Big Data and Bigger Firms: A Labor Market Channel *

Abhinav Gupta[†] **Naman Nishesh**[‡] **Elena Simintzi**[§]

July 29, 2024

PRELIMINARY AND INCOMPLETE: PLEASE DO NOT CIRCULATE WITHOUT PERMISSION

Abstract

This paper investigates the impact of individual-level output data on labor redistribution towards large firms by analyzing the disclosure of employee output information through GitHub, the world's largest software management platform. GitHub tracks and publicly displays real-time individual contributions. In 2016, a policy change enabled GitHub users to display their total contributions more accurately on their profiles. Following this update, employees with 1 standard deviation higher GitHub contributions witnessed a 5% increase in job transitions to large firms, predominantly at the expense of smaller companies. While productive individuals left small firms for senior roles in larger companies, the latter retained them through internal promotions. The departure of productive workers led to an overall reduction in employment growth and productivity for small firms with more productive employees prior to the shock. Our findings highlight the role of labor-related big data in amplifying the dominance of large firms in recent years.

*We are grateful to seminar participants at UNC Chapel Hill, University of Colorado at Boulder, Copenhagen Business School, BI Norwegian Business School, and Charles III University of Madrid for comments. All errors are our own.

[†]UNC Kenan-Flagler Business School (abhinav_gupta@kenan-flagler.unc.edu)

[‡]UNC Kenan-Flagler Business School (naman_nishesh@kenan-flagler.unc.edu)

[§]UNC Kenan-Flagler Business School and CEPR (elena_simintzi@kenan-flagler.unc.edu)

1 Introduction

The past two decades have witnessed an unprecedented surge in the prominence of large firms (Autor et al. (2020)), sparking a debate about the economic implications of this phenomenon (Grullon et al., 2019; De Loecker et al., 2020). Various studies have attributed this trend to changes in technology (Crouzet and Eberly, 2018; Autor et al., 2020; Hsieh and Rossi-Hansberg, 2023; Lashkari et al., 2018; Firooz et al., 2022), particularly emphasizing the role of online platforms and big data in enhancing the product market power of large firms (Veldkamp and Chung (2024)). Our paper shifts the focus to these technology-related trends in the labor market and examines how they may affect the relative importance of large versus small firms. Specifically, we argue that the abundance of data made available via web-based platforms, which enable employees to provide real-time output information, has implications for labor reallocation and, consequently, the growth of large versus small firms.

The rise of internet-based platforms has revolutionized employee recruitment, with an overwhelming majority of job seekers (over 80%) and employers (over 90%) now utilizing them (Smith (2015); Bradshaw (2024)). The direction of labor reallocation resulting from productivity signaling made possible through these platforms is not clear ex-ante. On the one hand, larger firms offer higher wages, on average, partially due to better worker quality and productivity (Brown and Medoff (1989); Oi and Idson (1999); Troske (1999)). Thus, working for a larger firm can provide a credible signal of employee quality in the absence of other employee-level information. The introduction of new signals means that employees may now successfully signal their quality at any firm, thereby diminishing the significance of job assignment and employer brand as quality indicators (Waldman (1984, 1990); Bernhardt and Scoones (1993)). Conversely, detailed signals may also enable larger firms to overcome adverse selection and recruit talented employees from smaller firms (Greenwald (2018); Gibbons and Katz (1991)). In addition, large firms, benefiting from economies of scale, can invest more heavily and better utilize such data (Brynjolfsson and McElheran (2016)) as compared to small firms.

To empirically study this question, we focus on GitHub, a software management and version control platform, used by over 100 million worldwide users and 200,000 firms in the United States alone.[1] Users and firms leverage GitHub to host their code and data in repositories, fostering collaborative work through contributions from multiple users. The GitHub employee data meet all three criteria for big data as defined by Goldstein et al. (2021): they are large in size[2], possess multiple dimensions[3], and have a complex layout. GitHub provides users and firms the flexibility to choose the visibility of their repositories — whether public, allowing anyone to view and download the code, or private, restricting access solely to users within the same firm. GitHub's unique feature is its public display of users' contributions, including time stamps, turning it into a hub for recruiters seeking potential employees.[4] The large dimensionality and extensive public use make GitHub an ideal starting point for studying big data in labor markets.

We identify the impact of information on individual output available on GitHub by exploiting a quasi-natural experiment that changed users' ability to signal their quality. Prior to May 2016, the GitHub contribution calendar only displayed public contributions. However, on May 19, 2016, GitHub introduced the option for users to include anonymized private contributions on their profiles. Given these private contributions typically refer to projects managed on Github by employers, the quantity of individuals' private contributions capture their productivity in their current firm. With this update, Github users could instantaneously reveal the number of previously hidden private contributions, without incurring any additional costs. Since revealing the number of private contributions could only improve the profile of an individual at no cost, increasing their outside option, individuals' rational response was to opt in and display their total contributions on their profiles.

[1] `https://github.com/about` and `https://enlyft.com/tech/products/github`.

[2] A compressed subset of GitHub repositories as of November 2020 was 21 TB in size `https://archiveprogram.github.com/arctic-vault/`

[3] Data can be organized by event, repository, organization, and user, among other dimensions

[4] See `https://arc.dev/hire-developers/github` or `https://www.toptal.com/github` as examples of how numerous third-party providers assist recruiters in sourcing candidates through GitHub.

We construct a novel database that integrates employee work history with individual output to implement our research design. Employee work history is sourced from LinkedIn, the largest online professional networking platform globally. We link this dataset to employees' GitHub profiles, which provide records of their daily contributions. Our final sample covers approximately 300,000 individuals and more than 36,000 firms in the U.S. We develop a person-specific measure of output by employing an AKM model (Abowd et al., 1999) on total pre-shock private contributions on GitHub. This is achieved by regressing total private contributions on individual fixed effect, firm fixed effects and time-varying controls, and using the individual component of contributions as our measure of employee output.

At the individual level, we examine whether improved information availability influences the reallocation of employees between firms. We find that individuals with 1 standard deviation higher productivity increased their mobility to larger firms (those employing over 1000 individuals), by 5% over the sample mean in our most robust specification, after the GitHub policy change. The coefficients remain unchanged regardless of granular dynamic employee and market controls, alleviating concerns of concurrent trends influencing our results. Additionally, our results are robust to using alternative AKM specifications and employee-productivity definitions. Importantly, the lack of pre-trends in the eight quarters preceding the change strengthens our confidence in attributing the observed impact to this specific event. Additionally, the absence of any significant effects on a placebo sample where individual output is based on public contributions unaffected by the shock, helps alleviate concerns regarding potential confounding factors. Another concern might be that only individuals looking to move jobs choose to reveal their private contributions, biasing our results. However, our results persist even within users who reveal their contributions and for whom we observe at least one pre-period private contribution, thereby minimizing the likelihood of such bias.

In contrast to the increasing flow of talent towards large firms after the shock, individuals with 1 standard deviation higher productivity are less likely to move to medium-sized firms (those employing between 50 and 1,000 employees) by 5% relative to the sample mean, and less likely to move to small firms (those employing 50 employees or less) by 6% relative to the sample mean. These results are also robust to a dynamic analysis, which shows there are no pre-trends prior to the GitHub policy change. Interestingly, when we estimate our baseline specification by firm size buckets, we find that talent flow increases almost monotonically across size buckets from -8% at the 11-50 size bucket to a +7% at firms with more than 10,000 employees. As such, the largest gains are captured by the largest employers in the economy, the so-called "superstar" firms.

Moreover, we examine where the talent flow to large firm originates from. Do we observe reallocation of talent between large firms or are these gains to large firms at the expense of smaller firms? Our results suggest that reallocation of talent within large firms drops, while it is small and medium-sized firms who lose their talent to large firms. In economic terms, the larger effects are seen amongst mobility from small firms, and less so from medium-sized firms: individuals with 1 standard deviation higher productivity are more likely to move from small (medium-sized) firms to large firms by 5% (4%) relative to the sample mean, following the shock. Such moves from small/med-sized to large firms are associated with career upgrades for talented employees, who move to more senior or higher-paying jobs in large firms. Large firms do not lose their talented employees, but they promote them, consistent with the fact that the outside option of talented employees increases after the policy. Overall, treated employees experience career upgrades following the GitHub policy change either by moving towards large firms or by getting promotions within large firms.

We next test the idea that the reduction of information asymmetry regarding employee quality due to GitHub should reduce the importance of traditional signals of employee quality. To this end, we consider three traditional signals: the work experience of the individual, having a degree from a highly-ranked university, and the network of universities large firms typically hire from. We find consistent results across all our measures that talented individuals with less experience, with degrees outside of the highly-ranked schools, and outside the traditional networks large firms typically hire from are more likely to move to large firms following the GitHub policy change. These results support the hypothesis that web-based platforms have led to a 'democratization' of screening in the labor markets, allowing talent to flow more freely across firms.

The flow of talent towards large firms can be a result of both demand and supply-side factors. On the one hand, employees may prefer to work for large firms as large firms tend to offer steeper career paths to their employees (Mueller et al. (2017); Di Porto et al. (2024)). To test this, we estimate our baseline specification for two subsets of firms: those with above-median average annual salary change and those with below-median average annual salary change. We observe that the coefficient estimates are larger for large firms characterized by steeper career growth as opposed to those with less steep career growth. Individuals with 1 standard deviation higher output are more likely to move to large firms characterized by steeper (less steep) career growth by 7% (2%) relative to the sample mean, following the shock. This potential for increased career growth may come with increased labor risk. We divide our sample by firm stability, defined as the probability of employees experiencing career downgrades when working for a firm. Our findings indicate that productive employees tend to move to firms with below-median stability following the shock, suggesting that talented employees prefer riskier jobs with increased career advancement opportunities.

On the other hand, large firms tend to invest more in technology to screen workers (Eckel and Yeaple (2017)), and are thus better able to discover talent once the GitHub signal becomes available. To see this, we split firms based on whether one of their Human Resource (HR) personnel uses GitHub or not. We then show that more talented individuals move to large firms with tech-savvy HR departments as opposed to large firms whose HR departments are not tech-savvy. Similarly, large firms already using the GitHub platform will be better versed to identify talent based on their contributions. To this end, we observe that talent flow to large firms is more pronounced when these firms are GitHub users as compared to large firms not on GitHub.

At the firm level, we show that heterogeneity in labor reallocation creates winners and losers. For this analysis, we define the treatment group as firms exhibiting above-median average employee productivity, with productivity determined by the individual component of private contributions during the pre-period. Small firms with above-median individual productivity experienced slower quarter-on-quarter employment growth after the shock, by 13% relative to the sample mean. Small firms with above-median individual productivity also experienced a decline in productivity, proxied by public GitHub contributions (which are not directly affected by GitHub's policy change). Specifically, productivity in small firms employing highly productive individuals pre-treatment, dropped by 10 to 12% post-announcement. Instead, large firms employing high-productivity talent pre-treatment experienced a 5% increase in quarterly employment growth over the mean and an 8 to 10% rise in public GitHub contributions over the mean following the policy change.

Finally, we aggregate the data at the industry level to evaluate the relationship between the GitHub private contributions and firm concentration. Although these results are based on correlations, they offer interesting insights into overarching trends. We test the correlation between an industry's pre-period GitHub contributions and the change in its average product and labor market concentration from 2011-2016 to 2016-2021. Our analysis reveals that industries with a higher average number of pre-shock private contributions (as of 2016) experienced a more substantial increase in labor market concentration compared to those with fewer such contributions. This finding aligns with our previous results, which indicate an increase in labor reallocation to larger firms due to the GitHub policy change. We also find that industries with higher pre-shock private contributions experienced a larger cross-sectional increase in product market concentration. Consequently, the dissemination of labor productivity information via online platforms may enhance large firms' capacity to identify and recruit the most productive employees from smaller firms, further amplifying their product market dominance.

Our paper makes significant contributions to three strands of the existing literature. A substantial body of literature has attributed the rise of large firms to the platform economy (Bessen (2020); Crouzet and Eberly (2018); Firooz et al. (2022); Brynjolfsson et al. (2023); Lashkari et al. (2018); Hsieh and Rossi-Hansberg (2023)) and big data (Farboodi et al. (2022); Tambe and Hitt (2012); Begenau et al. (2018); Brynjolfsson and McElheran (2016); Aghion et al. (2023)), arguing that these technologies provide product market advantages (Prat and Valletti (2022); Eeckhout and Veldkamp (2022)), superior returns on investment (Tambe and Hitt (2012)), and better forecasting abilities (Begenau et al. (2018)) to large firms. Our paper introduces a novel explanation, examining how big data, available via the platform economy, enhances large firms' ability to attract and retain top talent, thus contributing to their growth and dominance. Akcigit and Goldschlag (2023) observe a growing concentration of U.S. inventors within large firms and attribute this phenomenon to strategic talent hoarding. We propose an alternative explanation, suggesting that the increased concentration can also be explained by improved productivity signaling mechanisms due to the rise of big data.

Second, our paper contributes to the burgeoning literature on the impact of new data provision in labor markets. Existing papers on the impact of web-based platforms in labor markets have focused on the information content of employee review platforms. Several papers have studied the impact of review information on firm value, (Huang et al. (2020); Hales et al. (2018); Campbell and Shang (2022); Dube and Zhu (2021); Green et al. (2019); Sheng (2023)), ability to attract talent (Sockin and Sojourner (2023); Ma et al. (2023), and worker retention Pacelli et al. (2022). Our paper stands out as the first to investigate the impact of employee productivity, rather than firm quality information disclosure.

Finally, our paper contributes to the literature on the redistributive impacts of big data. Data-driven decision-making has been shown to enhance credit allocation to discriminated groups (Blattner and Nelson (2021); Di Maggio et al. (2022)) and improve investment decisions for less sophisticated investors (D'Acunto et al. (2019); Rossi and Utkus (2020); Coleman et al. (2022)). Conversely, big data can exacerbate biases in loan markets (Fuster et al. (2022); Foley et al. (2020)), and concentrate information on large firms (Farboodi et al. (2022)). While prior research has primarily focused on product and financial markets, our paper is the first to examine the reallocative impact of big data in labor markets.

2 Setting and Data

We use data from GitHub, the world's largest web-based platform in the software development ecosystem. GitHub's user base comprises of over 100 million developers–with more than 20 million in the U.S. alone. Additionally, over 200,000 U.S. firms, including 90% of the Fortune 100 companies use GitHub for software management. In this section, we provide a brief description of GitHub, followed by a summary of how we utilize GitHub's data to create a proxy for employee productivity.

2.1 GitHub: A Web-Platform for Software Management

GitHub is a web-based platform for version control and collaboration in software development projects. GitHub allows users–independent developers and firms alike–to store, manage, and share their data and code streamlined within projects known as *repositories*. By providing centralized online repositories for code storage and management, GitHub enables developers, who might be dispersed geographically, to work collaboratively on projects. It also helps stakeholders track modifications to repositories over time ensuring the integrity and coherence of project versions. Whenever a repository is modified, a snapshot of the repository's contents is automatically created and forever logged with a timestamp. With each change, along with the repository's snapshot, GitHub also records rich metadata: who made the change, what it consists of, and when exactly it was made. Before a proposed modification is integrated and the new version is released, it must, however, be approved by the repository owners (or designated managers).

An important feature of GitHub repositories is the flexibility of keeping them private or public. Public repositories, used for open-source projects, are accessible to all users, promoting collaboration and knowledge sharing within the entire developer community. Any GitHub user can view the contents, previous versions, contributors or contribution history, and even make changes (subject to approval by repository owners) to a public repository. In contrast, private repositories are restricted to specified users or teams (members invited by the repository owners or designated managers), providing a secure environment for proprietary or sensitive projects. While the content of private contributions is not publicly accessible, the requirement that each contribution must be approved by a repository manager for incorporation into a project makes these contributions difficult to manipulate. Today GitHub users can create an unlimited number of public or private repositories for free, in the past, only paid GitHub accounts (developers or firms paying a monthly subscription fee) could maintain private repositories.[5] Subsequently, until recently, private repositories were mostly favored by organizations seeking to protect intellectual property and maintain confidentiality while collaborating on development projects.

2.2 GitHub Contributions: A Measure of Productivity

In addition to facilitating version control and collaboration on software projects, GitHub provides users with individual profile pages akin to those found on traditional social media platforms. These user profile pages allow individuals to personalize their online presence by including their name, bio, location, affiliation with firms, and other relevant information. A sample GitHub profile is shown in Figure A.1. GitHub's user profile pages prominently feature the *contribution calendar*, offering a visual representation of a user's GitHub activity over time. This calendar serves as a dynamic record of a user's contributions (in the form of modifications) to GitHub repositories, providing a real-time high-frequency update on a user's work output.

[5] GitHub previously charged per repository, with prices ranging from $7 per month for up to 5 private repositories to $200 per month for up to 125 private repositories. In October 2015, they changed this policy to allow unlimited private repositories for $7 per month (Conti et al. (2021)). GitHub made private repositories free in 2019.

The precise records of users' GitHub contributions provide a way to quantify the output of software developers. GitHub hosts a vast and diverse user base comprising millions of software developers globally, resulting in extensive coverage of coding activity across various domains and projects. Moreover, the granularity and high-frequency of the contributions and the large-scale coverage of GitHub make it an appealing measure of individual output, and thereby of productivity of high-skilled professionals. Although we cannot quantify the importance of contributions (especially the private contributions whose content is not publicly available), all contributions have to be vetted by a repository manager before they are recorded on GitHub, which assures a minimum level of quality.

We validate the usability of GitHub contributions by comparing them to other commonly used measures of employee productivity and quality. Internet appendix Figure A.2 presents binned scatter plots of the logarithm of salary and publications against log contributions. Both plots display strong positive and linear correlations, indicating that total contributions effectively capture employee quality and productivity. Furthermore, GitHub contributions correlate with firm financing for startups (Conti et al., 2021), underscoring their value to sophisticated investors, such as Venture Capitalists (VCs), as a measure of human capital. Other studies McDermott and Hansen (2021); Holub and Thies (2023); El-Komboz and Fackler (2023) similarly use individual GitHub contributions to assess the productivity of software developers, thereby reinforcing our interpretation.

Unsurprisingly, recruiters and hiring managers increasingly rely on GitHub user activity to source and evaluate potential tech-sector employees. Figure A.3 presents examples of numerous third-party web platforms that help organizations in recruiting technical talent via GitHub. This shift towards leveraging GitHub contributions as a key metric in the recruitment process underscores the platform's growing significance as a talent marketplace. Motivated by these facts, we use GitHub as our laboratory to study the impact of web-based labor platforms on labor reallocation and utilize GitHub contributions as a measure of employee productivity.

2.3 GitHub Policy Shift

GitHub implemented a policy change that allowed users to reflect their GitHub contributions more accurately. Before May 19, 2016, users could only display contributions made to public GitHub repositories on their profiles. However, on May 19, 2016, GitHub introduced the option for users to include the number of (anonymized) private contributions to their existing contribution calendar. With a click of a button, users could instantaneously reveal their previously hidden private contributions, with no additional costs. While private contributions were always available internally to the firm, this policy change allowed users to credibly signal their within-firm productivity (verified by GitHub) to external employers for the first time. As an illustration, Figure A.4 shows how the policy changed users' GitHub contribution calendar. GitHub's new policy added a feature on users' profile pages to "turn on their private contributions" and include the number of (anonymized) private contributions in the calendar. We use users' pre-policy private contributions to construct an exogenously available measure of their output. It is important to note that we observe only the number, not the content, of private contributions, precluding the creation of a quality-adjusted contribution measure. However, the requirement that each contribution be approved by a repository manager for incorporation into a project helps ensure the quality of each contribution. Additionally, we verify in the previous section that the raw number of contributions correlates with other measures of productivity such as salary and publications.

2.4 Data Sources and Sample Construction

We construct a novel dataset using three main data sources: (i) user profiles from GitHub, (ii) GitHub contributions from GitHub API, and (iii) employment records from Revelio Labs. Revelio Labs continuously collects employees' online resumes from various websites (such as LinkedIn) to create one of the world's first universal HR database.

We obtain data on publicly sourced profiles of U.S.-based GitHub users from Humanpredictions, a public data aggregator that developed a proprietary database of over 150 million technical talent worldwide. This data includes users' names, location, coding languages, date of joining GitHub as well as their LinkedIn profile links. We further append this dataset with users' GitHub contributions from GitHub's official API (Application Protocol Interface)[6] The API provides the capability to programmatically access and retrieve all publicly available data on GitHub. We retrieve monthly-level public and private contributions for users from 2011 to 2021. Finally, we merge GitHub users with their employment details from LinkedIn, the largest global online professional networking platform. We obtain individual-level LinkedIn profiles from Revelio Labs. LinkedIn data provide comprehensive information regarding users' educational background (including programs pursued and graduation dates) as well as their employment history (listing firms, positions held, and tenure duration). The dataset also provides us with the direct LinkedIn URLs for each GitHub profile. These merges are made based on publicly available information, where users list both LinkedIn and GitHub URLs together. We are able to successfully merge over 1 million GitHub users with employment data from LinkedIn. Roughly 82% of the matches are obtained by directly merging the LinkedIn URLs in the two datasets. For the remaining matches, we merge on users' names and work history (employer name, start date of job, etc.).

We then construct a matched employer-employee panel detailing each employee's firm and title each quarter to track employee mobility. In case of an employee listing multiple jobs on LinkedIn at the same time (e.g., volunteering or part-time work along with a primary job), we retain only one primary job. We do so by excluding jobs where job titles include words related to volunteering or part-time employment activity. In case multiple jobs still persist, we retain the one that is listed higher on the user's LinkedIn profile (likely to be more important) and/or that corresponds to a higher seniority level.[7]

[6] We use GraphQL to query GitHub API using the endpoint: `https://api.github.com/graphql`. The documentation can be found at `https://docs.github.com/en/graphql/guides/forming-calls-with-graphql`.

[7] Revelio assigns each job a seniority score between 1-4 based on the title and job description among other details.

As a next step, we append the users' employment records with firm-level identifiers and attributes. For each job, the LinkedIn data provides a URL corresponding to the firm's LinkedIn page that has details on its size, year of incorporation, and industry among others. We use a snapshot of all LinkedIn firm pages as of 2017 to merge these details into our data. We derive the size and age of firms, close to the time of GitHub's policy change, using these appended details. Additionally, we also merge the firm URLs with an auxiliary dataset containing firm identifiers from FactSet (also provided by Revelio Labs). This dataset includes a firm's unique FactSet ID, its parent firm information, as well as its GVKEY in case the firm is publicly listed. For a small fraction of firms that remain unmatched, we also use firm data from Crunchbase to obtain the firm's size, age, and other identifiers.

In our empirical analysis, we utilize GitHub's policy change regarding users' contribution calendar and focus on its impact on their employment outcomes. Subsequently, we filter out a relevant set of GitHub users for our empirical setting from the merged sample in two steps. First, we restrict the sample to users who had joined GitHub and had made at least one GitHub contribution before the policy change on 19th May, 2016. And second, we limit the sample to users who had completed their highest educational degree before the policy change, namely we exclude users who were still studying at some point after May 2016. These filters reduce our sample size to approximately 380,000 US-based employees across 103,000 firms.

We next filter our data to create a consistent sample of employees where we can estimate individual component of productivity using the Abowd et al. (1999) (AKM) methodology. Table 2 Panel A presents the sample estimates. 93% of the raw sample are part of the connected set, representing the largest grouping of firms where employees move between the same organizations. After filtering to retain only firms with at least two employees observed in any given month, we retain 90% of the sample, enabling us to estimate regression coefficients for 89% of the sample. Further refinement involves focusing solely on employees employed at for-profit firms at the time of the policy change. This filtering process yields a final sample size of around 300,000 employees, constituting 80% of employees in the raw sample. Within this final sample, we identify approximately 36,000 firms, representing 35% of the raw sample, as the AKM requires discarding firms where multiple employees are not observed within the same role and month.

2.5 Summary Statistics

Table 1 provides summary statistics for our sample. Panel A presents cross-sectional distributions, showing that 80% of employees are male, 70% are white, and 40% hold software engineering roles. Internet Appendix Table A.1 offers further breakdowns by industry, coding language, and roles. The Software, Internet, and Information Technology sectors constitute over 45% of the sample. Figure A.2 presents the geographic distribution of our sample. The data reveals that most counties in the U.S. with significant technology clusters are represented, with New York, San Francisco, and Seattle comprising a quarter of the sample. On average, employees have eight years of experience and a salary of nearly ninety thousand dollars. Comparatively, the Bureau of Labor Statistics (BLS) data indicates that software engineers are 78% male, 60% white, with over 15% living in tech hubs like New York, Seattle, or San Francisco, earning an average salary close to 98,000 dollars.[8] Thus, our sample is representative of the median American software worker.

Panel B presents time-varying employee outcomes. The average probability of an employee moving is 7% per quarter, indicating an average tenure of 3.6 years, which aligns with the median tenure for employees in technical services (3.9 years) as reported by the BLS. Of those who move, 40% transition to large firms (over a thousand employees), 30% to medium firms (fifty to a thousand employees), and 30% to small firms (less than fifty employees). More than 55% of these moves result in a rank or salary increase, while 2% of employees are promoted within the same firm each quarter.

Panel C presents firm-level characteristics. In our sample, 15% of firms are large (over a

[8] Data obtained from Bureau of Labor Statistics table: `https://www.bls.gov/oes/tables.htm`

thousand employees), 37% are medium (fifty to a thousand employees), and 48% are small (less than fifty employees). However, as shown in Internet Appendix A.2, large firms employ more than half of the workers in our sample, similar to estimates from the Longitudinal Business Database (LBD), where large firms account for 16% of firms but employ 47% of the workforce. The average firm in our sample grows at 3% per quarter, comparable to 2.5% employee growth for all firms in the LBD. Most of these firms are active on GitHub, with the average firm seeing ten public contributions per employee per quarter.

2.6 Extracting Employee Productivity: AKM Model

We first create our measure of individual productivity based on individuals' output-GitHub contributions. The GitHub policy change revealed new information on the total number of employees' private contributions, encapsulating both individual and firm-level productivity. Hiring departments often gauge these contributions relative to peers, factoring in the influence of the specific firm environment on employee performance. To disentangle the individual component of productivity, independent of the firm influence of the firm they work for, we adopt the methodology proposed by Abowd et al. (1999) (AKM):

$$Log(PvtContributions)_{i,f,t} = \vartheta_i + \vartheta_f + \vartheta_t + \vartheta_k + X_{i,f,t} + \varepsilon_{i,f,t} \tag{1}$$

The dependent variable represents the logarithm of one plus the number of private contributions made by employee i, while employed at firm f during month t. We partition each employee's private contributions into individual, firm, and monthly components using employee (ϑ_i), firm (ϑ_f), role (ϑ_k) and month (ϑ_t) fixed effects. Additionally, $X_{i,f,t}$ includes controls for current job duration, and total work experience, ensuring that we account for variations attributable to different roles or employee seniority. Our regression analysis spans May 2011 to April 2016, precisely five years preceding the GitHub policy change. This time-frame enables us to discern how pre-existing productivity information influenced employee job outcomes. Specifically, we utilize the employee-specific private contributions over this period (ϑ_i) as our proxy for individual-level productivity information.

Table 2 presents the summary statistics for our AKM. Panel A shows that our connected set covers 93% of the employees in the raw sample. We further refine our sample by including only firms with multiple employees represented in any given year and focusing on employees who were working for profit-driven firms at the time of the shock, as detailed in Section 2.4. After applying these filters, we are able to calculate AKM estimates for 80% of the raw sample employees.Consequently, our AKM estimates do not entail significant attrition of employees in the sample. Panel B unveils that 67% of the total variation in contributions stems from differences in employee output, with an additional 8.5% attributed to disparities in average firm output. Thus, individual human capital explains 7.8 times the variation in contributions at the firm level. This finding aligns with prior research on innovation, which suggests that individual-level capabilities account for 5-10 times more variation in inventor patenting output compared to firm-level factors (Bhaskarabhatla et al. (2021)). This underscores how the GitHub policy change regarding employee private contributions can be thought of as a new signal on employee productivity, as inherent differences among employees account for the majority of variation in these contributions.

Traditionally, the AKM methodology has used the logarithm of wages as the dependent variable (Abowd et al., 1999). Recent literature has raised concerns about the validity of using the logarithm in case of count-like variables with several zeros and skewed distributions (Cohn et al. (2022); Chen and Roth (2024)). To address these concerns, we take two steps. First, we employ alternative AKM specifications, including continuous measures of contributions, the arcsine transformation of contributions, and categorical variables. Second, we demonstrate that our findings hold even without using the AKM framework by utilizing raw private contributions and the share of private contributions to define employees productivity. Our results remain consistent across all these approaches, reinforcing the robustness of our findings. Detailed discussions and the outcomes of these robustness checks are provided in Section 3, where we describe the employee-level results.

3 Employee Level Reallocation

3.1 Mobility to large firms

Next, we examine whether improved information on employee quality influences the reallocation of employees between firms. We carry out a differences-in-differences analysis where we compare the change in employee mobility for individuals with higher AKM-implied productivity after the GitHub policy change, as compared to those with lower productivity. Our baseline specification is as follows:

$$Y_{i,t} = \beta Productivity_i \times \mathbb{1}(Post_t) + \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{i,t} \tag{2}$$

where, the dependent variable Y is a binary indicator that equals 1 if employee i switches jobs in quarter t, and 0 otherwise. $Productivity_i$ represents the normalized version of the individual component of private GitHub contributions (ϑ_i), as obtained from Equation 2 in section 2.6. We normalize $Productivity_i$ by subtracting the sample mean and dividing by the standard deviation to enhance the interpretability of our coefficient estimates. The coefficient of interest, denoted as β, captures the interaction between the AKM-derived productivity estimate for each employee i and an indicator that switches on after GitHub's policy shift in May 2016. We include individual (α_i) and (δ_t) quarter fixed effects in all regressions to control for any time-invariant individual differences and time trends. To ensure the robustness of our findings, we sequentially add several controls to our model. First, we account for current salary and total employee experience to control for mobility changes due to seniority. We also include fixed effects that interact together the employee's cohort of graduation, most used coding language, first job role, and quarter fixed effects to account for time-varying job trends. Additionally, we control for industry (based on LinkedIn's 144 industry categories) interacted with geographical area (metropolitan service area) and time. This ensures that we are comparing employee transitions within the same market-quarter context. Our sample covers a period of six years: from April to June 2013 (three years before the shock) to April to June 2019 (three years after the shock). We cluster our standard errors at the individual employee level, which is the treatment level in our analysis.

We also estimate the dynamic version of the same regression as follows:

$$Y_{i,t} = \sum_{\tau=-8}^{12} \beta_\tau Productivity_i \times \mathbb{1}(\tau = t) + \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{i,t}$$

The model is similar to the one detailed in Equation 2, except we estimate a separate β_τ for each quarter τ relative to the shock. The omitted quarter is January-March 2014, one quarter before the shock. We test the parallel trends identifying assumption of the difference-in-differences methodology, by checking for any pre-trends in β_τ before the shock.

We follow Equation 2 to examine the divergent impact of the GitHub announcement on employees based on their historical private contributions. Table 3 presents our findings. Individuals with productivity estimates higher by 1 standard deviation augmented their likelihood of transitioning to larger firms (employing over 1000 individuals) by 4 to 5.7% above the sample mean following the shock, depending on the specification. Note our results remain statistically significant and consistent in magnitude across columns 1 to 5, as we progressively introduce controls from solely employee and time fixed effects to incorporating dynamic employee-level (salary and seniority), job-level (cohort-programming language-role-time), and market-level (industry-M.S.A-time) fixed effects.

The event study corresponding to column 5 Table 2, depicted in Figure 2, illustrates that employees with higher private contributions pre-treatment exhibit parallel trends with employees with lower private contributions. In other words, each pre-period coefficient is statistically indistinguishable from zero. Notably, there is a sudden and sustained surge in employee mobility among treated employees in the quarter following the GitHub announcement, indicative of a rapid change in information disclosure driving the observed outcomes.

Two primary challenges may undermine a causal interpretation of our estimates. First, concurrent changes in factors correlated with employee productivity could potentially yield similar results. However, as shown in Table 3, the coefficients remain robust to granular dynamic controls for both employees, firms, and markets. Additionally, we use a placebo sample of employee public contributions to further rule out the impact of unobservable employee characteristics. Unobserved employee traits should equally impact both public and private contributions. Therefore, if our results are driven by concurrent changes in unobserved individual-specific factors, we should observe comparable results when using public contributions as our independent variable. We present the results in Table 4, Panel A. We decompose pre-shock public contributions using the AKM methodology, similar to our baseline measure, to obtain the individual component of public contributions. We find no significant change, either economically or statistically, in the mobility of employees with 1 standard deviation higher public contributions-based productivity after the shock. This null result supports our conclusion that the observed effects are driven by signaling based on private contributions.

Moreover, there's a risk that individuals may endogenously opt to signal their pre-shock productivity when they anticipate moving to larger firms, potentially biasing our estimates upwards. To address this concern, we present results limited to active GitHub users. We test our findings across four alternative samples: users active on GitHub before the shock (with more than 10 pre-period total contributions), users whose employers were active on GitHub at the time of the shock (firms with at least 100 total contributions as of the shock date), users with at least one private contribution before the shock, and users with at least ten private contributions before the shock. Table 4, Panel B, presents the results. Although filtering on users reduces our sample size significantly, our results persist both in magnitude and statistical significance for all samples, diminishing the likelihood of such a bias driving our findings.

We also test the robustness of our results to different specification choices. Table A.3 presents these findings. First, as explained in section 2.6, we might be worried about the impact of using logarithm of one plus contributions as the dependent variable in the AKM calculations. We show that our results are robust to using other measures in the AKM calculation in Panel A, including using the arcsine transformation of contributions, categorical variables, and the raw contribution numbers. Panel B shows that the results persist even if we use the raw number of private contributions instead of AKM-implied productivity. The consistency across specifications reinforces the validity of our findings. Another concern might be that our results are capturing differences in working styles (e.g., making multiple small contributions versus several large ones) or identifying employees who were only productive in the few months before the shock rather than truly more productive employees. We address this by showing that our results are robust when using the share of private contributions over total contributions as dividing by total contributions helps control for these individual-level idiosyncrasies. A concern might stem from the usage of a continuous variable as our treatment in a diff-in-diff setting, as detailed by Callaway et al. (2024). We show our results are robust to using an indicator for non-zero productivity in Panel C. Panel D examines the sensitivity of our findings to the time period over which we estimate the AKM, Panel E explores the impact of alternate clustering methods, and Panel F tests the effect of more granular fixed effects.[9]

3.2 Mobility to medium and small firms

We next test for the impact of the GitHub announcement on the mobility of productive employees to small and medium-sized firms. We use a specification akin to Equation 2, except we now consider transitions to medium-sized firms (50 to 1000 employees) and small

[9]In Panel D, we estimate the AKM using contributions from the pre-sample period, up until the April-June quarter of 2014. Panel E presents results after clustering by industry-location, language, and role, employing double clustering at both the employee and firm levels. Panel F involves interacting all fixed effects, including cohort, role, language, location, industry, and time.

firms (less than 50 employees) as the outcome variables. The results, outlined in Table A.4, reveal that individuals with productivity estimates higher by 1 standard deviation reduced their likelihood of transitioning to small and medium firms by 6% and 4.8% of the sample mean post-shock, respectively. Event studies depicted in Figure 3 validate the absence of pre-trends in all of those outcomes pre-treatment, supporting the parallel trends assumption.

We further replicate our analysis by non-parametrically estimating our baseline specification across different firm size categories, as illustrated in Figure 4.[10]. Our findings indicate that the mobility of employees with 1 standard deviation higher productivity increases almost monotonically with firm size, ranging from -8% of the sample mean for firms with 11-50 employees to +7% for firms with more than 10,000 employees. An exception to this trend is observed for firms with fewer than 10 employees, where the decrease in mobility is economically small, at -2%, and statistically insignificant. This could be attributed to entrepreneurs starting their own firms within this category, making it less directly comparable to other categories consisting of employees joining established firms.

Overall, our findings indicate that the largest gains are captured by the largest employers in the economy, the so-called "superstar" firms. We confirm the robustness of our results with various definitions of superstar firms. Internet Appendix Table A.5 shows that the increase in mobility holds when defining superstars as the top 1% and 10% by market capitalization and revenue, as defined in Tambe et al. (2020). Moreover, internet appendix Table A.6 reveals that even within young firms, mobility increases to only "superstar" startups with over $500 million in funding. Thus, the GitHub policy change facilitates a redistribution of employees from small and medium-sized firms to the largest firms in the economy (irrespective of age).

3.3 Heterogeneity by initial firm size

The previous specification examines the changes in employee mobility following GitHub's policy change. Next, we aim to additionally identify the origin firms for these employees

[10] Table A.4 shows the regression counterparts of Figure 4.

to establish patterns of labor reallocation. Therefore, we analyze the post-GitHub policy change in the mobility of productive employees by firm size. Specifically, we estimate a triple-difference specification:

$$\begin{aligned}\mathbb{1}(MoveToLargeFirm)_{i,t} = \beta_1 Productivity_i \times \mathbb{1}(Post_t) + \\ \beta_2 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_{i(t-1)}) \\ + \beta_3 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Small_{i(t-1)}) + \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{i,t}\end{aligned} \tag{3}$$

The specification is an extension of Equation 2, incorporating a triple interaction involving three indicators: the productivity estimate of employee i derived from the AKM decomposition, an indicator that equals 1 for quarters following GitHub's policy change (April-June 2016), and indicators representing the size of the firm where the individual worked in the previous quarter. Here, *Mid* denotes an indicator variable that equals 1 for firms with 50-1000 employees, and *Small* represents an indicator for firms with fewer than 50 employees. While we also estimate the coefficients for double interaction terms between individual productivity and firm-size indicators, and firm-size indicators interacted with indicators for GitHub policy change, we do not report them in the specification or results for brevity. β_1 captures the impact of the policy change on more productive employees working at large firms, while β_2 and β_3 assess the additional impact on employees currently employed at mid-sized and small firms, respectively. This specification, thus, enables us to examine how the joint interaction of employee productivity, the GitHub policy change, and firm size influences the outcome of interest.

We also estimate the dynamic version of the Equation 3, to verify the absence of pre-trends, as follows:

$$
\begin{aligned}
\mathbb{1}(MoveToLargeFirm)_{i,t} = &\sum_{\tau=-8}^{12} \beta_{1,\tau} Productivity_i \times \mathbb{1}(\tau = t) \\
&+ \sum_{\tau=-8}^{12} \beta_{2,\tau} Productivity_i \times \mathbb{1}(\tau = t) \times \mathbb{1}(Mid_{i(t-1)}) \\
&+ \sum_{\tau=-8}^{12} \beta_{3,\tau} Productivity_i \times \mathbb{1}(\tau = t) \times \mathbb{1}(Small_{i(t-1)}) + \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{i,t}
\end{aligned}
$$

Table 6 presents the results. We find that mobility to large firms for individuals with productivity estimates higher by 1 standard deviation, who were currently employed at large firms, decreased by 12% of the sample mean, post-GitHub announcement. Conversely, the probability of individuals with 1 standard deviation higher productivity, currently employed at medium and small firms, increased by 4% and 5% of the sample mean post-shock, respectively. The absence of pre-trends, when plotting the triple-difference coefficient for both mid-sized small firms in Figure 5, underscores that we are capturing the impact of the change in GitHub policy. Our results hence indicate that productive employees currently employed at small and medium firms migrated to larger firms, leading to an overall re-allocation of talent following the GitHub policy change.

3.4 Impact on promotions

Our previous analysis has primarily focused on job changes as the outcome variable. Now, we aim to understand the impact of the GitHub policy change on promotions, encompassing both salary increases and title advancements, within and across firms. The results, presented in Table 7, dissect promotions into three distinct categories: those occurring when employees transition to large firms, those happening when employees transition to small and medium firms, and those occurring within the same firm. We maintain the baseline specification outlined in Equation 2 for odd columns and adopt Equation 3 for even columns. Additionally, we introduce firm interacted employee cohort fixed effects for within-firm promotions to ensure comparisons across employees within the same firm and with equal tenure. Our findings, summarized below, document the impact of the GitHub policy change on promotions.

Column 1 demonstrates that the likelihood of transitioning to a large firm with an increase in salary and title rose by 5.4% above the mean for individuals with 1 standard deviation higher productivity, following the GitHub announcement. Column 2 shows that this increase is mainly driven by employees from mid and small-sized firms. Specifically, the probability of moving from one large firm to another with promotion decreased by 9% of the sample mean for those individuals with 1 standard deviation higher productivity after the GitHub shock. Conversely, the probability of transitioning from small and medium firms to large firms with promotion increased by 5% of the sample mean for those with 1 standard deviation higher productivity post the GitHub announcement. These findings suggest that large firms attracted productive employees from small and medium firms by offering improved salary and title, after the GitHub change.

In contrast, Column 3 reveals that the probability of moving to a small or medium firm with an improvement in salary and title decreased by 6.1% of the sample mean for individuals with 1 standard deviation higher productivity, following the GitHub announcement. Column 2 reveals that this decrease is mainly attributed to employees from large firms. Specifically, the probability of moving from one large firm to small or medium firm with promotion decreased by 12.7% of the sample mean for those with 1 standard deviation higher productivity after the GitHub shock. Conversely, there was less than 1% change in probability of moving from one small and medium firm to another with promotion for those with 1 standard deviation higher productivity after the GitHub shock. These results suggest that small and medium firms were unable to attract highly productive employees from large firms with better title or pay.

Column 5 notably illustrates that current employers also react to the disclosure of new productivity information, showing a significant increase of 4.9% over the mean in the probability of within-firm promotions for employees with 1 standard deviation higher productivity post-shock. We also plot the dynamic event study for these results in Figure 6 and observe no pre-trends. Column 6 further demonstrates that this effect is particularly pronounced for large firms, which experience a post-shock increase of 6.2% over the mean in within-firm promotion probability for employees with 1 standard deviation higher productivity. In contrast, the same effect is 4.2% of the mean for small firms, diminished by the impact of productive employees transitioning to more senior positions at large firms following the shock. This increase in internal promotion probability aligns with existing theoretical models (e.g., Waldman (1984); Bernhardt and Scoones (1993)). These models suggest that firms are hesitant to promote talented employees without credible signals because such promotions could alert other firms to their quality, increasing the risk of losing them to competitors.

In summary, our findings indicate that the sudden disclosure of productivity information on GitHub was advantageous for productive employees, facilitating their career progression. While large firms demonstrated an ability to retain employees through internal promotions, small firms appeared to lose valuable talent to larger firms, which offered these employees better career opportunities.

4 Potential mechanisms

4.1 Heterogeneity by Individual Characteristics

Previous literature has highlighted significant information asymmetry regarding employee quality, suggesting that access to productivity information could help employers better screen candidates mitigating this asymmetry (Waldman (1984); Greenwald (2018)). We examine this mechanism by exploring the heterogeneity of our results across three types of alternate individual signals: work experience, degree from a high-ranked university, and education-based employment networks. We present these results in Table 8.

Columns 1 and 2 examine the heterogeneity of our results based on the total number of years of employee experience since graduation at the time of the shock. We find that more productive employees with less than median experience at the time of shock saw a 1.5 times larger increase in post-shock mobility to large firms compared to those with more than median experience. This suggests that productivity information from GitHub served as a substitute for work experience, enabling less experienced but productive employees to be employed by larger firms.

Columns 3 to 6 assess the heterogeneous impact of having a degree from a highly-ranked university on our results. We define elite universities as those within the top 30 ranked universities according to U.S. News and World Report (Columns 3 and 4) and QS Global University Rankings (Columns 5 and 6). Our findings indicate that productive employees without an elite university affiliation experienced a mobility increase to large firms that was 2.5 to 4 times greater than the change observed for employees with an elite university degree. Our results, so far, suggest that the increased availability of information on employee productivity may reduce reliance on hard information, such as work experience or university rankings, as a criterion for mobility to large firms.

Finally, columns 7 and 8 examine the differential impact of mobility based on the employment networks. We use the employees' schools at large firms as a proxy for such networks. Specifically, we define high-network schools as the top five recruitment destinations for all large firms in our sample. Our analysis shows that more productive employees from low-network schools experienced a 1.5 times greater increase in post-shock mobility to large firms compared to employees from high-network schools. These results suggest that the provision of productivity information reduces reliance on relationship-based information, enabling firms to recruit talented employees from outside their traditional recruitment networks.

Overall, our results support the hypothesis that web-based platforms have reduced reliance on both hard signals, such as experience and education, and network-based screening, leading to a "democratization" of labor markets. This has facilitated a less constrained flow of talent across firms, allowing employees to advance based on productivity rather than traditional metrics.

4.2 Heterogeneity by Firm Characteristics

The flow of talent towards large firms can result from both supply and demand-side factors. On the supply side, previous literature has documented that large firms offer steeper career paths than smaller firms (Mueller et al. (2017); Di Porto et al. (2024)). These steeper career paths, and the resultant higher wage inequality, have also been demonstrated to correlate with enhanced employee and firm productivity (Lemieux et al. (2009); Wallskog et al. (2024)). Hence, a preference for such steeper career paths may drive more productive employees toward larger firms. On the demand side, larger firms may have better IT resources (Eckel and Yeaple (2017)), enabling them to better utilize and recruit based on GitHub signals. We test these hypotheses using heterogeneity tests.

First, we classify firms based on their probability of offering salary increases. We calculate the average annual salary change for all employees in a firm using LinkedIn data for individuals with the same job roles as those in our GitHub sample during the sample period. Firms are then divided into two subsets: those with above-median average annual salary change and those with below-median average annual salary change. Table 9, columns 1-2, present the results. We observe that individuals with productivity estimates 1 standard deviation above the mean increased their likelihood of transitioning to larger firms with above-median salary growth by 6.8% above the sample mean following the GitHub policy change. In contrast, the increase was only 1.5% of the sample mean, and economically insignificant, for mobility to large firms with below-median salary growth. T-tests confirm that the difference between these two coefficients is significant at the 1% level. These results support the notion that employees prefer to join large firms due to the potential for faster career growth.

Second, we assess whether the increase in career growth potential increases job risk. Using the entire LinkedIn dataset for roles matching those in our GitHub sample during the sample period, we determine the probability of forced turnover for each firm. Forced turnover is defined as job separation with a gap of at least three months, accompanied by a decrease in salary or seniority. We categorize firms into two groups based on their forced turnover probability relative to the median for all firms. Columns 3-4, present the results. Individuals with productivity 1 standard deviation higher increased their mobility to larger firms with above-median forced turnover by 2.6% above the sample mean following the GitHub policy change. In contrast, the increase was 6.7% of the sample mean for mobility to large firms with below-median forced turnover. T-tests confirm that the difference between these two coefficients is significant at the 1% level. These findings show that the increase in steep career growth is accompanied by a shift to "riskier" jobs for more productive employees following the GitHub shock, suggesting that productive human capital prefers jobs with high upside.

Third, we examine the role for tech-savvy recruiters and firms' own GitHub usage. Columns 5-6 in Table 9 present the results, splitting firms based on whether their Human Resource (HR) personnel use GitHub. We find that individuals with productivity estimates 1 standard deviation above the mean increased their probability of moving to large firms with tech-savvy HR by 8.4% of the sample mean after the GitHub policy change. In contrast, the increase was only 1.3% of the sample mean for moves to large firms without any recruiters on GitHub. T-tests show these coefficients to be significantly different from each other at the 1% confidence level.

Columns 7-8 in Table 9 further split the sample based on whether firms themselves have created GitHub profiles. Large firms with a presence on GitHub may be more adept at utilizing these signals for recruitment. Consistent with this idea, we find that the probability of an employee with 1 standard deviation higher productivity moving to a large firm with its own GitHub account increased by 5.4% of the sample mean post-shock. In comparison, the increase was 3.6% of the sample mean for large firms not on GitHub. The difference between these two coefficients is statistically significant at the 2% confidence level. Together, these tests support the hypothesis that firms who invest more in, or possess better, screening technologies leverage new information from web-based platforms for recruiting talent.

5 Firm Level Impact

Section 3 demonstrates that high-productivity employees reallocated from small and medium-sized firms to large firms as a result of the GitHub productivity disclosure. It is crucial to assess whether this labor reallocation had an aggregate impact on firm outcomes. If small and medium firms are capable of replacing workers or internally redistributing projects, we would expect the disclosure to have no impact on firm growth or productivity. However, if these firms are constrained in their ability to hire and train talented employees, we would anticipate observing an aggregate impact.

We, thus, examine next the effect of the GitHub policy change on firm growth and productivity. To this end, we aggregate our data to the firm level and use a triple-differences specification. The first difference compares the changes in firms with a higher proportion of treated (high-productivity) employees with firms with a low proportion of high-productivity employees before and after the GitHub announcement. Subsequently, the second difference dissects this coefficient based on firm size, allowing us to discern the divergent impact of the GitHub policy change on treated small and medium-sized firms compared to treated large firms. We estimate the following regression equation:

$$\begin{aligned} FirmOutcome_{g,t} = {} & \beta_1 \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(Post_t) \\ & + \beta_2 \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_g) \\ & + \beta_3 \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(Post_t) \times \mathbb{1}(Small_g) + \alpha_g + \delta_t + \gamma_{g,t} + \varepsilon_{g,t} \end{aligned} \tag{4}$$

In our regression model, the dependent variable is an outcome variable for firm g at quarter t. The coefficient β_1 captures the interaction between $\mathbb{1}(Productivity_g > Median)$, an indicator that takes the value 1 if the average AKM-derived productivity estimate is greater than the median for firm g, and $\mathbb{1}(Post_t)$, an indicator that switches on after GitHub's policy shift in May 2016. Coefficients β_2 and β_3 further interact these variables with indicators for firm size. Our sample covers a period of six years: from April to June 2013 (three years before the shock) to April to June 2019 (three years after the shock). To ensure robustness of our analysis, we control for firm and time fixed effects, controlling for constant firm-level differences and time trends on our results. Additionally, we incorporate other relevant controls. In the case of employment-related outcomes, we control for lagged employment level, while for contribution-based productivity outcomes, we control for lagged total contribution stock and the median change in total contribution stock in the post-event period. We cluster our standard errors at the firm level.

We also estimate a dynamic version of the same regression as follows:

$$\begin{aligned} FirmOutcome_{g,t} = & \sum_{\tau=-8}^{12} \gamma_\tau \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(\tau = t) \\ & + \sum_{\tau=-8}^{12} \theta_\tau \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(Mid-SizedFirm_g) \\ & + \sum_{\tau=-8}^{12} \beta_\tau \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(SmallFirm_g) + \alpha_g + \delta_t + \varepsilon_{g,t} \end{aligned}$$

The model is similar to the one detailed in Equation 10, except we estimate a separate β_τ for each quarter τ relative to the shock. The omitted quarter is January-March 2014, one quarter before the shock. We test for the parallel trends assumption by checking for any pre-trends in β_τ before the shock.

We present the results for employee growth and productivity in Table 10. Columns 1 and 2 present the results for firm employment growth. We find that large firms with more than median individual productivity grew 4.8% faster than the mean after the GitHub announcement, depending on specification. Above-median productivity med-sized firms grew less by 2.5% than large firms following the policy shift, although the difference is not statistically significant. On the other hand, employment growth in above-median productivity small firms decreased by 12.9% of the mean after the announcement. These results remain consistent after controlling for location and industry time trends, as shown in columns 1 and 2. Figure 7 demonstrates that our results are not driven by differential pre-trends across firms. Overall, these findings support the notion that large firms benefited from the information revealed as they could discover new talent, while small firms experienced losses and were unable to replenish talent effectively.

Columns 3 to 6 use instead employee productivity as the outcome variable. Here, we utilize the public contributions for each employee, as these were not directly influenced by the shock. The results show an increase of 8.4% in the number of public contributions and 5% in public contributions per employee for large firms with productivity estimates exceeding the median post the GitHub announcement. We instead find statistically insignificant changes for med-sized firms. Notably, small firms with productivity estimates surpassing the median experienced a decrease of 12% in the number of contributions and 6.8% in contributions per employee after the GitHub shock. To visually represent these findings, Figure 8 displays the corresponding event studies. The absence of any discernible pre-trends before the shock is consistent with capturing the impact of the change in GitHub policy. In summary, these results highlight that large firms were able to enhance productivity, whereas small firms struggled to offset talent losses, resulting in an overall reduction in productivity.

6 Market Concentration

Sections 3 and 5 document the effects of GitHub productivity disclosure on individual reallocation to larger firms and the subsequent impact on the growth and productivity of treated firms by size. We now proceed to evaluate whether this reallocation has influenced broader industry dynamics, potentially leading to increased concentration in more affected industries. To address this, we aggregate our data to the industry level and investigate the correlation between changes in concentration, as measured by the Herfindahl-Hirschman Index (HHI), and the average number of private contributions in each industry. We estimate the following regression equation:

$$\Delta Concentration_h = \beta Productivity_h + \varepsilon_h \tag{5}$$

The dependent variable, $\Delta Concentration_h$, denotes the change in concentration for industry h over the period spanning five years before the shock (2011-2016) to five years after the shock (2016-2021). We investigate changes in both labor and product market concentration across various specifications. The coefficient β captures the correlation between the outcome and a measure of private contributions, $Productivity_h$. We present our findings for both the average of AKM-derived productivity and the proportion of employees with AKM-derived productivity greater than zero during the pre-shock period, within industry h. We conduct the analysis for both 144-LinkedIn industries and 4-digit SIC industry codes, clustering standard errors at the industry level. We abstain from asserting causality for these regressions, recognizing the difficulty of inferring causal relationships from aggregate correlations. Nevertheless, these correlations offer interesting insights into overarching trends, particularly concerning the observed increase in firm concentration over the past decade (Furman (2016); Gutiérrez and Philippon (2017); Grullon et al. (2019); Autor et al. (2023)).

Table 11 presents our results. Columns 1-2 and 3-4 test whether there was a greater increase in labor market HHI in industries with a higher proportion of GitHub private contributions, based on LinkedIn and SIC industry definitions, respectively. Industries with 1 standard deviation higher average AKM-implied productivity experienced a 34.6 (16.6) percentage point larger increase in labor concentration from 2016 to 2021, compared to 2011 to 2016, at the LinkedIn (SIC) industry level. Similarly, industries with a 1 percentage point higher ratio of pre-treatment productive employees saw a 45.5 (22.8) percentage point higher increase in labor concentration over a five-year period at the LinkedIn (SIC) industry level. Columns 5-6 present the same correlations at the product market level. We can only compute product market HHIs at the SIC industry level, as LinkedIn does not provide information on firm revenues. Industries with 1 standard deviation higher average productivity pre-treatment were associated with a 35.7 percentage point higher cross-sectional change in product market HHI, while industries with a 1 percentage point higher ratio of productive employees correlated with a 23.9 percentage point higher cross-sectional change in product market HHI.

These findings demonstrate that there were greater concentration changes in both labor and product markets within industries where more information was unveiled following the GitHub announcement.

7 Conclusion

This paper represents one of the first examinations of the impact of individual-productivity-related big data on labor distribution within the economy. Our investigation centers on GitHub, the world's largest repository of individual-level productivity information. Leveraging a quasi-natural experiment, we exploit a scenario where employees were exogenously enabled to disclose existing private contributions on their GitHub profiles.

Our findings unveil a significant paradigm shift: higher signaling ability by individuals triggered the reallocation of talented individuals toward larger firms. Smaller and medium-sized firms encountered challenges in both recruiting and retaining talented individuals post-shock. In contrast, large firms effectively retained existing talent through internal promotions and attracted new talent. Consistent with a reduction of information asymmetry on the labor market, the surge in productivity information mitigated reliance on traditional signals of hard information and existing networks, thereby democratizing opportunities for talent from lower-ranked, non-networked universities.

However, this talent redistribution had adverse effects on the growth and productivity of smaller firms. Further correlational evidence suggests that this talent reallocation led to an increase in both labor and product market concentration in industries with greater informational disclosure facilitated by GitHub. Our findings unearth a novel labor channel for big data, wherein increased employee productivity information reallocates labor from small to large firms, amplifying the significance of large firms in the economy.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2023): "A theory of falling growth and rising rents," *Review of Economic Studies*, 90, 2675–2702.

Akcigit, U. and N. Goldschlag (2023): "Where Have All the "Creative Talents" Gone? Employment Dynamics of US Inventors," Tech. rep., National Bureau of Economic Research.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): "The fall of the labor share and the rise of superstar firms," *The Quarterly Journal of Economics*, 135, 645–709.

Autor, D., C. Patterson, and J. Van Reenen (2023): "Local and national concentration trends in jobs and sales: The role of structural transformation," Tech. rep., National Bureau of Economic Research.

Begenau, J., M. Farboodi, and L. Veldkamp (2018): "Big data in finance and the growth of large firms," *Journal of Monetary Economics*, 97, 71–87.

Bernhardt, D. and D. Scoones (1993): "Promotion, turnover, and preemptive wage offers," *The American Economic Review*, 771–791.

Bessen, J. (2020): "Industry concentration and information technology," *The Journal of Law and Economics*, 63, 531–555.

Bhaskarabhatla, A., L. Cabral, D. Hegde, and T. Peeters (2021): "Are Inventors or Firms the Engines of Innovation?" *Management Science*, 67, 3899–3920.

Blattner, L. and S. Nelson (2021): "How costly is noise? Data and disparities in consumer credit," *arXiv preprint arXiv:2105.07554*.

Bradshaw, R. (2024): "Apollo Technical Research, Surprising Social Media Recruiting Statistics," `https://www.apollotechnical.com/social-media-recruiting-statistics/`.

Brown, C. and J. Medoff (1989): "The employer size-wage effect," *Journal of political Economy*, 97, 1027–1059.

Brynjolfsson, E., W. Jin, and X. Wang (2023): "Information technology, firm size, and industrial concentration," Tech. rep., National Bureau of Economic Research.

Brynjolfsson, E. and K. McElheran (2016): "The rapid adoption of data-driven decision-making," *American Economic Review*, 106, 133–139.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2024): "Difference-in-differences with a continuous treatment," Tech. rep., National Bureau of Economic Research.

Campbell, D. W. and R. Shang (2022): "Tone at the bottom: Measuring corporate misconduct risk from the text of employee reviews," *Management Science*, 68, 7034–7053.

Chen, J. and J. Roth (2024): "Logs with zeros? Some problems and solutions," *The Quarterly Journal of Economics*, 139, 891–936.

Cohn, J. B., Z. Liu, and M. I. Wardlaw (2022): "Count (and count-like) data in finance," *Journal of Financial Economics*, 146, 529–551.

Coleman, B., K. Merkley, and J. Pacelli (2022): "Human versus machine: A comparison of robo-analyst and traditional research analyst investment recommendations," *The Accounting Review*, 97, 221–244.

Conti, A., C. Peukert, and M. P. Roche (2021): "Beefing it up for your investor? open sourcing and startup funding: Evidence from github," *Open Sourcing and Startup Funding: Evidence from GitHub (August 25, 2021).*

Crouzet, N. and J. Eberly (2018): "Intangibles, investment, and efficiency," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 108, 426–431.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The rise of market power and the macroeconomic implications," *The Quarterly Journal of Economics*, 135, 561–644.

Di Maggio, M., D. Ratnadiwakara, and D. Carmichael (2022): "Invisible primes:

Fintech lending with alternative data," Tech. rep., National Bureau of Economic Research.

Di Porto, E., M. Pagano, V. Pezone, R. Saggio, and F. Schivardi (2024): "Careers and Wages in Family Firms: Evidence from Matched Employer-Employee Data," *Unpublished Manuscript*.

Dube, S. and C. Zhu (2021): "The disciplinary effect of social media: Evidence from firms' responses to Glassdoor reviews," *Journal of Accounting Research*, 59, 1783–1825.

D'Acunto, F., N. Prabhala, and A. G. Rossi (2019): "The promises and pitfalls of robo-advising," *The Review of Financial Studies*, 32, 1983–2020.

Eckel, C. and S. R. Yeaple (2017): "Too much of a good thing? Labor market imperfections as a source of exceptional exporter performance," Tech. rep., National Bureau of Economic Research.

Eeckhout, J. and L. Veldkamp (2022): "Data and market power," *NBER Working Paper*.

El-Komboz, L. A. and T. Fackler (2023): "Productivity Spillovers among Knowledge Workers in Agglomerations: Evidence from GitHub," *Available at SSRN 4600283*.

Farboodi, M., A. Matray, L. Veldkamp, and V. Venkateswaran (2022): "Where has all the data gone?" *The Review of Financial Studies*, 35, 3101–3138.

Firooz, H., Z. Liu, and Y. Wang (2022): "Automation and the rise of superstar firms," *Available at SSRN*.

Foley, C. F., A. Hurtado, A. Liberman, and A. Sepulveda (2020): "The effects of information on credit market competition: Evidence from credit cards," *Available at SSRN 3550904*.

Furman, J. (2016): "Benefits of competition and indicators of market power," *Council of Economic Advisers (Issue Brief), The White House, Washington*.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2022): "Predictably unequal? The effects of machine learning on credit markets," *The Journal of Finance*, 77, 5–47.

Gibbons, R. and L. F. Katz (1991): "Layoffs and lemons," *Journal of labor Economics*, 9, 351–380.

Goldstein, I., C. S. Spatt, and M. Ye (2021): "Big data in finance," *The Review of Financial Studies*, 34, 3213–3225.

Green, T. C., R. Huang, Q. Wen, and D. Zhou (2019): "Crowdsourced employer reviews and stock returns," *Journal of Financial Economics*, 134, 236–251.

Greenwald, B. C. (2018): *Adverse selection in the labor market*, Routledge.

Grullon, G., Y. Larkin, and R. Michaely (2019): "Are US industries becoming more concentrated?" *Review of Finance*, 23, 697–743.

Gutiérrez, G. and T. Philippon (2017): "Declining Competition and Investment in the US," Tech. rep., National Bureau of Economic Research.

Hales, J., J. R. Moon Jr, and L. A. Swenson (2018): "A new era of voluntary disclosure? Empirical evidence on how employee postings on social media relate to future corporate disclosures," *Accounting, Organizations and Society*, 68, 88–108.

Holub, F. and B. Thies (2023): "Air quality, high-skilled worker productivity and adaptation: Evidence from github," *High-Skilled Worker Productivity and Adaptation: Evidence from Github (March 6, 2023)*.

Hsieh, C.-T. and E. Rossi-Hansberg (2023): "The industrial revolution in services," *Journal of Political Economy Macroeconomics*, 1, 3–42.

Huang, K., M. Li, and S. Markov (2020): "What do employees know? Evidence from a social media platform," *The Accounting Review*, 95, 199–226.

Lashkari, D., A. Bauer, and J. Boussard (2018): "Information technology and returns to scale," *Available at SSRN 3458604*.

Lemieux, T., W. B. MacLeod, and D. Parent (2009): "Performance pay and wage inequality," *The Quarterly Journal of Economics*, 124, 1–49.

Ma, K. A., S. Y. Sheng, and H. Xie (2023): "Employer Reputation and the Labor Market: Evidence from Glassdoor. com and Dice. com," *Available at SSRN 4437565*.

McDermott, G. R. and B. Hansen (2021): "Labor reallocation and remote work during COVID-19: Real-time evidence from GitHub," Tech. rep., National Bureau of Economic Research.

Mueller, H. M., P. P. Ouimet, and E. Simintzi (2017): "Within-firm pay inequality," *The Review of Financial Studies*, 30, 3605–3635.

Oi, W. Y. and T. L. Idson (1999): "Firm size and wages," *Handbook of labor economics*, 3, 2165–2214.

Pacelli, J., T. Shi, and Y. Zou (2022): "Communicating corporate culture in labor markets: Evidence from job postings," *SSRN Electronic Journal.*

Prat, A. and T. Valletti (2022): "Attention oligopoly," *American Economic Journal: Microeconomics*, 14, 530–557.

Rossi, A. G. and S. P. Utkus (2020): "Who benefits from robo-advising? Evidence from machine learning," *Evidence from Machine Learning (March 10, 2020).*

Sheng, J. (2023): "Asset pricing in the information age: Employee expectations and stock returns," *Available at SSRN 3321275.*

Smith, A. (2015): "Pew Research, Searching for Work in the Digital Era," `https://www.pewresearch.org/internet/2015/11/19/searching-for-work-in-the-digital-era`.

Sockin, J. and A. Sojourner (2023): "What's the inside scoop? challenges in the supply and demand for information on employers," *Journal of Labor Economics*, 41, 1041–1079.

Tambe, P., L. Hitt, D. Rock, and E. Brynjolfsson (2020): "Digital capital and superstar firms," Tech. rep., National Bureau of Economic Research.

Tambe, P. and L. M. Hitt (2012): "The productivity of information technology investments: New evidence from IT labor data," *Information systems research*, 23, 599–617.

Troske, K. R. (1999): "Evidence on the employer size-wage premium from worker-establishment matched data," *Review of Economics and Statistics*, 81, 15–26.

Veldkamp, L. and C. Chung (2024): "Data and the aggregate economy," *Journal of*

Economic Literature, 62, 458–484.

Waldman, M. (1984): "Job assignments, signalling, and efficiency," *The RAND Journal of Economics*, 15, 255–267.

——— (1990): "Up-or-out contracts: A signaling perspective," *Journal of Labor Economics*, 8, 230–250.

Wallskog, M., N. Bloom, S. W. Ohlmacher, and C. Tello-Trillo (2024): "Within-Firm Pay Inequality and Productivity," Tech. rep., National Bureau of Economic Research.

Figures & Tables

Figure 1: **Our Sample GitHub Users by MSA**

Notes: This figure presents the distribution of our sample by employee's location (MSA) based on their first job.

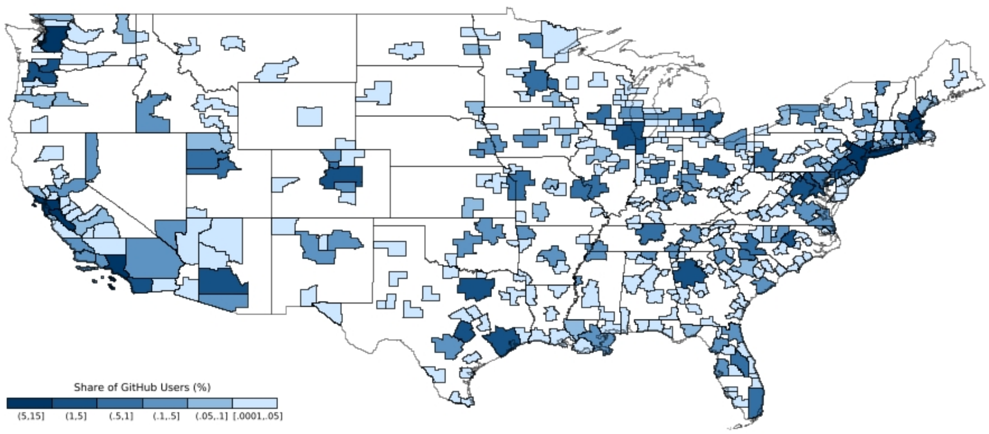


Figure 2: **GitHub Shock & Employee Mobility to Large Firms**

Notes: This figure presents an event study on employee mobility to large firms around the disclosure of employee productivity information on GitHub. The x-axis plots the quarter around the shock. The dashed line indicates the quarter April-June 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter immediately before April-June 2016 is the omitted quarter set equal to zero. On the y-axis, we plot the differential impact of the shock on more productive employees estimated by the β coefficients obtained from the equation:

$$\mathbb{1}(MoveToLargeFirm)_{i,t} = \sum_{\tau=-8}^{12} \beta_\tau Productivity_i \times \mathbb{1}(\tau = t) + \gamma X_{it} + \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with time indicator for each quarter. The dependent variable is an indicator that equals 100% when employee i switches job to a large firm (>1000 employees) in quarter t and is 0 otherwise. We control for time-varying salary and seniority as well as employee fixed effects, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Light blued lines show 95% confidence intervals for estimates.

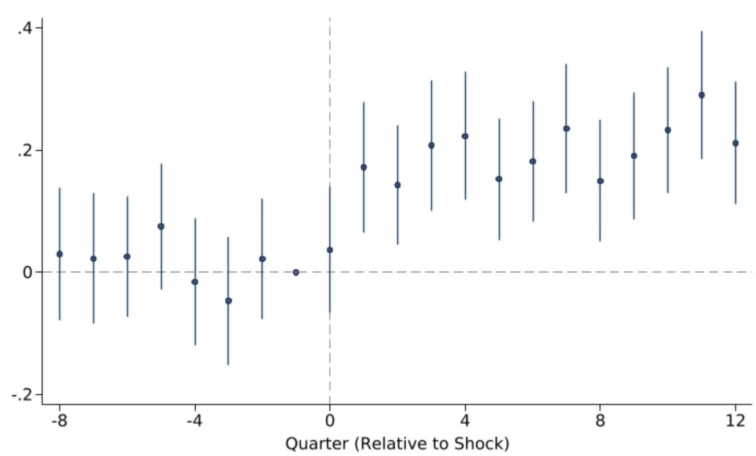


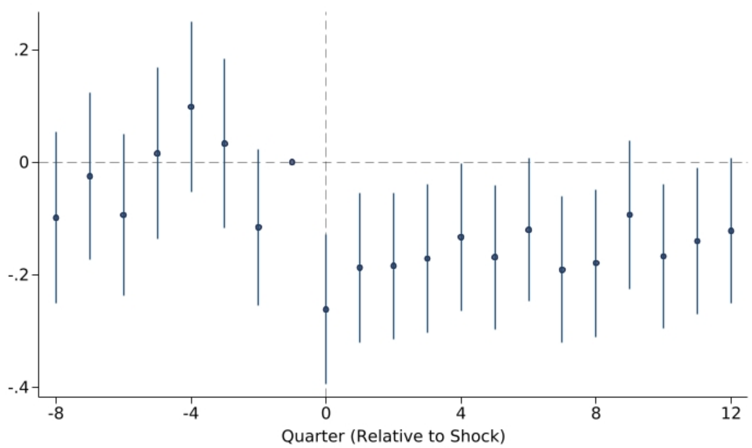
Figure 3: **GitHub Shock & Employee Mobility**

Notes: This figure presents event studies on employee mobility to small-sized and medium-sized firms around the disclosure of employee productivity information on GitHub. The x-axis plots the quarter around GitHub's policy change. The dashed line indicates the quarter April-June 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter immediately before April-June 2016 is the omitted quarter set equal to zero. On the y-axis, we plot the differential impact of the shock on more productive employees estimated by the β coefficients obtained from regression estimate:

$$\mathbb{1}(Move)_{i,t} = \sum_{\tau=-8}^{12} \beta_\tau Productivity_i \times \mathbb{1}(\tau = t) + \gamma X_{it} + \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with time indicator for each quarter. The dependent variable is an indicator that equals 100% when employee i switches job to a mid-sized firm (50 $<$ employees $<=$ 1000) for Panel (a), to a small firms ($<=$ 50 employee) for Panel (b), in quarter t and is 0 otherwise. We control for time-varying salary and seniority as well as employee fixed effects, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Light blued lines show 95% confidence intervals for estimates.

(a) Move to Mid-Sized Firms



(b) Move to Small Firms

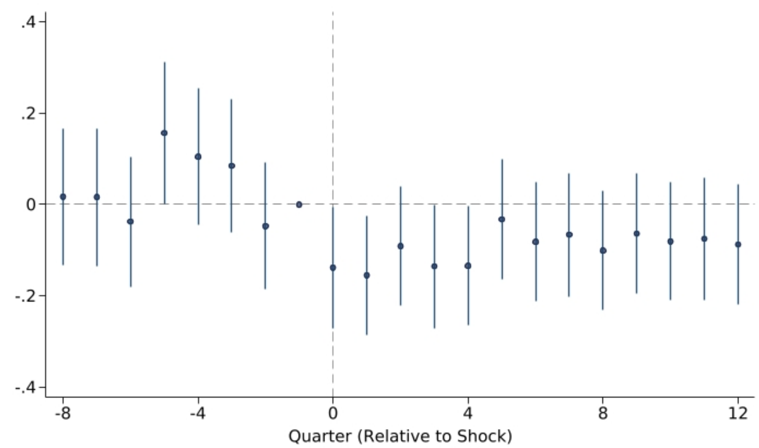


Figure 4: **GitHub Shock & Employee Mobility to Different Size Buckets**

Notes: This figure presents the economic magnitude of employee mobility to firms in different size buckets around the disclosure of employee productivity information on GitHub. On the y-axis, we plot the economic magnitude of the differential impact of the shock on more productive employees using the β coefficients obtained from regression estimate:

$$\mathbb{1}(Move)_{i,t} = \sum_{\tau=-8}^{12} \beta_\tau Productivity_i \times \mathbb{1}(\tau = t) + \gamma X_{it} + \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with time indicator for each quarter. The dependent variable is an indicator that equals 100% when employee i switches job to a firm in a given size bucket. We control for time-varying salary and seniority as well as employee fixed effects, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively.

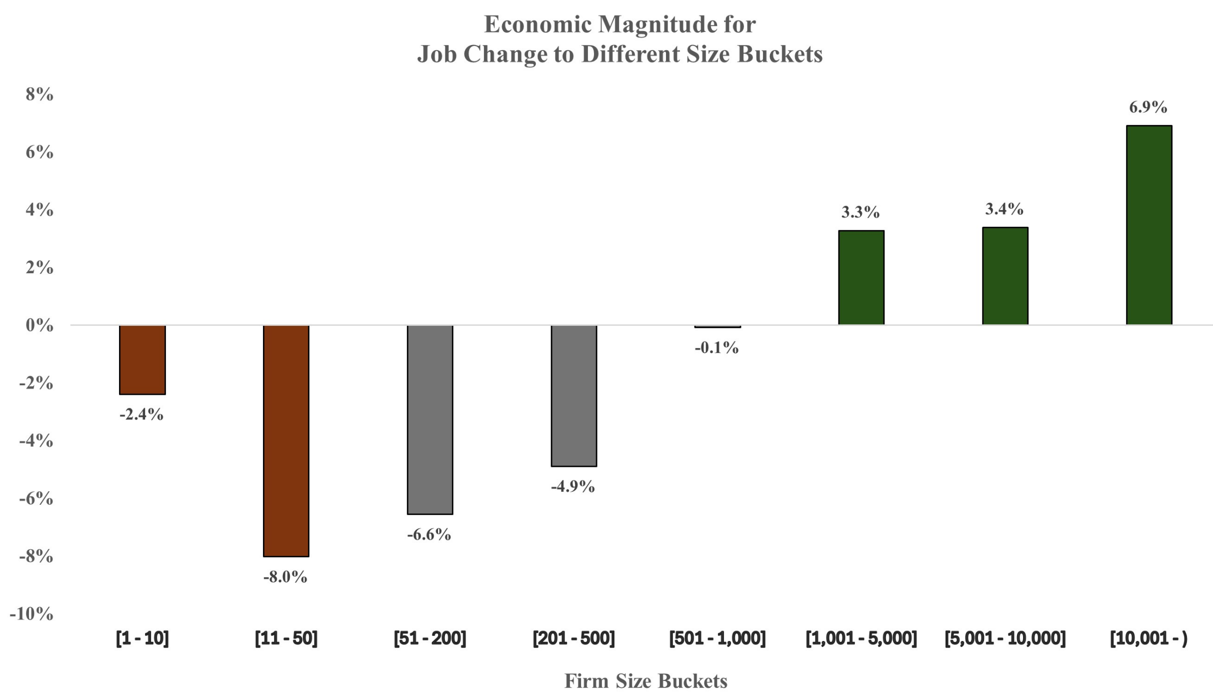


Figure 5: **GitHub Shock & Employee Mobility to Large Firms: Heterogeneity by Initial Firm Size**

Notes: This figure presents event studies on employee mobility to large firms from small and mid-sized firms around the disclosure of employee productivity information on GitHub. The x-axis plots the quarter around the shock. The dashed line indicates the quarter April-June 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter immediately before April-June 2016 is the omitted quarter set equal to zero. On the y-axis, we plot the differential impact of the shock on more productive employees, particularly at small and mid-sized firms, estimated by the β coefficients obtained from the equation:

$$\mathbb{1}(MoveToLargeFirm)_{i,t} = \sum_{\tau=-8}^{12} \gamma_\tau Productivity_i \times \mathbb{1}(\tau = t) + \sum_{\tau=-8}^{12} \beta_\tau Productivity_i \times \mathbb{1}(\tau = t) \times \mathbb{1}(Mid - Sized_{i(t-1)})$$
$$+ \sum_{\tau=-8}^{12} \beta_\tau^{'} Productivity_i \times \mathbb{1}(\tau = t) \times \mathbb{1}(Small_{i(t-1)}) + \alpha_i + \delta_t + \nu_{it} + \varepsilon_{i,t}$$

The independent variable is a triple interaction of three indicators: normalized productivity estimate of the employee i derived from the AKM decomposition, indicators for the size of the firm where user worked in the previous quarter, and indicator for time relative. Panel A plots interaction with *Mid*, an indicator variable that takes the value of 1 for firms with 50-1000 employees. Panel B plots interaction with *Small*, an indicator for firms with less than 50 employees. The dependent variable is an indicator that equals 100% if the employee switches job to a large firm ($>$ 1000 employees) in that quarter. We control for time-varying salary and seniority as well as employee fixed effects, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Light blued lines show 95% confidence intervals for estimates.

(a) Mid-Sized Firms (b) Small Firms

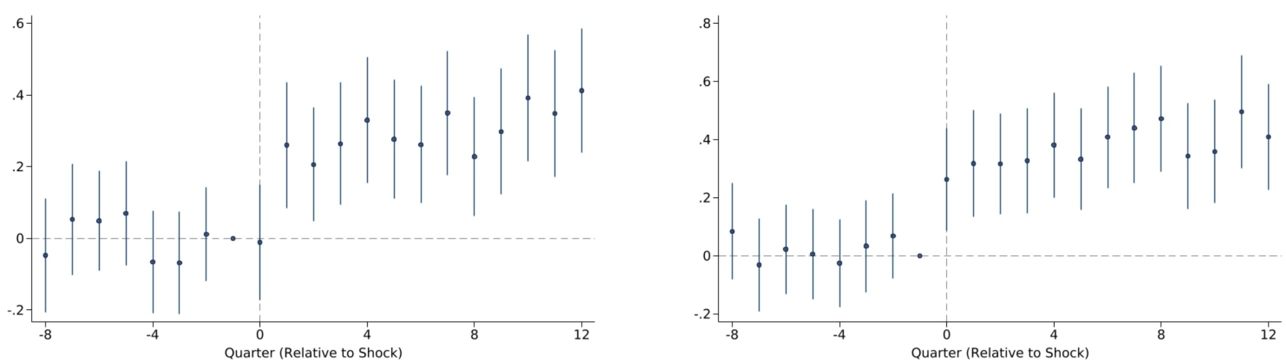


Figure 6: **GitHub Shock & Employee Promotions at Large Firms**

Notes: This figure presents an event study on employees' within-firm promotions at large firms around the disclosure of employee productivity information on GitHub. The x-axis plots the quarter around the shock. The dashed line indicates the quarter April-June 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter immediately before April-June 2016 is the omitted quarter set equal to zero. On the y-axis, we plot the differential impact of the shock on more productive employees, particularly at large firms, estimated by the β coefficients obtained from the equation:

$$\begin{aligned}\mathbb{1}(Within - FirmPromotion)_{i,t} = &\sum_{\tau=-8}^{12} \beta_\tau Productivity_i \times \mathbb{1}(\tau = t)\\ &+ \sum_{\tau=-8}^{12} \gamma_\tau Productivity_i \times \mathbb{1}(\tau = t) \times \mathbb{1}(Mid - Sized_{i(t-1)})\\ &+ \sum_{\tau=-8}^{12} \gamma_\tau^{'} Productivity_i \times \mathbb{1}(\tau = t) \times \mathbb{1}(Small_{i(t-1)}) + \alpha_i + \delta_t + \nu_{it} + \varepsilon_{i,t}\end{aligned}$$

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with time indicator for each quarter. The dependent variable is an indicator that equals 100% when employee i does not switch jobs and her salary or seniority improves in quarter t relative to quarter *t-1* and is 0 otherwise. We control for time-varying salary and seniority as well as employee fixed effects, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Light blued lines show 95% confidence intervals for estimates.

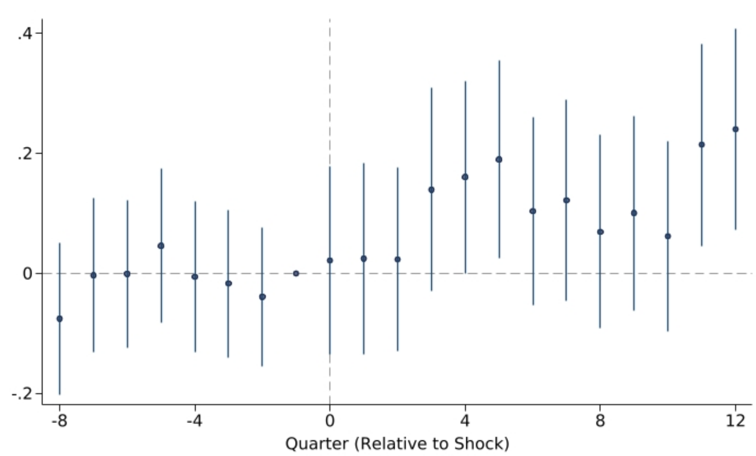


Figure 7: **GitHub Shock & Employment Growth at Small Firms**

Notes: This figure presents an event study on firm-level employment growth for small firms (<= 50 employees) around the disclosure of employee productivity information on GitHub. The x-axis plots the quarter around GitHub's policy change. The dashed line indicates the quarter April-June 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter immediately before April-June 2016 is the omitted quarter set equal to zero. On the y-axis, we plot the differential impact of the shock on employment growth of small firms with above-median employee productivity estimated by the β coefficients obtained from the following equation:

$$\begin{aligned} EmploymentGrowth_{g,t} = & \sum_{\tau=-8}^{12} \gamma_\tau \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(\tau = t) \\ & + \sum_{\tau=-8}^{12} \theta_\tau \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(MidSized_g) \\ & + \sum_{\tau=-8}^{12} \beta_\tau \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(Small_g) + \alpha_g + \delta_t + \varepsilon_{g,t} \end{aligned}$$

The independent variable is a triple interaction of three indicators: $\mathbb{1}(Productivity_g > Median)$ takes the value of 1 for if firm g has above median AKM based employed productivity, $\mathbb{1}(Small_g)$ takes the value of 1 if firm g is small (<= 50 employees), and $\mathbb{1}(\tau = t)$ takes the value 1 for time τ. The dependent variable is the QoQ employment change (in %) for firm g in quarter t. We control for the firm's lagged employment and include firm and industry $\times$ state interacted with quarter fixed effects. We cluster the standard errors at the firm level. Light blued lines show 95% confidence intervals for estimates.

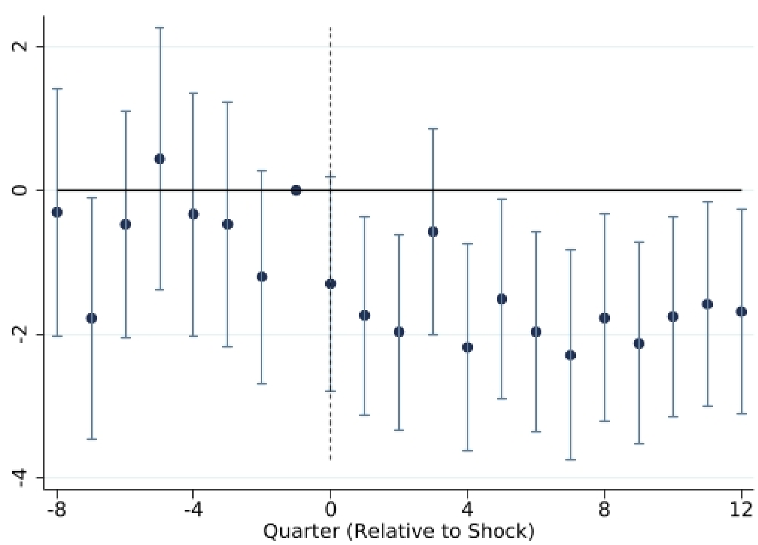


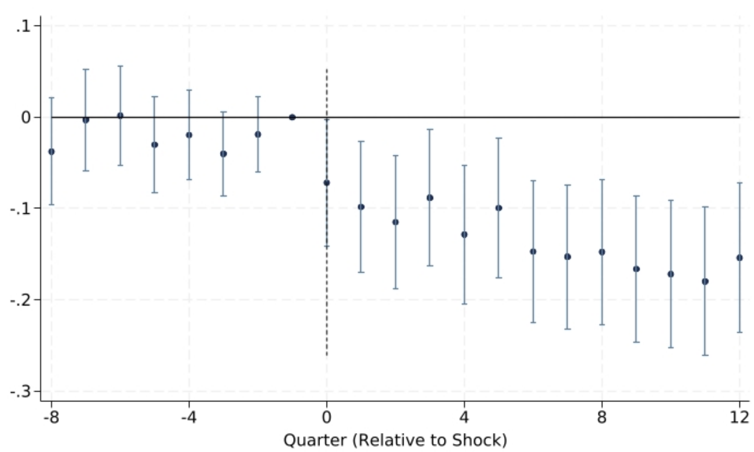
Figure 8: **GitHub Shock & Productivity at Small Firms**

Notes: This figure presents event studies on firm-level productivity of small firms (<= 50 employees) proxied by their public GitHub contributions around the disclosure of employee productivity information on GitHub. The x-axis plots the quarter around GitHub's policy change. The dashed line indicates the quarter April-June 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter immediately before April-June 2016 is the omitted quarter set equal to zero. On the y-axis, we plot the differential impact of the shock on public contributions of small firms with above-median employee productivity estimated by the β coefficients obtained from the following equation:

$$Contributions_{g,t} = \sum_{\tau=-8}^{12} \gamma_\tau \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(\tau = t) + \sum_{\tau=-8}^{12} \theta_\tau \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(MidSized_g)$$
$$+ \sum_{\tau=-8}^{12} \beta_\tau \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(Small_g) + \alpha_g + \delta_t + \varepsilon_{g,t}$$

The independent variable is a triple interaction of three indicators: $\mathbb{1}(Productivity_g¿Median)$ takes the value of 1 for if firm g has above median AKM based employee productivity, $\mathbb{1}(Small_g)$ takes the value of 1 if firm g is small (<= 50 employees), and $\mathbb{1}(\tau = t)$ takes the value 1 for time τ. The dependent variable is the inverse hyperbolic sine of total public contributions for panel (a), and the inverse hyperbolic sine of per-capita public contributions for panel (b), made by employees working at the firm g in quarter t. We control for the firm's lagged stock of total contributions as well as the median change in public contributions in the post-period. We also include firm and industry $\times$ state interacted with quarter fixed effects. We cluster the standard errors at the firm level. Light blued lines show 95% confidence intervals for estimates.

(a) Asinh(Public Contributions)



(b) Asinh(Public Contributions per Employee)

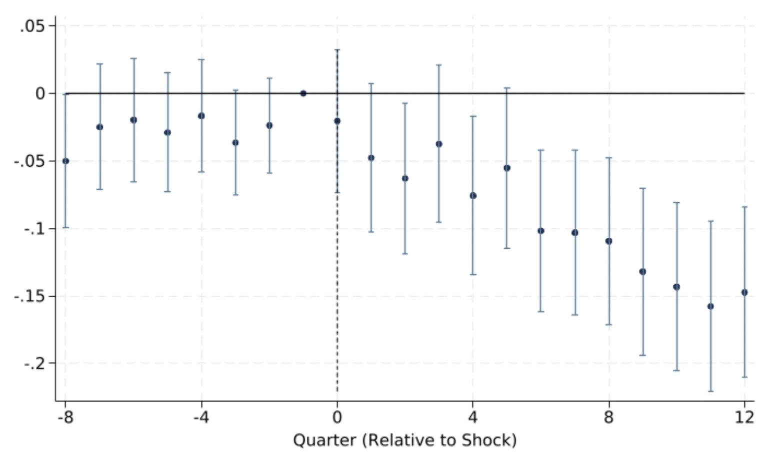


Table 1: **Descriptive statistics**

Notes: This table presents descriptive statistics for employee-level cross-sectional variables in Panel A, employee-level time-varying variables in Panel B, all in-sample firms in Panel C, and Compustat firms that have at least 10 employees in our sample in Panel D.

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	**Std. Dev.**	**P10**	**Median**	**P90**	**Obs.**
A: Employee Cross Sectional:						
$\mathbb{1}$(Male)	0.79	0.41	0	1	1	302,903
$\mathbb{1}$(White)	0.69	0.46	0	1	1	302,903
$\mathbb{1}$(Software Engg. Role)	0.37	0.48	0	0	1	302,903
$\mathbb{1}$(SF/NY/Seattle)	0.25	0.43	0	0	1	302,581
Experience (Years)	8.82	8.13	1	7	20	302,903
Initial Seniority	2.20	0.93	1	2	4	302,903
Initial Salary (in '000 $s)	87.34	26.85	56.73	84.88	116.66	302,903
Pre-Period Total Contributions (Log)	2.60	2.16	0	2.08	5.87	302,903
Pre-Period Public Contributions (Log)	2.35	1.91	0	1.95	5.17	302,903
Pre-Period Private Contributions (Log)	0.66	1.86	0	0	3.22	302,903
B: Employee Time Varying:						
$\mathbb{1}$(Job Change to Large Firms)	0.03	0.04	0	0	0.08	302,902
$\mathbb{1}$(Job Change to Mid-Size Firms)	0.02	0.04	0	0	0.08	302,902
$\mathbb{1}$(Job Change to Small Firms)	0.02	0.03	0	0	0.06	302,902
$\mathbb{1}$(Job Change)	0.07	0.07	0	0.05	0.17	302,903
$\mathbb{1}$(Promotion)	0.06	0.05	0	0.04	0.13	302,903
$\mathbb{1}$(Across-Firm Promotion)	0.04	0.05	0	0.04	0.10	302,903
$\mathbb{1}$(Within-Firm Promotion)	0.02	0.03	0	0	0.07	302,903
C: Firm Level (All Firms):						
$\mathbb{1}$(Large Firm) (Employees > 1000)	0.15	0.36	0	0	1	57,982
$\mathbb{1}$(Mid-size Firm) (50 < Employees <= 1000)	0.37	0.48	0	0	1	57,982
$\mathbb{1}$(Small Firm) (Employees <= 50)	0.48	0.50	0	0	1	57,982
Employment Change	0.03	0.09	-0.03	0.02	0.13	55,335
Public Contributions	46.49	739.07	0.17	4	65	50,308
Public Contributions/ Employee	10.26	107.47	0.13	1.75	17.26	50,307
D: Firm Level (Compustat Firms):						
Firm Size (in '000 Employees)	24.05	74.32	0.36	5.10	60.08	1,675
Daily Excess Return (in %)	-0.03	0.14	-0.18	-0.02	0.09	1,675

Table 2: **Sample & Variation breakdown for AKM Tests**

Notes: This table presents the details of sample construction (in Panel A) and variance decomposition (in Panel B) of the AKM (Abowd et al., 1999) tests done to estimate individual employee productivity from private GitHub contributions. We isolate employee productivity from historical private GitHub contributions, by running following regression specification:

$$Log(PvtContributions)_{i,f,t} = \alpha_i + \alpha_f + \alpha_t + X_{i,f,t} + \varepsilon_{i,f,t}$$

α_i, the employee fixed effect, represents the estimated employee *i's* productivity. The dependent variable is the logarithm of the number of private contributions made by employee i in month t. The regression is estimated on data spanning May 2011 to April 2016, exactly five years before GitHub policy change. We control for firm fixed effects, time fixed effects, job role, current job duration, and total work experience. Panel A shows the number of observations in raw sample, the connected set (the largest set of firms linked by employee mobility), filtered set (connected set with at least two employees at each firm-time), the sample for which productivity estimates were obtained, and the final sample obtained after excluding users working in non-profit industries at the time of GitHub's policy change. Panel B summarizes the breakdown of the variance in the dependent variable explained by the employee and firm components.

	(1)	(2)	(3)	(4)
Panel A: Summary Stats				
	Employees		**Parent Firms**	
Sample	N	% Raw Sample	N	% Raw Sample
Raw Sample	381,114		103,308	
Connected Set	356,106	93.4%	77,899	75.4%
Filtered Set	342,769	89.9%	36,358	35.2%
Productivity Estimates	338,523	88.8%	36,358	35.2%
Excluding Non-Profits	302,903	79.5%		

Panel B: Variation Breakdown	
	Var. Explained
Person F.E.	66.7%
Firm F.E.	8.5%
Residual	24.7%

Table 3: **GitHub Shock & Employee Mobility to Large Firms**

Notes: This table presents estimates on the impact of GitHub's disclosure of employee productivity information on employee mobility to large firms. We report the differential impact of GitHub's policy change on productive employees estimated by the β coefficient from the equation:

$$\mathbb{1}(MoveToLargeFirm)_{i,t} = \beta Productivity_i \times \mathbb{1}(Post_t) + \gamma X_{it} + \alpha_i + \theta_t + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with *Post*, and indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). The dependent variable is an indicator that equals 1 when employee i switches job to a large firm (>1000 employees) in quarter t and is 0 otherwise. All specifications include employee and quarter fixed effects. Columns 1 to 5 present estimates with varying controls. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)
Outcome:	**$\mathbb{1}$(Job Change to Large Firms)**				
Productivity × $\mathbb{1}$(Post)	0.0013***	0.0012***	0.0017***	0.0016***	0.0015***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Employee F.E.	Y	Y	Y	Y	Y
Time F.E.	Y	Y			
Salary & Seniority Controls		Y	Y	Y	Y
Cohort × Time F.E.			Y		
Cohort × Language × Role × Time F.E.				Y	Y
Industry × Location × Time F.E.					Y
Observations	6,804,386	6,804,386	6,804,166	3,811,484	3,199,173
# Unique Employees	302,781	302,781	302,772	170,880	143,446
R-squared	0.054	0.054	0.06	0.168	0.212
Y-Mean	0.031	0.031	0.031	0.032	0.033
Magnitude (%)	4.04	3.84	5.32	5.69	5.27

Table 4: **GitHub Shock & Employee Mobility to Large Firms: Placebo & Alternate Samples**

Notes: This table presents placebo and alternate sample tests for estimates reported in Table 3. Panel A presents a placebo test with the independent variable being an alternative productivity estimate computed by using employee's *public* contributions as the dependent variable in the AKM decomposition. Panel B estimates results on a subsample of active GitHub Users. Panel B1 only includes users with more than 10 total GitHub contributions during the pre-period (in the 12 quarters before May 2016). Panel B2, only includes users employed (at the time of shock) at firms with 100 or more total GitHub contributions during the pre-period. Panel B3 restricts the sample to users with non-zero private GitHub contributions during the pre-period. Panel B4 restricts the sample to users with at least 10 private GitHub contributions during the pre-period. All tests run tests corresponding to our baseline specification, Column 5 in Table 3. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	**Coefficient**	**Std. Error**	**Observations**	**R-squared**	**Y-Mean**	**Magnitude (%)**
Panel A: Placebo Test						
A1: Productivity based on Public Contributions						
Productivity × $\mathbb{1}$(Post)	-0.0004	(0.0003)	3,199,024	0.211	0.033	-1.4
Panel B: Within Active User Comparison						
B1: Pre Period Individual Total Contributions > 10						
Productivity × $\mathbb{1}$(Post)	0.0018***	(0.0002)	1,708,418	0.224	0.033	7.8
B2: Pre-period Firm Contributions >= 100						
Productivity × $\mathbb{1}$(Post)	0.0042***	(0.0002)	1,748,870	0.259	0.0025	12.56
B3: Pre-period Individual Private Contributions > 0						
Productivity × $\mathbb{1}$(Post)	0.0011***	(0.0003)	506,974	0.271	0.043	7.65
B4: Pre-period Individual Private Contributions >= 10						
Productivity × $\mathbb{1}$(Post)	0.0011***	(0.0003)	401,146	0.274	0.024	8.35

Table 5: **GitHub Shock & Employee Mobility to Medium & Small Firms**

Notes: This table presents estimates on the heterogeneous impact of GitHub's disclosure of employee productivity information on employee mobility based on destination firm size. We report the differential impact of GitHub's policy change on productive employees estimated by the β coefficient from the equation:

$$\mathbb{1}(MoveToLargeFirm)_{i,t} = \beta Productivity_i \times \mathbb{1}(Post_t) + \gamma X_{it} + \alpha_i + \theta_t + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with *Post*, and indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). The dependent variable is an indicator that equals 1 when employee i switches job in quarter t to a large firm (>1000 employees) for Column 1, mid-sized firm (50 < employees <= 1000) for Column 2, small firm (<= 50 employees) for Column 3, and to any firm for Column 4. We control for time-varying salary and seniority as well as employee fixed effects, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)
Outcome:		**$\mathbb{1}$(Job Change)**	
	Large Firm	**Medium Firm**	**Small Firm**
Productivity × $\mathbb{1}$(Post)	0.0015***	-0.0013***	-0.0012***
	(0.0002)	(0.0002)	(0.0002)
Salary & Seniority Controls	Y	Y	Y
Employee F.E.	Y	Y	Y
Cohort × Language × Role × Time F.E.	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y
Observations	3,199,173	3,199,173	3,199,173
# Unique Employees	143,446	143,446	143,446
R-squared	0.212	0.211	0.222
Y-Mean	0.033	0.03	0.023
Magnitude (%)	5.27	-4.83	-5.99

Table 6: **GitHub Shock & Employee Mobility to Large Firms: Heterogeneity by Initial Firm Size**

Notes: This table presents estimates on the heterogeneous impact of GitHub's disclosure of employee productivity information on employee mobility to large firms based on the *initial* firm size. We report the differential impact of GitHub's policy change on productive employees by initial firm type estimated by the β coefficient from the equation:

$$\mathbb{1}(MoveToLargeFirm)_{i,t} = \gamma X_{it} + \beta_1 Productivity_i \times \mathbb{1}(Post_t) + \beta_2 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_{i(t-1)}) + \beta_3 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Small)_{i(t-1)}) + \alpha_i + \delta_t + \varepsilon_{i,t}$$

The independent variable is a triple interaction of three indicators: normalized productivity estimate of employee i derived from the AKM decomposition, indicators for the size of the firm where employee i worked in the previous quarter, and indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). *Mid* is an indicator variable that takes the value of 1 for firms with 50-1000 employees, and *Small* is an indicator for firms with less than 50 employees. The dependent variable is an indicator that equals 1 when employee i switches job to a large firm (>1000 employees) in quarter t and is 0 otherwise. We do not report other double interactions for brevity. Columns 1 to 5 present estimates with varying controls. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Significance levels: **($p < 0.05$), ***($p < 0.01$).

	(1)	(2)	(3)	(4)	(5)
Outcome:	**𝟙(Job Change to Large Firms)**				
Productivity × 𝟙(Post)	-0.0037***	-0.0039***	-0.0029***	-0.0025***	-0.0025***
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
Productivity × 𝟙(Post) × 𝟙(Mid)	0.0045***	0.0047***	0.0042***	0.0031***	0.0030***
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0005)
Productivity × 𝟙(Post) × 𝟙(Small)	0.0050***	0.0053***	0.0046***	0.0041***	0.0040***
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0005)
Employee F.E.	Y	Y	Y	Y	Y
Time F.E.	Y	Y			
Salary & Seniority Controls		Y	Y	Y	Y
Cohort × Time F.E.			Y		
Cohort × Language × Role × Time F.E.				Y	Y
Industry × Location × Time F.E.					Y
Observations	6,455,765	6,455,765	6,455,545	3,596,761	3,022,682
# Unique Employees	300,779	300,779	300,770	169,629	142,754
R-squared	0.069	0.069	0.075	0.184	0.227
Y-Mean	0.03	0.03	0.03	0.031	0.032
Magnitude (Small) (%)	4.56	4.48	5.62	5.99	5.42

Table 7: **GitHub Shock & Employee Promotion**

Notes: This table presents estimates on the impact of GitHub's disclosure of employee productivity information on employee career advances. We report the differential impact of GitHub's policy change on productive employees by initial firm type estimated by the β coefficient from the equation:

$$\mathbb{1}(Promotion)_{i,t} = \gamma X_{it} + \beta_1 Productivity_i \times \mathbb{1}(Post_t) + \beta_2 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_{i(t-1)}) \\ + \beta_3 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Small)_{i(t-1)}) + \alpha_i + \delta_t + \varepsilon_{i,t}$$

The independent variable is an interaction of two indicators: normalized productivity estimate of the employee i derived from the AKM decomposition, and an indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). Even-numbered columns additionally interact this term with indicators for the size of the firm where user worked in the previous quarter. *Mid* is an indicator variable that takes the value of 1 for firms with 50-1000 employees, and *Small* is an indicator for firms with less than 50 employees. The dependent variable is an indicator variable that is 1 if there is advancement (salary or job-seniority increase from previous quarter), and the employee switches job to a large firm (> 1000 employees) for Columns 1 and 2, switches job to a small or mid-sized firm (<= 1000 employees) for Columns 3 and 4, remains at the same firm for columns 5 and 6. We do not report other double interactions for brevity. We control for time-varying salary and seniority as well as employee fixed effects, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. In columns 5 and 6, we also control for time-varying firm cohort representing an interaction of employee's firm and the year of joining the firm. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:			$\mathbb{1}$**(Promotion)**			
	Move to Large		*Move to Small & Med.*		*Within-Firm*	
Productivity × $\mathbb{1}$(Post)	0.0011***	-0.0018***	-0.0020***	-0.0039***	0.0008***	0.0011**
	(0.0001)	(0.0003)	(0.0002)	(0.0006)	(0.0002)	(0.0005)
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Mid)		0.0023***		0.0032***		0.0002
		(0.0004)		(0.0007)		(0.0006)
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Small)		0.0028***		0.0042***		-0.0004
		(0.0004)		(0.0007)		(0.0006)
Employee F.E.	Y	Y	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y	Y	Y
Salary & Seniority Controls	Y	Y	Y	Y	Y	Y
Cohort × Language × Role × Time F.E.	Y	Y	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y	Y	Y
Firm Cohort					Y	Y
Observations	3,199,173	3,022,682	3,199,173	3,022,682	2,860,181	2,828,288
# Unique Employees	143,446	142,754	143,446	142,754	142,827	142,521
R-squared	0.204	0.216	0.212	0.22	0.269	0.271
Y-Mean	0.024	0.023	0.036	0.035	0.019	0.019
Magnitude (%) (Main)	5.42	-9	-6.16	-12.7	4.96	6.25
Magnitude (%) (Small)		5.11		0.84		4.19

Table 8: **GitHub Shock & Employee Mobility to Large Firms: Heterogeneity by Employee Type**

Notes: This table presents estimates on the heterogeneous impact of GitHub's disclosure of employee productivity information based on employee-type. We report the differential impact of GitHub's policy change on productive employees, particularly those susceptible to higher information asymmetry, estimated by the β coefficients from the equation:

$$\begin{aligned}\mathbb{1}(MoveToLargeFirm)_{i,t} &= \beta_1 Productivity_i \times \mathbb{1}(Post_t) \\ &+ \beta_2 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Group_i) + \gamma X_{it} + \alpha_i + \theta_t + \gamma_{it} + \varepsilon_{i,t}\end{aligned}$$

The independent variable is a triple interaction of three indicators: the employee's normalized productivity estimate derived from the AKM decomposition, indicators for employee group, and indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). Columns 1-2 include triple interaction for $\mathbb{1}$(Experience $<$ Median), which takes the value of 1 for employees that have less than median work experience at the time of shock. Columns 3-4 (5-6) include triple interactions with $\mathbb{1}$(Non-Elite School) that takes the value 1 if the employee did not attend a school ranked among the top 30 in the US (world) by US News & Ranking (QS University Rankings) in 2016. Columns 7-8 include triple interaction with $\mathbb{1}$(Low-Network School) that takes the value of 1 if the employee did not attend a school that is among the top 5 universities any large firm hired from in the pre-period. All specifications include salary and seniority controls as well as employee fixed effects and employee's cohort interacted with language, role, and quarter fixed effects. In even-numbered columns, we also control for location and industry interacted with time. Cohort refers to the employee's final education year.Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p $<$ 0.05), ***(p $<$ 0.01).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	**$\mathbb{1}$(Job Change to Large Firms)**							
Productivity $\times$ $\mathbb{1}$(Post)	0.0013***	0.0012***	0.0007	0.0006	0.0005	0.0004	0.0015***	0.0014***
	(0.0002)	(0.0002)	(0.0005)	(0.0006)	(0.0006)	(0.0007)	(0.0001)	(0.0002)
Productivity $\times$ $\mathbb{1}$(Post) $\times$ $\mathbb{1}$(Experience $<$ Median)	0.0005*	0.0006**						
	(0.0003)	(0.0003)						
Productivity $\times$ $\mathbb{1}$(Post) $\times$ $\mathbb{1}$(Non-Elite School - US)			0.0009*	0.0010*				
			(0.0005)	(0.0006)				
Productivity $\times$ $\mathbb{1}$(Post) $\times$ $\mathbb{1}$(Non-Elite School - Global)					0.0011*	0.0012*		
					(0.0006)	(0.0007)		
Productivity $\times$ $\mathbb{1}$(Post) $\times$ $\mathbb{1}$(Low-Network School)							0.0006*	0.0007*
							(0.0003)	(0.0004)
Employee F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Salary & Seniority Controls	Y	Y	Y	Y	Y	Y	Y	Y
Cohort $\times$ Language $\times$ Role $\times$ Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Industry $\times$ Location $\times$ Time F.E.		Y		Y		Y		Y
Observations	4,397,956	3,112,896	4,397,956	3,112,896	4,397,956	3,112,896	4,397,956	3,112,896
# Unique Employees	196,369	139,680	196,369	139,680	196,369	139,680	196,369	139,680
R-squared	0.092	0.213	0.092	0.213	0.092	0.213	0.091	0.212

Table 9: **GitHub Shock & Employee Mobility to Large Firms: Heterogeneity by Destination Firm Characteristics**

Notes: This table presents estimates on the heterogeneous impact of GitHub's disclosure of employee productivity information based on destination firm characteristics. We report the differential impact of GitHub's policy change on productive employees, particularly those susceptible to higher information asymmetry, estimated by the β coefficients from the equation:

$$\mathbb{1}(MoveToLargeFirm)_{i,t} = \beta Productivity_i \times \mathbb{1}(Post_t) + \gamma X_{it} + \alpha_i + \theta_t + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is an interaction of the employee's normalized productivity estimate derived from the AKM decomposition and an indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). The dependent variable is mobility to a subset of large firms (> 1,000 employees) and differs in each column based on the subset. In columns 1-2, the subset is based on steepness in career growth defined by whether the firm had an above-median average annual salary change in the pre-period. In columns 3-4, the subset is based on job stability defined by whether the firm had below-median average forced turnovers in the pre-shock period. In columns 5-6, the subset is based on whether any Human Resource professional at the firm used GitHub during the pre-period, and in columns 7-8, the subset is based on whether the firm and its employees use GitHub (derived from firm-level private contributions). We control for time-varying salary and seniority as well as employee fixed effects, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. In columns 5 and 6, we also control for time-varying firm cohort representing an interaction of employee's firm and the year of joining the firm. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	**$\mathbb{1}$(Job Change to Large Firms with)**							
	Steep Career Growth		**High Stability**		**Tech-savvy Recruiters**		**GitHub Usage**	
	Yes	No	Yes	No	Yes	No	Yes	No
Productivity × $\mathbb{1}$(Post)	0.0014***	0.0001	0.0002**	0.0014***	0.0014***	0.0002*	0.0014***	0.0001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)
Employee F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Salary & Seniority Controls	Y	Y	Y	Y	Y	Y	Y	Y
Cohort × Language × Role × Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3,192,202	3,192,202	3,194,104	3,194,104	3,199,173	3,199,173	3,199,173	3,199,173
# Unique Employees	143,445	143,445	143,445	143,445	143,446	143,446	143,446	143,446
R-Squared	0.2	0.221	0.207	0.205	0.189	0.23	0.204	0.266
Y-Mean	0.024	0.007	0.008	0.024	0.018	0.015	0.03	0.003
Magnitude (%)	6.81	1.56	2.56	6.65	8.43	1.31	5.44	3.6
T-test ($H_0 : \beta_{Yes} = \beta_{No}$)	(0.0000)***		(0.0000)***		(0.0000)***		(0.0197)**	

Table 10: **GitHub Shock & Firm Outcomes**

Notes: This table presents estimates on the impact of GitHub's disclosure of employee productivity information on firm employment and productivity. We report the differential impact of GitHub's policy change on small and mid-sized firms with more productive employees estimated by the β coefficient from the equation:

$$FirmOutcome_{g,t} = \beta_1 \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(Post_t) + \beta_2 \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_g) + \beta_3 \mathbb{1}(Productivity_g > Median) \times \mathbb{1}(Post_t) \times \mathbb{1}(Small_g) + \alpha_g + \delta_t + \varepsilon_{g,t}$$

The independent variable is a triple interaction of three indicators: $\mathbb{1}(Productivity_g > Median)$ takes the value of 1 if firm g has above median AKM-based employee productivity, indicators for firm size, and indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). *Mid* is an indicator variable that takes the value of 1 for firms with 50-1000 employees, and *Small* is an indicator for firms with less than 50 employees. In columns 1-2, the dependent variable is the QoQ change in employment. In columns 3-4 and 5-6, the dependent variable is the inverse hyperbolic sine of firm's public contributions and firm's per-capita public contributions respectively. We do not report other double interactions for brevity. All specifications include firm, quarter, and industry $\times$ state interacted with quarter fixed effects. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	**Employment**		**Productivity**			
	QoQ Emp. Growth		*Asinh (Contr)*		*Asinh(Contr/ Emp)*	
$\mathbb{1}$(Productivity > Median) × $\mathbb{1}$(Post)	0.0025*	0.0034**	0.0841***	0.1029***	0.0499***	0.0505**
	(0.0013)	(0.0016)	(0.0254)	(0.0282)	(0.0180)	(0.0197)
$\mathbb{1}$(Productivity > Median) × $\mathbb{1}$(Post) × $\mathbb{1}$(Mid)	-0.0033	-0.0018	0.0295	0.0371	0.011	0.0275
	(0.0021)	(0.0025)	(0.0305)	(0.0337)	(0.0219)	(0.0239)
$\mathbb{1}$(Productivity > Median) × $\mathbb{1}$(Post) × $\mathbb{1}$(Small)	-0.0124***	-0.0125***	-0.1087***	-0.1199***	-0.0748***	-0.0678***
	(0.0027)	(0.0030)	(0.0298)	(0.0329)	(0.0222)	(0.0242)
Controls	Y	Y	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y	Y	Y
Time F.E.	Y		Y		Y	
Industry × Location × Time F.E.		Y		Y		Y
Observations	1,192,173	1,110,207	927,161	860,044	927,161	860,044
# Unique Firms	54,839	51,144	47,859	44,600	47,859	44,600
R-squared	0.168	0.222	0.636	0.667	0.525	0.564
Y-Mean	0.071	0.071	1.569	1.598	1.007	1.014
Magnitude- Base (%)	3.47	4.76	5.36	6.44	4.96	4.98
Magnitude - Small (%)	-13.98	-12.88	-1.57	-1.06	-2.47	-1.71

Table 11: **Industry Concentration**

Notes: This table presents correlational results on the change in industry-level concentration and the average productivity of GitHub employees within the industry. In columns 1-4, the dependent variable is the change in the Herfindahl-Hirschman Index (HHI) for employment five years before and after GitHub's policy change. In columns 5-6, the dependent variable is the change in industry's product market concentration computed using 5-year changes in the HHI based on firms' revenue. In columns 1-2, the industry and employment figures correspond to LinkedIn while in columns 3-6, the data is based on Compustat firms/industries. The independent variable is the average AKM-based employee productivity (fraction of employees with normalized productivity greater than zero) within the industry. Standard errors are clustered at the industry level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01)

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	Change in Industry Concentration (HHI) (in %)					
	Labor Market				Product Market	
	All Firms (LinkedIn)		Compustat		Compustat	
Average Productivity	34.556**		16.577*		35.752***	
	(15.048)		(9.826)		(13.794)	
Fraction of Productive Employees		46.478*		22.745**		23.917**
		(26.906)		(9.801)		(10.424)
Observations	143	143	335	335	335	335
R-squared	0.050	0.026	0.014	0.016	0.063	0.017

A.1 Appendix

Figure A.1: **Sample GitHub Profile and Contribution Calendar**

Notes: This figure presents a sample GitHub profile and the user's contribution calendar.

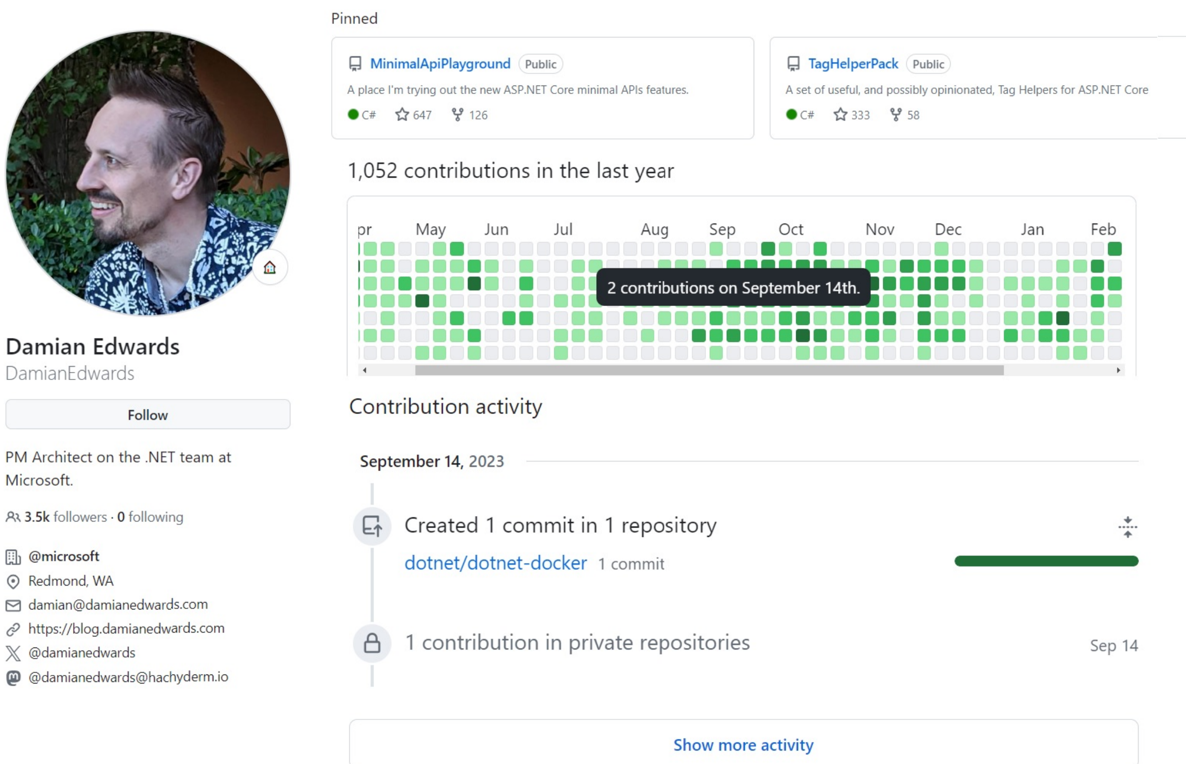
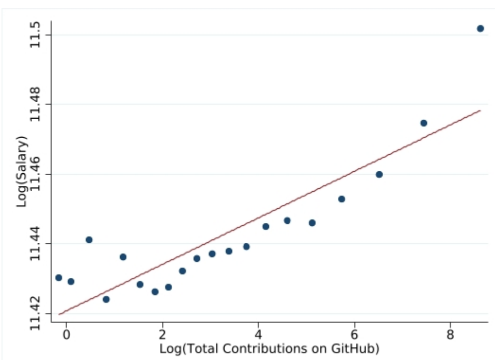
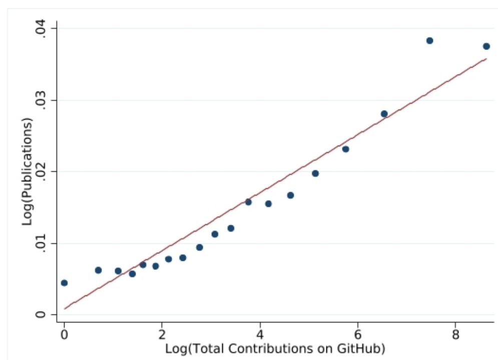


Figure A.2: **Association between Employees' GitHub Contributions, Salaries, and Publications**

Notes: This figure presents the scatter plot of employees' GitHub contributions and salary (in panel A) and publications (in Panel B). On the y-axis, we plot the logarithm of employees' average salary in Panel A and the logarithm of the number of articles authored by the employee on Google Scholar, DevTo, and ORCiD in Panel B. On the x-axis, we plot the logarithm of the number of total GitHub contributions by the employees. We collapse the data into equal-sized bins and plot the mean value of the variables in each bin.



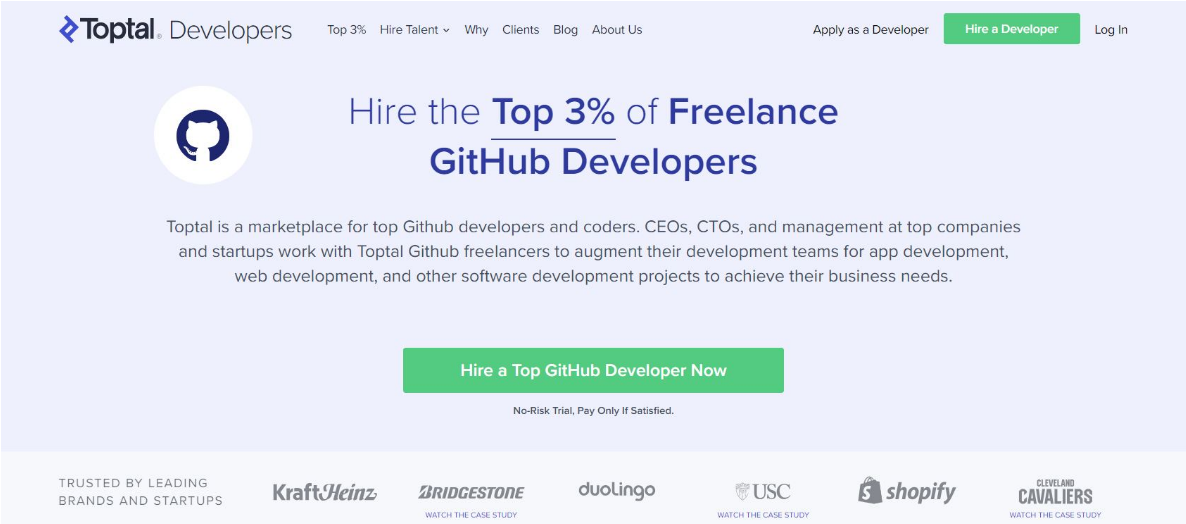
Panel A



Panel B

Figure A.3: **GitHub as a Source of Technical Talent**

Notes: This figure presents snapshots of third-party web platforms that assist firms in sourcing potential employees using GitHub.



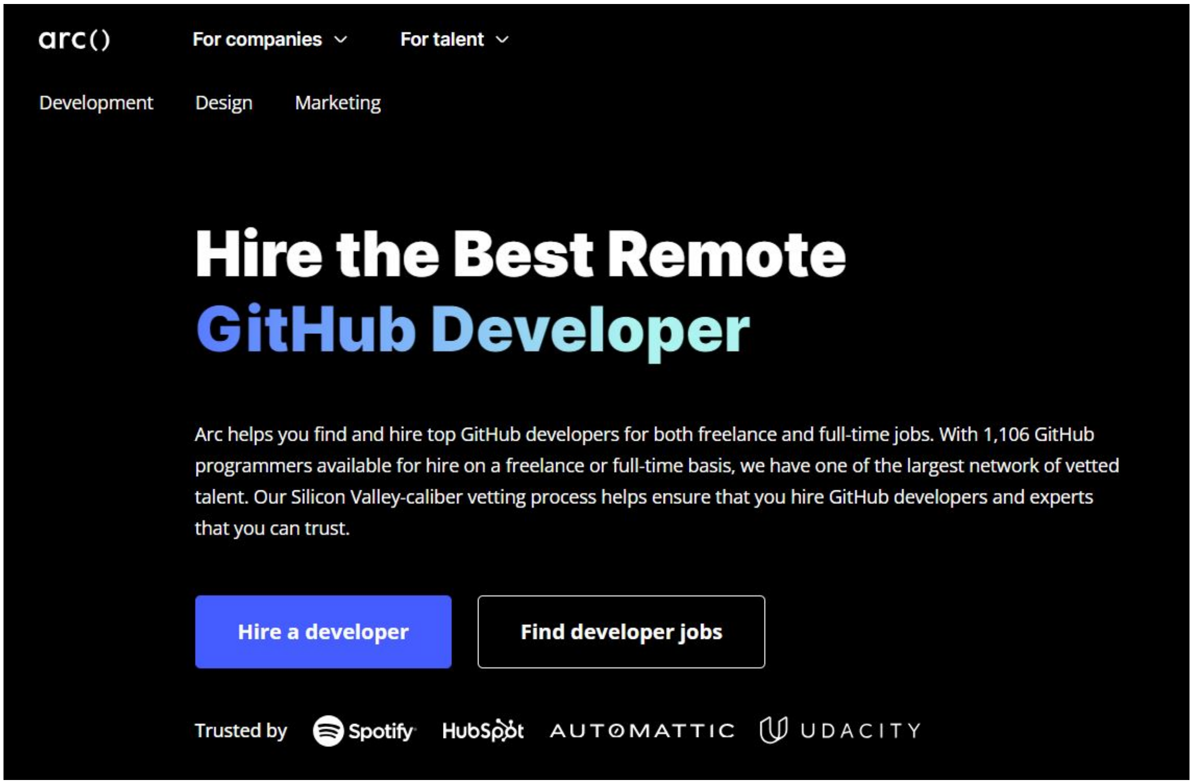


Figure A.4: **GitHub Calendar Before and After Policy Change**

Notes: This figure provides an illustration of the change in a user's GitHub contribution calendar before and after GitHub's policy change on May 19, 2016. Prior to May 19, 2016, users' contribution calendar only showcased the number of contributions made by users towards public repositories on GitHub. Following the policy change, users could also include the (anonymized) number of private contributions in their calendar, subsequently increasing their total number of contributions and making their contribution calendar more active.

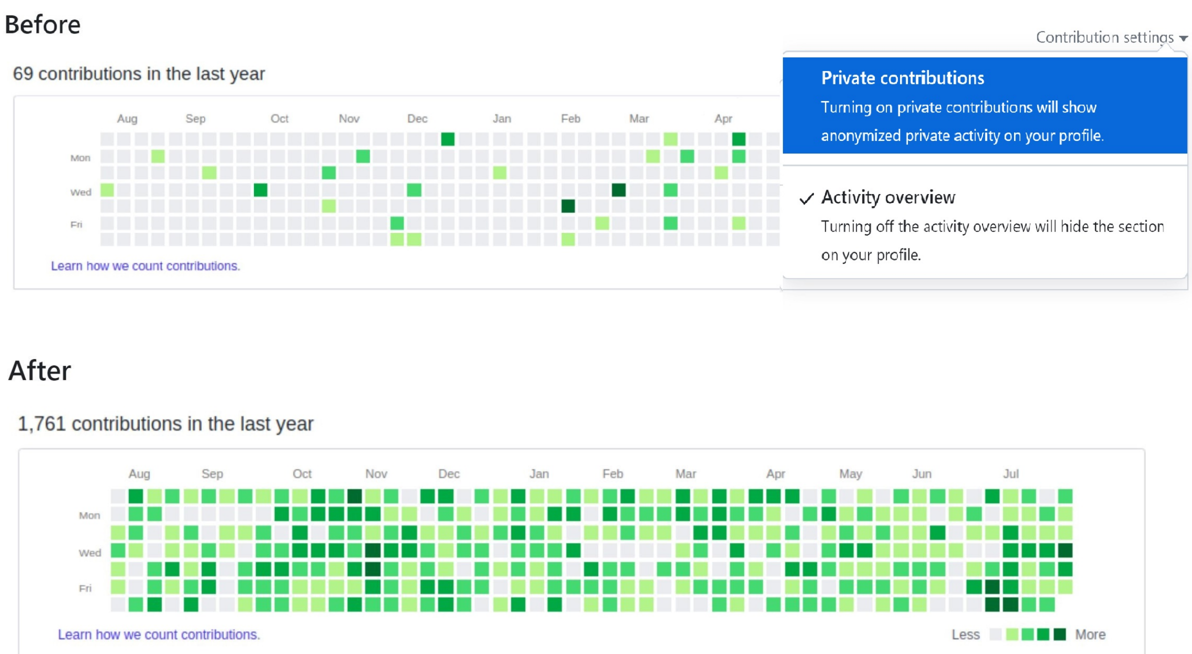


Table A.1: **Sample Characteristics**

Notes: This table presents the distribution of employees across five major industries, locations, coding languages, and roles. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively.

Panel A: Top 5 Industries

Industry	**Share of Users (%)**
Computer Software	16.77
Internet	15.17
Information Technology and Services	14.04
Financial Services	4.88
Marketing and Advertising	3.57
Total	54.43

Panel B: Top 5 Locations

MSA	**Share of Users (%)**
New York-Northern New Jersey-Long Island	11.94
Seattle-Tacoma-Bellevue	7.65
San Francisco-Oakland-Fremont	7.41
San Jose-Sunnyvale-Santa Clara	5.85
Boston-Cambridge-Quincy	5.29
Total	38.14

Panel C: Top 5 Coding Languages

Language	**Share of Users (%)**
Javascript	26.09
Python	14.03
Java	12.35
Ruby	10.18
C#	5
Total	67.65

Panel D: Top 5 Roles

Role	**Share of Users (%)**
Software Engineer - Applications	16.66
Software Engineer - Systems	15.7
Full Stack Developer	5.15
Software Developer	3.71
Web Developer	3.26
Total	44.48

Table A.2: **Distribution by Firm Size**

Notes: This table presents the distribution of firms by their size buckets and the subsequent classification of firms into small, medium, or large category.

	Firm Size Buckets from LinkedIn							
	[1 - 10]	[11 - 50]	[51 - 200]	[201 - 500]	[501 - 1,000]	[1,001 - 5,000]	[5,001 - 10,000]	[10,001 -)
Share (% All Firms)	19.82	28.25	21.92	9.34	5.39	8.15	2.4	4.73
Percentile	10th	25th	50th	75th	85th	90th	95th	99th
Classification	**Small (Size <= 50)**		**Mid (50 < Size <= 1,000)**			**Large (Size > 1,000)**		
Share (% All Firms)	48.07		36.65			15.28		
Share (% All Employees)	19.35		30.23			50.42		

Table A.3: **GitHub Shock & Employee Mobility to Large Firms: Robustness**

Notes: This table presents robustness tests for estimates reported in Table 3. In Panel A, we use alternative measures of productivity from alternative AKMs estimated by treating the dependent variable differently. In part A1 (A3), we use the inverse hyperbolic sine (raw value) of user's private contributions as the dependent variable. In part A2, we use five buckets of private contributions (0, [1, 9], [10,99], [100,999], and [1000+)) as input in the AKM. In Panel B, we use different measures of employee productivity computed without estimating the AKM. We define productivity as the logarithm of private contributions (private contributions as a share of total contributions) in part B1 (B2). Panel C presents results from estimates using a binary dummy indicator for employees' productivity. $\mathbb{1}(Productivity > 0)$ is a dummy variable that takes the value of 1 for employees whose normalized productivity is greater than zero and 0 for the remaining employees. In Panel E, we use the original AKM specification but restrict the AKM sample until April-June quarter in 2014. We then estimate employee mobility in the quarters after April-June 2014. In Panel E, we report our main results by clustering our standard errors alternatively. In part E1 (E2) (E3) (E4), we cluster at the level of industry $\times$ location (employee's coding language) (employee's initial role) (firm and employee both). Finally, in Panel F, we control for more granular fixed effects by interacting employee's cohort, role, language, location, and industry with time. All panels run tests corresponding to our baseline specification, Column 5 in Table 3. In panels A, B, C, D, and F, standard errors are clusters at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)	(6)
	Coefficient	**Std. Error**	**Observations**	**R-squared**	**Y-Mean**	**Magnitude (%)**
Panel A: Alternate AKM Specifications						
A1: AKM with Asine Function						
Productivity × 𝟙(Post)	0.0015***	(0.0002)	3,199,173	0.212	0.033	4.64
A2: AKM with Categorial Variables						
Productivity × 𝟙(Post)	0.0016***	(0.0002)	3,199,173	0.212	0.033	4.67
A3: AKM with Raw Contribution Numbers						
Productivity × 𝟙(Post)	0.0008***	(0.0001)	3,199,173	0.211	0.033	2.5
Panel B: Alternate Productivity Definition						
B1: Log(Private Contributions)						
Productivity × 𝟙(Post)	0.0011***	(0.0001)	3,199,173	0.212	0.033	3.63
B2: Share Private Contributions						
Productivity × 𝟙(Post)	0.0040***	(0.0008)	3,199,173	0.212	0.033	13.66
Panel C: Dummy Productivity Definition						
C1: Productivity defined as Dummy Indicator Variable						
𝟙(Productivity > 0) × 𝟙(Post)	0.0066***	(0.0005)	3,199,173	0.212	0.033	22.92
Panel D: Alternate AKM Sample Period						
D1: Pre Sample AKM (2011-14)						
Productivity × 𝟙(Post)	0.0011***	(0.0002)	1,680,047	0.223	0.03	4.2
Panel E: Alternate Clustering						
E1: Cluster at Industry × Location Level						
Productivity × 𝟙(Post)	0.0015***	(0.0002)	3,199,173	0.212	0.032	5.27
E2: Cluster at Language Level						
Productivity × 𝟙(Post)	0.0015***	(0.0002)	3,199,173	0.212	0.032	5.27
E3: Cluster at Role Level						
Productivity × 𝟙(Post)	0.0015***	(0.0001)	3,199,173	0.212	0.032	5.27
E4: Double Cluster at Employee & Firm						
Productivity × 𝟙(Post)	0.0015***	(0.0004)	3,197,424	0.212	0.034	5.27
Panel F: More Granular Controls						
F1: Control for Cohort × Role × Language × Location × Industry × Time						
Productivity × 𝟙(Post)	0.0017***	(0.0004)	745,659	0.37	0.036	5.71

Table A.4: **GitHub Shock & Employee Mobility to Different Firm-Size Buckets**

Notes: This table presents estimates on the heterogeneous impact of GitHub's disclosure of employee productivity information on employee mobility based on destination firm size. We report the differential impact of GitHub's policy change on productive employees estimated by the β coefficient from the equation:

$$\mathbb{1}(Move)_{i,t} = \beta Productivity_i \times \mathbb{1}(Post_t) + \gamma X_{it} + \alpha_i + \theta_t + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with *Post*, and indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). The dependent variable is a dummy variable that equals 1 when employee i switches job in quarter t to a firm size as indicated in various columns. We control for employee fixed effects, salary controls, seniority controls, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	**𝟙(Job Change)**							
Firm Size:	[1 - 10]	[11 - 50]	[51 - 200]	[201 - 500]	[501 - 1,000]	[1,001 - 5,000]	[5,001 - 10,000]	[10,001 -)
Productivity × 𝟙(Post)	-0.0002	-0.0010***	-0.0009***	-0.0004***	0.0000	0.0003***	0.0001**	0.0011***
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Employee F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Salary & Seniority Controls	Y	Y	Y	Y	Y	Y	Y	Y
Cohort × Language × Role × Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3,199,173	3,199,173	3,199,173	3,199,173	3,199,173	3,199,173	3,199,173	3,199,173
R-Squared	0.221	0.21	0.206	0.206	0.205	0.205	0.2	0.205
Y-Mean	0.008	0.015	0.016	0.008	0.006	0.01	0.005	0.018
Magnitude (%)	-2.4	-8.03	-6.56	-4.9	-0.09	3.26	3.37	6.9

Table A.5: **GitHub Shock & Employee Mobility to Superstar Firms**

Notes: This table presents estimates on the impact of GitHub's disclosure of employee productivity information on employee mobility to star startups. We report the differential impact of GitHub's policy change on productive employees estimated by the β coefficient from the equation:

$$\mathbb{1}(MoveToSuperstar)_{i,t} = \beta Productivity_i \times \mathbb{1}(Post_t) + \gamma X_{it} + \alpha_i + \theta_t + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with *Post*, and indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). The dependent variable is an indicator that equals 1 when employee i switches job to a *superstar firm* in quarter t and is 0 otherwise. In each column, we define superstar firms differently based on the average market capitalization or average revenue of the firm in the three years before GitHub's policy change. We control for employee fixed effects, salary controls, seniority controls, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	**$\mathbb{1}$(Mobility to Superstar Firms)**							
	Within-Industry Top 10%		*Within-Industry Top 1%*		*Across Industry Top 10%*		*Across Industry Top 1%*	
	Market Cap	Revenue	Market Cap	Revenue	Market Cap	Revenue	Market Cap	Revenue
Productivity × $\mathbb{1}$(Post)	0.0008***	0.0008***	0.0005***	0.0004***	0.0008***	0.0008***	0.0005***	0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Employee F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Salary & Seniority Controls	Y	Y	Y	Y	Y	Y	Y	Y
Cohort × Language × Role × Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3,219,087	3,219,087	3,219,087	3,219,087	3,219,087	3,219,087	3,219,087	3,219,087
# Unique Employees	143,447	143,447	143,447	143,447	143,447	143,447	143,447	143,447
R-Squared	0.191	0.194	0.175	0.18	0.193	0.196	0.18	0.188
Y-Mean	0.017	0.018	0.007	0.006	0.017	0.017	0.009	0.006
Magnitude (%)	5.35	4.84	8.61	7.58	5.38	5.3	6.93	4.9

Table A.6: **GitHub Shock & Employee Mobility to Star Startups**

Notes: This table presents estimates on the impact of GitHub's disclosure of employee productivity information on employee mobility to star startups. We report the differential impact of GitHub's policy change on productive employees estimated by the β coefficient from the equation:

$$\mathbb{1}(MoveToStarStartup)_{i,t} = \beta Productivity_i \times \mathbb{1}(Post_t) + \gamma X_{it} + \alpha_i + \theta_t + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with *Post*, and indicator that takes the value of 1 for quarters after GitHub's policy change (April-June 2016). The dependent variable is an indicator that equals 1 when employee i switches job to a *star startup* in quarter t and is 0 otherwise. We define a firm as a startup if the firm's age at the time of GitHub's policy change was less than ten years. Subsequently, as indicated across different columns, we define star startups differently based on the final venture funding round or the total funding raised before GitHub's policy change. We control for employee fixed effects, salary controls, seniority controls, employee's cohort, role, and language interacted with time, and location and industry interacted with time in all specifications. Cohort represents an employee's final education year. Language is the employee's most preferred coding language. Role, industry, and location correspond to an employee's initial job role, MSA, and industry respectively. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)
Outcome:	**$\mathbb{1}$(Mobility to Star Startups)**				
	Funding > Series B	*Funding > Series D*	*Funding > \$ 100 Mn*	*Funding > \$ 500 Mn*	*Funding > \$ 1 Bn*
Productivity × $\mathbb{1}$(Post)	-0.0002**	0.0001	0.0001	0.0001***	0.0001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Employee F.E.	Y	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y	Y
Salary & Seniority Controls	Y	Y	Y	Y	Y
Cohort × Language × Role × Time F.E.	Y	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y	Y
Observations	3,145,396	3,145,396	3,145,396	3,145,396	3,145,396
# Unique Employees	143,437	143,437	143,437	143,437	143,437
R-Squared	0.173	0.168	0.166	0.155	0.151
Y-Mean	0.005	0.003	0.003	0.001	0.001
Magnitude (%)	-5.03	3.07	2.75	13.08	14.77
10th percentile Firm Size	[11 - 50]	[11-50]	[51 - 200]	[201 - 500]	[501 - 1,000]
Median Firm Size	[51 - 200]	[201 - 500]	[201 - 500]	[1,001 - 5,000]	[1,001 - 5,000]
90th percentile Firm Size	[501 - 1,000]	[1,001 - 5,000]	[1,001 - 5,000]	[10,001 -)	[10,001 -)