Big Data and Bigger Firms: A Labor Market Channel*

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

October 21, 2024

Abstract

This paper studies the impact of employee output information disclosure through GitHub on labor reallocation towards large firms. GitHub, which is the world's largest software management platform, tracks and publicly displays real-time individual contributions. In 2016, a policy change enabled GitHub users to display their contributions more accurately on their profiles. Following this update, employees with 1 standard deviation higher GitHub contributions witnessed a 5.7% 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 NBER Megafirms and the Economy (Fall 2024), UNC Chapel Hill, University of Colarado at Boulder, Copenhagen Business School, BI Norwegian Business School, Charles III University of Madrid, Chinese University of Hong Kong, University of Southern California, and Duke University 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

Informational asymmetry, where one party in a transaction holds more knowledge than the other, has long been a fundamental friction in economic models (Akerlof, 1970; Spence, 1973). However, the advent of big data has begun to challenge traditional boundaries of information control (Veldkamp and Chung, 2024; Jiang and Li, 2024). This is especially true for labor markets, where firms hold exclusive information on their employee performance and potential (Farber and Gibbons, 1996; Altonji and Pierret, 2001; Schönberg, 2007; Kahn, 2013). The rise of internet-based platforms has fundamentally transformed the recruitment process, with over 80% of job seekers and more than 90% of employers now relying on these platforms as their primary tool for hiring decisions (Smith, 2015; Bradshaw, 2024). These platforms provide new, detailed information on individual worker output, which serves as a signal of employee productivity, mitigating informational disparities. We explore how this shift influences labor reallocation and what it means for firm outcomes.

The impact of productivity signals on labor market dynamics is a priori unclear. Several studies argue that larger firms attract more productive employees due to their reputational advantages, which signal productivity in the absence of performance measures (Waldman, 1984, 1990; Bernhardt and Scoones, 1993). In this view, enhanced information could enable smaller firms to compete for top talent, especially if they can adopt new technologies more easily due to fewer organizational constraints. Conversely, in other studies, adverse selection prevents productive workers at smaller firms from moving to larger ones (Greenwald, 2018; Gibbons and Katz, 1991). In this case, stronger productivity signals might further enhance large firms' ability to recruit top talent. Additionally, large firms may leverage economies of scale to invest in data-driven recruitment strategies more effectively (Brynjolfsson and McElheran, 2016). The net effect of these productivity signals on labor market dynamics and firm hiring behavior remains an open question, necessitating further empirical exploration.

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. GitHub provides contributors 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.[2] The large dimensionality and extensive public use make GitHub an ideal starting point for studying how big data[3] hosted by technology platforms impacts 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.

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

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

[2] 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.

[3] GitHub data meet all three criteria for big data as defined by Goldstein et al. (2021): they are large in size (a compressed subset of GitHub repositories as of November 2020 was 21 TB in size `https://archiveprogram.github.com/arctic-vault/`), possess multiple dimensions, and have a complex layout.

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 information on employee-specific output impacts 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 5.7% 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. 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. Moreover, 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 4% relative to the sample mean, and less likely to move to small firms (those employing 50 employees or less) by 5% 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 -7% 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% (3%) 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. In contrast, 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. After the GitHub policy change, treated employees show a significant overall rise in productivity. Hence, the GitHub signal appears to facilitate both career advancements and enhanced employee performance.

We next test the idea that the reduction of information asymmetry regarding employee quality through 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, years of schooling, and having a degree from a highly-ranked university. We find consistent results across all our measures that talented individuals with less experience, with less schooling, and with degrees outside of the highly-ranked universities 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. We further show that our effects are mitigated in the presence of limited individual mobility due to greater enforceability of non-compete agreements.

The flow of talent towards large firms can be a result of firm-demand factors or labor-supply preferences. On the firm demand side, 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 test 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.

On the labor supply side, 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. This potential for increased career growth may come with increased labor risk. To this end, 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.

Finally, we explore whether productive workers are moving to larger firms with higher growth potential. While this could reflect both high firm demand for talent as well as worker preferences, it is critical for understanding the efficiency of talent flows to larger firms. Our results indicate that workers tend to move to firms with higher ex-ante growth, both in terms of employment and sales, consistent with the idea that workers are attracted to firms with a greater need for talent and strong growth prospects.

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 AKM-based 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. We also observe a higher likelihood of left-tail outcomes, as the probability of small treated firms exiting the market increased by 38% of the mean after the shock. 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 3 to 4% post-announcement. Instead, large firms employing high-productivity talent pre-treatment experienced a 4% increase in quarterly employment growth over the mean and a 5 to 6% rise in public GitHub contributions following the policy change.

The labor market reallocation towards large firms has also implications for aggregate industry dynamics. When aggregating our data at the industry level, we find that industries with more GitHub contributions before the shock (2013-2016) saw a greater increase in national labor market concentration in the five years after the shock (average concentration over 2016 to 2021) compared to the five years before it (average concentration over 2011 to 2016). Although these results are based on correlations, they offer interesting insights into the overarching trends and are consistent with a broad consensus that national industry-level labor concentration has risen in the U.S. since 1980 (Autor et al., 2023).

Our paper makes significant contributions to four strands of the existing literature. A substantial body of literature has attributed the rise of large firms to technological trends such as the rise of the platform economy (Crouzet and Eberly, 2018; Lashkari et al., 2018; Bessen, 2020; Firooz et al., 2022; Brynjolfsson et al., 2023; Hsieh and Rossi-Hansberg, 2023) and big data (Brynjolfsson and McElheran, 2016; Farboodi et al., 2022; Aghion et al., 2023), arguing that these technologies provide product market advantages (Prat and Valletti, 2022;

Eeckhout and Veldkamp, 2022) and superior returns on investment (Tambe and Hitt, 2012) to large firms. Our paper introduces a novel labor-market-based explanation, examining how big data, available via platform technologies, enhance large firms' ability to attract and retain top talent, thus contributing to their growth and dominance.

Second, a growing body of research examines how performance metrics affects labor outcomes. Rockoff et al. (2012) study the impact of teacher performance metrics on job separations in teachers labor markets in New York, Floyd et al. (2024) analyze the impact of grade disclosures on MBA graduates' job placements, and Hager et al. (2024) explore how citation metrics impacts assortative matching in academia. Our paper makes two key contributions to the literature. We explore how a new, previously unobserved signal of individual performance drives labor reallocation from smaller to larger firms, suggesting that new information availability on the labor market contributes to the increasing dominance of larger firms in the U.S. economy. Additionally, we leverage granular, person-day level data from GitHub, offering a fresh and expansive view of U.S. technology workers, enabling us to analyze these effects at an unprecedented scale.

Third, our paper contributes to the existing literature on labor reallocation across firms of different sizes within the economy. Prior research has extensively documented how regulatory policies influence the distribution of talent between established companies and startups. For instance, Acemoglu et al. (2018) examine R&D taxation, Pagano and Picariello (2023) study unemployment insurance, Araujo et al. (2023); Baghai et al. (2024) analyze bankruptcy, Jeffers (2024) explores the impact of non-compete agreements, and Gupta (2023) investigates the effects of immigration policies. Beyond regulatory factors, large scale trends and strategic behavior also play a critical role in shaping talent allocation. For example, Bernard et al. (2006); Bloom et al. (2015); Hombert and Matray (2018); Bena and Simintzi (2024) analyze how Chinese import competition and access to offshore labor impacted labor and technology reallocation across firms, Bernstein et al. (2023) document flight of employees towards more established startups during the COVID downturn, and Akcigit and Goldschlag (2023) look into strategic hoarding of inventors by large firms. However, relatively little attention has been given to the role the introduction of big data in labor markets has played. Our study fills this gap by exploring how the emergence of big data on employee performance influences the shift of talent towards larger firms.

Finally, our paper adds to the literature on re-distributive impacts of big data. The evidence regarding the effects of these technologies has been mixed. 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 may lower the cost of capital for larger firms (Begenau et al., 2018). While prior research has primarily focused on financial markets, our paper is the first to examine the impact of big data on 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. While 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.[4] Subsequently, until recently, private repositories were mostly held by organizations seeking to protect intellectual property and maintain confidentiality while collaborating on development projects.

[4]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.

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 B.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.

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 B.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 similarly use individual GitHub contributions to assess the productivity of software developers, thereby reinforcing our interpretation (Holub and Thies, 2023; El-Komboz and Fackler, 2023; Conti et al., 2023).

Unsurprisingly, recruiters and hiring managers increasingly rely on GitHub user activity to source and evaluate potential tech-sector employees. Figure B.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 B.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 a 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) as described in the Internet Appendix, Section B.1. 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.[5]

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.

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

These filters result in an initial sample size of 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 initial 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 are able 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 initial 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 1 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 nine 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.[6]

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 thousand employees), 37% are medium (fifty to a thousand employees), and 48% are small (less than fifty employees). However, as shown in Internet Appendix Table 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 benchmark 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

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

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, role, 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 summary statistics for the AKM. Panel A shows that our connected set covers 93% of the employees in the initial sample. We are able to calculate AKM estimates for 80% of the initial sample employees, after applying filters as detailed in Section 2.4. 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). We also test the persistence of individual productivity estimates from the pre-shock period (2011–2016) to the post-shock period (2016–2021). Figure A.2 presents these results, indicating that the majority of employees remain within similar productivity quintiles before and after the shock. This persistence suggests that our estimates capture inherent differences in individual quality, rather than any attempt by employees to manipulate these metrics. The results highlight that the GitHub policy change regarding employee private contributions can be interpreted as a new signal of employee productivity, with inherent and persistent differences among employees accounting for the majority of variation in these contributions.

A key concern when using the AKM design is whether firm-worker matching accounts for most of the variation in productivity. Following the approach of Card et al. (2013), we explore this possibility. Figure A.1 presents an event study analysis that examines the impact of job transitions on productivity during the pre-shock period (2011-2016), which is used to estimate productivity. The outcome of interest is the logarithm of employees' private contributions, measured over five quarters before and after a job move. Job transitions are categorized based on the firm-fixed effects from the AKM model for both the origin and destination firms. The data reveal that moving between low- and high-productivity firms has a significant effect on productivity. Workers moving between quartile 1 firms maintain relatively low but stable productivity, while those who move from quartile 1 to quartile 4 firms experience substantial productivity gains. Conversely, workers moving within quartile 4 firms see little change, but those transitioning to quartile 1 firms suffer considerable productivity losses. These results align with a simple model of additive worker and firm effects, suggesting limited influence of better firm-worker matching, which would typically predict productivity increases with any job transition. Additionally, if match effects were important, a fully saturated match-effects model—including a separate dummy for each worker-firm pair—would explain significantly more variation than our baseline AKM model with separate firm and worker fixed effects. However, including these additional fixed effects only raises the adjusted R^2 from 73.5% to 79%, a 7.5% improvement over the mean, smaller than the 11-15% increase found in Card et al. (2013).

To address recent 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), we further perform two analyses. First, we employ alternative AKM specifications, including continuous measures of contributions and categorical variables. Second, we demonstrate that our findings hold even without using the AKM framework by utilizing 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

In our baseline analysis, we examine whether improved information on employee quality influences the reallocation of employees between firms. We carry out a difference-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:

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

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 $Post_t$ – an indicator that switches on after GitHub's policy shift in May 2016 (i.e., after the second quarter of 2016). We include individual (α_i) and (δ_t) quarter fixed effects 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 employee's current salary and seniority (X_{it}). 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 of employee's first job (based on LinkedIn's 144 industry categories) interacted with the geographical area (metropolitan statistical area) of employee's first job 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 the third quarter of 2013 (July to September quarter, three years before the shock) to the second quarter of 2019 (April to June quarter, three years after the shock). Because our independent variable, $Productivity_i$, is itself estimated from the AKM regression, we use bootstrap estimation to derive our standard errors. We bootstrap within employee clusters, which is the treatment level in our analysis, and perform 100 iterations to obtain standard errors.[7]

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 6% 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 time-varying employee-level controls (salary and seniority), dynamic job-level (cohort-programming language-role-time), and market-level (industry-M.S.A-time) fixed effects. We also test whether the moves to large firms are accompanied by changes in industry and location, as shown in Table A.8. We find that the probability of moving to a larger firm increases by 6.5% over the sample mean when

[7]We perform bootstrapping in two stages. In the first stage, we randomly filter and estimate an AKM sample representing a random 80% of the total employees in our connected set to get one set of productivity estimates. We repeat this exercise for 100 iterations and obtain 100 estimates of employee productivity. Then, in the second stage, we run our individual mobility tests using each of the 100 productivity estimates and take the average value of the corresponding standard errors as our bootstrapped standard error.

accompanied by a location change, compared to a 4.8% increase when the move occurs within the same location. Furthermore, the likelihood of transitioning to a large firm across both similar and different industries rises by 5-6% over the sample mean. These findings support the idea that enhanced signaling plays a more crucial role in cross-location moves, as it helps compensate for the absence of localized networks.

We also estimate the dynamic version of Equation 2 as follows:

$$Y_{i,t} = \sum_{\tau=-8}^{12} \beta_\tau Productivity_i \times \mathbb{1}(\tau = t) + \lambda X_{it} + \alpha_i + \gamma_{it} + \varepsilon_{i,t} \tag{3}$$

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 the first quarter of 2016 (January to March 2016), 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.

The event study corresponding to Table 3 Column 5, 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 quarters following the GitHub announcement, indicative of a rapid change in information disclosure driving the observed outcomes.

3.1.1 Identification challenges

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 (those unaffected by the shock) to further rule out the impact of unobservable employee characteristics. We exclude any public contributions made by an employee to their own repository, as these may represent personal activity. 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.

Additionally, we test if our results are influenced by trends correlated with the 2016 presidential elections or other local economic characteristics. If our results were driven by the elections or differences in local economic conditions, we would expect to find different impacts in areas based on their political leanings or economic characteristics. However, Table A.9 shows no differential impact on counties that were more Republican, white, richer, or more unequal.

As a final validation, we analyze changes in firm stock prices following the GitHub announcement. This analysis helps isolate the immediate daily impact of the GitHub announcement, controlling for any longer-term changes.[8] Using CompuStat data from 1,679 firms, we categorize firms two ways: by their average productivity (above or below median), and by size (with indicators for the top and bottom quartiles).[9] Table A.11 shows a 65 basis point decline in returns for small firms (bottom quartile) with high GitHub private contributions (above median), indicating market anticipation of talent loss. No significant

[8] Bloomberg does not report any major tech-related news announcements from 17 May 2016 to 20 May 2016.

[9] The bottom quartile firms have fewer than 1,200 employees while the top quartile firms have more than 18,000 employees.

impact is observed for larger firms (top quartile) or smaller firms without GitHub exposure (bottom quartile in size and below median average productivity). We rerun the same analysis on 100 randomly selected dates one year before the date of the actual policy change. Figure A.4 shows that we observe similar or more negative coefficients in less than five percent of the cases.

Moreover, there's a concern 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 limit our sample to active GitHub users. We test our findings across three alternative samples: users active on GitHub before the shock (with more than 10 pre-period total contributions), 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. A related concern is that certain firms may not rely on private contributions, or somehow forbid employees of signaling their private contributions on GitHub. Panel C restricts the sample to the subset of firms where employee private contributions can be observed (firms with at least positive or 10 private contributions as of the shock date). Our results are much stronger within this sub-sample despite a reduction in sample size. A final concern is that transitory shocks could simultaneously affect both individual productivity and mobility. To address this issue, we estimate employee productivity separately for periods before and after the GitHub shock. As shown in Figure A.2, individual productivity remains highly persistent, with the majority of employees staying within the same productivity quintiles across both periods. Furthermore, Table A.4 narrows our analysis to users whose productivity levels are consistent in both the pre- and post-shock periods. Our findings remain robust, even when the sample is restricted to individuals exhibiting stable private contribution productivity throughout the study period.

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). 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 also 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 a binary indicator for productivity in Panel C.[10] Panel D examines the sensitivity of our findings to the time period over which we estimate the AKM, Panel E tests the effect of more granular fixed effects, and Panel F shows that our results are robust to estimating our coefficients on a matched treatment and control sample.[11]

3.2 Mobility to medium and small firms

We next examine 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,

[10] Callaway et al. (2024) argue that the parallel trends assumption may not rule out selection bias in difference-in-differences designs with a continuous treatment variable. Subsequently, in Figure A.3, we show that our parallel trends assumption is robust to using a dummy indicator for treatment.

[11] In Panel D, we estimate the AKM using contributions from the pre-sample period, up until the the second quarter (April to June) of 2014. Panel E involves interacting all fixed effects, including cohort, role, language, location, industry, and time. We match treated employees (those with positive normalized productivity estimates) with control employees (having negative normalized productivity estimates) on gender, race, industry, location, experience, salary, and seniority in Panel F1. We additionally match on public contributions in Panel F2.

except we now consider transitions to medium-sized firms (50 to 1000 employees) and small firms (less than 50 employees) as the outcome variables. The results, outlined in Table 5, reveal that individuals with productivity estimates higher by 1 standard deviation reduced their likelihood of transitioning to small and medium firms by 5% and 3.7% 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.[12]. Our findings indicate that the mobility of employees with 1 standard deviation higher productivity increases almost monotonically with firm size, ranging from -7% 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 -1%, and statistically insignificant. This last category could be driven by both employees transitioning to startups and individuals founding their own firms. In Table A.13, we explicitly examine whether there is any shift in the probability of entrepreneurship. Our findings reveal an 11% decrease in the overall rate of new firm formation among employees whose productivity is one standard deviation above the average following the shock, supporting the idea of labor reallocation away from startups. In contrast, we observe a modest 5% increase in entrepreneurial activity among employees at small firms, where the likelihood of transitioning to entrepreneurship rises for those with similarly high productivity. This suggests a smaller, but notable, reallocation from small firms toward entrepreneurship as a consequence of the shock.

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.6 shows that the increase in mobility holds when defining superstars as the top 1% and 10% by market capitalization and

[12] Table A.5 shows the regression counterparts of Figure 4.

revenue, as defined in Tambe et al. (2020). Moreover, internet appendix Table A.7 reveals that even within young firms, mobility increases to only "superstar" startups with over $500 million in funding.[13] 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 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:

$$\begin{aligned} \mathbb{1}(MoveToLargeFirm)_{i,t} = \beta_1 Productivity_i \times \mathbb{1}(Post_t) + \\ \beta_2 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_{i(t-1)}) \\ + \beta_3 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Small_{i(t-1)}) + \lambda X_{it} + \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{i,t} \end{aligned} \tag{4}$$

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, and indicators representing the size of the firm where the individual worked in the previous quarter. Here, $\mathbb{1}(Mid)$ denotes an indicator variable that equals 1 for firms with 50-1000 employees, and $\mathbb{1}(Small)$ 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

[13] Startups with at least $500 Million have a median size of 1000-5000 employees.

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 Equation 4, to verify the absence of pre-trends, as follows:

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

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 8% 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 3% 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 4 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 6% 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 8% 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 5.2% of the sample mean for individuals with 1 standard deviation higher productivity, following the GitHub announcement. Column 4 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 15% of the sample mean for those with 1 standard deviation higher productivity after the GitHub shock. Conversely, there was less than 2.3% 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 5.9% 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 8.4% over the mean in within-firm promotion probability for employees with 1 standard deviation higher productivity. In contrast, the same effect is 5.1% 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.

We also explore whether these career upgrades are associated with a change in employee productivity. We test this hypothesis in Table A.12, where we check if treated employees experienced a larger change in productivity compared to others following the shock. We utilize the public contributions for each employee, as these were not directly influenced by the shock. We exclude any contributions made by an employee to their own repository, as these may represent personal activity. We find a 9% increase over the mean in public contributions for employees with productivity 1 standard deviation higher in pre-shock private contributions after the GitHub shock. We find that these results are driven by changes in the probability of experiencing very high (top decile) contributions, reducing concerns that we are capturing bias in logarithm due to count like variables. These results align with the notion that the shock facilitated both career advancements and enhanced employee performance.

In summary, our findings indicate that the sudden disclosure of productivity information on GitHub was advantageous for productive employees, facilitating their productivity and 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 the significant information asymmetry surrounding employee quality (Greenwald, 2018), suggesting that in the absence of direct productivity indicators, firms may rely on hard signals, such as experience and educational background, when making hiring decisions (Spence, 1973; Waldman, 1984). We examine this mechanism by exploring the heterogeneity of our results across three types of alternate individual signals: work experience, years of schooling, and having degree from a high-ranked university. We present these results in Table 8.

Column 1 examines 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 (i.e., 7 years) at the time of shock saw a 1.7 times larger increase in post-shock mobility to large firms compared to those with more than median experience.

Columns 2 examines the differential impact of mobility based on the number of years of schooling. Our analysis shows that more productive employees with below median schooling tenure experienced a 2 times greater increase in post-shock mobility to large firms compared to employees with above median schooling.

Column 3 assesses the heterogeneous impact of having a degree from a highly-ranked university on our results. We define elite universities as those within the top 20 ranked universities according to U.S. News and World Report. Our findings indicate that productive employees without an elite university affiliation experienced a mobility increase to large firms that was 2.2 times the change observed for employees with an elite university degree. Overall, our results support the hypothesis that web-based platforms have reduced reliance on hard signals, such as experience and education leading to a "democratization" of screening in the labor market. This has facilitated a less constrained flow of talent across firms, allowing employees to advance based on productivity rather than traditional metrics.

We also test the heterogeneity based on the enforcement of non-compete agreements. Previous research has demonstrated that non-competes significantly restrict employee mobility (Jeffers, 2024). Hence, we expect a stronger impact of signaling on mobility in states where non-competes are less enforceable.[14] Column 4 presents our findings. As anticipated, the data reveals a substantially larger increase in mobility for employees residing in states without enforceable non-compete agreements, consistent with the hypothesis that we capture the impact of signaling on labor mobility.

4.2 Heterogeneity by Firm Characteristics

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

[14] We use the US government's official archives to get state-level variation in NCs as of 2016. Link: `https://obamawhitehouse.archives.gov/sites/default/files/competition/state-by-statenoncompetesexplainer_unembargoedfinal.pdf`

toward larger firms. We test these hypotheses using heterogeneity tests.

First, we examine the role for tech-savvy recruiters and firms' own GitHub usage. Columns 1-2 in Table 9A 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 8.5% of the sample mean after the GitHub policy change. In contrast, the increase was only 2.2% 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 3-4 in Table 9A 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 post-shock increased by 5.8% of the sample mean. In comparison, the increase was 4.7% of the sample mean for large firms not on GitHub. The difference between these two coefficients is statistically significant at the 10% 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.

Second, 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 pre-treatment sample period (quarter 3 of 2013 to quarter 2 of 2016). 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 9A, columns 5-6, present the results. We find that individuals with productivity 1 standard deviation above the mean were 7.2% more likely (compared to the sample average) to move to large firms with steeper salary growth after the GitHub policy change. In contrast, the increase was only 2.2% of the sample mean 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.

We also 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 pre-treatment sample period (quarter 3 of 2013 to quarter 2 of 2016), 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 in our sample. Columns 7-8, present the results. Individuals with 1 standard deviation higher productivity increased their mobility to larger firms with above-median forced turnover by 6.6% above the sample mean following the GitHub policy change. In contrast, there was a statistically insignificant decrease of 0.3% 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.

Finally, we explore whether productive workers move to larger firms due to their higher growth potential. While this could reflect both high firm demand for talent and worker preferences, it is critical for understanding the efficiency of talent flows to larger firms. Table 9B presents our results. We first split our sample based on firm's ex-ante growth in columns 1-2. We estimate the firms average employment growth based on the entire LinkedIn data during the pre-period and create indicators for move to large firms with above-median and below-median employment growth. Individuals with 1 standard deviation higher productivity increased their mobility to larger firms with above-median employment growth by 7.9% above the sample mean following the GitHub policy change. In contrast, there was an increase of 5.1% of the sample mean for mobility to large firms with below-median growth. T-tests confirm that the difference between these two coefficients is significant at the 1% level. Similarly, in columns 3-4, the subset splits the sample of large firms into those with above-median and below-median Compustat based sales growth in the pre-period. We find that the probability of an employee with 1 standard deviation higher productivity moving to a large firm with above-median sales growth increased by 8.6% of the sample mean. In comparison, the increase was 3.9% of the sample mean for large firms with below median sales growth. The difference between these two coefficients is statistically significant at the 1% level.

Our results are consistent with both firm-demand and labor-supply mechanisms. Productive workers move to large firms better able to exploit the GitHub data, and with higher upside potential. Overall this results in the reallocation of these productive workers to firms with ex-ante higher growth prospects.

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 basic difference-in-differences setup compares the changes in firms with an ex-ante higher proportion of high-productivity (treated) employees with firms with a low proportion of high-productivity employees, before and after the GitHub announcement. Subsequently, the third 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:

$$\begin{aligned} FirmOutcome_{j,t} = &\beta_1 \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(Post_t) \\ &+ \beta_2 \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_j) \\ &+ \beta_3 \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(Post_t) \times \mathbb{1}(Small_j) + \lambda X_{jt} + \alpha_j + \gamma_{j,t} + \varepsilon_{j,t} \end{aligned} \tag{5}$$

In our regression model, the dependent variable is an outcome variable for firm j at quarter t. The coefficient β_1 captures the interaction between $\mathbb{1}(Productivity_j > Median)$, an indicator that takes the value 1 if the average AKM-derived productivity estimate in the pre-period is greater than the median for firm j, and $\mathbb{1}(Post_t)$, an indicator that switches on after GitHub's policy shift in May 2016 (i.e., after quarter 2 of 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 the third quarter of 2013 (three years before the shock) to the second quarter of 2019 (three years after the shock). To ensure robustness of our analysis, we control for firm and industry-state-time fixed effects, controlling for constant firm-level differences and market level 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 pre-period contribution interacted with $\mathbb{1}(Post_t)$. We cluster our standard errors at the firm level.

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

$$\begin{aligned}
FirmOutcome_{j,t} = &\sum_{\tau=-8}^{12} \gamma_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \\
&+ \sum_{\tau=-8}^{12} \theta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(Mid-SizedFirm_j) \\
&+ \sum_{\tau=-8}^{12} \beta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(SmallFirm_j) + \lambda X_{jt} + \alpha_j + \gamma_{jt} + \varepsilon_{j,t}
\end{aligned} \tag{6}$$

The model is similar to the one detailed in Equation 5, except we estimate a separate β_τ for each quarter τ relative to the shock. The omitted quarter is the first quarter of 2016, 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% faster than the mean after the GitHub announcement. Above-median productivity med-sized firms grew less by 2% 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 13.5% of the mean after the announcement. Column 2 creates a proxy indicator for firm exit, defined as a firm experiencing a headcount reduction of more than half compared to the previous quarter. Our findings show no significant change in the probability of closures for large or medium firms with above-median productivity levels post-GitHub shock. However, small firms with above-median productivity witnessed a substantial 38% increase over the mean in the probability of closures, after the GitHub announcement.[15] 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

[15] The magnitude appears large because the pre-period mean of closures is very low (0.6%), particularly due to large firms that are much less likely to exit. The pre-period mean of exits among small firms is 1%, suggesting that the shock leads to about 22.5% increase in their probability to exit.

effectively.[16]

Columns 3 to 6 instead use employee productivity as the outcome variable. Here, we utilize the public contributions for each employee, as these were not directly influenced by the shock. We exclude any contributions made by an employee to their own repository, as these may represent personal activity. The results show the number of public contributions increased by 5.6% for large firm with productivity estimates exceeding the median post the GitHub announcement. There was a 4 to 6% (though statistically insignificant in one specification) increase in public contributions per employee for large firms with above-median productivity levels post-GitHub shock. We find statistically indistinguishable changes for med-sized firms. Notably, the entire positive impact is negated for small firms with above-median productivity, which witnessed a 3% reduction in the number of contributions and a 4% reduction in the number of contributions per employee after the GitHub shock. We find that these results are driven by changes in the probability of experiencing very high (top decile) contributions, reducing concerns that we are capturing bias in logarithm due to count like variables. 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. We also test our results using different cut-offs for high GitHub activity. Table A.10 confirms robustness across alternate cut-offs, including the top quartiles and quintiles.

5.1 Industry-level implications

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

[16] In un-reported analysis we find that the results in Column 1 are robust to dropping firms which exited from the sample. We also find that a maximum of 1.5% of our sample is aqui-hires (where more than 80%) of the team moves to a new firm. Our results are also robust to excluding acqui-hire deals from the regression analysis.

of most impacted 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 explore this, we aggregate our data to the industry level and examine the correlation between changes in concentration and number of private contributions in each industry.[17] The change in concentration is measured by the difference in Herfindahl-Hirschman Index (HHI) in any industry, over two periods: five years before the shock (2011–2016) and five years after the shock (2016–2021). We use the average number of private contributions in each industry as our independent variable.

Although these correlations do not establish causal relationships, they provide valuable insights into broader trends, especially regarding the notable increase in industry concentration over the past decade (Autor et al., 2023). Table A.14 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 share of GitHub employees with private contributions experienced a 6.4% (14.0%) larger increase in labor concentration from 2016 to 2021, compared to 2011 to 2016, at the LinkedIn (SIC 3-digit) industry level. Similarly, industries with a 1 standard deviation higher average number of private contributions saw a 6.7% (13.7%) higher increase in labor concentration over a five-year period at the LinkedIn (SIC 3-digit) industry level. These findings demonstrate that there were greater concentration changes in labor markets within industries where more information was unveiled following the GitHub announcement.

[17] There has been considerable debate on the aggregate trends with regards to local labor market concentration trends. However, our analysis of LinkedIn data, filtered to focus on workers in the same firms and roles as those in our sample, reveals that the labor market for these workers operates at a national level, rather than being confined to local regions. Indeed, we examined the location transitions of workers who changed firms during the three years prior to the GitHub shock (2013-2016). 35% of these individuals relocated to different states as part of their job change.

6 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. Overall productivity of treated employees also improved after the information disclosure on individual quality. Consistent with a reduction of information asymmetry on the labor market, the surge in productivity information mitigated reliance on traditional signals of hard information, thereby democratizing opportunities for talent.

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 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. The overall welfare consequences ultimately depend on the relative importance of the productivity gains versus the costs associated with increased market concentration. Balancing these two factors offers a compelling direction for future research.

References

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

ACEMOGLU, D., U. AKCIGIT, H. ALP, N. BLOOM, AND W. KERR (2018): "Innovation, Reallocation, and Growth," *American Economic Review*, 108, 3450–91.

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.

AKERLOF, G. A. (1970): "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 84, 488–500.

ALTONJI, J. G. AND C. R. PIERRET (2001): "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116, 313–350.

ARAUJO, A., R. FERREIRA, S. LAGARAS, F. MORAES, J. PONTICELLI, AND M. TSOUTSOURA (2023): "The labor effects of judicial bias in bankruptcy," *Journal of Financial Economics*, 150.

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.

BAGHAI, R. P., R. C. SILVA, AND L. YE (2024): "Teams and Bankruptcy," *The Review of Financial Studies*, 37, 2855–2902.

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.

BENA, J. AND E. SIMINTZI (2024): "Machines could not compete with Chinese labor: Evidence from U.S. firms' innovation," *Working Paper*.

BERNARD, A. B., J. B. JENSEN, AND P. K. SCHOTT (2006): "Survival of the best fit:

Exposure to low-wage countries and the (uneven) growth of U.S. manufacturing plants," *Journal of International Economics*, 68, 219–237.

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

Bernstein, S., R. R. Townsend, and T. Xu (2023): "Flight to Safety: How Economic Downturns Affect Talent Flows to Startups," *The Review of Financial Studies*, 37, 837–881.

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*.

Bloom, N., M. Draca, and J. Van Reenen (2015): "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity," *The Review of Economic Studies*, 83, 87–117.

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

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.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of

West German Wage Inequality*," *The Quarterly Journal of Economics*, 128, 967–1015.

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., V. Gupta, J. Guzman, and M. P. Roche (2023): "Incentivizing Innovation in Open Source: Evidence from the GitHub Sponsors Program," Tech. rep., National Bureau of Economic Research.

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.

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*.

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*.

Farber, H. S. and R. Gibbons (1996): "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111, 1007–1047.

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*.

Floyd, E., S. Tomar, and D. J. Lee (2024): "Making the Grade (But Not Disclosing It): How Withholding Grades Affects Student Behavior and Employment," *Management Science*, 70, 2497–2517.

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*.

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.

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

Gupta, A. (2023): "Labor Mobility, Entrepreneurship, and Firm Monopsony: Evidence from Immigration Wait-Lines," *Entrepreneurship, and Firm Monopsony: Evidence from Immigration Wait-Lines (May 16, 2023)*.

Hager, S., C. Schwarz, and F. Waldinger (2024): "Measuring Science: Performance Metrics and the Allocation of Talent," *American Economic Review (Forthcoming).*

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).*

Hombert, J. and A. Matray (2018): "Can Innovation Help U.S. Manufacturing Firms Escape Import Competition from China?" *The Journal of Finance*, 73.

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

Jeffers, J. S. (2024): "The impact of restricting labor mobility on corporate investment and entrepreneurship," *The Review of Financial Studies*, 37, 1–44.

Jiang, W. and T. Li (2024): "Corporate Governance Meets Data and Technology," *Foundations and Trends in Finance*, 14, 61–136.

Kahn, L. B. (2013): "Asymmetric Information between Employers," *American Economic Journal: Applied Economics*, 5, 165–205.

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.

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

Pagano, M. and L. Picariello (2023): "Talent discovery, layoff risk and unemployment insurance," *European Economic Review.*

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

Rockoff, J. E., D. O. Staiger, T. J. Kane, and E. S. Taylor (2012): "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools," *American Economic Review*, 102, 3184–3213.

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

Schönberg, U. (2007): "Testing for Asymmetric Employer Learning," *Journal of Labor Economics*, 25, 651–691.

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.

Spence, M. (1973): "Job Market Signaling," *The Quarterly Journal of Economics*, 87, 355–374.

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.

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.

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

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

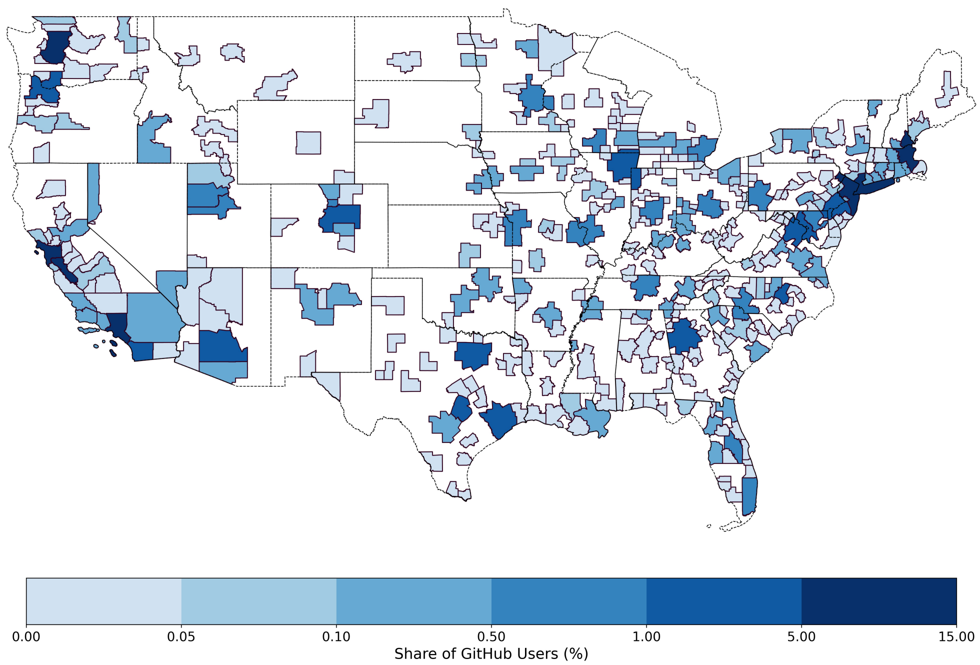


Figure 2: **GitHub Shock & Employee Mobility to Large Firms**

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 2 (April-June) of 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter 1 of 2016 (Jan-March) 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:

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

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with the time indicator for each quarter. 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. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Light blued lines show 95% confidence intervals for estimates.

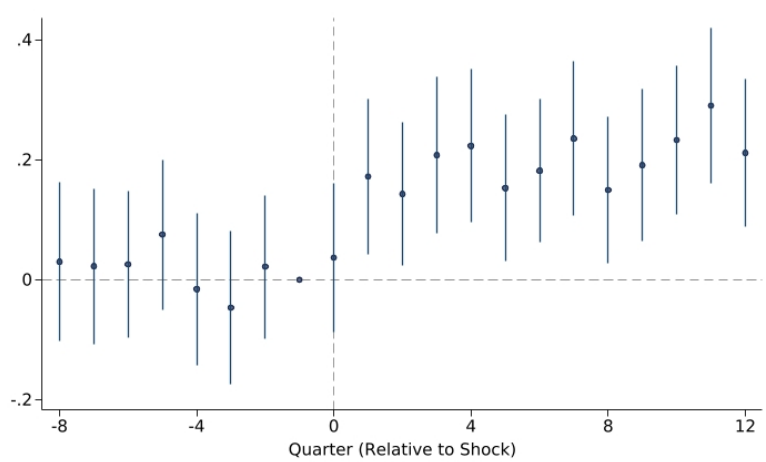


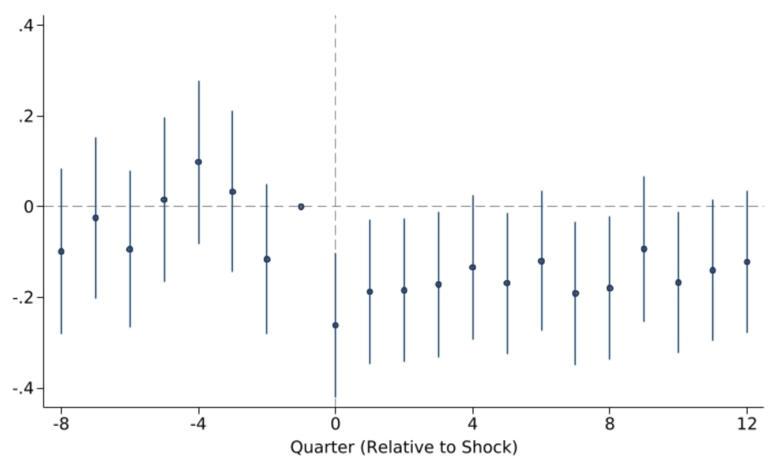
Figure 3: **GitHub Shock & Employee Mobility**

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 2 (April-June) of 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter 1 of 2016 (Jan-March) 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:

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

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with the time indicator for each quarter. The dependent variable is an indicator that equals 1 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. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Light blued lines show 95% confidence intervals for estimates.

(a) Move to Mid-Sized Firms



(b) Move to Small Firms

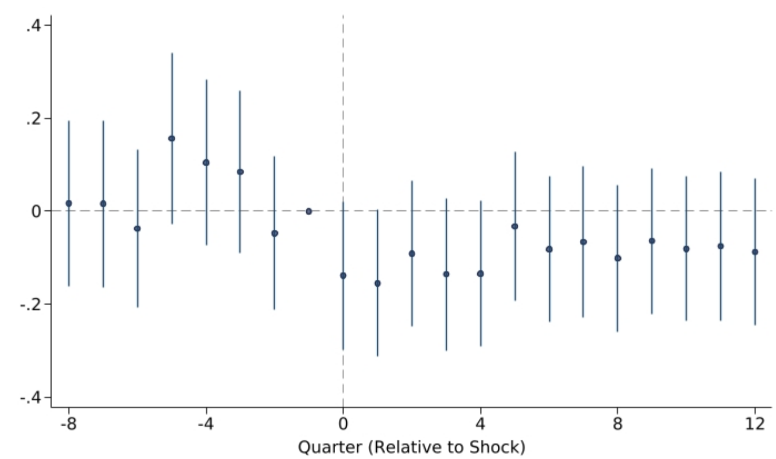


Figure 4: **GitHub Shock & Employee Mobility to Different Size Buckets**

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:

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

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with $\mathbb{1}(Post)$, an indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 2016). The dependent variable is an indicator that equals 1 when employee i switches job to a firm in a given size bucket. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Light blued lines show 95% confidence intervals for estimates.

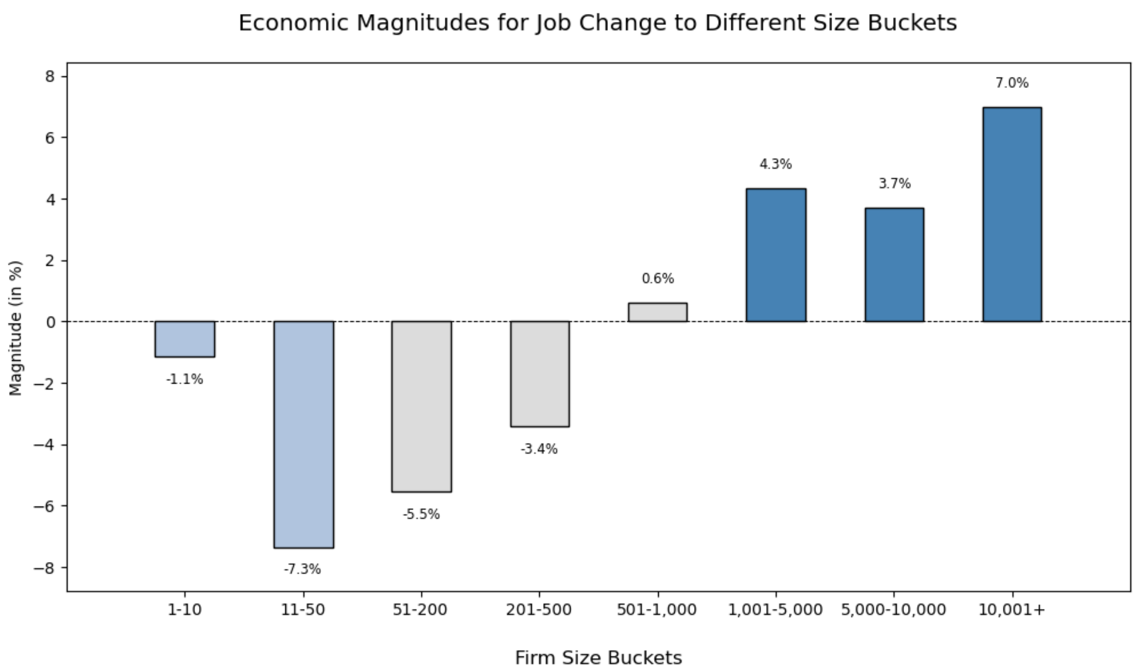


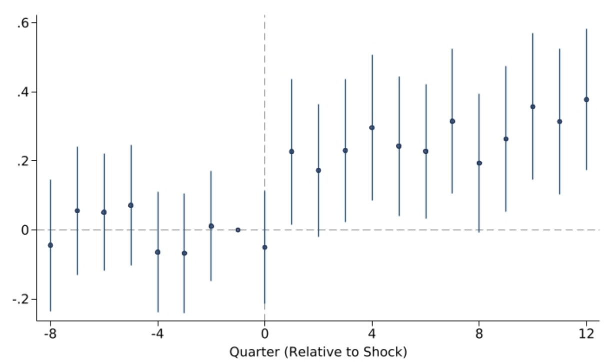
Figure 5: **GitHub Shock & Employee Mobility to Large Firms: Heterogeneity by Initial Firm Size**

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 2 (April-June) of 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter 1 of 2016 (Jan-March) 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:

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

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 the employee worked in the previous quarter, and the time indicator for each quarter. We also include indicator variables $\mathbb{1}(Mid)$ and $\mathbb{1}(Small)$ as well as their double interactions with *Productivity* and time, but do not show them in the above equation for brevity. Panel A plots interaction with $\mathbb{1}(Mid)$, an indicator variable that takes the value of 1 for firms with 50-1000 employees. Panel B plots the interaction with $\mathbb{1}(Small)$, an indicator for firms with less than 50 employees. The dependent variable is an indicator that equals 1 if the employee switches job to a large firm ($>$ 1000 employees) in that quarter. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Light blued lines show 95% confidence intervals for estimates.

(a) Mid-Sized Firms



(b) Small Firms

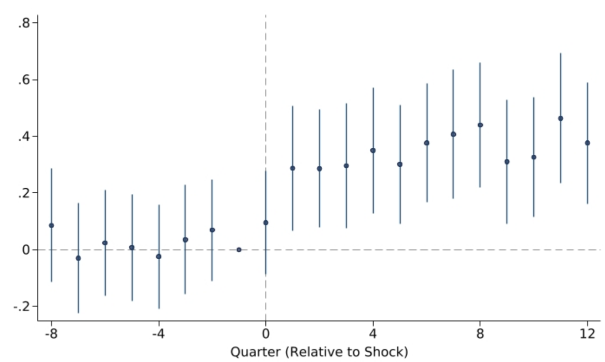


Figure 6: **GitHub Shock & Employee Promotions at Large Firms**

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 2 (April-June) of 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter 1 of 2016 (Jan-March) 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:

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

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with the time indicator for each quarter. We also include indicator variables $\mathbb{1}(Mid)$ and $\mathbb{1}(Small)$ as well as their double interactions with *Productivity* and time, but do not show them in the above equation for brevity. The dependent variable is an indicator that equals 1 when employee i does not switch jobs and her salary or seniority improves in quarter t compared to quarter *t-1* and is 0 otherwise. α_i represents employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. δ_{ft} is firm cohort fixed effects representing an interaction of the employee's current firm and the year of joining the firm. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Light blued lines show 95% confidence intervals for estimates.

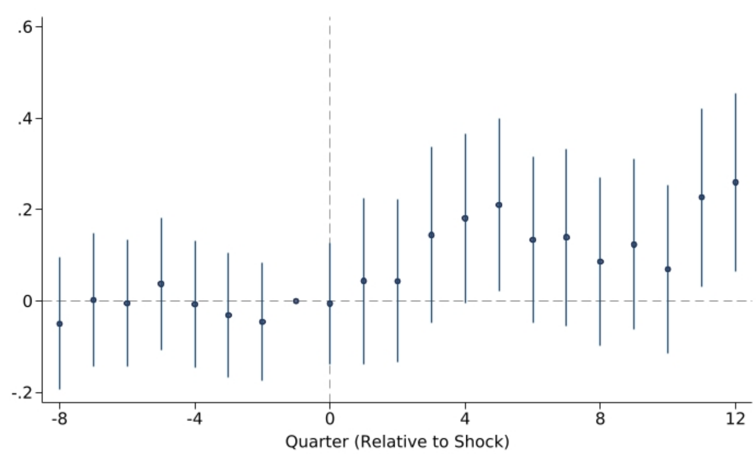


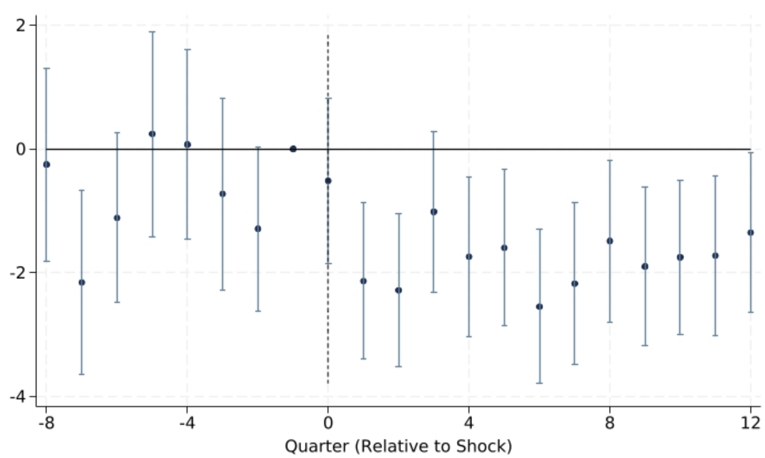
Figure 7: **GitHub Shock & Employment Growth at Small Firms**

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 2 (April-June) of 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter 1 of 2016 (Jan-March) 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:

$$
\begin{aligned}
EmploymentGrowth_{j,t} = & \sum_{\tau=-8}^{12} \theta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \\
& + \sum_{\tau=-8}^{12} \delta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(MidSized_j) \\
& + \sum_{\tau=-8}^{12} \beta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(Small_j) + \lambda X_{jt} + \alpha_j + \gamma_{jt} + \varepsilon_{j,t}
\end{aligned}
$$

The independent variable is a triple interaction of three binary indicators: $\mathbb{1}(Productivity_j > Median)$ that takes the value of 1 if firm j's average AKM-based employee productivity is above median, $\mathbb{1}(Small_j)$ that takes the value of 1 if firm j is small (<= 50 employees), and $\mathbb{1}(\tau = t)$ that is the time indicator for each quarter. We also include the double interaction of indicator variables *MidSized* and $\mathbb{1}(Small)$ with the indicator $\mathbb{1}(Productivity_j > Median)$ and time, but do not show them in the above equation for brevity. In Panel (a), the dependent variable is the employment change (in %) for firm j in quarter t with respect to quarter *t-1*. In Panel (b), the dependent variable is a dummy that takes the value of 1 if the employment at the firm drops by more than 50% in that quarter and 0 otherwise. X_{jt} represents firm's time-varying lagged employment. α_j is firm fixed effects. γ_{jt} represents the firm's industry $\times$ state interacted with time fixed effects. We cluster the standard errors at the firm level. Light blued lines show 95% confidence intervals for estimates.

(a) Employment Growth



(b) $\mathbb{1}$(Employment Reduction > 50%)

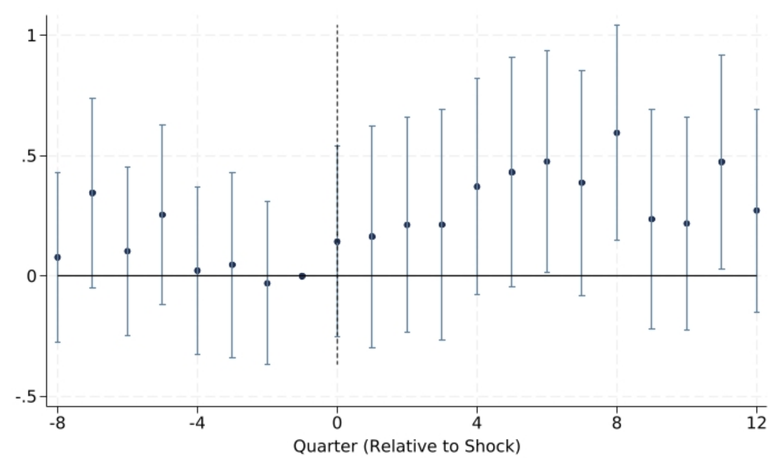


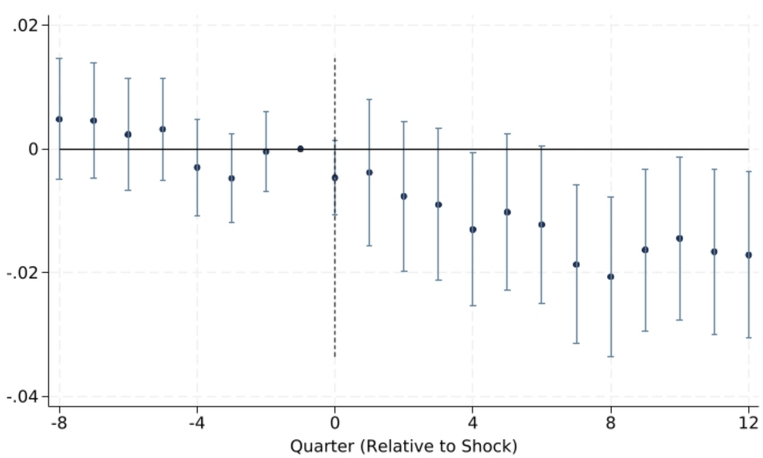
Figure 8: **GitHub Shock & Productivity at Small Firms**

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 2 (April-June) of 2016, during which GitHub introduced the policy that allowed users to display their productivity information more accurately. The quarter 1 of 2016 (Jan-March) 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:

$$\mathbb{1}(Contributions_{j,t}) = \sum_{\tau=-8}^{12} \theta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) + \sum_{\tau=-8}^{12} \delta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(MidSized_j)$$
$$+ \sum_{\tau=-8}^{12} \beta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(Small_j) + \lambda X_{jt} + \alpha_j + \gamma_{jt} + \varepsilon_{j,t}$$

The independent variable is a triple interaction of three binary indicators: $\mathbb{1}(Productivity_j > Median)$ that takes the value of 1 if firm j's average AKM-based employee productivity is above median, $\mathbb{1}(Small_j)$ that takes the value of 1 if firm j is small (<= 50 employees), and $\mathbb{1}(\tau = t)$ that is the time indicator for each quarter. We also include the double interaction of indicator variables *MidSized* and $\mathbb{1}(Small)$ with the indicator $\mathbb{1}(Productivity_j > Median)$ and time, but do not show them in the above equation for brevity. The dependent variable in Panel (a) (Panel (b)) is a binary indicator that takes the value of 1 if the firm j's total public contributions (total public contributions per employee) in the quarter is in the top decile and is 0 otherwise. Firm contributions only include public contributions made by incumbent employees towards repositories not owned by the employees themselves. X_{jt} represents firm's time-varying lagged stock of total contributions as well as the median pre-period public contribution interacted with an indicator for quarters after GitHub's policy change. α_j is firm fixed effects. γ_{jt} represents the firm's industry $\times$ state interacted with time fixed effects. We cluster the standard errors at the firm level. Light blued lines show 95% confidence intervals for estimates.

(a) Public Contributions



(b) Public Contributions per Employee

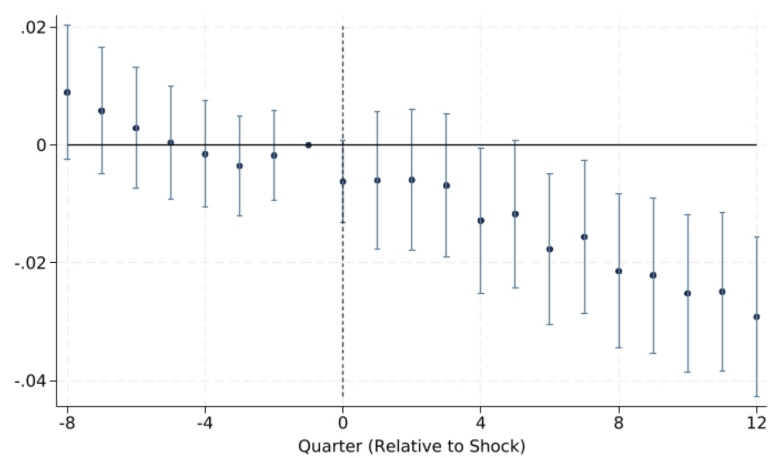


Table 1: **Descriptive Statistics**

This table presents descriptive statistics for employee-level variables in Panel A, and firm-level variables in Panel B.

Panel A: Employee Level Summary						
	Mean	**Std. Dev.**	**P10**	**Median**	**P90**	**Obs.**
$\mathbb{1}$(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
$\mathbb{1}$(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
Panel B: Firm Level Summary						
	Mean	**Std. Dev.**	**P10**	**Median**	**P90**	**Obs.**
$\mathbb{1}$(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	50,307
Public Contributions/Employee	10.26	107.47	0.13	1.75	17.26	50,307

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

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 the following regression specification:

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

ϑ_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 employee's current job duration, and total work experience. We also include firm fixed effects (ϑ_f), time fixed effects (ϑ_t), and employee's role fixed effects (ϑ_k). Panel A shows the number of observations in our initial sample, the connected set (the largest set of firms linked by employee mobility), the filtered set (connected set with at least two employees in each firm-time observation), the sample for which productivity estimates were obtained, and the final sample remaining 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	% Initial Sample	N	% Initial Sample
Initial 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**

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 $\mathbb{1}(Post)$, an indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 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. Columns 1 to 5 present estimates with varying controls and fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **($p < 0.05$), ***($p < 0.01$).

	(1)	(2)	(3)	(4)	(5)
Outcome:	**𝟙(Job Change to Large Firms)**				
Productivity × 𝟙(Post)	0.0012***	0.0012***	0.0017***	0.0017***	0.0017***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Employee F.E.	Y	Y	Y	Y	Y
Time F.E.	Y	Y	Y	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 (%)	3.98	3.8	5.6	6.09	5.72

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

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. Public contributions only include contributions made towards repositories not owned by the employees themselves. Panel B estimates results on a subsample of active GitHub Users. Panel B1 only includes users with at least 10 total GitHub contributions during the pre-period (in the 12 quarters before May 2016). Panel B2 restricts the sample to users with non-zero private GitHub contributions during the pre-period. Panel B3 restricts the sample to users with at least 10 private GitHub contributions during the pre-period. Panel C1 (C2) excludes users who, during the pre-shock period, work at firms that have zero (less than 10) private contributions on GitHub in our sample. All tests correspond to our baseline specification, Column 5 in Table 3. Standard errors are bootstrapped within employee clusters using 100 repetitions. 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)	-0.0000	(0.0002)	3,182,288	0.212	0.033	-0.02
Panel B: Within Active User Comparison						
B1: Pre-Period Individual Total Contributions ≥ 10						
Productivity × $\mathbb{1}$(Post)	0.0019***	(0.0002)	1,708,418	0.224	0.033	8.2
B2: Pre-period Individual Private Contributions > 0						
Productivity × $\mathbb{1}$(Post)	0.0012***	(0.0004)	506,974	0.271	0.043	8.56
B3: Pre-period Individual Private Contributions ≥ 10						
Productivity × $\mathbb{1}$(Post)	0.0013***	(0.0003)	401,146	0.274	0.024	9.58
Panel C: Within Active Firm Comparison						
C1: Firm Private Contributions > 0						
Productivity × $\mathbb{1}$(Post)	0.0038***	(0.0003)	2,078,974	0.215	0.034	11.2
C2: Firm Private Contributions ≥ 10						
Productivity × $\mathbb{1}$(Post)	0.0041***	(0.0003)	1,953,461	0.216	0.033	12.12

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

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:

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

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with $\mathbb{1}(Post)$, an indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 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, and small firm (<= 50 employees) for Column 3. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. 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)	0.0017***	-0.0010***	-0.0010***
	(0.0002)	(0.0002)	(0.0002)
Employee F.E.	Y	Y	Y
Time F.E.	Y	Y	Y
Salary & Seniority Controls	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	0.029	0.023
Magnitude (%)	5.72	-3.74	-5.08

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

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 *origin* firm size. We report the differential impact of GitHub's policy change on productive employees by origin firm type estimated by the β coefficient from the equation:

$$\begin{aligned}\mathbb{1}(MoveToLargeFirm)_{i,t} = \beta_1 Productivity_i \times \mathbb{1}(Post_t) + \beta_2 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_{i(t-1)}) \\ + \beta_3 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Small)_{i(t-1)}) + \lambda X_{it} + \alpha_i + \theta_t + \gamma_{it} + \varepsilon_{i,t}\end{aligned}$$

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 origin firm, i.e. where employee i worked in the previous quarter, and an indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 2016). $\mathbb{1}(Mid)$ is an indicator variable that takes the value of 1 for firms with 50-1000 employees and $\mathbb{1}(Small)$ 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 also estimate coefficients for indicator variables $\mathbb{1}(Mid)$ and $\mathbb{1}(Small)$ as well as their double interactions with *Productivity* and $\mathbb{1}(Post)$, but do not report them for brevity. Columns 1 to 5 present estimates with varying controls and fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)
Outcome:	**$\mathbb{1}$(Job Change to Large Firms)**				
Productivity × $\mathbb{1}$(Post)	-0.0039***	-0.0042***	-0.0028***	-0.0022***	-0.0022***
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0005)
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Mid)	0.0049***	0.0051***	0.0043***	0.0031***	0.0030***
	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0006)
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Small)	0.0054***	0.0056***	0.0046***	0.0039***	0.0037***
	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0006)
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 (Mid) (%)	3.27	3.14	4.86	3.29	3.01
Magnitude (Small) (%)	4.81	4.73	5.95	6.21	5.46

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

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:

$$\mathbb{1}(Promotion)_{i,f,t} = \beta_1 Productivity_i \times \mathbb{1}(Post_t) + \beta_2 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_{i(t-1)}) \\ + \beta_3 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(Small)_{i(t-1)}) + \alpha_i + \gamma_{it} + \delta_{ft} + \varepsilon_{i,f,t}$$

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 (i.e., after quarter 2 of 2016). Even-numbered columns additionally interact this term with indicators for the size of the firm where the employee worked in the previous quarter. $\mathbb{1}(Mid)$ is an indicator variable that takes the value of 1 for firms with 50-1000 employees, and $\mathbb{1}(Small)$ is an indicator for firms with less than 50 employees. The dependent variable is an indicator variable that is 1 if there is a promotion (salary or job-seniority increase compared to the previous quarter) and the employee switches job to a large firm (> 1000 employees) in columns 1 and 2, switches job to a small or mid-sized firm (<= 1000 employees) in columns 3 and 4, remains at the same firm in columns 5 and 6. We also estimate coefficients for indicator variables $\mathbb{1}(Mid)$ and $\mathbb{1}(Small)$ as well as their double interactions with *Productivity* and $\mathbb{1}(Post)$, but do not report them for brevity. α_i represents employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. In columns 5 and 6, we also control for δ_{ft}, firm cohort representing an interaction of the employee's current firm and the year of joining the firm. Standard errors are bootstrapped within employee clusters using 100 repetitions. 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 x $\mathbb{1}$(Post)	0.0012***	-0.0016***	-0.0016***	-0.0045***	0.0010***	0.0014**
	(0.0001)	(0.0003)	(0.0002)	(0.0005)	(0.0002)	(0.0005)
Productivity x $\mathbb{1}$(Post) x $\mathbb{1}$(Mid)		0.0023***		0.0042***		0.0001
		(0.0005)		(0.0007)		(0.0008)
Productivity x $\mathbb{1}$(Post) x $\mathbb{1}$(Small)		0.0026***		0.0052***		-0.0005
		(0.0005)		(0.0008)		(0.0007)
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 x Language x Role x Time F.E.	Y	Y	Y	Y	Y	Y
Industry x Location x 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	2,939,059	2,918,178
# Unique Employees	143,446	142,754	143,446	142,754	143,022	142,797
R-squared	0.204	0.216	0.212	0.22	0.274	0.275
Y-Mean	0.024	0.023	0.036	0.035	0.019	0.019
Magnitude (Main) (%)	5.92	-7.9	-5.19	-14.9	5.9	8.43
Magnitude (Small) (%)		5.23		2.34		5.1

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

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 and limited mobility, estimated by the β coefficients from the equation:

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

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 (i.e., after quarter 2 of 2016). Columns 1 includes 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. Column 2 includes interaction with $\mathbb{1}$(Schooling < Median), which takes the value of 1 for employees whose years of school attended is less than median and 0 otherwise. Column 3 includes triple interaction with $\mathbb{1}$(Non-Elite School) that takes the value 1 if the employee did not attend a school ranked among the top 20 in the US by US News & Ranking in 2016. Column 4 includes triple interaction with $\mathbb{1}$(Less NCA Enforceability) that takes the value of 1 if the employee's state at the time of shock did not not have any statute governing NCAs. We also estimate coefficients for double interaction of *Group* and $\mathbb{1}(Post)$, but do not report them for brevity. We control for employees' time-varying salary and seniority. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)
Outcome:				
Productivity × $\mathbb{1}$(Post)	0.0012***	0.0016***	0.0008*	0.0012***
	(0.0003)	(0.0002)	(0.0005)	(0.0003)
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Experience < Median)	0.0008**			
	(0.0004)			
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Schooling < Median)		0.0016*		
		(0.0008)		
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Non-Elite School)			0.0010*	
			(0.0006)	
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Less NCA Enforceability)				0.0006*
				(0.0004)
Employee F.E.	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y
Salary & Seniority Controls	Y	Y	Y	Y
Cohort × Language × Role × Time F.E.	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y
Observations	3,199,173	3,199,173	3,199,173	3,199,173
# Unique Employees	143,446	143,446	143,446	143,446
R-Squared	0.212	0.211	0.212	0.212
Y-Mean	0.033	0.033	0.033	0.033
Magnitude - Main (%)	4.23	5.45	2.79	4.14
Magnitude - Interaction (%)	7.12	10.91	6.08	6.35

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

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 estimated by the β coefficients from the equation:

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

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 (i.e., after quarter 2 of 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 whether any Human Resource professional at the firm used GitHub during the pre-period. In columns 3-4, the subset is based on whether the firm and its employees use GitHub (derived from firm-level private contributions). In columns 5-6, 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. Finally, in columns 7-8, the subset is based on job stability defined by whether the firm had below-median average forced turnovers in the pre-shock period. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. 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)**							
	Tech-savvy Recruiters		**GitHub Usage**		**Steep Career Growth**		**High Stability**	
	Yes	No	Yes	No	Yes	No	Yes	No
Productivity × $\mathbb{1}$(Post)	0.0014***	0.0003***	0.0015***	0.0001***	0.0015***	0.0001*	0.0000	0.0017***
	(0.0001)	(0.0001)	(0.0002)	(0.0000)	(0.0002)	(0.0001)	(0.0001)	(0.0002)
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	3,199,173	3,199,173	3,199,173	3,199,173	3,192,168	3,192,168	3,194,104	3,194,104
# Unique Employees	143,446	143,446	143,446	143,446	143,445	143,445	143,445	143,445
R-Squared	0.189	0.23	0.204	0.266	0.2	0.221	0.224	0.207
Y-Mean	0.019	0.015	0.03	0.003	0.024	0.007	0.003	0.029
Magnitude (%)	**8.49**	**2.19**	**5.82**	**4.7**	**7.16**	**2.19**	**-0.35**	**6.6**
T-test ($H_0 : \beta_{Yes} = \beta_{No}$)	(0.0000)***		(0.0604)*		(0.0000)***		(0.0000)***	

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

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 estimated by the β coefficients from the equation:

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

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 (i.e., after quarter 2 of 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 whether average employment growth at the firm in the pre-period is high (above-median) or low (below-median). Similarly, in columns 3-4, the subset splits the sample of large firms into those with high (above-median) or low (below-median) sales growth. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)
Outcome:	**$\mathbb{1}$(Job Change to Large Firms with)**			
	High Employment Growth		**High Sales Growth**	
	Yes	No	Yes	No
Productivity × $\mathbb{1}$(Post)	0.0008***	0.0009***	0.0010***	0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Employee F.E.	Y	Y	Y	Y
Salary & Seniority Controls	Y	Y	Y	Y
Cohort × Language × Role × Time F.E.	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y
Observations	3,191,534	3,191,534	3,152,320	3,152,320
# Unique Employees	143,445	143,445	143,441	143,441
R-Squared	0.18	0.221	0.19	0.217
Y-Mean	0.011	0.02	0.014	0.005
Magnitude (%)	**7.99**	**5.11**	**8.63**	**3.92**
T-test ($H_0 : \beta_{Yes} = \beta_{No}$)	(0.0027)***		(0.0001)***	

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

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 pre-treatment estimated by the β coefficient from the equation:

$$FirmOutcome_{j,t} = \beta_1 \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(Post_t) + \beta_2 \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(Post_t) \times \mathbb{1}(Mid_j) + \beta_3 \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(Post_t) \times \mathbb{1}(Small_j) + \lambda X_{jt} + \alpha_j + \gamma_{jt} + \varepsilon_{j,t}$$

The independent variable is a triple interaction of three indicators: $\mathbb{1}(Productivity_j > Median)$ that takes the value of 1 if firm j's average AKM-based employee productivity is above median, an indicator for firm size, and an indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 2016). $\mathbb{1}(Mid)$ is an indicator variable that takes the value of 1 for firms with 50-1000 employees, and $\mathbb{1}(Small)$ is an indicator for firms with less than 50 employees. In column 1, the dependent variable is the percentage change in employment compared to the previous quarter. In column 2, the dependent variable is a binary indicator which takes the value of 1 if the employment at the firm falls by more than 50% compared to the previous quarter and 0 otherwise. In columns 3-4 and 5-6, the dependent variable is the firm's public contributions and per-capita public contributions, respectively. Firm public contributions only include contributions made by incumbent employees towards repositories not owned by the employees themselves. In columns 3 and 5, the dependent variable is the logarithm of one added to the firm public contributions variable. In columns 4 and 6, the dependent variable takes the value of 1 if the firm contribution variable in the quarter is in the top decile and 0 otherwise. We also estimate coefficients for double interaction of indicator variables $\mathbb{1}(Mid)$ and $\mathbb{1}(Small)$ with $\mathbb{1}(Post)$, but do not report them for brevity. X_{jt} represents firm's time-varying controls (lagged employment in columns 1-2 and lagged stock of total contributions as well as pre-period median public contributions interacted with $\mathbb{1}(Post)$ in columns 3-6). α_j is firm fixed effects. γ_{jt} represents the firm's industry $\times$ state interacted with time fixed effects. Standard errors are clustered at the firm level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	**Employment**		**Public Contributions**		**Per-capita Contributions**	
	QoQ Emp. Growth	*Emp Reduction > 50%*	*Log(1+Y)*	*$\mathbb{1}$(Top Decile)*	*Log(1+Y)*	*$\mathbb{1}$(Top Decile)*
$\mathbb{1}$(Productivity > Median) × $\mathbb{1}$(Post)	0.0028*	-0.0005	0.0694***	0.0083*	0.0370**	0.0065
	(0.0016)	(0.0004)	(0.0252)	(0.0049)	(0.0162)	(0.0041)
$\mathbb{1}$(Productivity > Median) × $\mathbb{1}$(Post) × $\mathbb{1}$(Mid)	-0.0018	-0.0005	0.0038	-0.0033	0.0002	-0.0021
	(0.0024)	(0.0006)	(0.0301)	(0.0058)	(0.0199)	(0.0051)
$\mathbb{1}$(Productivity > Median) × $\mathbb{1}$(Post) × $\mathbb{1}$(Small)	-0.0125***	0.0024***	-0.1061***	-0.0149***	-0.0668***	-0.0160***
	(0.0028)	(0.0008)	(0.0293)	(0.0056)	(0.0203)	(0.0052)
Controls	Y	Y	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y	Y	Y
Observations	1,106,564	1,106,564	878,160	878,160	878,160	878,160
# Unique Firms	51,038	51,038	46,471	46,471	46,471	46,471
R-squared	0.232	0.13	0.725	0.58	0.647	0.497
Y-Mean	0.072	0.005	1.248	0.102	0.766	0.1
Magnitude- Base (%)	3.95	-11.25	5.56	8.11	4.83	6.55
Magnitude - Small (%)	-13.45	38.89	-2.94	-6.46	-3.89	-9.53

Internet Appendix A

A.1 Figures

Figure A.1: **Test for Endogenous Mobility in AKM Sample: Event Study around Moves**

This figure presents the evolution of employee private contributions on GitHub around job moves in the AKM sample (May 2011 - April 2016), before GitHub's policy change. The outcome of interest is the logarithm of employee's private contributions in the five quarters before and after they move. The analysis is restricted to movers who were employed at the origin firm for at least 6 months prior to the move, and remain employed at the new destination firm for at least 6 months following the move. Moves are then classified based on the unobserved firm fixed component derived from the AKM corresponding to both the origin as well as the destination firm. For instance, a move "1 to 2" signifies that an employee moves from an origin firm whose firm fixed component, derived from the AKM, lies in quartile 1 of the distribution to a destination firm whose firm fixed component falls in quartile 2. For simplicity, the figure only plots moves away from firms either in the top or bottom quartile.

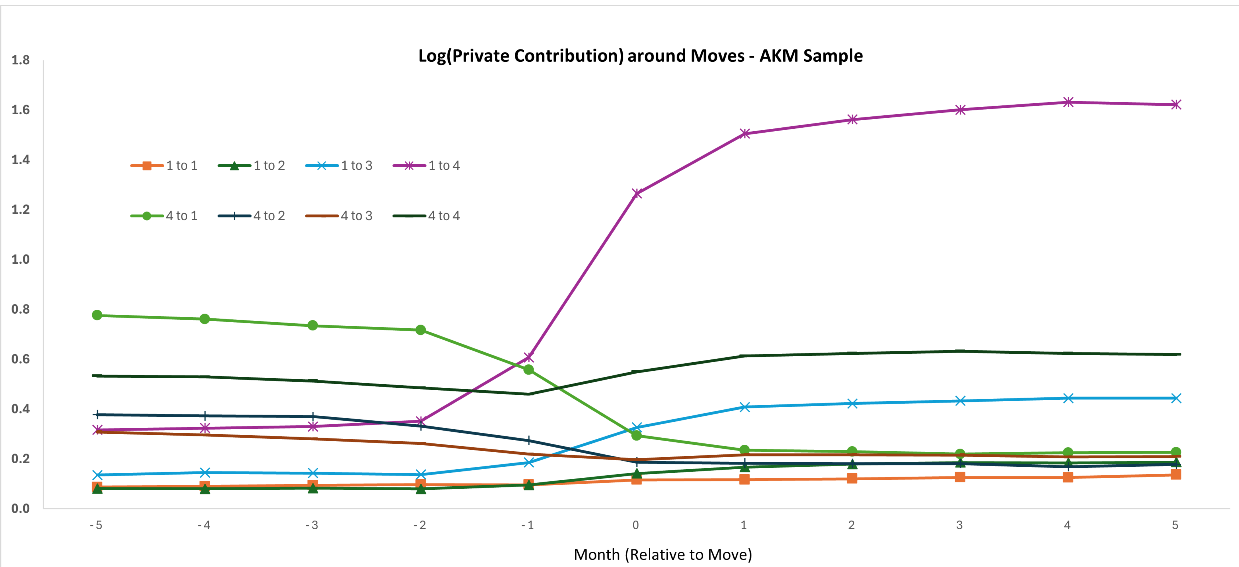
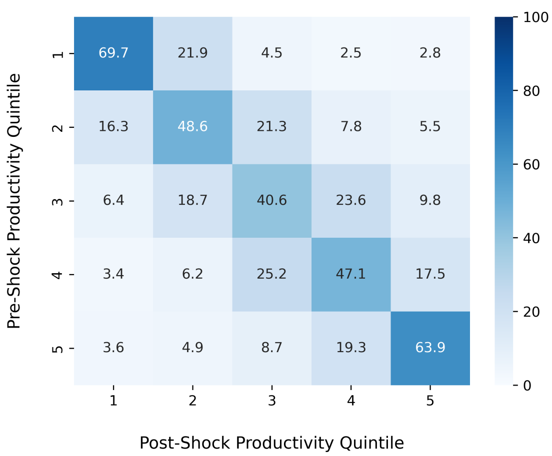
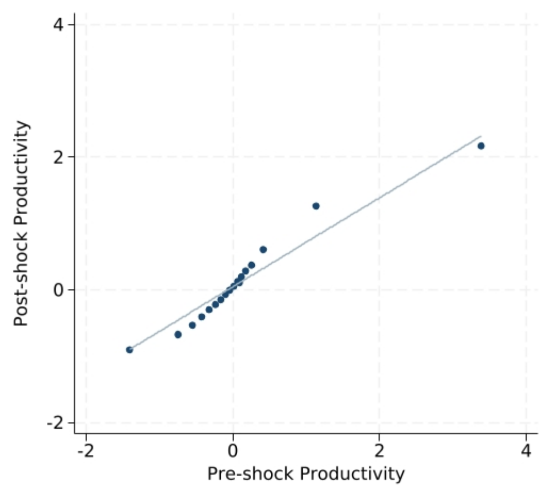


Figure A.2: **Persistence of Employee Productivity: Evidence from AKM Decomposition Before and After Shock**

This figure presents the relationship between employee productivity estimated from two different AKM samples - before and after GitHub's policy change. Panel A shows a heat map on the probability of transition of employees across productivity quintiles derived from the two AKMs. The pre-shock productivity quintile sorts employees into 5 buckets based on the productivity estimates from the pre-shock AKM decomposition (AKM sample spanning May 2011 - April 2016). The post-shock productivity quintile sorts employees similarly using AKM decomposition, but in the post-shock period (AKM sample spanning May 2016 - April 2021). A cell (row, column) represents the percentage of people in the pre-shock productivity quintile "row" who transition to post-shock quintile "column". For instance, cell (1, 2) indicates that 21.9% employees, who were initially in quintile 1 based on their productivity estimates from pre-shock AKM move to quintile 2 of productivity when estimates are derived from AKM in the post-period. Panel B shows the binscatter plot of the two productivity estimates after controlling for employees' industry, location, and coding language. The entire analysis is restricted to employees for whom productivity estimates could be derived from the AKM decomposition in both pre- and post-shock periods.



Panel A



Panel B

Figure A.3: **GitHub Shock & Employee Mobility to Large Firms: Robustness with Dummy Treatment Indicator**

This figure presents an event study on employee mobility to large firms around the disclosure of employee productivity information on GitHub. The event study replicates the setup in Figure 2 but uses a dummy measure of *Productivity* as the dependent variable. The dummy indicator takes the value of 1 if the employee's normalized AKM-based productivity is positive and takes the value of 0 otherwise. The standard errors are clustered at the employee level. The rest of the variables and empirical setup remain unchanged.

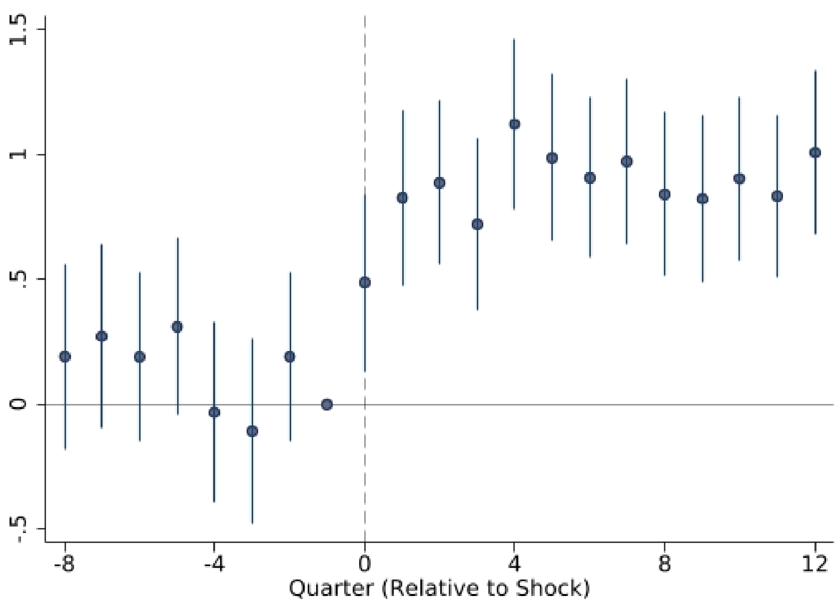
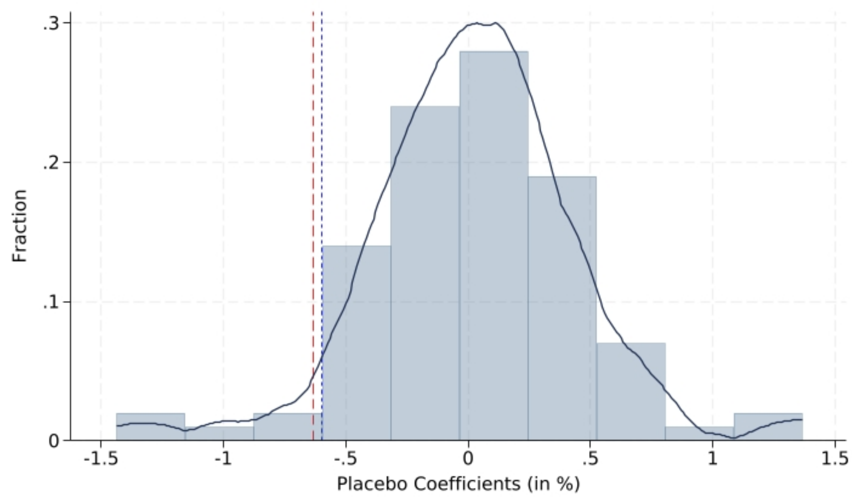


Figure A.4: **Firm Market Reaction: Placebo Test**

The figure below plots the histogram of estimated coefficients from 100 trials of a falsification exercise using randomly selected placebo dates. In each trial, we randomly select a placebo event date between 19th May 2015 and 18th May 2016 (from one year before the actual date of GitHub's policy change) and obtain a placebo coefficient by running the regression estimate corresponding to column 2 of Table A.11. The blue line indicates the 5th percentile of the distribution and the red line plots the the baseline coefficient estimate from Column 2 of Table A.11.



Tables

Table A.1: **Sample Characteristics**

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**

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**

This table presents robustness tests for estimates reported in column 5 of 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 binary dummy indicators for employees' productivity. In C1, $\mathbb{1}(Productivity > 0)$ 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 C2, C3, and C4, the dummy variable takes the value of 1 if the employee's productivity is in the top decile, top quartile, and top tercile respectively. In Panel D, we use the original AKM specification but restrict the AKM sample until quarter 2, i.e., April-June quarter of 2014. We then estimate employee mobility in the quarters after quarter 2 (April-June) of 2014. In Panel E, we control for more granular fixed effects by interacting employee's cohort, role, language, location, and industry with time. Finally, in Panel F, we run the test on a matched sample. Employees with positive normalized productivity estimates are matched with employees with non-positive estimates based on demographic characteristics (gender, white or non-white, whether in software engineering, whether located in SF/NY/Seattle, experience, salary, and seniority) in F1 and demographic characteristics along with public contributions in F2. All panels run tests corresponding to our baseline specification, Column 5 in Table 3. Standard errors are bootstrapped within employee clusters using 100 repetitions in Panels A, D, E, and F and are clustered at the employee level in Panels B and C. 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: Alternate AKM Specifications						
A1: AKM with Asine Function						
Productivity × $\mathbb{1}$(Post)	0.0017***	(0.0002)	3,199,173	0.212	0.033	5.77
A2: AKM with Categorical Variables						
Productivity × $\mathbb{1}$(Post)	0.0017***	(0.0002)	3,199,173	0.212	0.033	5.77
Panel B: Alternate Productivity Definition						
B1: Log(Private Contributions)						
Productivity × $\mathbb{1}$(Post)	0.0011***	(0.0001)	3,199,173	0.212	0.033	7.49
B2: Share Private Contributions						
Productivity × $\mathbb{1}$(Post)	0.0042***	(0.0008)	3,199,173	0.212	0.033	5.79
Panel C: Dummy Productivity Definition						
C1: Normalized Productivity Estimate > 0						
$\mathbb{1}$(Productivity > 0) × $\mathbb{1}$(Post)	0.0067***	(0.0005)	3,199,173	0.212	0.033	20.21
C2: Productivity in Top Decile						
$\mathbb{1}$(Top Decile Productivity) × $\mathbb{1}$(Post)	0.0068***	(0.0006)	3,199,173	0.212	0.033	20.55
C3: Productivity in Top Quartile						
$\mathbb{1}$(Top Quartile Productivity) × $\mathbb{1}$(Post)	0.0068***	(0.0005)	3,199,173	0.212	0.033	20.47
C4: Productivity in Top Tercile						
$\mathbb{1}$(Top Tercile Productivity) × $\mathbb{1}$(Post)	0.0052***	(0.0005)	3,199,173	0.212	0.033	15.61
Panel D: Alternate AKM Sample Period						
D1: Pre Sample AKM (2011-14)						
Productivity × $\mathbb{1}$(Post)	0.0011***	(0.0002)	2,014,005	0.219	0.031	4.09
Panel E: More Granular Controls						
E1: Control for Cohort × Role × Language × Location × Industry × Time						
Productivity × $\mathbb{1}$(Post)	0.0020***	(0.0005)	745,659	0.37	0.036	6.78
Panel F: Matched Sample						
F1: Match on Individual Demographics						
Productivity × $\mathbb{1}$(Post)	0.0017***	(0.0002)	1,444,081	0.247	0.033	7.67
F2: Match on Individual Demographics and Public Contributions						
Productivity × $\mathbb{1}$(Post)	0.0017***	(0.0002)	1,405,593	0.246	0.032	7.28

Table A.4: **GitHub Shock & Employee Mobility to Large Firms: Robustness by Excluding Users with Change in Pre- and Post-Productivity Estimates**

This table presents robustness tests for estimates reported in column 5 of Table 3 by excluding users with sizeable changes in productivity estimates derived from the AKM decomposition in the pre- and post-shock period. Change in productivity estimates is defined as the change in productivity quintile in Panel A of Figure A.2. Column 1 (2) (3) presents robustness by limiting the sample to employees whose absolute productivity quintile change, i.e. $\Delta Prod$, is not more than three (two) (one). Column 4 only retains employees who fall in the same productivity quintile in the pre- and post-period. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)
Outcome:	**$\mathbb{1}$(Job Change to Large Firms)**			
	Filtering Employees based on Change in Productivity Quintile ($\Delta Prod$)			
	$\Delta Prod <= 3$	$\Delta Prod <= 2$	$\Delta Prod <= 1$	$\Delta Prod = 0$
Productivity × $\mathbb{1}$(Post)	0.0018***	0.0020***	0.0023***	0.0029***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Employee F.E.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Cohort × Language × Role × Time F.E.	Yes	Yes	Yes	Yes
Industry × Location × Time F.E.	Yes	Yes	Yes	Yes
Observations	2,941,306	2,825,413	2,529,563	1,488,920
# Unique Employees	131,095	125,900	112,593	65,952
R-Squared	0.213	0.215	0.22	0.245
Y-Mean	0.035	0.035	0.036	0.035
Magnitude (%)	6.07	6.5	7.63	11.93

Table A.5: **GitHub Shock & Employee Mobility to Different Firm-Size Buckets**

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:

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

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with $\mathbb{1}(Post)$, an indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 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. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. 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)**				
Firm Size:	[1 - 10]	[11 - 50]	[51 - 200]	[201 - 500]	[501 - 1,000]	[1,001 - 5,000]	[5,001 - 10,000]	[10,001 -)
Productivity × $\mathbb{1}$(Post)	-0.0001	-0.0009***	-0.0007***	-0.0002**	0.0000	0.0004***	0.0002**	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	0.015	0.008	0.006	0.01	0.005	0.018
Magnitude (%)	-1.13	-7.35	-5.54	-3.41	0.62	4.35	3.69	7.01

Table A.6: **GitHub Shock & Employee Mobility to Superstar Firms**

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

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

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with $\mathbb{1}(Post)$, an indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 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. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. 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)	0.0009***	0.0009***	0.0005***	0.0004***	0.0010***	0.0009***	0.0006***	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	0.007
Magnitude (%)	6.34	5.54	8.9	7.84	6.32	5.93	7.66	6.11

Table A.7: **GitHub Shock & Employee Mobility to Star Startups**

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:

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

The independent variable is the normalized productivity estimate of the employee i derived from the AKM decomposition interacted with $\mathbb{1}(Post)$, an indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 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. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. 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)	(0.0000)
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 (%)	-4.53	3.76	2.49	12.95	13.87
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 -)

Table A.8: **GitHub Shock & Employee Mobility to Different Industry and Location**

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, estimated by the β coefficients from the equation:

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

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 (i.e., after quarter 2 of 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 whether the move is associated with a change in the employee's location. In columns 3-4, the subset is based on whether the industry categorization changes with the new job. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)
Outcome:	**$\mathbb{1}$(Job Change to Large Firms with)**			
	Location Change		**Industry Change**	
	Yes	No	Yes	No
Productivity × $\mathbb{1}$(Post)	0.0010***	0.0006***	0.0015***	0.0002***
	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Employee F.E.	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y
Salary & Seniority Controls	Y	Y	Y	Y
Cohort × Language × Role × Time F.E.	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y
Observations	3,199,173	3,199,173	3,198,865	3,198,865
# Unique Employees	143,446	143,446	143,446	143,446
R-Squared	0.213	0.204	0.214	0.19
Y-Mean	0.018	0.015	0.029	0.004
Magnitude (%)	**6.52**	**4.8**	**5.83**	**5.17**
T-test ($H_0 : \beta_{Yes} = \beta_{No}$)	(0.0719)*		(0.6133)	

Table A.9: **GitHub Shock & Employee Mobility: Heterogeneity by Local Economic and Demographic Characteristics**

This table presents estimates on any heterogeneous impact of GitHub's disclosure of employee productivity information based on local economic and demographic characteristics. We report the differential impact of GitHub's policy change on productive employees, estimated by the β coefficients from the equation:

$$\mathbb{1}(MoveToLargeFirm)_{i,t} = \beta_1 Productivity_i \times \mathbb{1}(Post_t) + \beta_2 Productivity_i \times \mathbb{1}(Post_t) \times \mathbb{1}(MSAIndicator_i) + \lambda X_{it} + \alpha_i + \gamma_{it} + \varepsilon_{i,t}$$

The independent variable is a triple interaction of three indicators: the employee's normalized productivity estimate derived from the AKM decomposition, indicators for employee's initial location (MSA), and indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 2016). Each column includes a triple interaction with a dummy indicator that takes the value of 1 if the employee's initial MSA has an above-median value of the economic/demographic characteristic described in the column at the time of GitHub's policy change. X_{it} represents employee's time-varying salary and seniority controls. α_i is employee fixed effects. γ_{it} represents employee's cohort, role, and language interacted with time fixed effects as well as employee's location and industry interacted with time fixed effects. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)	(5)
Outcome:	**$\mathbb{1}$(Job Change to Large Firms)**				
	MSA Indicator - Above Median				
	Republican Vote Share	*Share of White Population*	*Median House Value*	*Change in House Value*	*Gini Index*
Productivity × $\mathbb{1}$(Post)	0.0017***	0.0017***	0.0014**	0.0018***	0.0018***
	(0.0002)	(0.0002)	(0.0007)	(0.0003)	(0.0004)
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(MSA Indicator)	-0.0012	-0.0003	0.0002	-0.0003	-0.0001
	(0.0008)	(0.0005)	(0.0007)	(0.0003)	(0.0004)
Employee F.E.	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Cohort × Language × Role × Time F.E.	Yes	Yes	Yes	Yes	Yes
Industry × Location × Time F.E.	Yes	Yes	Yes	Yes	Yes
Observations	3,132,014	3,143,808	3,143,808	3,143,808	3,143,808
# Unique Employees	140,433	140,966	140,966	140,966	140,966
R-Squared	0.209	0.21	0.21	0.21	0.21
Y-Mean	0.033	0.033	0.033	0.033	0.033

Table A.10: **GitHub Shock & Firm Productivity: Robustness**

This table presents robustness tests for estimates reported in columns 3-6 of Table 10 using different specifications for the dependent variable. In columns 1 and 3 (2 and 4), the dependent variable takes the value of 1 if the firm contribution variable in a quarter lies is in the top quartile (top quintile) of all public contributions in that quarter and 0 otherwise. Firm contributions only include public contributions made by incumbent employees towards repositories not owned by the employees themselves. We also estimate coefficients for double interaction of indicator variables $\mathbb{1}(Mid)$ and $\mathbb{1}(Small)$ with $\mathbb{1}(Post)$, but do not report them for brevity. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)
Outcome:	**Public Contributions**		**Public Contributions/Employees**	
	$\mathbb{1}$*(Top Quartile)*	$\mathbb{1}$*(Top Quintile)*	$\mathbb{1}$*(Top Quartile)*	$\mathbb{1}$*(Top Quintile)*
$\mathbb{1}$(Productivity > Median) × $\mathbb{1}$(Post)	0.0156***	0.0130**	0.0094	0.0076
	(0.0060)	(0.0057)	(0.0063)	(0.0058)
$\mathbb{1}$(Productivity > Median) × $\mathbb{1}$(Post) × $\mathbb{1}$(Mid)	0.0057	0.0054	0.0061	0.0021
	(0.0072)	(0.0069)	(0.0075)	(0.0069)
$\mathbb{1}$(Productivity > Median) × $\mathbb{1}$(Post) × $\mathbb{1}$(Small)	-0.0223***	-0.0214***	-0.0162**	-0.0142**
	(0.0071)	(0.0068)	(0.0075)	(0.0069)
Controls	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y
Industry × Location × Time F.E.	Y	Y	Y	Y
Observations	878,160	878,160	878,160	878,160
# Unique Firms	46,471	46,471	46,471	46,471
R-squared	0.587	0.587	0.52	0.513
Y-Mean	0.25	0.202	0.249	0.198
Magnitude (Main) (%)	6.23	6.47	3.78	3.82
Magnitude (Small) (%)	-2.69	-4.15	-2.72	-3.35

Table A.11: **GitHub Shock & Firm Market Reaction**

This table presents estimates on the impact of GitHub's disclosure of employee productivity information on firm value for Compustat firms. The dependent variable is the firm's abnormal return using the Fama-French 3-factor model. The event window starts 252 days (one calendar year) before GitHub's policy change and ends on the day of the change. The independent variable Event Date takes the value 1 on the day of the policy change and 0 otherwise. High-productive (Low-productive) firms have above-median (below-median) AKM-based average employee productivity. Small (Large) firms comprise the bottom (top) 25th percentile firms in terms of employee size.

	(1)	(2)	(3)	(4)	(5)
	All Firms	High Productive		Low Productive	
		Small	Large	Small	Large
Event Date	-0.0017***	-0.0065**	0.0016	-0.0033	0.0004
	(0.0006)	(0.0028)	(0.0021)	(0.0029)	(0.0022)
#Firms	1679	112	106	136	57

Table A.12: **GitHub Shock & Individual Productivity**

This table presents estimates on the impact of GitHub's disclosure of employee productivity information on employees' productivity measured by their public contributions. Public contributions only include contributions made towards repositories not owned by the employees themselves. In column 1, the dependent variable takes the value of 1 added to the log of employees' GitHub contributions. In columns 2 (3) (4), the dependent variable takes the value of 1 if an employee's GitHub contributions in a quarter fall in the top quartile (top quintile) (top decile) of all public contributions in that quarter and 0 otherwise. All columns include employee's time-varying controls (lagged stock of contributions as well as pre-period median public contributions interacted with $\mathbb{1}(Post)$). Other independent variables, controls, and fixed effects correspond to our main specification, i.e., column 5 of Table 3. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)
	Public Contributions			
Outcome:	*Log(1+Y)*	$\mathbb{1}$*(Top Quartile)*	$\mathbb{1}$*(Top Quintile)*	$\mathbb{1}$*(Top Decile)*
Productivity × $\mathbb{1}$(Post)	0.0541***	0.0172***	0.0163***	0.0109***
	(0.0020)	(0.0006)	(0.0006)	(0.0005)
Employee F.E.	Yes	Yes	Yes	Yes
Contribution Stock & Trend Controls	Yes	Yes	Yes	Yes
Salary & Seniority Controls	Yes	Yes	Yes	Yes
Cohort × Language × Role × Time F.E.	Yes	Yes	Yes	Yes
Industry × Location × Time F.E.	Yes	Yes	Yes	Yes
Observations	3,218,835	3,218,835	3,218,835	3,218,835
# Unique Employees	143,420	143,420	143,420	143,420
R-Squared	0.668	0.573	0.572	0.556
Y-Mean	0.669	0.287	0.246	0.137
Magnitude (%)	9.28	6.88	7.6	9.14

Table A.13: **GitHub Shock & New Firm Formation**

This table presents estimates on the impact of GitHub's disclosure of employee productivity on new firm formation. The dependent variable takes the value of 1 if an employee switches job to start a new firm (or becomes a founding member) in the given quarter and 0 otherwise. The independent variable is a triple interaction of three indicators: normalized productivity estimate of the employee derived from the AKM decomposition, indicators for the size of the origin firm, i.e., where the employee worked in the previous quarter, and an indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 2016). $\mathbb{1}(Mid)$ is an indicator variable that takes the value of 1 for firms with 50-1000 employees and $\mathbb{1}(Small)$ is an indicator for firms with less than 50 employees. We also estimate coefficients for indicator variables $\mathbb{1}(Mid)$ and $\mathbb{1}(Small)$ as well as their double interactions with *Productivity* and $\mathbb{1}(Post)$, but do not report them for brevity. Cohort represents an employee's graduation year corresponding to the last degree. Language is the employee's most preferred coding language. Role, industry, and location represent the employee's first job role, MSA, and industry in our sample period respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)
Outcome:	**Move to Entrepreneurship**	
Productivity × $\mathbb{1}$(Post)	-0.0001	-0.0004*
	(0.0001)	(0.0002)
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Mid)		0.0003
		(0.0003)
Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Small)		0.0006*
		(0.0003)
Employee F.E.	Yes	Yes
Controls	Yes	Yes
Cohort × Language × Role × Time F.E.	Yes	Yes
Industry × Location × Time F.E.	Yes	Yes
Observations	3,219,087	3,040,193
# Unique Employees	143,447	142,952
R-Squared	0.216	0.23
Y-Mean	0.004	0.004
Magnitude (Main) (%)	-2.71	-11.69
Magnitude (Mid) (%)		-3.23
Magnitude (Small) (%)		5.03

Table A.14: **Industry Labor Market Concentration**

This table presents correlations between the change in industry-level concentration around GitHub's policy change and the average productivity of GitHub employees within the industry. We estimate the following equation:

$$\Delta Concentration_h = \beta Productivity_h + \varepsilon_h \tag{A.2}$$

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). The coefficient β captures the correlation between the outcome and a measure of private contributions, $Productivity_h$. In columns 1-2, the industry and employment figures are sourced from LinkedIn (144-LinkedIn industries) while in columns 3-4, the data is based on Compustat firms (SIC 3-digit industries). In columns 1 and 3 (2 and 4), the independent variable is the normalized share of employees with positive private contributions (normalized average number of private contributions) at the time of GitHub's policy change within the industry. Standard errors are clustered at the LinkedIn industry level in columns 1-2 and SIC 3-digit industry level in columns 3-4. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

	(1)	(2)	(3)	(4)
Outcome:	Change in Industry Labor Market Concentration (ΔHHI)			
	All Firms (LinkedIn)		Compustat Firms (SIC 3-Digit)	
Share Emp with Private Contr > 0	0.064**		0.140***	
	(0.029)		(0.002)	
Average Private Contr		0.067**		0.137***
		(0.027)		(0.013)
Observations	143	143	221	221
R-squared	0.098	0.107	0.133	0.128

Internet Appendix B

B.1 Measuring GitHub Contributions

In this study, we rely on two primary data sources to measure GitHub contributions: the GitHub Archive and the GitHub API. Each source provides distinct advantages and limitations in tracking user productivity.

The GitHub Archive (`https://www.gharchive.org/`), a third-party repository, records and stores all public events on GitHub, making daily snapshots of public contributions available for download. This dataset offers comprehensive historical coverage of public contributions that cannot be altered retroactively, ensuring data integrity over time. However, it is limited to public contributions, as it does not include private activity on the platform.

In contrast, the GitHub API (`https://api.github.com/graphql`), which can be accessed via live queries, provides the most up-to-date data on both public and private contributions, making it an essential source for capturing the full range of user activity. The API is subject to rate limits, constraining the volume of data that can be retrieved at any given time. We use the GitHub API as our primary source of data as it the only source to provide private contributions, critical to the identification in our paper.

A concern is that the API reflects the current state of user contributions, which may have been amended after the original event, and may not fully align with historical records. To assess the extent of this potential discrepancy, we compare public contributions reported by the GitHub Archive (at the time they were made) to those retrieved from the GitHub API (as of today). Our findings, presented in Figure B.5 Panel A, indicate that the discrepancy between these two data sources is minimal, accounting for less than 2% of total contributions, with 95% confidence. Additionally, Panel B shows no significant difference in this discrepancy between users who switched jobs (movers) and those who did not (non-movers), suggesting that strategic manipulation of contribution records is unlikely. Overall, while the GitHub API allows us to capture both public and private contributions, the GitHub Archive provides a valuable benchmark for verifying the stability of public contribution data over time. The observed discrepancies between these sources are small, suggesting that measurement error is unlikely to materially affect our analysis.

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

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

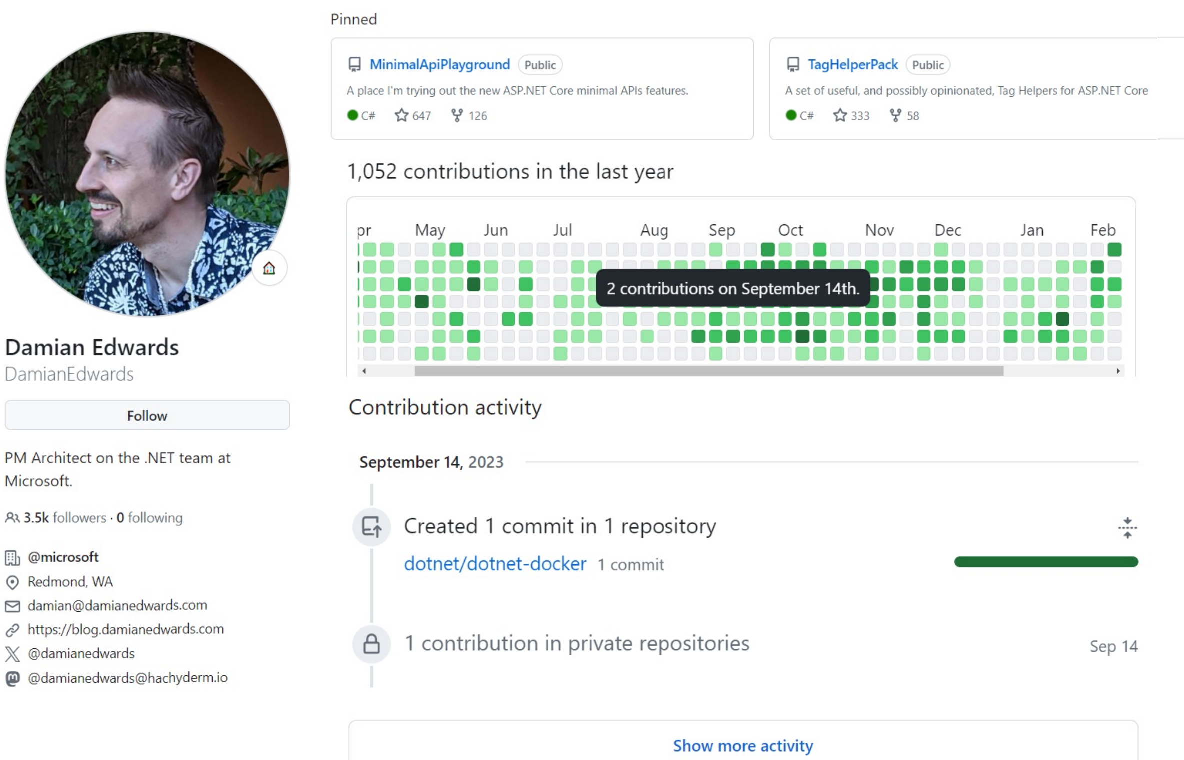
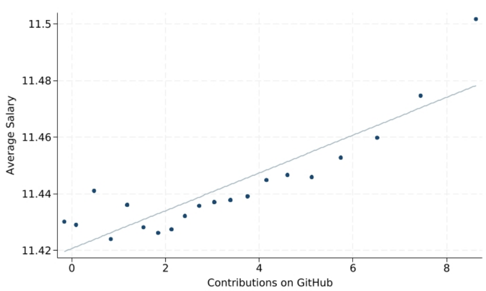
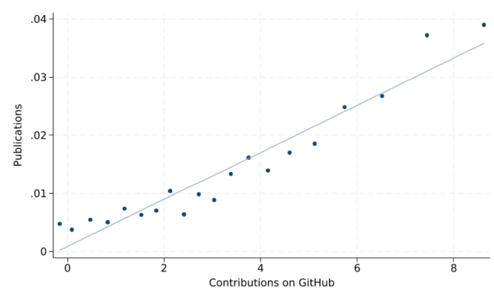


Figure B.2: **Association between Employees' GitHub Contributions, Salaries, and Publications**

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.



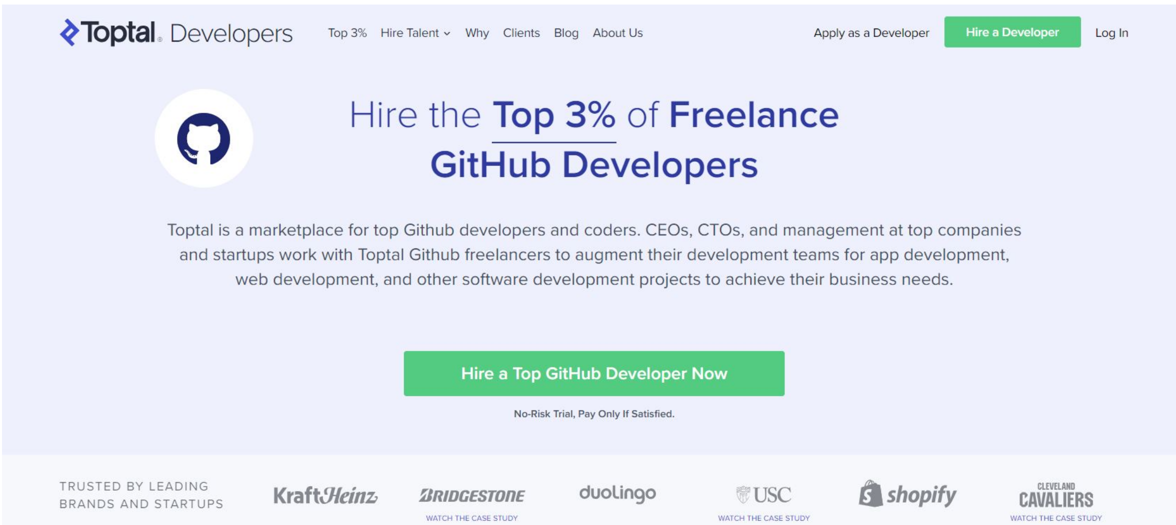
Panel A



Panel B

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

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



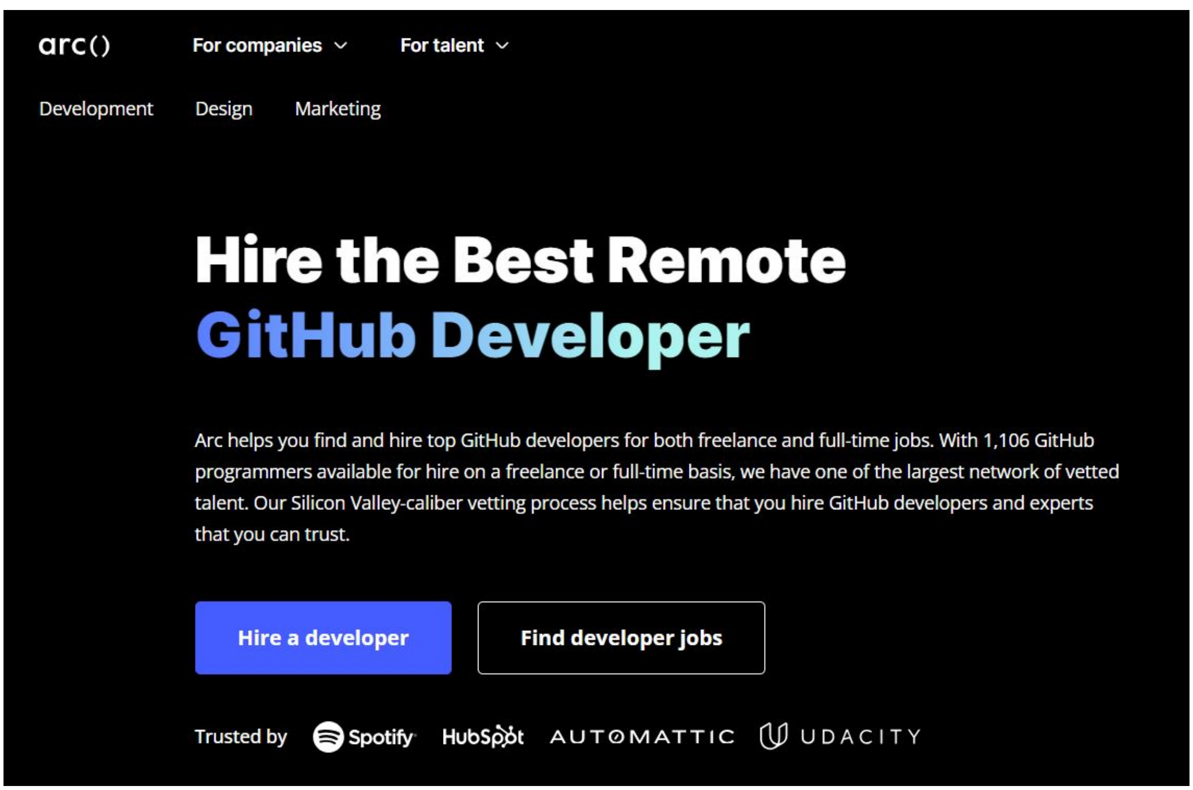


Figure B.4: **GitHub Calendar Before and After Policy Change**

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.

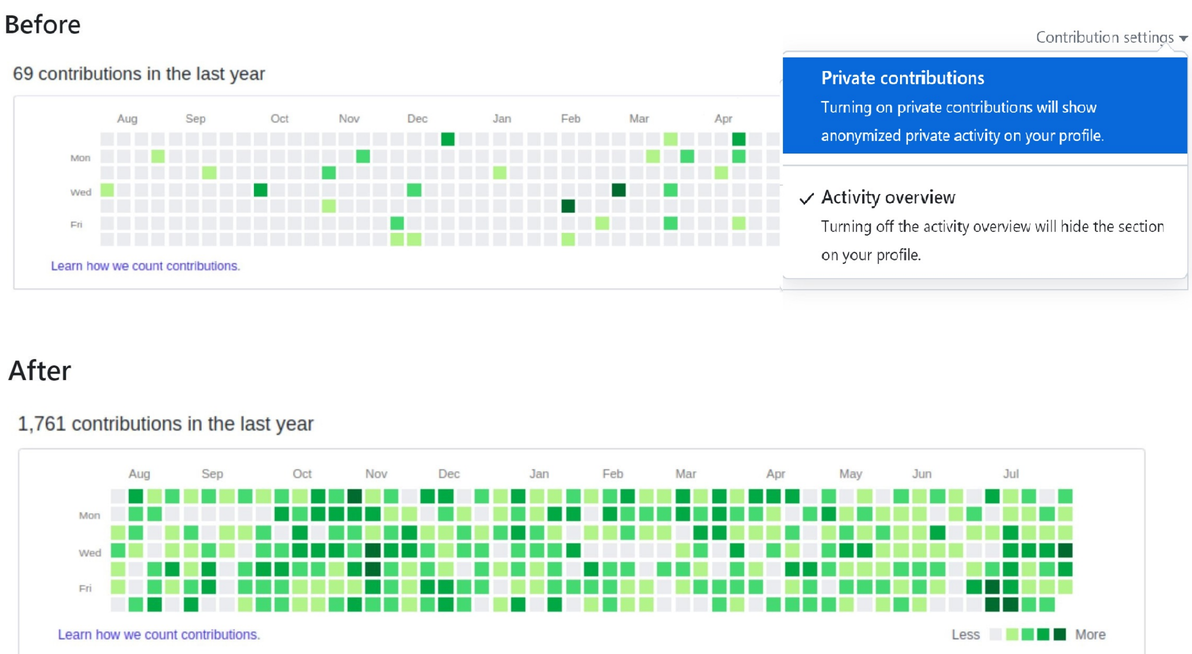
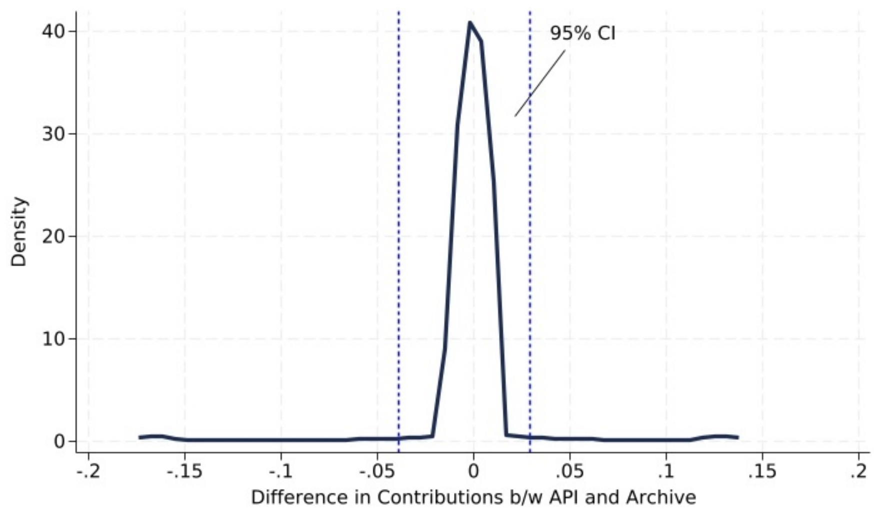
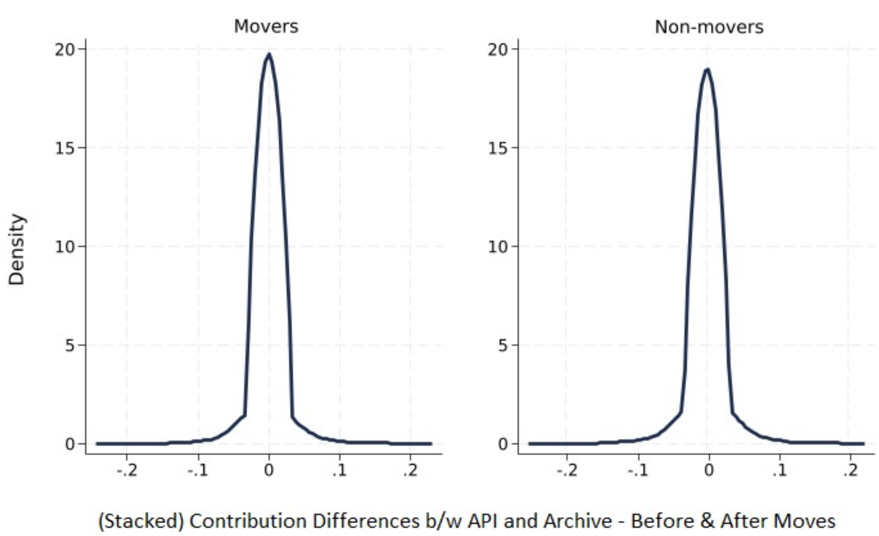


Figure B.5: **Discrepancy in Public GitHub Contributions**

This figure presents the distribution (density) of differences in GitHub contributions recorded on the day and hour they were originally made (timestamped and stored in GitHub Archive) versus those visible today (provided by GitHub API). GitHub contributions in the analysis are restricted to public pull request contributions which can be fetched and matched from both GitHub Archive and API. Panel A plots the kernel density of percentage point differences in the share of contributions made by an employee in a given quarter (share compared to their total contributions over our sample) computed using the Archive data and the API data. The dotted lines represent the 95% confidence interval. Panel B plots the density separately for movers and non-movers four quarters before and after move in a matched sample: movers matched to non-movers within the same origin firm.



Panel A



Panel B