17
Education	course; coursera; example; examples; exercise; exercises; freecodecamp; helloworld; homework; learning; nowgithubstarter; programmingassignment; repdata; peerassessment; test	0.85
Data	data; database	0.48
Other		13.06

Notes: Table reports keywords assigned to project type clusters. Projects may be assigned to multiple clusters. Keywords search is conducted in project descriptions; 24.73% of projects feature non-empty project descriptions. *Sources:* GHTorrent, own calculations.

Table A.10: Model specification

Model class:	OLS				LPM	NB	PPML
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	log	ihs	ihs	ihs	dummy	count	count
Sample:	full	full	geo	change	full	full	full
Job mover × job search	0.1326***	0.1646***	0.1654***	0.1384**	0.0711***	0.4983***	0.1358***
	(0.0119)	(0.0141)	(0.0142)	(0.0548)	(0.0039)	(0.0280)	(0.0521)
Job mover × post move	-0.0851***	-0.1036***	-0.1021***	-0.2804***	-0.0307***	-0.2690***	-0.1707**
	(0.0159)	(0.0190)	(0.0190)	(0.0849)	(0.0056)	(0.0453)	(0.0670)
User FE	×	×	×	×	×	×	×
Month FE	×	×	×	×	×	×	×
Experience FE	×	×	×	×	×	×	×
Adjusted R^2	0.35803	0.35945	0.35958	0.43305	0.34000		
Observations	1,946,413	1,946,413	1,941,317	76,797	1,946,413	1,630,215	1,630,215
# User FE	22,896	22,896	22,838	885	22,896	22,896	22,896

Notes: Results from estimation of Equation 2 for different model classes, outcome transformations, and sample definitions. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, own calculations.

Table A.11: Project ownership and initial forks

IHS(single commits)	project owner (1) own	project owner (2) non-own	(3) no initial forks
Job mover × job search	0.1310***	0.0428***	0.1440***
	(0.0138)	(0.0080)	(0.0149)
Job mover × post move	-0.1157***	0.0024	-0.1088***
	(0.0182)	(0.0097)	(0.0194)
User FE	×	×	×
Month FE	×	×	×
Experience FE	×	×	×
Adjusted R^2	0.33534	0.32483	0.32440
Observations	1,946,413	1,946,413	1,946,413
Users	22,896	22,896	22,896

Notes: Results from estimation of Equation 2 by repository ownership and without initial fork projects. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, own calculations.

Table A.12: Heterogeneity by project types (keywords)

IHS(single commits)	(1) education	(2) data	(3) website	(4) code	(5) files	(6) other
Job mover × job search	0.0030	0.0000	0.0135***	0.0307***	0.0154***	0.1097***
	(0.0024)	(0.0027)	(0.0043)	(0.0069)	(0.0037)	(0.0123)
Job mover × post move	-0.0091***	-0.0089***	-0.0049	-0.0335***	-0.0044	-0.0641***
	(0.0035)	(0.0031)	(0.0056)	(0.0090)	(0.0045)	(0.0163)
User FE	×	×	×	×	×	×
Month FE	×	×	×	×	×	×
Experience FE	×	×	×	×	×	×
Adjusted R^2	0.09276	0.10952	0.15628	0.16827	0.21257	0.31379
Observations	1,946,413	1,946,413	1,946,413	1,946,413	1,946,413	1,946,413
Users	22,896	22,896	22,896	22,896	22,896	22,896

Notes: Results from estimation of Equation 2 for different project types, according to keyword-based method. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, own calculations.

Table A.13: Event study coefficients

IHS(single commits)	(1)	(2)	(3)
Job mover × event_time = -21	0.0126	0.0025	0.0017
	(0.0206)	(0.0217)	(0.0215)
Job mover × event_time = -20	0.0178	0.0283	0.0326
	(0.0208)	(0.0217)	(0.0215)
Job mover × event_time = -19	-0.0397**	-0.0016	0.0042
	(0.0196)	(0.0208)	(0.0207)
Job mover × event_time = -18	-0.0555***	-0.0084	-0.0066
	(0.0193)	(0.0204)	(0.0203)
Job mover × event_time = -17	-0.0328*	-0.0167	-0.0145
	(0.0168)	(0.0178)	(0.0176)
Job mover × event_time = -15	0.1771***	0.0574***	0.0557***
	(0.0188)	(0.0197)	(0.0196)
Job mover × event_time = -14	0.5110***	0.1608***	0.1596***
	(0.0239)	(0.0251)	(0.0252)
Job mover × event_time = -13	0.5415***	0.1787***	0.1807***
	(0.0243)	(0.0251)	(0.0251)
Job mover × event_time = -12	0.6329***	0.2443***	0.2455***
	(0.0277)	(0.0282)	(0.0282)
Job mover × event_time = -11	0.5882***	0.1942***	0.1996***
	(0.0271)	(0.0276)	(0.0278)
Job mover × event_time = -10	0.5708***	0.1640***	0.1675***
	(0.0268)	(0.0272)	(0.0273)
Job mover × event_time = -9	0.4677***	0.1141***	0.1221***
	(0.0264)	(0.0270)	(0.0269)
Job mover × event_time = -8	0.4538***	0.1290***	0.1377***
	(0.0273)	(0.0278)	(0.0277)
Job mover × event_time = -7	0.4278***	0.1339***	0.1475***
	(0.0273)	(0.0278)	(0.0278)
Job mover × event_time = -6	0.4627***	0.1440***	0.1630***
	(0.0287)	(0.0293)	(0.0295)
Job mover × event_time = -5	0.4658***	0.1158***	0.1318***
	(0.0278)	(0.0284)	(0.0285)
Job mover × event_time = -4	0.3806***	0.0759***	0.0967***
	(0.0274)	(0.0276)	(0.0278)
Job mover × event_time = -3	0.3846***	0.0388	0.0654**
	(0.0265)	(0.0272)	(0.0272)
Job mover × event_time = -2	0.3617***	0.0416	0.0690**
	(0.0264)	(0.0271)	(0.0273)
Job mover × event_time = -1	0.4193***	0.0331	0.0738***
	(0.0275)	(0.0283)	(0.0285)
Job mover × event_time = 0	-0.0184	-0.1128***	-0.0799***
	(0.0225)	(0.0237)	(0.0242)
Job mover × event_time = 1	0.1672***	-0.2069***	-0.0380
	(0.0357)	(0.0363)	(0.0360)
Job mover × event_time = 2	0.1323***	-0.2101***	-0.0355
	(0.0391)	(0.0397)	(0.0394)
Job mover × event_time = 3	-0.0117	-0.3078***	-0.1291***
	(0.0379)	(0.0383)	(0.0380)
Job mover × event_time = 4	-0.0196	-0.2641***	-0.0780**
	(0.0338)	(0.0342)	(0.0340)
Job mover × event_time = 5	-0.0234	-0.2527***	-0.0621*
	(0.0364)	(0.0371)	(0.0367)
Job mover × event_time = 6	0.0134	-0.2151***	-0.0197
	(0.0386)	(0.0386)	(0.0381)
Job mover × event_time = 7	-0.3461***	-0.2582***	-0.0785***
	(0.0311)	(0.0309)	(0.0303)
Job mover × event_time = 8	-0.3202***	-0.2582***	-0.0671**
	(0.0302)	(0.0298)	(0.0295)
Job mover × event_time = 9	-0.2907***	-0.2614***	-0.0634**
	(0.0320)	(0.0316)	(0.0313)
Job mover × event_time = 10	-0.3573***	-0.2762***	-0.0725**
	(0.0312)	(0.0310)	(0.0307)
User FE	×	×	×
Month FE		×	×
Experience FE			×
Adjusted R^2	0.28992	0.30870	0.35963
Observations	1,946,413	1,946,413	1,946,413
Users	22,896	22,896	22,896

Notes: Estimates for $t_j \times \text{JobChanger}_i$ based on Equation 1 with user and calendar month fixed effects. The outcome is IHS-transformed commits to single-authored projects. The reference month is $t = -16$. Bars show 95% confidence intervals. Standard errors are clustered at the user level. * p > 0.01, ** p > 0.05, and *** p > 0.1. *Sources:* GHTorrent, own calculations.

Table A.14: Job search period

Job search period:	(1) $[-15,-9]$	(2) $[-15,-6]$	(3) $[-15,-3]$	(4) $[-15,0]$
Job mover $\times$ job search	0.1947***	0.1836***	0.1768***	0.1646***
	(0.0154)	(0.0144)	(0.0141)	(0.0141)
Job mover $\times$ uncertain	0.1423***	0.1308***	0.0925***	
	(0.0161)	(0.0178)	(0.0203)	
Job mover $\times$ post move	-0.1099***	-0.1099***	-0.1100***	-0.1036***
	(0.0184)	(0.0184)	(0.0184)	(0.0190)
User FE	$\times$	$\times$	$\times$	$\times$
Month FE	$\times$	$\times$	$\times$	$\times$
Experience FE	$\times$	$\times$	$\times$	$\times$
Adjusted R^2	0.35946	0.35946	0.35946	0.35945
Observations	1,946,413	1,946,413	1,946,413	1,946,413
Users	22,896	22,896	22,896	22,896
Relation to baseline	+3.01 p.p.	+1.90 p.p.	+1.22 p.p.	baseline
	+18.3 %	+11.5 %	+7.4 %	baseline

Notes: Results from estimation of Equation 2 for different definitions of the job serach period. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, own calculations.

Table A.15: International movers

IHS(single commits)	international		inter-continental	
	(1) yes	(2) no	(3) yes	(4) no
Job mover $\times$ job search	0.2027***	0.1474***	0.2335***	0.1483***
	(0.0263)	(0.0167)	(0.0336)	(0.0155)
Job mover $\times$ post move	-0.0812**	-0.1124***	-0.1057**	-0.1031***
	(0.0342)	(0.0228)	(0.0435)	(0.0211)
User FE	$\times$	$\times$	$\times$	$\times$
Month FE	$\times$	$\times$	$\times$	$\times$
Experience FE	$\times$	$\times$	$\times$	$\times$
Adjusted R^2	0.36811	0.35640	0.36273	0.35907
Observations	562,982	1,383,431	366,271	1,580,142
Users	6,598	16,298	4,305	18,591

Notes: Results from estimation of Equation 2 with IHS-transformed number of commits to (non-)international and (non-)inter-continental single-authored projects. Upward income group moves are defined as moves from developing to developed countries. Upward moves in GDP per capita are based on current 2021 PPP USD. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, own calculations.

Table A.16: Upward movers

IHS(single commits)	GDP p. c.		income class	
	(1) other	(2) up	(3) other	(4) up
Job mover × job search	0.1622***	0.1821***	0.1610***	0.2381***
	(0.0149)	(0.0437)	(0.0146)	(0.0512)
Job mover × post move	-0.1034***	-0.1038*	-0.1038***	-0.0949
	(0.0199)	(0.0627)	(0.0195)	(0.0755)
User FE	×	×	×	×
Month FE	×	×	×	×
Experience FE	×	×	×	×
Adjusted R^2	0.36073	0.34025	0.35980	0.33293
Observations	1,776,167	170,246	1,854,956	91,457
Users	20,829	2,067	21,763	1,133

Notes: Results from estimation of Equation 2 with IHS-transformed number of commits to (non-)upward single-authored projects in terms of GDP p.c. and income class, respectively. Upward income group moves are defined as moves from developing to developed countries. Upward moves in GDP per capita are based on current 2021 PPP USD. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, own calculations.

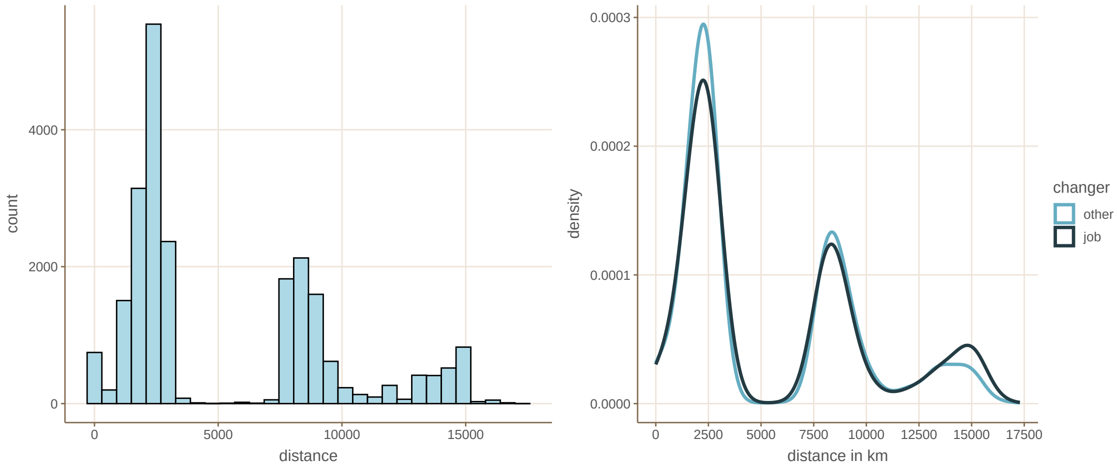
Table A.17: Affiliation

IHS(single commits)	destination						origin			
	median		big tech		academia		median		adademia	
	(1) below	(2) above	(3) no	(4) yes	(5) no	(6) yes	(7) below	(8) above	(9) no	(10) yes
Job mover × job search	0.1770***	0.0068	0.1740***	0.1556***	0.1535***	0.3158***	0.1636***	0.2339**	0.1565***	0.1511***
	(0.0145)	(0.0526)	(0.0212)	(0.0176)	(0.0146)	(0.0459)	(0.0142)	(0.1052)	(0.0519)	(0.0472)
Job mover × post move	-0.0955***	-0.1556**	-0.1001***	-0.0895***	-0.1199***	0.1547**	-0.1038***	-0.0610	-0.1755**	-0.1593**
	(0.0195)	(0.0668)	(0.0296)	(0.0232)	(0.0194)	(0.0717)	(0.0191)	(0.1320)	(0.0758)	(0.0721)
User FE	×	×	×	×	×	×	×	×	×	×
Month FE	×	×	×	×	×	×	×	×	×	×
Experience FE	×	×	×	×	×	×	×	×	×	×
Adjusted R^2	0.35927	0.36002	0.36084	0.35832	0.35823	0.36154	0.35933	0.35989	0.35999	0.36103
Observations	1,900,195	1,369,596	1,553,857	1,715,934	1,900,917	1,368,874	1,935,568	1,334,223	1,361,217	1,368,330
Users	22,387	16,194	18,374	20,207	22,378	16,203	22,767	15,814	16,130	16,212

Notes: Results from estimation of Equation 2 with IHS-transformed number of commits to single-authored projects. Median split refers to median size of affiliation in terms of users in the full *GHTorrent* sample. Big tech refers to Google, Amazon, Meta, Apple and Microsoft. Academia refers to students and university affiliations. Destination (origin) refers to users' affiliation before (after) the affiliation change. Experience is measured as months since the first commit at move month. Robust standard errors are clustered at the user level. * $p > 0.01$, ** $p > 0.05$, and *** $p > 0.1$. *Sources:* GHTorrent, own calculations.

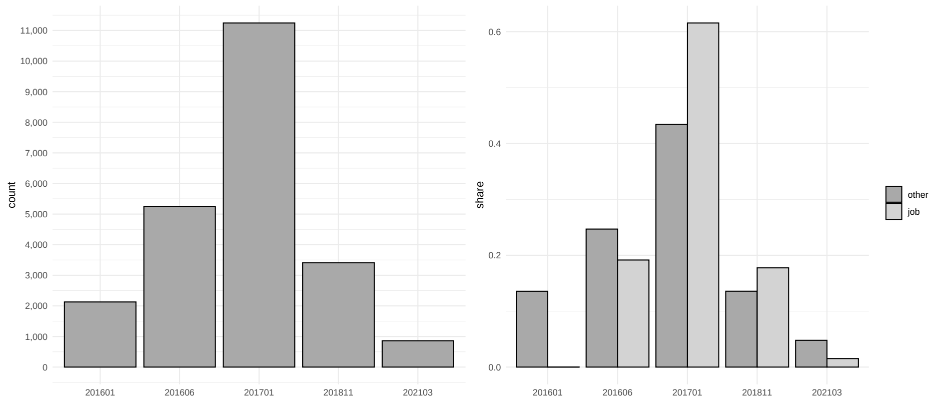
A.2 Figures

Figure A.1: Distribution of move distances



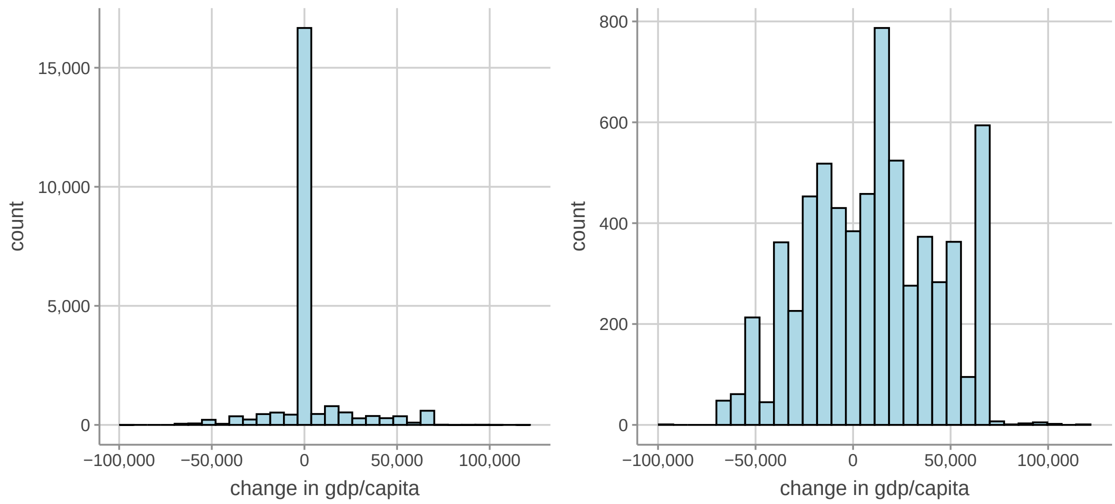
Notes: Histogram on the left shows the distribution of move distances. Estimates on the right show kernel densities for job movers and other movers. *Sources:* GHTorrent, own calculations.

Figure A.2: Distribution of moves across time



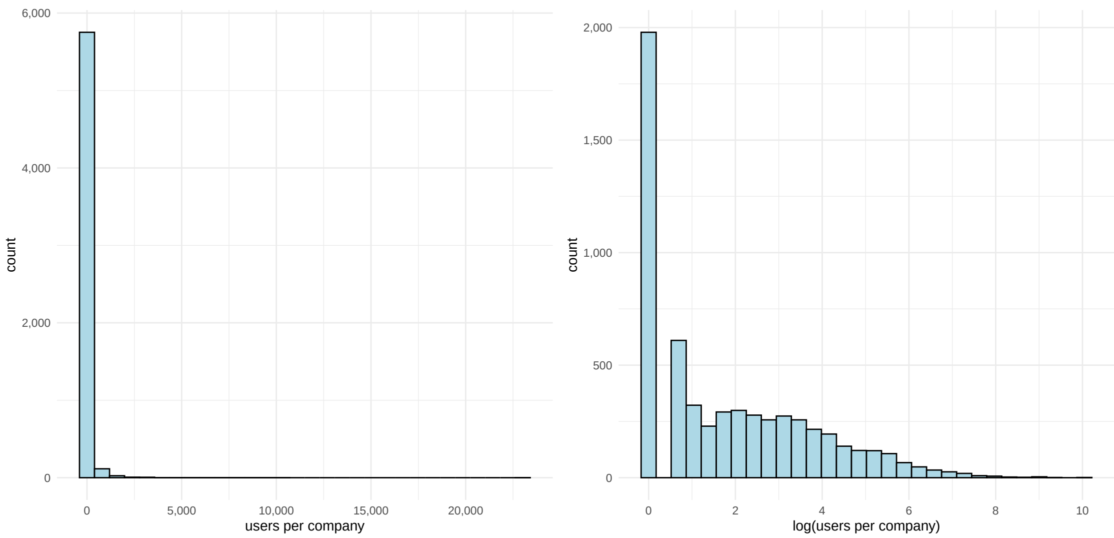
Notes: Histogram on the left shows the distribution of moves across data snapshots. Shares on the right depict the distribution of moves across data snapshots for job movers (dark gray) and other movers (light gray). *Sources:* GHTorrent, own calculations.

Figure A.3: Distribution of income changes



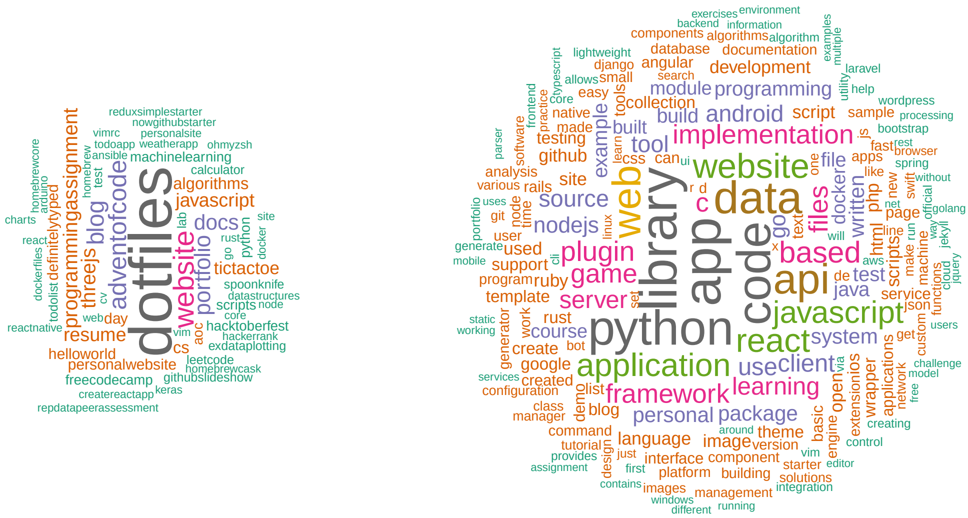
Notes: Histograms depict the distribution of national per capita GDP changes of movers in the full sample (left) and the international sample (right). GDP is measured in current 2021 PPP USD. *Sources:* GHTorrent, World Development Indicators, own calculations.

Figure A.4: Distribution of affiliation size



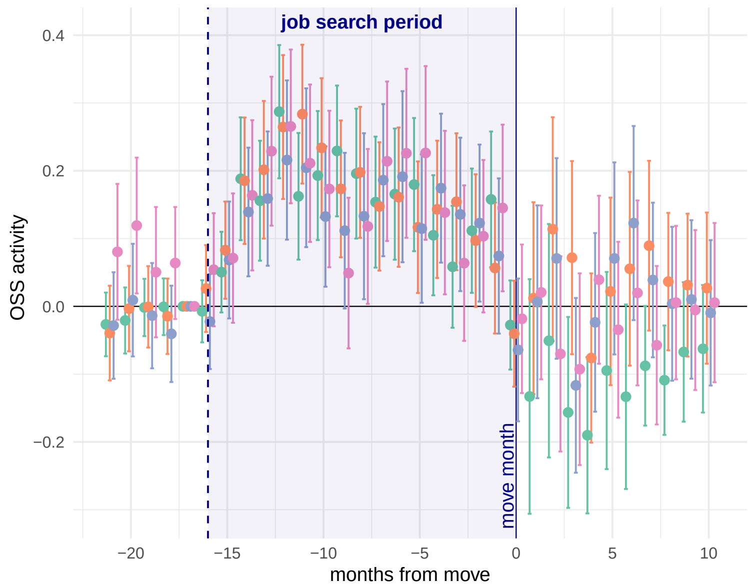
Notes: Histograms depict the distribution of affiliations with respect to the number of affiliated users in the full *GHTorrent* sample as counts (left) and after logarithmic transformation (right). Note that string-based merging of affiliations is likely imperfect, especially for small firms, which leads to a downward bias of firm size. *Sources:* GHTorrent, own calculations.

Figure A.5: Frequent words in project names and descriptions



Notes: Word clouds show frequently occurring words in single projects of movers. Word size and color represent word frequency in project titles (left) and descriptions (right). Frequency limits are set at 50 (titles) and 100 (descriptions). We remove English stop words, numbers, punctuation, URLs, white space, and the words *project*, *repository/repo*, *simple*, and *using*. *Sources:* GHTorrent, own calculations.

Figure A.6: Heterogeneity by user popularity

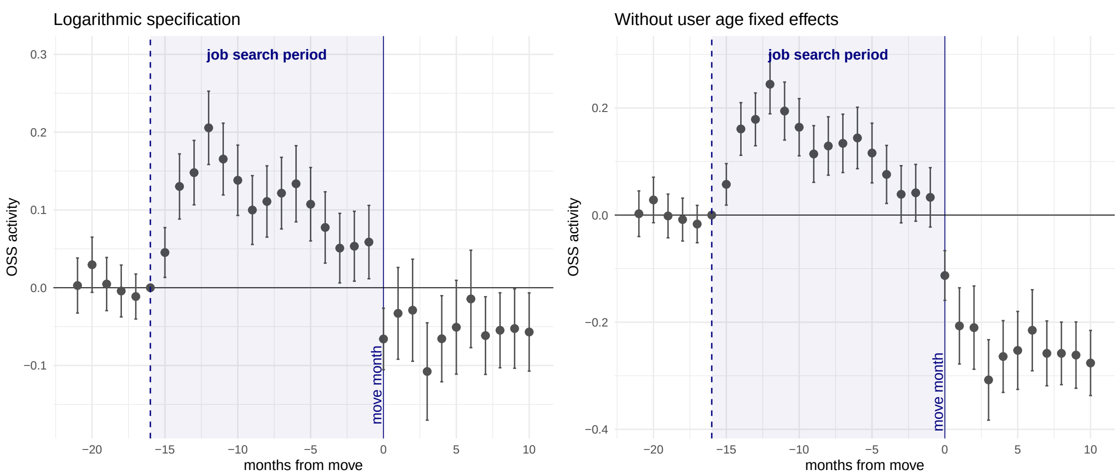


Notes: Estimates for $t_j \times \text{JobChanger}_i$ based on Equation 1 with user, experience and calendar month fixed effects. The outcome is IHS-transformed commits to single-authored projects in the respective follower quartile (1st quartile: green; 2nd quartile: orange; 3rd quartile: blue and 4th quartile: purple.). The reference month is $t = -16$. Bars show 95% confidence intervals. Standard errors are clustered at the user level. *Sources:* GHTorrent, own calculations.

Figure A.7: Heterogeneity by project age

Notes: Estimates for $t_j \times \text{JobChanger}_i$ based on Equation 1 with user and calendar month fixed effects. The outcome is IHS-transformed commits to single-authored new (orange) and old (green) projects. New projects are defined as projects with the date of the first commit in the month under consideration. The reference month is $t = -16$. Bars show 95% confidence intervals. Standard errors are clustered at the user level. *Sources:* GHTorrent, own calculations.

Figure A.8: Event study model robustness



Notes: Estimates for $t_j \times \text{JobChanger}_i$ based on Equation 1 with user and calendar month fixed effects. The outcome is logarithmically transformed using $ln(y+1)$ in the left panel and IHS-transformed commits to single-authored projects in the right panel. The reference month is $t = -16$. Bars show 95% confidence intervals. Standard errors are clustered at the user level. *Sources:* GHTorrent, own calculations.