# Big Data and Bigger Firms: A Labor Market Channel *

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

August 15, 2025

**Abstract**

This paper studies the impact of employee productivity information disclosure on labor reallocation across firms. GitHub, the world's largest software management platform, publicly tracks individual contributions. A 2016 policy change enhanced user GitHub contribution visibility, leading employees with 1 standard deviation higher contributions to experience a 5.2% increase in mobility to large firms, often at the expense of smaller firms. While top talent from small firms secured senior roles at larger firms, the latter retained them through internal promotions. This shift reduced employment growth and productivity in affected small firms, highlighting how labor-related big data reinforces large firms' dominance.

*We are grateful to seminar participants at the American Finance Association 2025, Midwest Finance Association 2025, Labor and Finance Group 2025, SFS Cavalcade 2025, Western Finance Association 2025, FSU Truist Beach Conference 2025, NBER Big Data, Artificial Intelligence, and Financial Economics (Fall 2024), NBER Corporate Finance (Fall 2024), NBER Megafirms and the Economy (Fall 2024), UNC Chapel Hill, University of Colorado at Boulder, 2024 Athens Winter Economics Meeting, Copenhagen Business School, BI Norwegian Business School, Charles III University of Madrid, Chinese University of Hong Kong, Duke University, Federal Reserve Bank of Atlanta, Oxford University, University of Cambridge (Economics), University of Illinois, University of Iowa, University of Southern California, and Warwick Business School. 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, CEPR, and ECGI (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, driven by a sharp increase in both the quantity and speed of data availability (Goldstein et al., 2021; Jiang and Li, 2024). This is especially true for labor markets, where firms hold exclusive information on their employee performance and potential (Waldman, 1984; Greenwald, 1986; Gibbons and Katz, 1991; 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, and real-time information on individual worker output, helping disclose employee productivity, mitigating informational disparities. This paper investigates how transparency in worker productivity influences labor market reallocation, examining whether it consolidates the dominance of large firms or fosters business dynamism by enhancing the ability of smaller firms to attract talent, and exploring the implications for relative firm performance.

The impact of productivity information on labor market dynamics is unclear a priori. Several studies argue that larger firms attract more productive employees due to their reputational advantages, which serve as proxies for individual ability in the absence of direct performance measures (Waldman, 1984, 1990; Bernhardt and Scoones, 1993). Greater access to productivity data could reduce these informational asymmetries, enabling smaller firms to compete for top talent. However, other models propose that in a frictionless labor market, the most productive workers should sort into the largest firms (Lucas, 1978; Kremer, 1993). In practice, adverse selection, where incomplete information hides the quality of workers at smaller companies, prevents such sorting (Greenwald, 1986; Gibbons and Katz, 1991). Therefore, greater transparency might amplify the advantages of large

firms, particularly as they leverage economies of scale to implement data-driven recruitment strategies (Brynjolfsson and McElheran, 2016). As a result, the net effect of productivity disclosure on labor reallocation and firm hiring behavior remains unclear.

To empirically study this question, we focus on GitHub, a software management and version control platform, used by over 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: 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 user contributions, including time stamps, turning it into a hub for recruiters seeking potential employees.[2] These contributions contain meaningful signals of employee productivity and quality, showing strong correlations with established productivity proxies at both the individual (career advancement and publications) and the firm level (organizational capital and Tobin's Q). 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 reveal 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. We validate that private contributions meaningfully correlate with other proxies for productivity used in the literature. We

[1] `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.

further show that private contributions significantly predict individual and firm-level outcomes—even after controlling for public contributions—indicating they contain additional information on productivity.

With this update, GitHub users could instantly 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. This setting represents a special case of the Spence (1973) model, where signaling is costless and perfectly reveals employee quality. Under these assumptions, a separating equilibrium arises in which high types always signal their quality, while low types are indifferent between not signaling or sending out a distinct "low-quality" signal.

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

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 thousand individuals), by 5.2% over the sample mean in our most robust

specification, after the GitHub policy change. This effect is primarily driven by more productive employees increasing their mobility to larger firms, rather than by a decline in movement among less productive employees. 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. The persistence of the change over the next twelve quarters suggests that information disclosure permanently shaped these employees' longer-term career trajectories rather than merely altering the timing of joining larger firms. Overall, these results are consistent with models suggesting that adverse frictions can prevent productive workers from sorting into large firms (Greenwald, 1986; Gibbons and Katz, 1991), and in the absence of such frictions, the most productive workers should sort into the largest and most productive firms (Lucas, 1978; Kremer, 1993).

A key concern is that employees with a higher number of private contributions are inherently different, and our results capture time trends for productive employees during this period. To address this, we validate that our findings are specific to the visibility of GitHub private contributions rather than general productivity trends. We employ placebo tests using alternative productivity measures, such as public GitHub contributions, the number of starred repositories, employee salaries, and publication records, all of which were consistently visible before and after the shock. The absence of significant effects across these placebo outcomes supports the interpretation that our results reflect the revelation of new information. Additionally, we document an immediate decline in stock prices for small firms with high exposure to GitHub private contributions on the day of the shock, further isolating the effect of the GitHub policy change from any concurrent trends.

Additionally, we address potential biases from self-selection, where only individuals seeking new employment might choose to reveal private contributions. Importantly, our results persist when restricted to users who reveal their private contributions and have at

least one or ten pre-period private contributions, reducing the likelihood of such biases. These findings demonstrate that the effects are evident at the intensive margin and are not solely driven by the extensive margin. Also, an analysis within only the set of job movers (namely holding selection into moving jobs constant) reveals that individuals with more private contributions are disproportionately more likely to transition to larger firms following the policy change.

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 fifty and one thousand employees) or small firms (those employing fifty employees or less) by 3.69 to 3.74% 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 -6% at the eleven to fifty size bucket to a +6.6% at firms with more than ten thousand 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 among 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 4.8% (3.2%) 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 ability to reveal productivity on GitHub 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 metrics of employee quality. To this end, we consider three traditional metrics: 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.

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 productivity information becomes available. To test this, we split firms based on whether one of their Human Resources (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.

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.

Next, we test if this heterogeneity in labor reallocation has firm-level implications. 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 17% 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 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, the productivity rank in small firms employing highly productive individuals prior to treatment decreased by 6% of the mean after the announcement.

The labor market reallocation towards large firms also has 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. in the recent years (Autor et al., 2023). Our results, hence, underscore a critical trade-off inherent in the growing proliferation of productivity-related big data in the economy. On one hand, greater transparency in labor markets facilitates the assortative allocation of talent to the most productive firms, boosting overall productivity and efficiency. On the other hand, this transparency risks accelerating market consolidation, diminishing competition, and constraining the growth of smaller firms.

Our paper contributes to the literature on the growing concentration of talent in large firms. Between 1982 and 2012, the share of US employment in large mature firms rose from 40% to 50%, while employment in small young firms declined sharply (Haltiwanger, 2015). This change coincided with a decrease in the responsiveness of firm employment to productivity shocks (Decker et al., 2020), and a decrease in the overall labor share of superstar firms (Autor et al., 2020). The pattern is particularly pronounced among high-skilled workers: for example, more than 80% of inventors are employed by large firms (Giroud et al., 2024).

While prior work has linked the rise of large firms to the use of big data - improving firm productivity (Bessen, 2020; Aghion et al., 2023; Brynjolfsson et al., 2023) and product quality (Eeckhout and Veldkamp, 2022; Tambe and Hitt, 2012) - its effects on labor flows and the concentration of talent within larger firms remains an open question, with conflicting predictions in the current literature. On one hand, large firms retain talent through wage premia (Akcigit and Goldschlag, 2023), over-hire skilled workers via R&D subsidies (Acemoglu et al., 2018), and acquire human capital through M&A activity (Tate and Yang, 2023). On the other hand, top talent frequently exits large firms to found startups (Gompers et al., 2005), and firms often struggle to retain inventors after going public (Bernstein, 2015). These contrasting dynamics raise a central question: when information frictions about worker productivity are resolved, does top talent gravitate toward large firms—or away from them? We provide new evidence by studying a setting in which individual-level

granular performance data become publicly available.

We also contribute to research on how performance metrics influence labor allocation. Previous work shows that reducing information asymmetry can influence labor market outcomes in specific institutional settings, such as increased turnover among low-rated public school teachers (Rockoff et al., 2012), changes in MBA job tenure (Floyd et al., 2024), and the movement of high-performing academics to top departments (Hager et al., 2024). However, it remains unclear whether improved performance metrics reallocate talent toward startups or toward larger firms, and if this reallocation generates broader firm-level or macroeconomic effects? Answering this question requires a setting in which individual-level productivity information was revealed for a large share of employees within a sector. We leverage a unique shock that made such data publicly available for a broad cross-section of U.S. tech workers.

# 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 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 facilitating within-team collaboration on project development.

[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. Our main analysis uses productivity data from May 2011 to April 2016, and our results are robust to restricting the sample to May 2011 to April 2014, ensuring that changes to GitHub's pricing policy are not driving our findings.

## 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. 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 the productivity of high-skilled professionals.

We validate this measure showing that GitHub contributions correlate with other commonly used measures of employee productivity and quality. Internet Appendix Figure B.2 presents binned scatter plots of the number of career advancements and publications against employees' percentile rank in terms of contributions.[5] Both plots show strong positive and linear correlations, indicating that total contributions effectively capture employee quality and productivity.

In addition to individual-level outcomes, GitHub contributions are also strongly associated with firm-level human capital metrics. Internet Appendix Figure B.3 shows that firm-level GitHub contributions correlate with organizational capital (as defined by Eisfeldt and Papanikolaou (2013)) and Tobin's Q of the firm. These patterns are consistent with findings from Conti et al. (2021), who highlight that these contributions correlate

[5]We use percentile ranks instead of log(1+x) throughout the paper to address the challenges posed by the prevalence of zeros and highly skewed distributions in the data, as discussed in Chen and Roth (2024).

with startup financing, underscoring their importance to sophisticated investors such as venture capitalists (VCs) as a measure of human capital. Other studies also leverage individual GitHub contributions to evaluate software developer productivity, reinforcing our interpretation (Holub and Thies, 2023; El-Komboz and Fackler, 2023; Conti et al., 2023).

While these correlations hold when using overall GitHub contributions, it is important to examine private contributions more closely, given our focus on changes in their visibility. In particular, we assess whether private contributions meaningfully capture employee productivity. There is reason to believe they do: private commits must be approved by a repository manager before being recorded, ensuring a minimum standard of quality. We also test this empirically.

We begin by comparing private and public contributions in Figure B.4, which reveals a strong positive linear correlation between the two. However, despite this correlation, private contributions contain substantial additional information beyond what is reflected in public activity. Specifically, public contributions explain only 4.5% of the variation in private contributions.

Next, we examine whether this additional information is related to productivity. Specifically, we test whether private contributions provide incremental information about human capital by analyzing their relationship with both individual and firm-level outcomes, controlling for public contributions. As shown in Internet Appendix Table B.1, private contributions significantly predict individual career advancement and publication outcomes, as well as firm-level measures such as organizational capital and Tobin's Q. These findings indicate that private contributions capture meaningful dimensions of both employee productivity and firm value.

## 2.3 GitHub and Tech-Sector Recruitment

Given that GitHub activity is predictive of career outcomes and firm-level performance, as we document above, it is likely that recruiters use these signals to identify and screen talent.

Indeed, a growing body of industry blogs and hiring guides, starting in the early 2010s, discuss how recruiters can evaluate developers based on their GitHub profiles. For instance, Walton (2014) outlines heuristics for identifying qualified developers via GitHub, while Cooke (2016) offers a step-by-step guide for recruiters to find talent on the platform. Notably, these sources highlight GitHub activity as a proxy for an individual's quality and effort. In addition to such blogs, numerous third-party web platforms further emerged in 2010s to explicitly help organizations in recruiting technical talent via GitHub. Figure B.6 provides a collage of third-party web platforms, discussion forums, and hiring resources contemporaneous to our period of study, illustrating GitHub's growing importance in recruitment discourse during our sample.

Complementing this qualitative evidence, in Figure B.5, we also document trends in the presence of HR professionals on GitHub using our matched GitHub–LinkedIn dataset. The fraction of firms with at least one HR professional on GitHub rose steadily through 2010s. This rise is especially pronounced among large firms, where the fraction of firms with HR professionals on GitHub is highest and increased roughly fourfold (from 2% to 8%) between 2009 and 2019.[6] These patterns suggest that GitHub had already become embedded in recruitment practices at many firms, particularly larger employers with more formalized HR processes.

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.

## 2.4 GitHub Policy Shift

GitHub implemented a policy change on May 19, 2016, that allowed users to present a more complete picture of their contributions. Prior to this date, users could only display

[6]These figures likely understate the true number of HR professionals using GitHub. Our estimates are based on a matched sample of users who have publicly linked their GitHub and LinkedIn profiles. While software developers often showcase their GitHub activity in professional settings, recruiters are potentially less likely to link to GitHub from their LinkedIn pages or personal websites.

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.[7] 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 disclose their within-firm productivity (verified by GitHub) to external employers for the first time. GitHub stated that the motivation for this change was to provide a more accurate representation of a user's overall contributions by including work performed in private repositories. As an illustration, Figure B.7 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 exploit users' pre-policy private contributions as an exogenous measure of their output. Since these contributions were made before users could anticipate the policy change, they are unlikely to reflect any strategic adjustments in behavior in response to the policy. The only other major change to GitHub's contribution metrics during this period occurred in November 2016, when GitHub began counting code reviews as contributions. However, this change only applied to activity going forward and did not retroactively affect pre-2016 contributions–the period we use to measure employee output.

Finally, it is important to emphasize that, prior to this policy change, users had no credible way to disclose their number of private contributions. By enabling users to publicly display this information, GitHub introduced a new third-party certification mechanism for private contribution productivity.

[7] `https://github.blog/news-insights/product-news/more-contributions-on-your-profile/` gives the URL of the official GitHub announcement.

## 2.5 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.[8]

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 May 19, 2016. And second, we limit the sample to users who had completed their highest educational degree before the policy change, namely we exclude users who were still studying at some point after May 2016. These filters 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. Our initial sample includes approximately

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

380,000 individuals with employment history from 2011 to 2016—the period used to measure productivity—out of 1 million employees whose profiles are matched from LinkedIn to GitHub. Of this sample, 93% belong to the connected set, the largest group of firms where employees transition 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 initial sample, as the AKM requires discarding firms where multiple employees are not observed within the same role and month.

## 2.6 Summary Statistics

Table 1 provides summary statistics for our sample. Panel A presents cross-sectional distributions, showing that 79% of employees are male, and 69% are white. Internet Appendix Table A.1 offers further breakdowns by industry, roles, coding language, and location. Panel A indicates that the Software, Internet, and Information Technology sectors constitute over 45% of the sample. Panel B classifies employees by occupation using two-digit O-NET categories, revealing that over 75% of employees hold roles identified as involving computer science and mathematics. Many of the remaining roles also involve substantial coding components-for instance, design roles (e.g. computer-aided designers), management positions (e.g. Chief Technology Officers), business and operations roles (e.g. analysts), and positions in architecture and engineering. We find that more than 80% of employees hold extensive coding roles.[9] Panel C extends this analysis using the top eight digit O-NET codes. Overall, 42% of job titles are software engineers, while other commonly observed roles are also closely related to computer science, including web developers, IT managers, data

[9]We input all O*NET role technology skill descriptions at the 4-digit occupation level into ChatGPT and asked it to classify whether the roles required extensive coding skills.

scientists, and computer systems engineers. Panel D uses LinkedIn role classifications, which provide the most detailed controls and are employed throughout the rest of the paper. The most frequent roles in this classification are also related to software engineering.

Internet Appendix Table A.1 Panel E and Figure 1 present 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 87,000 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.[10]

We next study 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 B 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.

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

## 2.7 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 work for, we adopt the methodology proposed by Abowd et al. (1999) (AKM):

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

The dependent variable represents the percentile rank of 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 ($\vartheta_i$), firm ($\vartheta_f$), role ($\vartheta_k$) and month ($\vartheta_t$) 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 ($\vartheta_i$) as our proxy for individual-level productivity information.

Private contributions, our main dependent variable in the AKM model, are a count variable with many zeros and a highly skewed distribution. Following (Chen and Roth, 2024), we use the percentile rank of private contributions as the dependent variable to avoid the limitations of a $log(1+x)$ transformation. We also conduct further robustness checks. First, we employ AKM specifications that incorporate alternate categorical variables, raw number of contributions, and the logarithm of one plus contributions as the dependent variable.

Second, we verify our results outside the AKM framework by defining employee productivity through private contributions, or the share of private to total contributions. Across all approaches, our results remain consistent, reinforcing the robustness of our findings. Section 3 provides a detailed discussion and the outcomes of these robustness checks in the context of employee-level results.

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 2.5. Panel B unveils that 65.7% of the total variation in contributions stems from differences in employee output, with an additional 6.6% attributed to disparities in average firm output. Thus, individual human capital explains almost 10 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.1 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. The results highlight that the GitHub policy change regarding employee private contributions can be interpreted as a new metric 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.2 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 percentile rank of employees' private contributions, measured over six 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 70.4% to 74.5%, representing a 5.8% improvement over the mean. This increase is considerably smaller than the 11–15% improvement found in Card et al. (2013), who ignore firm-worker matching, due to the small magnitude of their estimates.

# 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:

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

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 ($\vartheta_i$), as obtained from Equation 2 in Section 2.7. 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 $\beta$, 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 ($\alpha_i$) and ($\delta_t$) 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 ($X_{i,t}$). First, we account for employee's current salary and seniority. We also include fixed effects that interact together the employee's cohort of graduation, most used coding language, first job role, and quarter fixed effects to account for time-varying job trends. Additionally, we control for industry 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.[11]

[11]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. Studies using AKM-style estimates and similar sample sizes typically bootstrap with 50 to 100 iterations (e.g., Bennedsen et al. (2019); Bhaskarabhatla et al. (2021); Best et al. (2023)). Standard errors in our main results remain virtually unchanged (differing by less than 1%) when we use 1,000 iterations instead of 100.

Table 3 presents our findings. Individuals with productivity estimates higher by 1 standard deviation increased their likelihood of transitioning to larger firms (employing more than one thousand 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. These results are therefore consistent with models that suggest productive workers cannot sort into large firms due to adverse frictions (Greenwald, 1986; Gibbons and Katz, 1991) and, in the absence of such informational frictions, the most productive workers should tend to sort into the largest and most productive firms (Lucas, 1978; Kremer, 1993).

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

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

The model is similar to the one detailed in Equation 2, except we estimate a separate $\beta_\tau$ for each quarter $\tau$ 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 $\beta_\tau$ 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 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. We

observe that this change persists for up to three years after the announcement. If these employees were already likely to join larger firms, and the effect simply shifted the timing, we would expect the impact to reverse quickly. Instead, the sustained increase is more consistent with the notion that information disclosure has shaped employees' longer-term career trajectories.

We also test whether the moves to large firms are accompanied by changes in industry and location, as shown in Table A.3. We find that the probability of moving to a larger firm increases by 6% over the sample mean when accompanied by a location change, compared to a 4.3% 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 4-5% over the sample mean. These findings support the idea that enhanced information disclosure plays a more crucial role in cross-location moves, as it helps compensate for the absence of localized networks.

A final question is whether the observed impact is driven primarily by more productive employees reallocating to larger firms, or by a reduction in the movement of less productive employees to these firms. To examine this, we divide the sample into binary treatment and control groups based on employee productivity. Figure A.4 presents the event study using this binary treatment, showing a 1 percentage point increase in the likelihood that treated employees move to larger firms-approximately 20% of the sample mean. We then separately plot the estimates for the treatment and control groups in Figure A.5. The results reveal both an increase in mobility among more productive employees and a decline among less productive ones. However, the effect for more productive employees is roughly four times greater, suggesting that this channel is the main driver of the overall result.

### 3.1.1 Identification challenges

Two challenges could undermine the causal interpretation of our estimates. First, concurrent changes in factors correlated with employee productivity may produce similar results,

confounding our analysis. Second, individuals might endogenously choose to display their pre-shock productivity when anticipating to move, potentially biasing our estimates. Relatedly, it is valuable to assess whether our results reflect differences between individuals with and without contributions, or broader effects on the extensive margin. We address these concerns below.

**Omitted Variable Bias**

First, we investigate whether our findings might simply reflect time-varying trends among more productive employees by conducting falsification tests using measures of employee productivity that were visible prior to the shock. If such trends were the primary drivers of the results, they should affect all productive employees, regardless of whether productivity is assessed through private contributions or other observable traits. Conversely, if the observed effects are attributable to new information revealed by the shock, differential impacts should be evident only among employees with higher productivity as measured by private contributions. To test this hypothesis, we use four alternative definitions of observable productivity: public contributions on GitHub (which were visible before the shock), the number of stars received for work on public GitHub repositories (capturing repository quality), a proxy for employee salary, and the total number of publications authored by each employee.

The results of these falsification tests are presented in Table 4, Panel A. Column 1 presents our baseline results with private contributions for comparison. Column 2 examines public contributions by employees, excluding those made to their own repositories to avoid conflating professional productivity with personal activity. We decompose pre-shock public contributions using the AKM methodology, consistent with our baseline approach, to isolate the individual component of public contributions. Column 3 employs the percentile rank of the total number of stars received on employees' GitHub public repositories during the pre-shock period as a placebo productivity measure. Column 4 incorporates the logarithm

of employees' imputed salaries[12], while Column 5 uses the percentile rank of the total number of articles authored by employees on platforms such as Google Scholar, DevTo, and ORCiD during the pre-period. Across all these specifications, we find no significant economic or statistical changes in the mobility of employees with one standard deviation higher productivity based on these placebo measures after the shock. These null results strengthen our conclusion that the observed effects are driven by new information revealed through private contributions.

Additionally, we test if our results are influenced by trends correlated with the 2016 presidential elections. 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.4 shows that our results are similar for states irrespective of their political leaning, and that there was no differential impact on counties that were more white, richer, or more unequal, as a result of the shock.

As a final validation against concurrent trends, we analyze changes in firm stock prices immediately following the GitHub announcement.[13] Using CompuStat data from 1,675 firms, we categorize firms along two dimensions: average employee productivity (top or bottom quartile) and firm size (with indicators for the top and bottom quartiles).[14] Table A.5 shows a 71-basis-point decline in stock returns for firms that are in the bottom quartile by size but the top quartile by average employee productivity, suggesting that the market anticipates potential talent loss in these firms. In contrast, no significant impact is observed for firms in the top size quartile or for those in the bottom quartile by both size and productivity. To further validate our findings, we repeat the analysis using 100 randomly selected dates within the year preceding the policy change. As shown in Figure A.3, we observe similar or more

[12]We use salary as imputed by Revelio, based on the employee's role, level, firm, and location. This approach aligns closely with our empirical goal, as external employers are similarly limited to observable characteristics and typically do not know an employee's exact salary.

[13]Bloomberg does not report any major tech-related news announcements between 17 May 2016 and 20 May 2016.

[14]Firms in the bottom quartile have fewer than twelve hundred employees—closely aligning with the one thousand employee threshold used to distinguish medium-sized from large firms in other analyses in the paper—while firms in the top quartile have more than eighteen thousand employees.

negative coefficients in fewer than five percent of the cases, reinforcing the conclusion that the observed effects are driven by the GitHub announcement rather than random fluctuations.

**Selection**

Next, we address the potential bias arising from movers differentially displaying their contributions. For example, individuals may only show their private contributions when they are willing to move jobs. Alternatively, firms may introduce policies not allowing their workers to display their private contributions after the GitHub policy change. To address those possibilities, we re-estimate our analysis in samples where such selection concerns are shut down. To this end, we consider three alternative samples: only users who display their private contributions, users employed at firms with a track-record of private contributions, and a sample of only moves. The results of these analyses are presented in Table 4, Panel B.

We begin by testing our results on the sub-sample of users with non-zero private contributions. This within-sample analysis eliminates potential bias by individuals showing their contributions when intending to move. Columns 1 and 2 restrict the sample to users with at least one and ten private GitHub contributions during the pre-period, respectively. While these restrictions substantially reduce the sample size, our results remain robust in both magnitude and statistical significance. These results also suggest that our findings are not solely driven by extensive margin differences between employees with and without private contributions but also reflect better outcomes–on the intensive margin–for those with higher private contributions within the group of contributors.

A related concern is that some firms may not utilize private contributions or may restrict employees from disclosing private contributions on GitHub. To address this, Columns 3 and 4 focus on firms where private contributions are observable (i.e., firms with at least one or ten private contributions as of the shock date). The results are even stronger within this sub-sample, despite the reduced sample size.

Next, we restrict our sample to movers and study their differential likelihood of

transitioning to large firms based on private contributions after the shock. If only movers revealing their private contributions was driving our baseline results, we would expect all movers to uniformly display private contributions, resulting in a null effect in this analysis. However, if private contributions are responsible for the transitions toward larger firms, a significant effect should still emerge within this restricted sample. Column 5 confirms this hypothesis: following the shock, movers with higher private contributions were significantly more likely to transition to larger firms.

A final concern is that transitory individual shocks could simultaneously affect both individual productivity and mobility. To address this, we estimate employee productivity separately for the periods before and after the GitHub shock. As illustrated in Figure A.1, individual productivity remains highly persistent, with the majority of employees staying within the same productivity quintiles across both periods. Additionally, Table A.6 narrows the analysis to users with consistent productivity levels across the pre- and post-shock periods. Even with this restriction, our findings remain robust, confirming that the observed effects are not driven by transitory fluctuations in productivity.

**Other Robustness**

We also test the robustness of our results to different specification choices. Table A.7 presents these findings. First, as explained in Section 2.7, we might be worried about the functional form of our 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 categorical variables, the raw contribution numbers, logarithm of one added to the raw contributions, and the logarithm of private contributions with zero contributions explicitly coded as zero (as recommended by Chen and Roth (2024)). Further, Panel B1 shows that our results are robust to defining employee productivity simply as the percentile rank of employees in terms of their pre-period private contributions without carrying out AKM decomposition. 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 difference-in-differences setting, as detailed by Callaway et al. (2024). We show our results are robust to using binary indicators for productivity in Panel C.[15] Panel D examines the sensitivity of our findings to the time period over which we estimate the AKM. Panel E shows that the results are robust to using more granular fixed effects—constructed by interacting all controls used in the main specification—and to controlling for public productivity signals such as public contributions, repository stars, and total publications. Panel F confirms the findings using alternative sample definitions, including a matched sample and one that excludes Microsoft employees (due to Microsoft's acquisition of GitHub in 2018).[16]

## 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, except we now consider transitions to medium-sized firms (fifty to thousand employees) and small firms (less than fifty 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 3.69% and 3.74% 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

[15] 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.4, we show that our parallel trends assumption is robust to using a dummy indicator for treatment.

[16] 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 E1 involves interacting all fixed effects, including cohort, role, language, location, industry, and time. Panel E2 controls for public contributions, stars, and publications. 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. We exclude Microsoft in Panel F3

assumption.

We further replicate our analysis by non-parametrically estimating our baseline specification across different firm size categories, as illustrated in Figure 4.[17]. Our findings indicate that the mobility of employees with 1 standard deviation higher productivity increases almost monotonically with firm size, ranging from -6% of the sample mean for firms with eleven to fifty employees to +6.8% for firms with more than ten thousand employees. An exception to this trend is observed in firms with fewer than ten employees, where the increase in mobility is economically small (0.4% of the sample mean) and statistically insignificant. This null result may arise because many of these firms consist solely of founders, rather than regular employees.

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.9 shows that the increase in mobility holds when defining superstars as the top 1% and 10% by market capitalization and revenue, as defined in Tambe et al. (2020). Moreover, Internet Appendix Table A.10 reveals that even within young firms, mobility increases to only "superstar" startups with over $500 million in funding.[18] 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

[17]Table A.8 shows the regression counterparts of Figure 4.

[18]Startups with at least $500 Million have a median size of thousand to five thousand employees.

triple-difference specification:

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

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 fifty to thousand employees, and $\mathbb{1}(Small)$ represents an indicator for firms with fewer than fifty 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. $\beta_1$ captures the impact of the policy change on more productive employees working at large firms, while $\beta_2$ and $\beta_3$ assess the additional impact on employees currently employed at mid-sized and small firms, respectively. This specification, thus, enables us to examine how the joint interaction of employee productivity, the GitHub policy change, and firm size influences the outcome of interest.

We also estimate the dynamic version of Equation 4, to verify the absence of pre-trends, as follows:

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

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 7% 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 3.2% and 4.9% of the sample mean post-shock, respectively. The absence of pre-trends, when plotting the triple-difference coefficient for both mid-sized and 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 5.8% 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 6.9% of the sample mean for those individuals with 1 standard deviation higher productivity after the GitHub shock. Conversely, the probability of transitioning from small and medium firms to large firms with promotion increased by 5.1% 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 4.5% 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 11.2% of the sample mean for those with 1 standard deviation higher productivity after the GitHub shock. Conversely, there was a 2.2% 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 5.8% over the mean in the probability of within-firm promotions for employees with 1 standard deviation higher productivity post-shock. We 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 7.5% 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 because doing so could signal their quality to other firms, increasing the risk of poaching. However, following the GitHub policy change, information asymmetry regarding employee quality is reduced, limiting firms' ability to obscure talent by withholding promotions. In fact, since talented workers now receive more outside offers due to GitHub disclosures, firms promote them as a retention strategy.

## 3.5 Impact on productivity

We also explore whether these career upgrades are associated with a change in employee productivity. We test this hypothesis in Table A.11, 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 2% increase in percentile rank 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. 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, 1986), suggesting that in the absence of direct productivity indicators, firms may rely on hard metrics, 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 metrics: 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.

Column 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 2.7 times the change observed for employees with an elite university degree. These results are consistent with two complementary mechanisms: first, that the availability of productivity information reduces reliance on traditional signals such as elite college credentials; and second, that it enables firms to identify and hire talented employees outside

of their existing recruitment networks.

Overall, our results support the hypothesis that web-based platforms have reduced reliance on hard metrics, such as experience and education, and network-based screening, 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.

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

First, we examine the role of tech-savvy recruiters. Using our matched GitHub–LinkedIn dataset, we identify HR professionals with a presence on GitHub. Figure B.5 shows that by 2016, over 6% of large firms had at least one recruiter active on the platform. We divide the sample of large firms based on whether their number of GitHub-active recruiters in 2016 is above or below the median. Columns 1–2 of Table 9A present the results. 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 7.7% of the sample mean after the GitHub policy change. In contrast, the increase was only 2.1% 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.

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 3-4, present the results. We find that individuals with productivity 1 standard deviation above the mean were 6.5% 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 5-6, present the results. Individuals with 1 standard deviation higher productivity increased their mobility to larger firms with above-median forced turnover by 6.1% above the sample mean following the GitHub policy change. In contrast, there was a statistically insignificant decrease of 0.6% 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% above the sample mean following the GitHub policy change. In contrast, there was an increase of 4.8% 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 5% 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% 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:

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

In our regression model, the dependent variable is an outcome variable for firm $j$ at quarter $t$. The coefficient $\beta_1$ 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 $\beta_2$ and $\beta_3$ 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, depreciated annually at 15% in line with standard practices in the productivity/innovation literature (e.g., for patent stock measures). We cluster our standard errors at the firm level.

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

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

The model is similar to the one detailed in Equation 5, except we estimate a separate $\beta_\tau$ for each quarter $\tau$ 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 $\beta_\tau$ 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 small firms with more than median individual productivity grew 17% slower than the mean after the GitHub announcement.[19]

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 that small firms with above-median productivity witnessed a substantial 56% increase over the mean in the probability of closures, after the GitHub announcement.[20] Figure 7 demonstrates that our results are not driven by differential pre-trends across firms.

[19] 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.

[20] The magnitude appears large because the pre-period mean of closures is very low (0.47%), particularly due to large firms that are much less likely to exit. The pre-period mean of exits among small firms is 0.82%, suggesting that the shock leads to about 32% increase in their probability to exit.

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 indicate a decrease in public contributions, with a 6% reduction in the percentile rank and a 20% decrease in the probability of achieving top-decile contributions, compared to the mean, among small firms with productivity estimates above the median following the GitHub announcement. Similarly, public contributions per employee saw a 6% decrease in percentile rank and a 23% reduction in the probability of reaching the top decrease, compared to the mean, for small firms with above-median productivity levels after the GitHub shock. 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.

The absence of strong positive effects after the GitHub shock for large firms with higher GitHub contributions aligns with the idea that all large firms benefited from the influx of talented labor, regardless of their own GitHub usage. Instead, small firms reliant on GitHub lost critical talent and struggled to compensate for these losses, resulting in an overall decline in both growth and productivity.

A final question concerns the magnitudes of these effects. At the employee level, our results show that the shock increased the reallocation of productive employees from small to large firms by approximately 5% of the mean mobility of the employees. However, how this translates at the firm level is less clear. To explore this, we repeat the analysis at the firm level in Internet Appendix Table A.12. Following the shock, small firms with more than the median number of productive employees saw the share of productive employees fall by about 1 percentage point per quarter. This represents roughly one-third of their average share before the shock. This decline was driven by the movement of employees in the top productivity quintile. Thus, the results in this section underscore the critical role of the most productive employees within firms. A reduction of about one-third in their share corresponds

to a decline in firm growth by 17% of the mean and productivity by approximately 6% of the mean.

## 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 of most impacted firms by size. Next, we examine the relationship between talent reallocation and industry dynamics, specifically whether employment concentration changed in the industries most affected by the Github disclosure policy. To explore this, we aggregate our data to the industry level and examine the correlation between changes in concentration and the number of private contributions in each industry.[21] 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.13 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

[21] There has been considerable debate on the aggregate trends with regard 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.

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.

# 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: the ability to better disclose productivity 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 metrics 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.

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

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

Bennedsen, M., M. Tsoutsoura, and D. Wolfenzon (2019): "Drivers of effort: Evidence from employee absenteeism," *Journal of Financial Economics*, 133, 658–684.

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

Bernstein, S. (2015): "Does Going Public Affect Innovation?" *The Journal of Finance*, 70, 1365–1403.

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

Best, M. C., J. Hjort, and D. Szakonyi (2023): "Individuals and organizations as

sources of state effectiveness," *American Economic Review*, 113, 2121–2167.

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

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.

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)*.

Cooke, K. (2016): "Recruiting and Sourcing Developers on GitHub: The Complete Guide," `https://beamery.com/resources/blogs/recruiting-and-sourcing-developers-on-github-the-complete-guide`.

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2020): "Changing

Business Dynamism and Productivity: Shocks versus Responsiveness," *American Economic Review*, 110, 3952–90.

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*.

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*.

Eisfeldt, A. L. and D. Papanikolaou (2013): "Organization capital and the cross-section of expected returns," *The Journal of Finance*, 68, 1365–1406.

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

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.

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

Giroud, X., E. Liu, and H. Mueller (2024): "Innovation Spillovers across U.S. Tech Clusters," Working Paper 32677, National Bureau of Economic Research.

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

Gompers, P., J. Lerner, and D. Scharfstein (2005): "Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999," *The Journal of Finance*, 60, 577–614.

Greenwald, B. C. (1986): "Adverse Selection in the Labour Market," *The Review of*

*Economic Studies*, 53, 325–347.

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

Haltiwanger, J. (2015): "Top Ten Signs of Declining Business Dynamism and Entrepreneurship in the U.S." *Working Paper*.

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)*.

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.

Kremer, M. (1993): "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, 108, 551–575.

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

Lucas, R. E. (1978): "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9, 508–523.

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

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.

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.

Tate, G. and L. Yang (2023): "The Human Factor in Acquisitions: Cross-industry Labor Mobility and Corporate Diversification," *The Review of Financial Studies*, 37, 45–88.

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

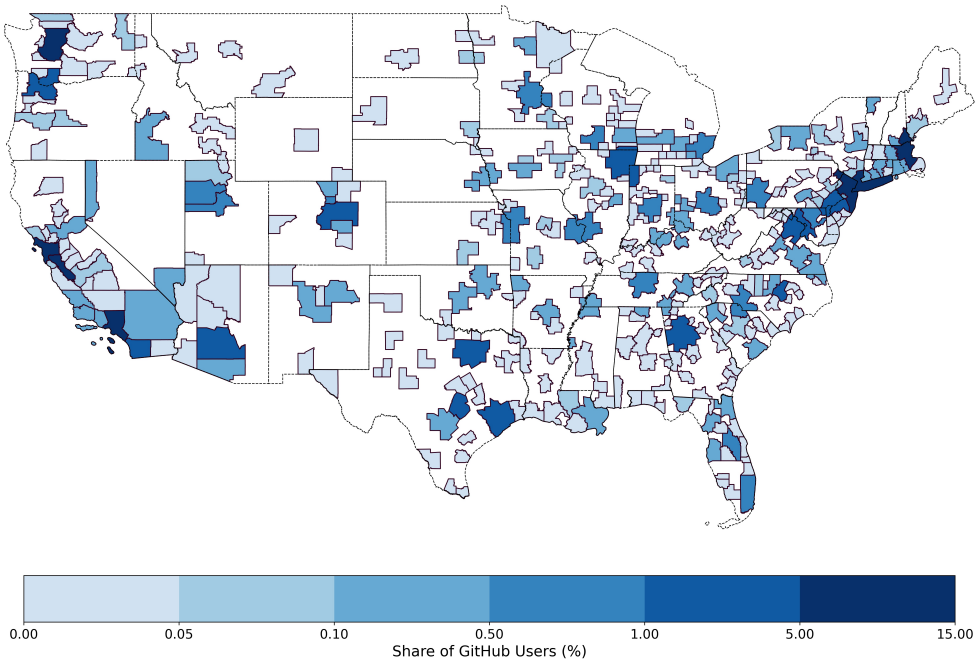
——— (1990): "Up-or-out contracts: A signaling perspective," *Journal of Labor Economics*, 8, 230–250.

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.

Walton, P. (2014): "How to Find Qualified Developers," `https://philipwalton.com/articles/how-to-find-qualified-developers/`.

# 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.

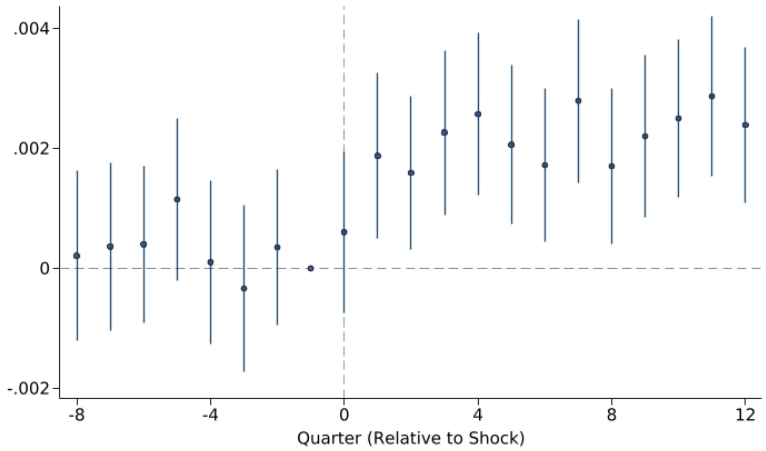


## Figure 2: **GitHub Shock & Employee Mobility to Large Firms**

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 $\beta$ coefficients obtained from the equation:

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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of industry, location, and time. 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**

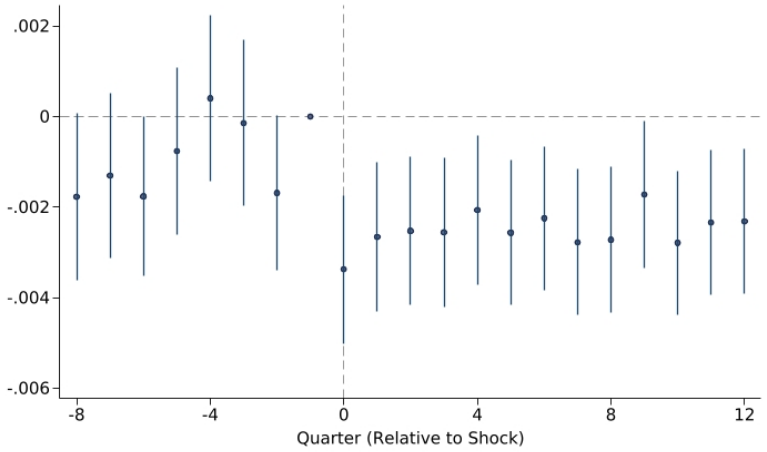
This figure presents event studies on employee mobility to small-sized and medium-sized firms around the disclosure of employee productivity information on GitHub. The x-axis plots the quarter around GitHub's policy change. The dashed line indicates the quarter 2 (April-June) of 2016, during which GitHub 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 $\beta$ coefficients obtained from regression estimate:

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

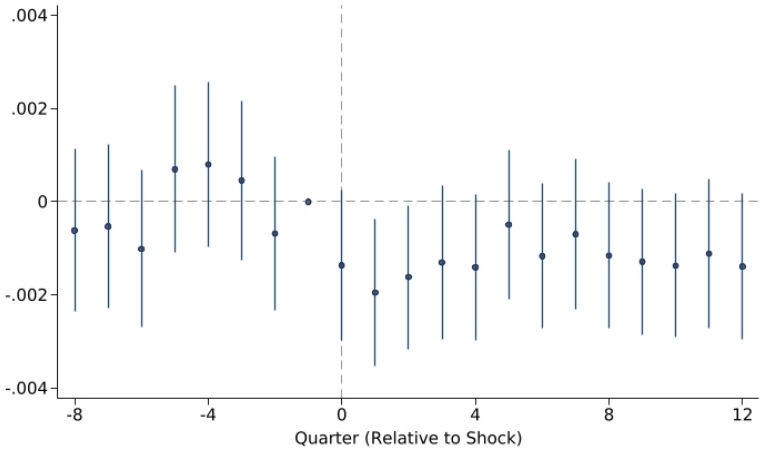
The independent variable is the normalized productivity estimate of the employee $i$ derived from the AKM decomposition interacted with the time indicator for each quarter. The dependent variable is an indicator that equals 1 when employee $i$ switches job to a mid-sized firm (50 < employees <= 1000) for Panel (a), to a small firms (<= 50 employee) for Panel (b), in quarter $t$ and is 0 otherwise. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of industry, location, and time. 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, along with its 95% confidence interval, of the differential impact of the GitHub policy shock on more productive employees using the $\beta$ coefficients obtained from regression estimate:

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

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. We divide the coefficient estimate $\beta$ by the pre-period mean of the outcome variable and multiply the estimates by 100 to get the economic magnitude in percentage terms. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of industry, location, and time. 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.



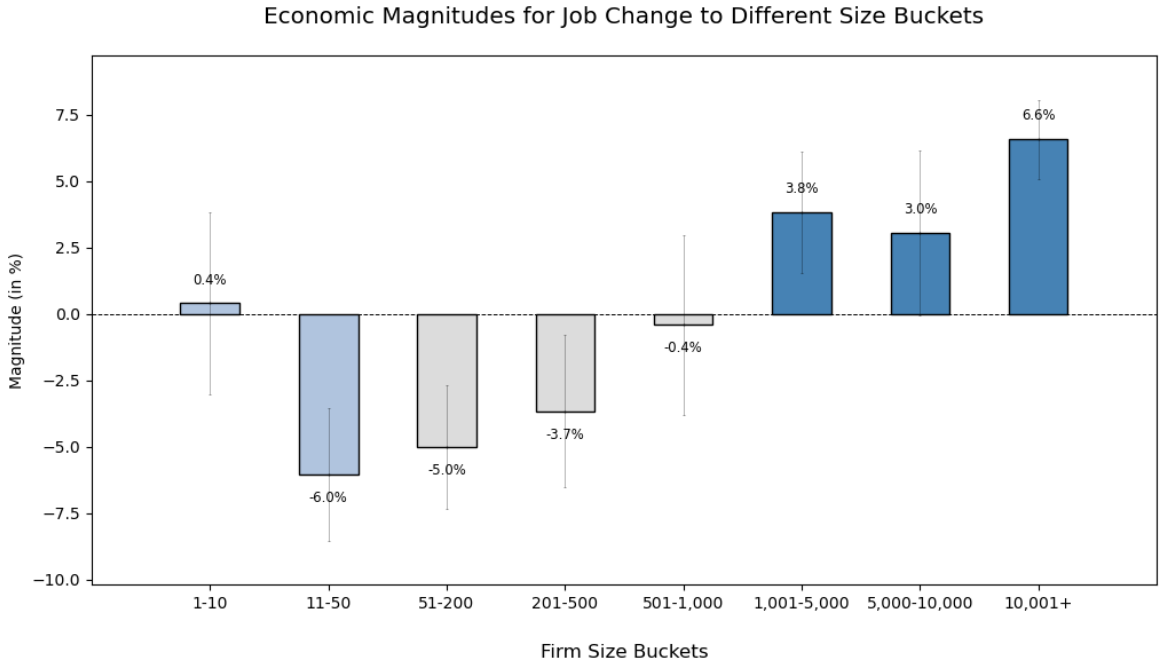


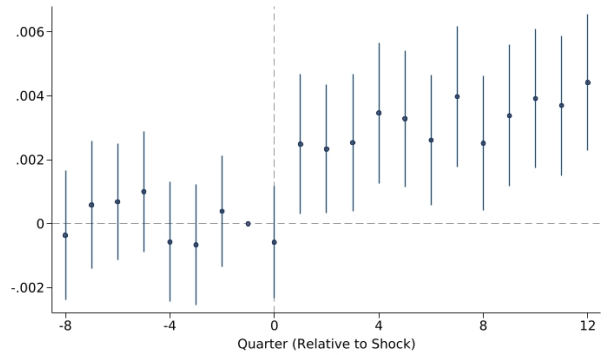
Figure 5: **GitHub Shock & Employee Mobility to Large Firms: Heterogeneity by Initial Firm Size**

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 $\beta$ coefficients obtained from the equation:

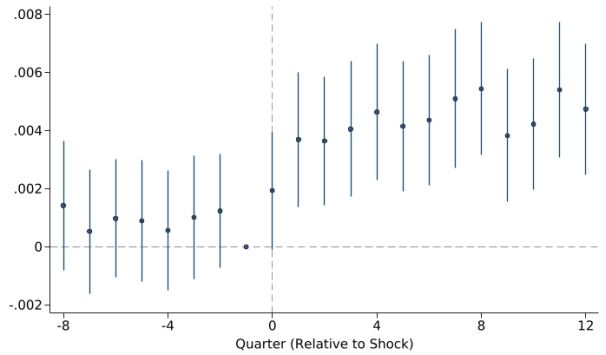
$$\mathbb{1}(MoveToLargeFirm)_{i,t} = \sum_{\tau=-8}^{12} \delta_\tau Productivity_i \times \mathbb{1}(\tau = t) + \sum_{\tau=-8}^{12} \beta_\tau Productivity_i \times \mathbb{1}(\tau = t) \times \mathbb{1}(Mid - Sized_{i(t-1)})$$
$$+ \sum_{\tau=-8}^{12} \beta_\tau^{'} Productivity_i \times \mathbb{1}(\tau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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of industry, location, and time. 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

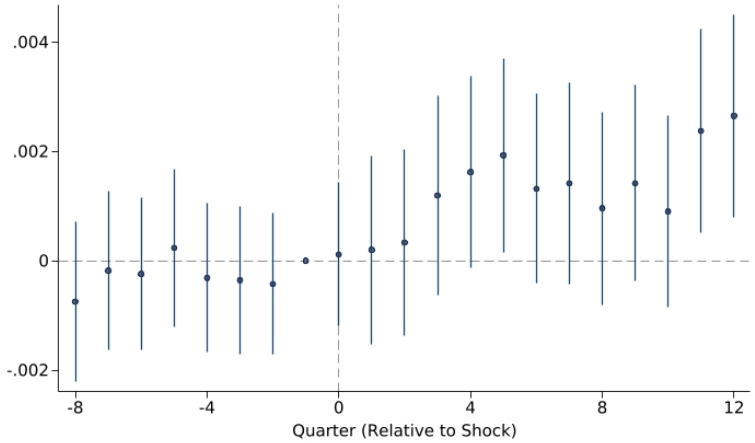


## Figure 6: **GitHub Shock & Employee Promotions at Large Firms**

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 $\beta$ coefficients obtained from the equation:

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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. $\theta_f$ is firm cohort fixed effects representing an interaction of the employee's firm as of the previous quarter 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.



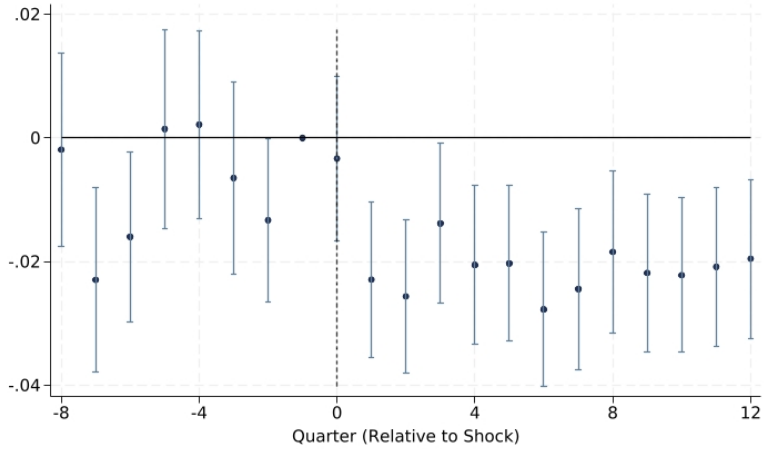
# Figure 7: **GitHub Shock & Employment Growth at Small Firms**

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 $\beta$ coefficients obtained from the following equation:

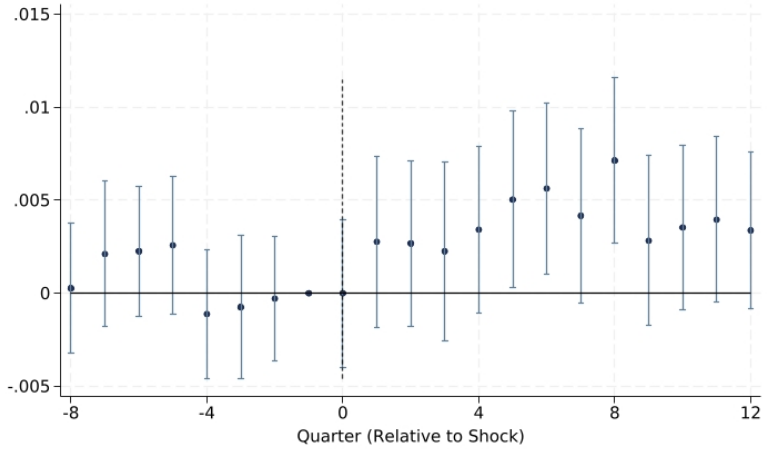
$$\begin{aligned} EmploymentGrowth_{j,t} = & \sum_{\tau=-8}^{12} \theta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \\ & + \sum_{\tau=-8}^{12} \delta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(MidSized_j) \\ & + \sum_{\tau=-8}^{12}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

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. $\alpha_j$ is firm fixed effects. $\delta_t$ denotes time fixed effects. $X_{j,t}$ represents firm's time-varying lagged employment as well as the firm's industry × 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) 𝟙(Employment Reduction > 50%)



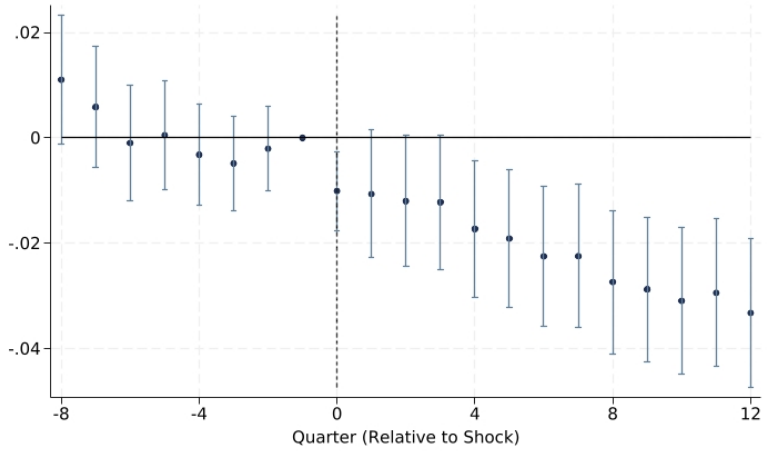
# Figure 8: **GitHub Shock & Productivity at Small Firms**

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 $\beta$ coefficients obtained from the following equation:

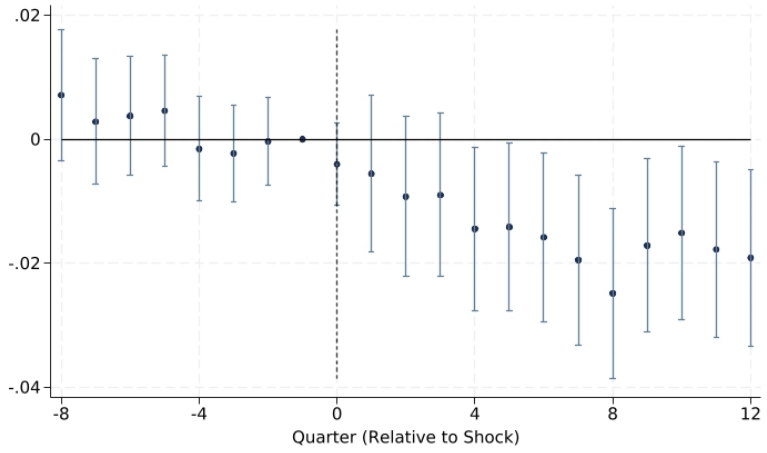
$$\mathbb{1}(Contributions_{j,t}) = \sum_{\tau=-8}^{12} \theta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) + \sum_{\tau=-8}^{12} \delta_\tau \mathbb{1}(Productivity_j > Median) \times \mathbb{1}(\tau = t) \times \mathbb{1}(MidSized_j)$$
$$+ \sum_{\tau=-8}^{12} \beta_\tau \mathbb{1}(Productivity_j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

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. $\alpha_j$ is firm fixed effects. $\delta_t$ denotes time fixed effects. $X_{j,t}$ represents firm's time-varying lagged stock of total contributions (depreciated at 15% annually) as well as the firm's industry × 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.** |
| 𝟙(Male) | 0.79 | 0.41 | 0 | 1 | 1 | 302,903 |
| 𝟙(White) | 0.69 | 0.46 | 0 | 1 | 1 | 302,903 |
| 𝟙(Extensive Coding Role) | 0.81 | 0.39 | 0 | 1 | 1 | 302,872 |
| 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) | 3.27 | 2.43 | 0 | 2.89 | 6.98 | 302,903 |
| Pre-Period Public Contributions (Log) | 2.27 | 2.34 | 0 | 1.61 | 5.70 | 302,903 |
| Pre-Period Private Contributions (Log) | 0.96 | 2.40 | 0 | 0 | 5.78 | 302,903 |
| 𝟙(Job Change to Large Firms) | 0.03 | 0.04 | 0 | 0 | 0.08 | 302,902 |
| 𝟙(Job Change to Mid-Size Firms) | 0.02 | 0.04 | 0 | 0 | 0.08 | 302,902 |
| 𝟙(Job Change to Small Firms) | 0.02 | 0.03 | 0 | 0 | 0.06 | 302,902 |
| 𝟙(Job Change) | 0.07 | 0.07 | 0 | 0.05 | 0.17 | 302,903 |
| 𝟙(Promotion) | 0.06 | 0.05 | 0 | 0.04 | 0.13 | 302,903 |
| 𝟙(Across-Firm Promotion) | 0.04 | 0.05 | 0 | 0.04 | 0.10 | 302,903 |
| 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.** |
| 𝟙(Large Firm) (Employees > 1000) | 0.15 | 0.36 | 0 | 0 | 1 | 57,982 |
| 𝟙(Mid-size Firm) (50 < Employees <= 1000) | 0.37 | 0.48 | 0 | 0 | 1 | 57,982 |
| 𝟙(Small Firm) (Employees <= 50) | 0.48 | 0.50 | 0 | 0 | 1 | 57,982 |
| QoQ Employment Growth | 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:

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

$\vartheta_i$, the employee fixed effect, represents the estimated employee *i's* productivity. The dependent variable is the percentile rank of the employee $i$ based on the number of private contributions made by them 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 ($\vartheta_f$), time fixed effects ($\vartheta_t$), and employee's role fixed effects ($\vartheta_k$). 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. | 65.7% |
| Firm F.E. | 6.6% |
| Residual | 27.8% |

# 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 $\beta$ coefficient from the equation:

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

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) | 0.0014*** | 0.0013*** | 0.0019*** | 0.0018*** | 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 | 0.169 | 0.212 |
| Y-Mean | 0.031 | 0.031 | 0.031 | 0.032 | 0.033 |
| Magnitude (%) | 4.35 | 4.11 | 5.98 | 5.6 | 5.23 |

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

This table presents placebo and alternate sample tests for our main result. All panels run tests corresponding to our baseline specification, Column 5 in Table 3, included here as Column 1 in Panel A for comparison. Panel A presents a placebo test with alternative productive measures. In Column 2, the independent variable is an alternative productivity estimate computed by using employee's percentile rank based on pre-period *public* contributions as the dependent variable in the AKM decomposition. Public contributions only include contributions made towards repositories not owned by the employees themselves. In column 3, employee's percentile rank based on the total number of stars received on their GitHub public repositories during the pre-period constitutes the placebo productivity measure. In column 4, the dependent variable is the logarithm of employee's pre-period average salary. Column 5 includes the employee's percentile rank based on total number of articles authored by the employee on Google Scholar, DevTo, and ORCiD in the pre-period as the independent variable. Panel B estimates results on a subsample of active GitHub users and firms. Column 1 only includes users with non-zero private GitHub contributions during the pre-period. Column 2 restricts the sample to users with at least 10 private GitHub contributions during the pre-period. Column 3 (4) excludes users who, during the pre-shock period, work at firms that have zero (less than 10) private contributions in our sample. Column 5 includes robustness on the baseline specification when data are organized at a job-move level (every observation being a job move). In this column, the time-varying fixed effects include an indicator of $\mathbb{1}(Post)$ as the time variable, instead of quarters as in the baseline. In Columns 1-2 (1-5) of Panel A (Panel B), standard errors are bootstrapped within employee clusters using 100 repetitions. In Columns 3-5 of Panel A, standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).



**Panel A: Placebo Measures of Productivity**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| **Outcome:** | **𝟙(Job Change to Large Firms)** | | | | |
| *Measure* | *Baseline: Pvt Contr* | *Public Contr* | *Pct(Starred Repos)* | *Log(Salary)* | *Pct(Publications)* |
| Placebo Productivity × 𝟙(Post) | 0.0017*** | 0.0000 | 0.0000 | -0.0004 | 0.0000 |
| | (0.0002) | (0.0003) | (0.0000) | (0.0009) | (0.0000) |
| Employee F.E. | Y | Y | Y | Y | Y |
| 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 | 3,199,173 | 3,180,560 | 3,198,931 | 3,198,931 | 3,198,931 |
| # Unique Employees | 143,446 | 142,616 | 143,419 | 143,419 | 143,419 |
| R-Squared | 0.212 | 0.212 | 0.211 | 0.211 | 0.211 |
| Y-Mean | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 |
| Magnitude (%) | 5.72 | 0.05 | -0.62 | -0.29 | 0.66 |

| Panel B: Sub-samples based on User/Firm Activity & Job Moves | | | | | |
|---|---|---|---|---|---|
| **Outcome:** | **𝟙(Job Change to Large Firms)** | | | | |
| *Sample Selection* | *Individual Pvt Contr > 0* | *Individual Pvt Contr >= 10* | *Firm Pvt Contr > 0* | *Firm Pvt Contr >= 10* | *Moves* |
| Productivity × 𝟙(Post) | 0.0012*** | 0.0013*** | 0.0033*** | 0.0035*** | 0.0105*** |
| | (0.0004) | (0.0005) | (0.0002) | (0.0002) | (0.0027) |
| Employee F.E. | Y | Y | Y | Y | Y |
| Controls | Y | Y | Y | Y | Y |
| Cohort × Language × Role × Quarter F.E. | Y | Y | Y | Y | |
| Industry × Location × Quarter F.E. | Y | Y | Y | Y | |
| Observations | 506,974 | 401,146 | 2,239,975 | 2,101,636 | 360,466 |
| # Unique Employees | 22,818 | 18,001 | 99,469 | 93,299 | 130,411 |
| R-Squared | 0.271 | 0.274 | 0.218 | 0.219 | 0.491 |
| Y-Mean | 0.026 | 0.024 | 0.037 | 0.038 | 0.372 |
| Magnitude (%) | 4.77 | 5.42 | 8.87 | 9.21 | 2.81 |

# 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 $\beta$ coefficient from the equation:

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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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:** | | **𝟙(Job Change)** | |
| | **Large Firm** | **Medium Firm** | **Small Firm** |
| Productivity × 𝟙(Post) | 0.0017*** | -0.0011*** | -0.0009*** |
| | (0.0002) | (0.0002) | (0.0002) |
| Employee 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 (%) | 5.23 | -3.74 | -3.69 |

# 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 $\beta$ coefficient from the equation:

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

The independent variable is a triple interaction of three indicators: normalized productivity estimate of employee $i$ derived from the AKM decomposition, indicators for the size of the 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:** | **𝟙(Job Change to Large Firms)** | | | | |
| Productivity × 𝟙(Post) | -0.0038*** | -0.0040*** | -0.0028*** | -0.0023*** | -0.0023*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0004) | (0.0005) |
| Productivity × 𝟙(Post) × 𝟙(Mid) | 0.0050*** | 0.0052*** | 0.0045*** | 0.0034*** | 0.0033*** |
| | (0.0004) | (0.0004) | (0.0004) | (0.0005) | (0.0006) |
| Productivity × 𝟙(Post) × 𝟙(Small) | 0.0054*** | 0.0056*** | 0.0048*** | 0.0040*** | 0.0038*** |
| | (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 (Main) (%) | -12.77 | -13.55 | -9.53 | -7.6 | -7.05 |
| Magnitude (Mid) (%) | 3.96 | 3.8 | 5.7 | 3.51 | 3.21 |
| Magnitude (Small) (%) | 5.3 | 5.16 | 6.58 | 5.56 | 4.86 |

# 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 $\beta$ coefficient from the equation:

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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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 $\theta_f$, 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:** | | | **𝟙(Promotion)** | | | |
| | *Move to Large* | | *Move to Small & Med.* | | *Within-Firm* | |
| Productivity × 𝟙(Post) | 0.0014*** | -0.0016*** | -0.0016*** | -0.0039*** | 0.0010*** | 0.0014** |
| | (0.0001) | (0.0004) | (0.0003) | (0.0006) | (0.0003) | (0.0006) |
| Productivity × 𝟙(Post) × 𝟙(Mid) | | 0.0026*** | | 0.0039*** | | 0.0001 |
| | | (0.0005) | | (0.0007) | | (0.0007) |
| Productivity × 𝟙(Post) × 𝟙(Small) | | 0.0028*** | | 0.0047*** | | -0.0005 |
| | | (0.0005) | | (0.0008) | | (0.0007) |
| Employee F.E. | Y | Y | Y | Y | Y | Y |
| Salary & Seniority Controls | Y | Y | Y | Y | Y | Y |
| Cohort × Language × Role × Time F.E. | Y | Y | Y | Y | Y | Y |
| Industry × Location × Time F.E. | Y | Y | Y | Y | Y | Y |
| Firm Cohort | | | | | Y | Y |
| Observations | 3,199,173 | 3,022,682 | 3,199,173 | 3,022,682 | 2,939,059 | 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 (%) | 5.76 | -6.91 | -4.52 | -11.16 | 5.84 | 7.5 |
| Magnitude - Small (%) | | 5.14 | | 2.21 | | 5.11 |

## 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 $\beta$ coefficients from the equation:

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

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. We also estimate coefficients for double interaction of *Group* and $\mathbb{1}(Post)$, but do not report them for brevity. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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 to Large Firm)** | | |
| | **$\mathbb{1}$(Heterogeneity)** | | |
| | *Experience < Median* | Schooling < Median | *Non-Elite School* |
| Productivity × $\mathbb{1}$(Post) | 0.0013*** | 0.0017*** | 0.0007 |
| | (0.0003) | (0.0002) | (0.0006) |
| Productivity × $\mathbb{1}$(Post) × $\mathbb{1}$(Heterogeneity) | 0.0009*** | 0.0016* | 0.0012* |
| | (0.0004) | (0.0009) | (0.0006) |
| Employee 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 | 0.212 | 0.212 |
| Y-Mean | 0.033 | 0.033 | 0.033 |
| Magnitude - Main (%) | 3.77 | 4.97 | 2.04 |
| Magnitude - Interaction (%) | 6.58 | 9.88 | 5.61 |

# 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 $\beta$ coefficients from the equation:

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

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 steepness in career growth defined by whether the firm had an above-median average annual salary change in the pre-period. Finally, in columns 5-6, the subset is based on job stability defined by whether the firm had below-median average forced turnovers in the pre-shock period. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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) |
|---|---|---|---|---|---|---|
| **Outcome:** | **𝟙(Job Change to Large Firms with)** | | | | | |
| | **Tech-savvy Recruiters** | | **Steep Career Growth** | | **High Stability** | |
| | Yes | No | Yes | No | Yes | No |
| Productivity × 𝟙(Post) | 0.0014*** | 0.0003*** | 0.0016*** | 0.0001 | -0.0000 | 0.0018*** |
| | (0.0002) | (0.0001) | (0.0002) | (0.0001) | (0.0001) | (0.0002) |
| Employee F.E. | Y | Y | Y | Y | Y | Y |
| Salary & Seniority Controls | Y | Y | Y | Y | Y | Y |
| Cohort × Language × Role × Time F.E. | Y | Y | Y | Y | Y | Y |
| Industry × Location × Time F.E. | Y | Y | Y | Y | Y | Y |
| Observations | 3,199,173 | 3,199,173 | 3,192,168 | 3,192,168 | 3,194,104 | 3,194,104 |
| # Unique Employees | 143,446 | 143,446 | 143,445 | 143,445 | 143,445 | 143,445 |
| R-Squared | 0.189 | 0.23 | 0.2 | 0.221 | 0.224 | 0.207 |
| Y-Mean | 0.019 | 0.015 | 0.024 | 0.007 | 0.003 | 0.029 |
| Magnitude (%) | 7.69 | 2.09 | 6.48 | 2.02 | -0.65 | 6.08 |
| T-test ($H_0 : \beta_{Yes} = \beta_{No}$) | (0.0000)*** | | (0.0000)*** | | (0.0001)*** | |

# 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 $\beta$ coefficients from the equation:

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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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:** | **𝟙(Job Change to Large Firms with)** | | | |
| | **High Employment Growth** | | **High Sales Growth** | |
| | Yes | No | Yes | No |
| Productivity × 𝟙(Post) | 0.0008*** | 0.0010*** | 0.0011*** | 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 (%) | 7.03 | 4.85 | 8.02 | 3.86 |
| T-test ($H_0 : \beta_{Yes} = \beta_{No}$) | (0.0129)** | | (0.0003)*** | |

# 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 $\beta$ coefficient from the equation:

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

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 firm's percentile rank based on its public contribution variable in that quarter. 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. $\alpha_j$ is firm fixed effects. $\delta_t$ denotes time fixed effects. $X_{j,t}$ represents firm's time-varying controls (lagged employment in columns 1-2 and lagged stock of total contributions, depreciated annually at 15%, in columns 3-6) as well as the firm's industry × 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).

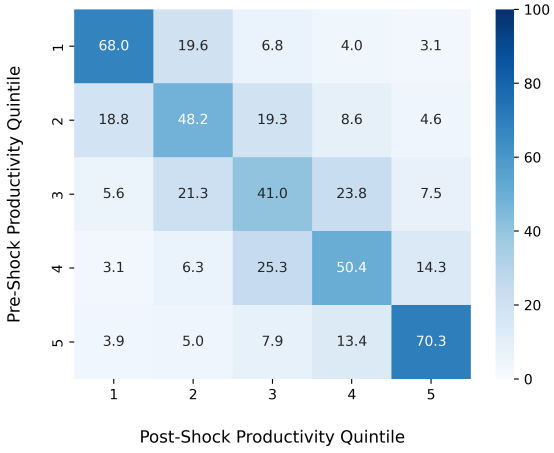
| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Outcome:** | **Employment** | | **Public Contributions** | | **Per-capita Contributions** | |
| | *QoQ Emp. Growth* | *Emp Reduction >= 50%* | *Pct(Y)* | *𝟙(Top Decile)* | *Pct(Y)* | *𝟙(Top Decile)* |
| 𝟙(Productivity > Median) × 𝟙(Post) | 0.0027 | -0.0007* | -0.7686 | -0.0032 | -0.7849 | -0.0003 |
| | (0.0016) | (0.0004) | (0.5405) | (0.0051) | (0.5300) | (0.0041) |
| 𝟙(Productivity > Median) × 𝟙(Post) × 𝟙(Mid) | -0.0023 | 0.0000 | -0.8271 | -0.0161*** | -0.9467 | -0.0165*** |
| | (0.0024) | (0.0006) | (0.6691) | (0.0061) | (0.6620) | (0.0052) |
| 𝟙(Productivity > Median) × 𝟙(Post) × 𝟙(Small) | -0.0149*** | 0.0034*** | -1.7186*** | -0.0175*** | -1.5779** | -0.0222*** |
| | (0.0028) | (0.0008) | (0.6606) | (0.0059) | (0.6646) | (0.0054) |
| 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 | 756,475 | 756,475 | 756,475 | 756,475 |
| # Unique Firms | 51,038 | 51,038 | 41,539 | 41,539 | 41,539 | 41,539 |
| R-squared | 0.232 | 0.13 | 0.619 | 0.583 | 0.577 | 0.498 |
| Y-Mean | 0.072 | 0.005 | 38.928 | 0.105 | 38.918 | 0.098 |
| Magnitude - Small (%) | -17.11 | 56.27 | -6.39 | -19.65 | -6.07 | -22.94 |

# Internet Appendix A

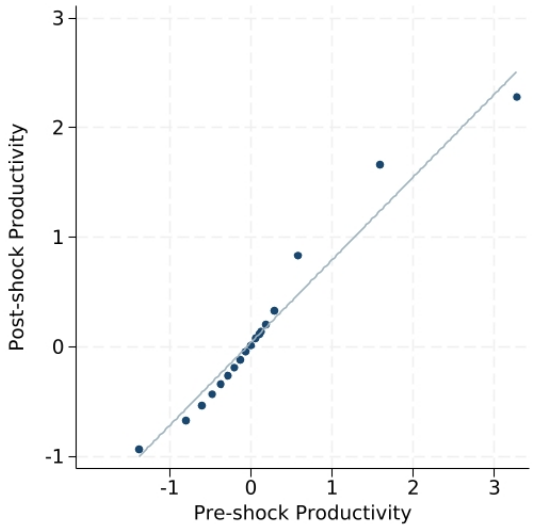
## A.1 Figures

### Figure A.1: **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.



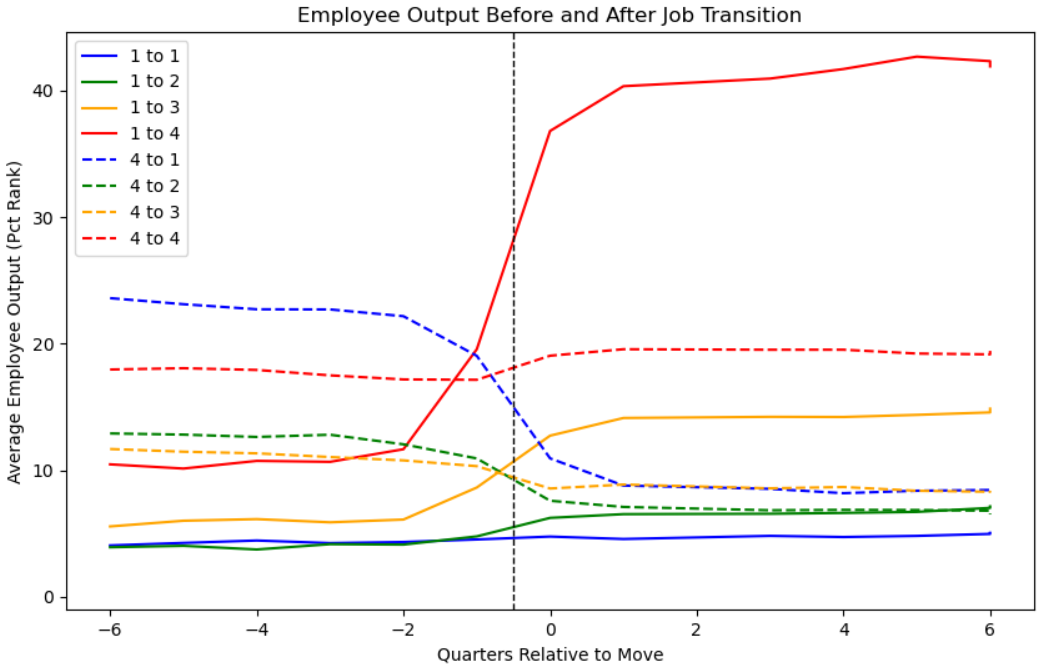

Panel A




Panel B

## Figure A.2: **Test for Endogenous Mobility in AKM Sample: Event Study around Moves**

This figure presents the evolution of employee output on GitHub around job moves in the AKM sample (May 2011 - April 2016), before GitHub's policy change. The outcome of interest is the percentile rank of employee's private contributions in the six 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 firm quality using 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.3: **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.5. The blue lines indicates the 5th and 10th percentile of the distribution and the red line plots the baseline coefficient estimate from Column 2 of Table A.5.

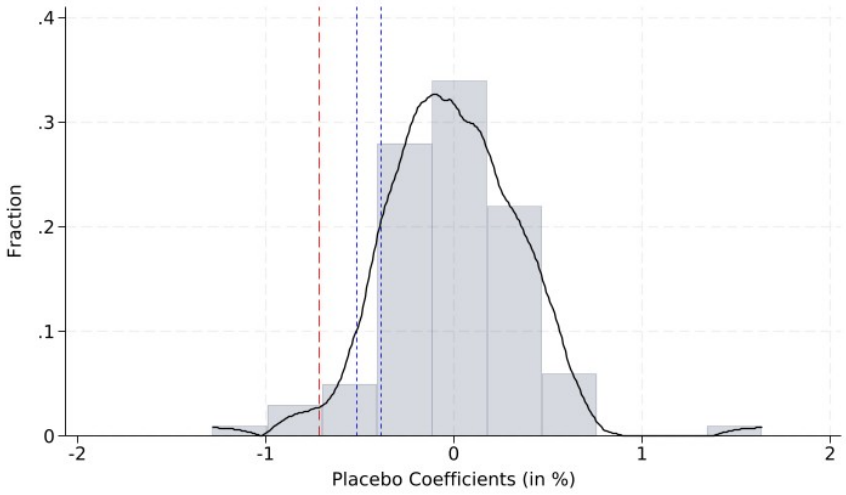


Figure A.4: **GitHub Shock & Employee Mobility to Large Firms: Robustness with Dummy Treatment Indicator**

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.

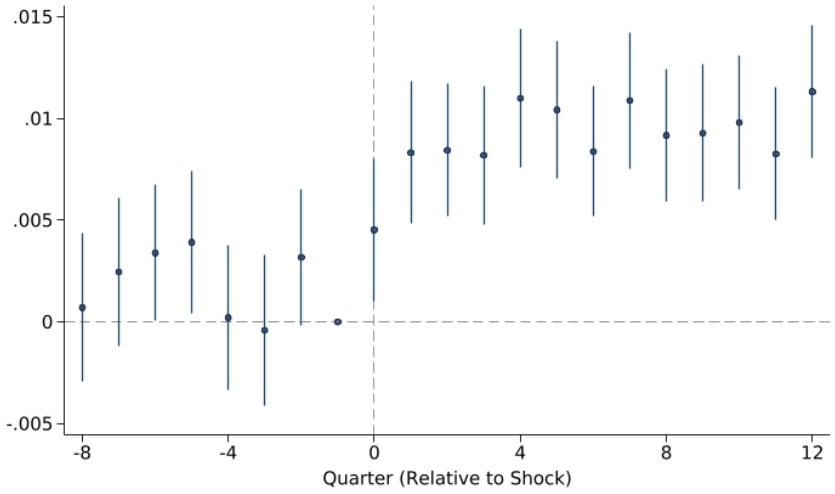
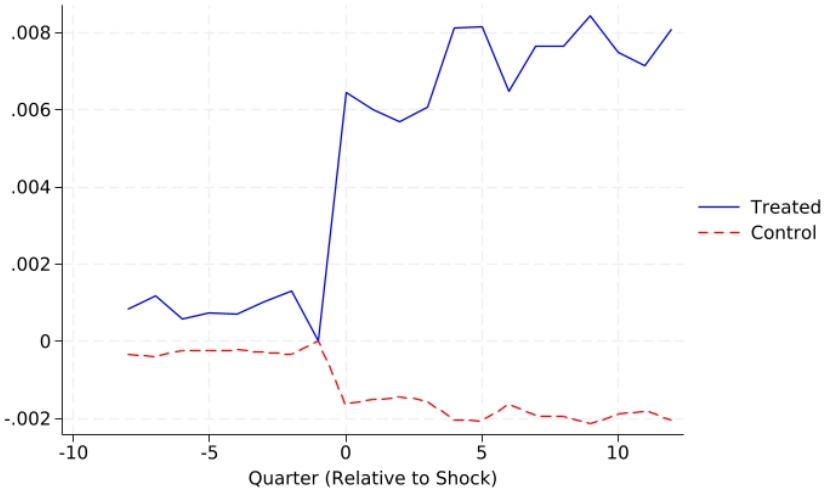


Figure A.5: **GitHub Shock & Employee Mobility to Large Firms: Breakdown by Treated and Control Group**

This figure presents the differential effect of the disclosure of employee productivity information on GitHub on mobility to large firms, comparing more productive and less productive employees. As in Figure A.4, treated employees are those with positive normalized AKM-based productivity, while the rest are classified as control. We begin by residualizing the probability of moving to a large firm by regressing the move indicator on time-varying employee salary and seniority controls, along with time fixed effects. We then plot the quarterly averages of these residuals for treated (blue) and control (red) employees, relative to 2016Q1—the quarter of the policy change. The estimates are normalized to zero in quarter $t = -1$ (2016Q1), which serves as the omitted baseline period.



# Tables

## Table A.1: **Sample Characteristics**

This table presents the distribution of employees across five major industries, O*NET-SOC occupational classifications, LinkedIn-based job roles, location, and coding languages. Language is the employee's most preferred coding language. Industry, O*NET-SOC occupational classifications, role, and location correspond to an employee's job at the time of GitHub's policy change.

**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 Two-digit O*NET Codes**

| **Role** | **Share of Users (%)** |
|---|---|
| 15 - Computer and Mathematical | 75.12 |
| 27 - Arts, Design, Entertainment, Sports, & Media | 7.16 |
| 11 - Management | 5.35 |
| 13 - Business and Financial Operations | 4.36 |
| 17 - Architecture and Engineering | 3.74 |
| Total | 95.73 |

**Panel C: Top 5 Eight-digit O*NET Codes**

| **Language** | **Share of Users (%)** |
|---|---|
| 15-1252.00 - Software Developers | 42.3 |
| 15-1254.00 - Web Developers | 4.62 |
| 15-1299.09 - Information Technology Project Managers | 4.14 |
| 15-2051.00 - Data Scientists | 3.41 |
| 15-1299.08 - Computer Systems Engineers/Architects | 3.32 |
| Total | 57.79 |

**Panel D: Top 5 LinkedIn Roles**

| **Role** | **Share of Users (%)** |
|---|---|
| Software Engineer - Applications | 17.84 |
| Software Engineer - Systems | 12.75 |
| Full Stack Developer | 3.92 |
| Software Developer | 2.87 |
| Web Developer | 2.47 |
| Total | 39.85 |

**Panel E: 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 F: 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 |

# 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 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 $\beta$ coefficients from the equation:

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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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:** | **𝟙(Job Change to Large Firms with)** | | | |
| | **Location Change** | | **Industry Change** | |
| | Yes | No | Yes | No |
| Productivity × 𝟙(Post) | 0.0011*** | 0.0007*** | 0.0015*** | 0.0002*** |
| | (0.0001) | (0.0001) | (0.0002) | (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 | 3,199,173 | 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 (%) | 5.97 | 4.37 | 5.36 | 4.44 |
| T-test ($H_0 : \beta_{Yes} = \beta_{No}$) | (0.0749)* | | (0.4714) | |

# Table A.4: **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 $\beta$ coefficients from the equation:

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

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, and indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 2016). Column 1 includes triple interaction with dummy indicators based on how the user's state voted in the presidential elections of 2016. Columns 2-5 include a triple interaction with a dummy indicator that takes the value of 1 if the employee's MSA has an above-median value of the economic/demographic characteristic: share of white population (median house value) (change in house value) (gini index) in column 2 (3) (4) (5). $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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) | 0.0020*** | 0.0018*** | 0.0012* | 0.0019*** | 0.0019*** |
| | (0.0002) | (0.0002) | (0.0006) | (0.0003) | (0.0004) |
| Productivity × 𝟙(Post) | -0.0009 | | | | |
| × 𝟙(Always Red) | (0.0006) | | | | |
| Productivity × 𝟙(Post) | -0.0009 | | | | |
| × 𝟙(Swinged Red in 2016) | (0.0005) | | | | |
| Productivity × 𝟙(Post) | -0.0007 | | | | |
| × 𝟙(Swinged Blue in 2016) | (0.0007) | | | | |
| Productivity × 𝟙(Post) | | -0.0003 | | | |
| × 𝟙(Above-Median Share of White Population) | | (0.0005) | | | |
| Productivity × 𝟙(Post) | | | 0.0006 | | |
| × 𝟙(Above-Median House Value) | | | (0.0006) | | |
| Productivity × 𝟙(Post) | | | | -0.0003 | |
| × 𝟙(Above-Median Change in House Value) | | | | (0.0003) | |
| Productivity × 𝟙(Post) | | | | | -0.0002 |
| × 𝟙(Above-Median Gini Index) | | | | | (0.0004) |
| Employee F.E. | Y | Y | Y | Y | Y |
| 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 | 3,178,126 | 3,143,808 | 3,143,808 | 3,143,808 | 3,143,808 |
| # Unique Employees | 142,513 | 140,966 | 140,966 | 140,966 | 140,966 |
| R-Squared | 0.212 | 0.21 | 0.21 | 0.21 | 0.21 |
| Y-Mean | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 |

## Table A.5: **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 AKM-based average employee productivity that lies in the top quartile (bottom quartile). 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*** | -0.0071*** | 0.0016 | -0.0033 | -0.0002 |
| | (0.0006) | (0.0024) | (0.0022) | (0.0029) | (0.0025) |
| #Firms | 1675 | 106 | 103 | 136 | 56 |

## Table A.6: **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.1. 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) | 0.0019*** | 0.0020*** | 0.0023*** | 0.0028*** |
| | (0.0002) | (0.0002) | (0.0002) | (0.0003) |
| Employee F.E. | Y | Y | Y | Y |
| 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 | 2,938,500 | 2,823,538 | 2,535,125 | 1,553,657 |
| # Unique Employees | 130,984 | 125,827 | 112,864 | 69,072 |
| R-Squared | 0.213 | 0.215 | 0.219 | 0.244 |
| Y-Mean | 0.035 | 0.035 | 0.035 | 0.034 |
| Magnitude (%) | 5.59 | 5.8 | 6.43 | 8.15 |

# Table A.7: **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, we use five buckets of private contributions (0, [1, 9], [10,99], [100,999], and [1000+)) as input in the AKM. In A2, we use the raw counts of private contributions as the dependent variable in the AKM. In A3, we use the logarithm of one added to private contributions as the dependent variable, and in A4 we use the logarithm of private contributions if contributions are greater than zero and 0 otherwise. In Panel B, we use different measures of employee productivity computed without estimating the AKM. We define productivity as the employees' rank based on their private contributions (employees' 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. Panel E presents results with more granular controls. In E1, we control for employee's cohort, role, language, location, and industry interacted with time fixed effects. In E2, we control for employees' public contributions-based productivity, number of stars received on GitHub, and number of publications in the pre-period. Finally, in Panel F, we run the test on alternative samples. In F1 and F2, run the test on a matched subsample of employees. We match employees with positive normalized productivity estimates 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. In F3, we exclude employees who have ever worked at Microsoft, the firm that acquired GitHub in 2018. 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 Categorical Variables* | | | | | | |
| Productivity × 𝟙(Post) | 0.0016*** | (0.0002) | 3,199,173 | 0.212 | 0.033 | 5.02 |
| *A2: AKM with Raw Contribution Numbers* | | | | | | |
| Productivity × 𝟙(Post) | 0.0008*** | (0.0001) | 3,199,173 | 0.211 | 0.033 | 2.73 |
| *A3: AKM with Logarithm of (1+Y)* | | | | | | |
| Productivity × 𝟙(Post) | 0.0017*** | (0.0002) | 3,199,173 | 0.212 | 0.033 | 4.98 |
| *A4: AKM with Logarithm where Y_min = 0* | | | | | | |
| Productivity × 𝟙(Post) | 0.0016*** | (0.0002) | 3,199,173 | 0.212 | 0.033 | 4.89 |
| **Panel B: Alternate Productivity Definition** | | | | | | |
| *B1: Pct(Private Contributions)* | | | | | | |
| Productivity × 𝟙(Post) | 0.0020*** | (0.0002) | 3,199,173 | 0.212 | 0.033 | 6.08 |
| *B2: Share Private Contributions* | | | | | | |
| Productivity × 𝟙(Post) | 0.0040*** | (0.0008) | 3,199,173 | 0.212 | 0.033 | 5.79 |
| **Panel C: Dummy Productivity Definition** | | | | | | |
| *C1: Normalized Productivity Estimate > 0* | | | | | | |
| 𝟙(Productivity > 0) × 𝟙(Post) | 0.0071*** | (0.0005) | 3,199,173 | 0.212 | 0.033 | 21.35 |
| *C2: Productivity in Top Decile* | | | | | | |
| 𝟙(Top Decile Productivity) × 𝟙(Post) | 0.0057*** | (0.0006) | 3,199,173 | 0.212 | 0.033 | 17.15 |
| *C3: Productivity in Top Quartile* | | | | | | |
| 𝟙(Top Quartile Productivity) × 𝟙(Post) | 0.0068*** | (0.0005) | 3,199,173 | 0.212 | 0.033 | 20.39 |
| *C4: Productivity in Top Tercile* | | | | | | |
| 𝟙(Top Tercile Productivity) × 𝟙(Post) | 0.0046*** | (0.0005) | 3,199,173 | 0.212 | 0.033 | 13.72 |
| **Panel D: Alternate AKM Sample Period** | | | | | | |
| *D1: Pre Sample AKM (2011-14)* | | | | | | |
| Productivity × 𝟙(Post) | 0.0015*** | (0.0003) | 2,014,005 | 0.219 | 0.031 | 4.68 |
| **Panel E: More Controls** | | | | | | |
| *E1: Control for Cohort × Role × Language × Location × Industry × Time* | | | | | | |
| Productivity × 𝟙(Post) | 0.0023*** | (0.0005) | 745,659 | 0.37 | 0.036 | 6.35 |
| *E2: Control for Public Contributions, Stars, and Publications* | | | | | | |
| Productivity × 𝟙(Post) | 0.0020*** | (0.0002) | 3,180,434 | 0.211 | 0.033 | 5.92 |
| **Panel F: Alternate Employee Sample** | | | | | | |
| *F1: Sample Matched on Individual Demographics* | | | | | | |
| Productivity × 𝟙(Post) | 0.0019*** | (0.0003) | 1,241,915 | 0.254 | 0.03 | 6.18 |
| *F2: Sample Matched on Individual Demographics and Public Contributions* | | | | | | |
| Productivity × 𝟙(Post) | 0.0017*** | (0.0003) | 1,405,593 | 0.25 | 0.031 | 6.37 |
| *F3: Sample Excluding Microsoft Employees* | | | | | | |
| Productivity × 𝟙(Post) | 0.0017*** | (0.0002) | 3,035,617 | 0.217 | 0.032 | 5.13 |

# Table A.8: **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 $\beta$ coefficient from the equation:

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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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:** | **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 × 𝟙(Post) | 0.0000 | -0.0009*** | -0.0008*** | -0.0003*** | -0.0000 | 0.0004*** | 0.0001* | 0.0012*** |
| | (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 |
| 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 (%) | 0.4 | -6.04 | -5.02 | -3.67 | -0.41 | 3.84 | 3.05 | 6.57 |

# Table A.9: **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 $\beta$ coefficient from the equation:

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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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:** | **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 × 𝟙(Post) | 0.0010*** | 0.0009*** | 0.0006*** | 0.0005*** | 0.0010*** | 0.0009*** | 0.0007*** | 0.0004*** |
| | (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 |
| 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 (%) | 5.82 | 5.05 | 8.2 | 7.56 | 5.73 | 5.48 | 7.19 | 5.56 |

# Table A.10: **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 $\beta$ coefficient from the equation:

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

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. $\alpha_i$ denotes employee fixed effects. $X_{i,t}$ includes time-varying employee controls – current salary and seniority. We also include fixed effects for the interaction of employee cohort, role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. 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(Mobility to Star Startups)** | | | | |
| | *Funding > Series B* | *Funding > Series D* | *Funding > $ 100 Mn* | *Funding > $ 500 Mn* | *Funding > $ 1 Bn* |
| Individual Contributions × 𝟙(Post) | -0.0003*** | 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 |
| Salary & Seniority Controls | Y | Y | Y | Y | Y |
| Cohort × Language × Role × Time F.E. | Y | Y | Y | Y | Y |
| Industry × Location × Time F.E. | Y | Y | Y | Y | Y |
| Observations | 3,145,396 | 3,145,396 | 3,145,396 | 3,145,396 | 3,145,396 |
| # Unique Employees | 143,437 | 143,437 | 143,437 | 143,437 | 143,437 |
| R-Squared | 0.173 | 0.168 | 0.166 | 0.155 | 0.151 |
| Y-Mean | 0.005 | 0.003 | 0.003 | 0.001 | 0.001 |
| Magnitude (%) | -5.04 | 2.8 | 1.99 | 11.49 | 12.38 |
| 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.11: **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 is the percentile rank of user's public contributions in the given quarter. 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:** | *Pct(Y)* | *$\mathbb{1}$(Top Quartile)* | *$\mathbb{1}$(Top Quintile)* | *$\mathbb{1}$(Top Decile)* |
| Productivity × $\mathbb{1}$(Post) | 0.5088*** | 0.0058*** | 0.0063*** | 0.0052*** |
| | (0.0658) | (0.0007) | (0.0007) | (0.0007) |
| Employee F.E. | Y | Y | Y | Y |
| Contribution Stock & Trend Controls | 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 | 3,075,320 | 3,075,320 | 3,075,320 | 3,075,320 |
| # Unique Employees | 143,324 | 143,324 | 143,324 | 143,324 |
| R-Squared | 0.614 | 0.586 | 0.585 | 0.571 |
| Y-Mean | 27.742 | 0.292 | 0.247 | 0.138 |
| Magnitude (%) | 1.83 | 1.99 | 2.56 | 3.78 |

## Table A.12: **GitHub Shock & Share of Productive Employees at Small Firms**

This table presents estimates on the impact of GitHub's disclosure of employee productivity information on the share of productive employees at firms. The sample, setup, and the independent variables remain the same as in Table 10. The dependent variable is the share of productive employees at the firm-quarter level. The employment share of productive employees is defined as the number of productive employees divided by the total number of employees (from LinkedIn) in the given firm-quarter. In column 1, we classify employees as productive if their normalized productivity estimate from the AKM is positive. We define employees as productive in column 2 (3) (4) if their normalized productivity estimate lies in the top tercile (quartile) (quintile). We control for firms' time-varying lagged employment as well as firm fixed effects and industry × 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) |
|---|---|---|---|---|
| **Outcome:** | **Share of Productive Employees** | | | |
| | *Productivity > 0* | *Top Tercile Productivity* | *Top Quartile Productivity* | *Top Quintile Productivity* |
| 𝟙(Productivity > Median) × 𝟙(Post) | -0.0009 | -0.0007 | -0.0012 | -0.0009 |
| | (0.0011) | (0.0011) | (0.0011) | (0.0011) |
| 𝟙(Productivity > Median) × 𝟙(Post) × 𝟙(Mid) | -0.0056*** | -0.0066*** | -0.0057*** | -0.0055*** |
| | (0.0014) | (0.0015) | (0.0015) | (0.0014) |
| 𝟙(Productivity > Median) × 𝟙(Post) × 𝟙(Small) | -0.0120*** | -0.0147*** | -0.0123*** | -0.0118*** |
| | (0.0017) | (0.0019) | (0.0018) | (0.0017) |
| Controls | Y | Y | Y | Y |
| Firm F.E. | Y | Y | Y | Y |
| Industry × Location × Time F.E. | Y | Y | Y | Y |
| Observations | 1,106,564 | 1,106,564 | 1,106,564 | 1,106,564 |
| # Unique Firms | 51,038 | 51,038 | 51,038 | 51,038 |
| R-squared | 0.791 | 0.793 | 0.793 | 0.791 |
| Y-Mean | 0.04 | 0.054 | 0.045 | 0.04 |
| Magnitude- Base (%) | -2.15 | -1.35 | -2.58 | -2.32 |
| Magnitude - Small (%) | -32.41 | -28.74 | -29.58 | -32.24 |

## Table A.13: **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 $\beta$ 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.

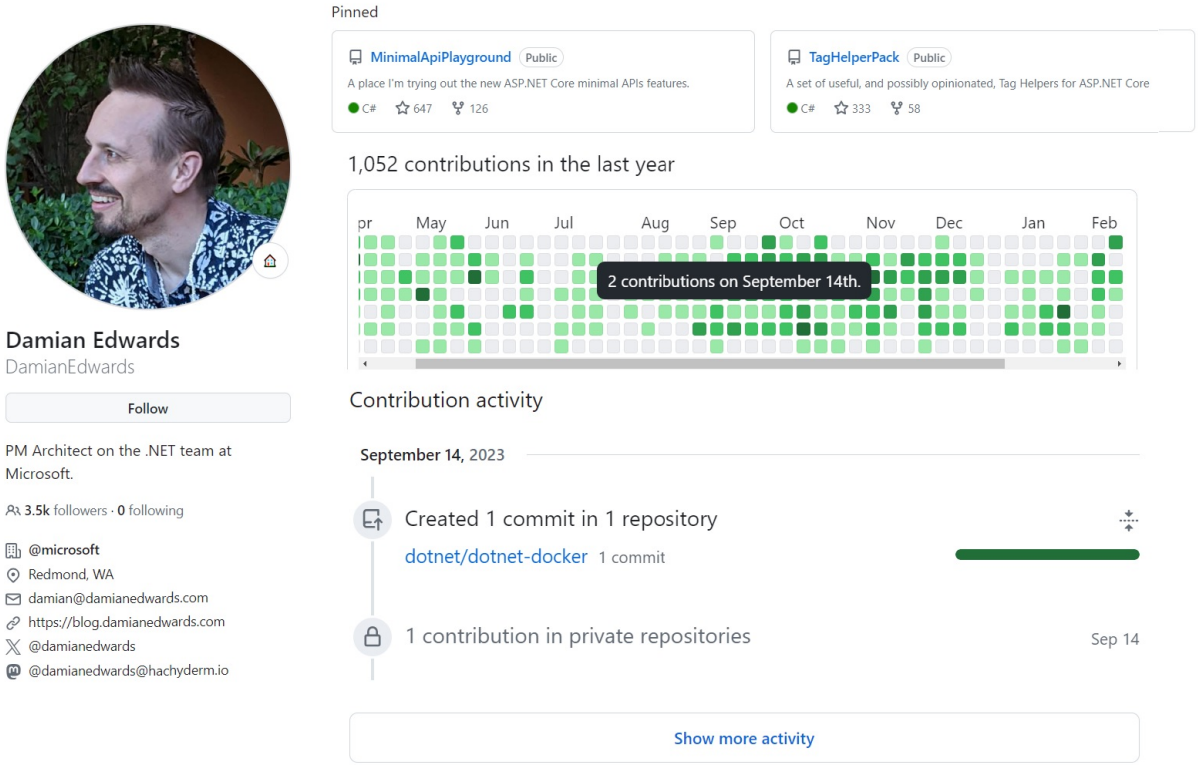
The GitHub Archive (`https://www.gharchive.org/`), a third-party repository, records and stores all public events on GitHub, making daily snapshots of public contributions available for download. This dataset offers comprehensive historical coverage of public contributions that cannot be altered retroactively, 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.8 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.

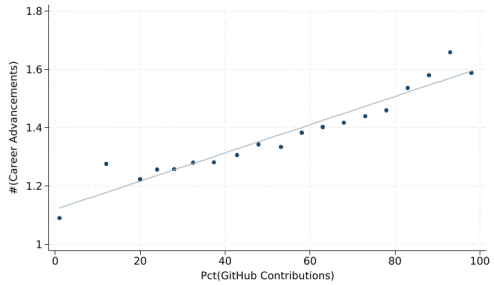
Figure B.1: **Sample GitHub Profile and Contribution Calendar**

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

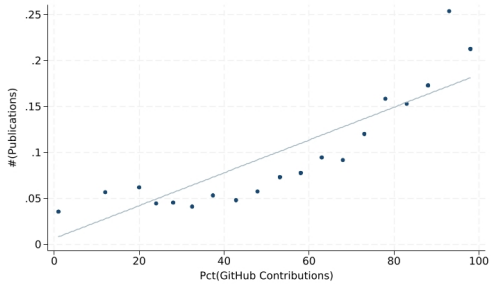


## Figure B.2: **Association between Employees' GitHub Contributions, Career Advancements, and Publications**

This figure presents the scatter plot of employees' GitHub contributions and career advancements (in panel A) and publications (in Panel B). On the y-axis, we plot the number of career advancements in Panel A and the number of articles authored by the employee on Google Scholar, DevTo, and ORCiD in Panel B. On the x-axis, we plot the percentile rank of the employee in terms of their total GitHub contributions. Contributions exclude public contributions made towards repositories owned by the employees themselves. We collapse the data into equal-sized bins and plot the mean value of the variables in each bin.



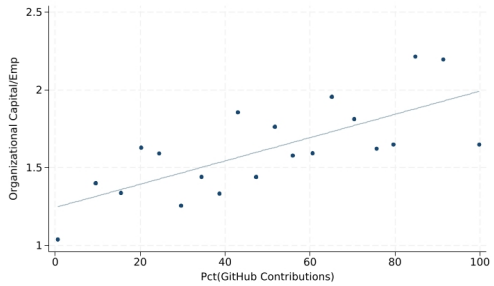

Panel A



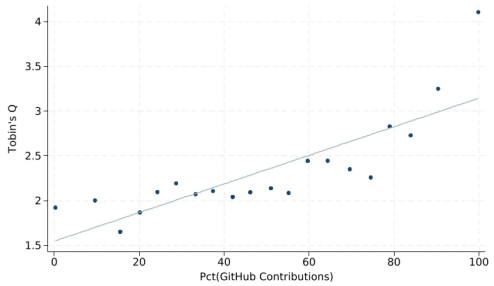

Panel B

## Figure B.3: **Association between Firms' GitHub Contributions, Organizational Capital, and Tobin's Q**

This figure presents the scatter plot of firms' GitHub contributions and organization capital per employee (in panel A) and Tobin's Q (in Panel B). The y-axis plots the average of the firm values over the six years spanning our sample period (2013-2018). Firms' organizational capital measure is based on Eisfeldt and Papanikolaou (2013). The x-axis plots the percentile rank of firms in terms of their total GitHub contributions. Contributions exclude public contributions made towards repositories owned by the employees themselves. We collapse the data into equal-sized bins and plot the mean value of the variables in each bin.



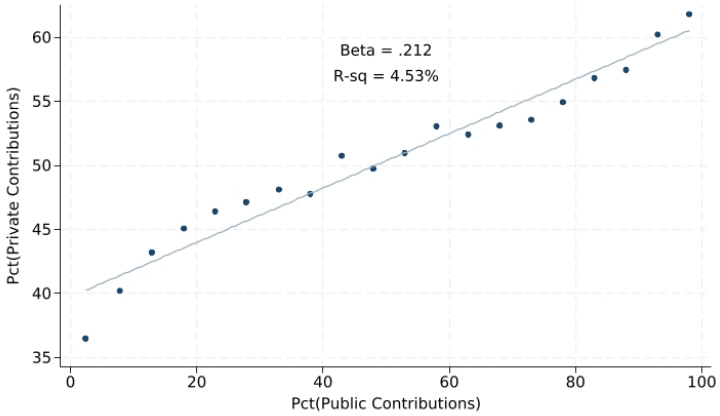

Panel A




Panel B

## Figure B.4: **Association between Employees' Public and Private Contributions**

This figure presents the scatter plot of employees' private and public GitHub contributions. The y-axis plots employees' percentile rank based on their total private contributions and the x-axis plots the percentile rank in terms of their total public contributions. We collapse the data into equal-sized bins and plot the mean value of the variables in each bin. The sample includes employees with non-zero total public and private contributions.



## Figure B.5: **Presence of Recruiters (HR Professionals) on GitHub**

This figure plots the fraction of firms with at least one Human Resource (HR) professional on GitHub over time (2009-2019) by firm size. The solid dark blue line represents the percentage of large firms (> 1,000 employees) that have at least on HR professional on GitHub, while the dashed grey line plots the statistic for mid-sized (50 < employees <= 1,000) and the dash and dot light blue line corresponds to the small firms (<= 50 employees). The numbers are based on the matched sample of employees within the U.S. with publicly-linked GitHub and LinkedIn profiles.

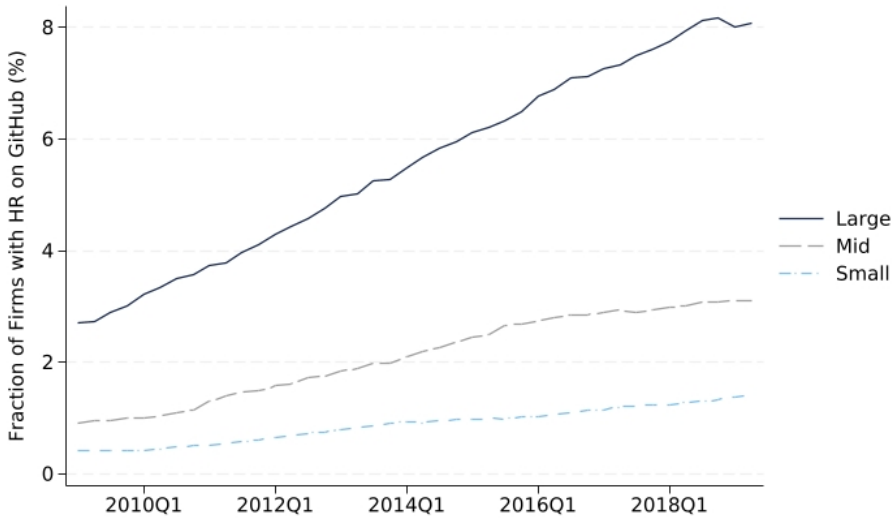
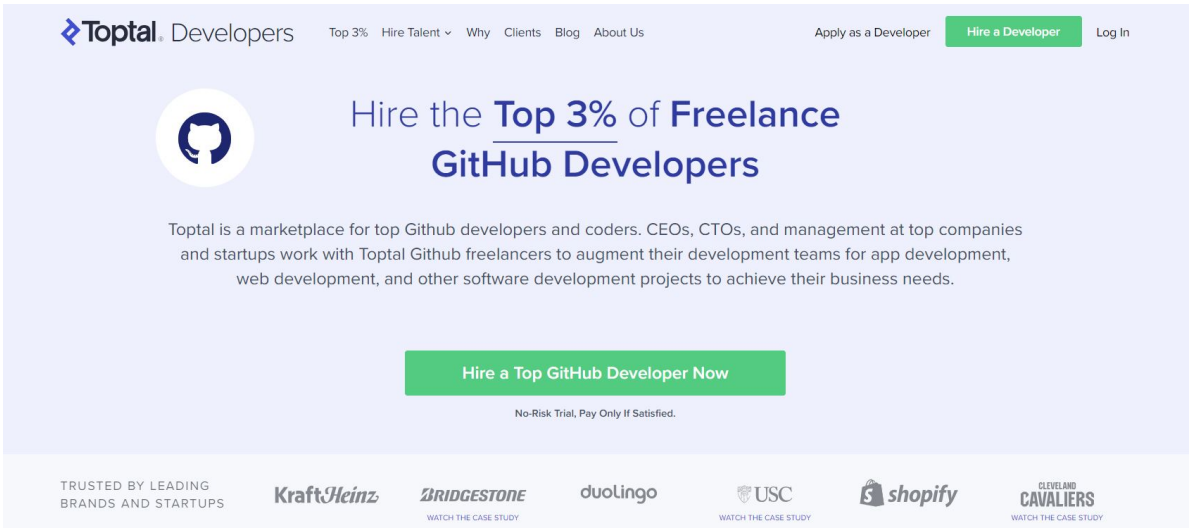
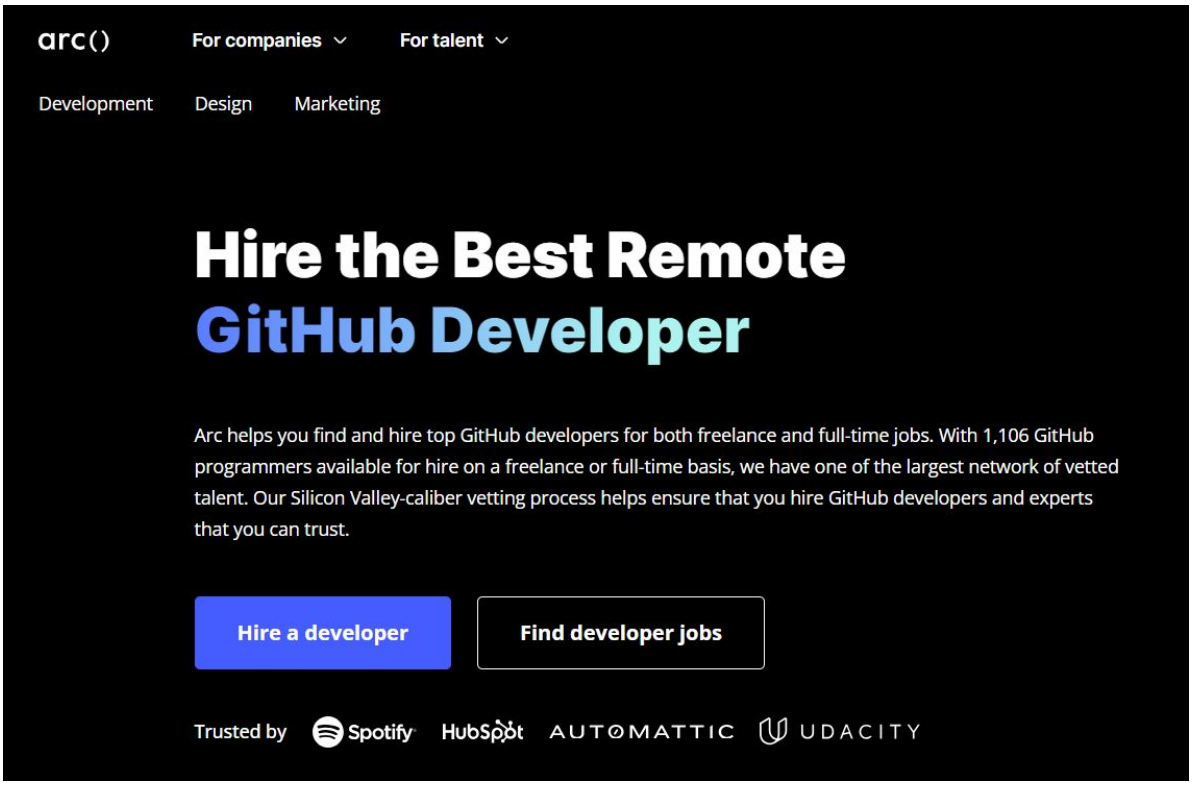


Figure B.6: **GitHub as a Source of Technical Talent**

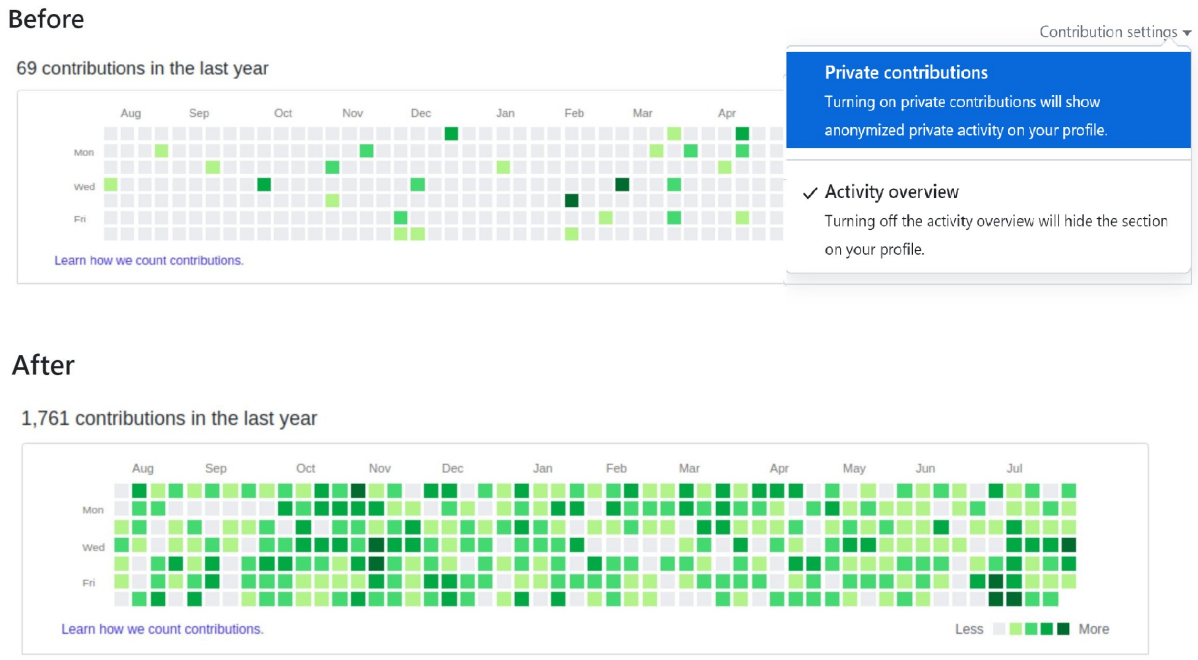
This figure presents snapshots of third-party web platforms that assist firms in sourcing potential employees using GitHub.



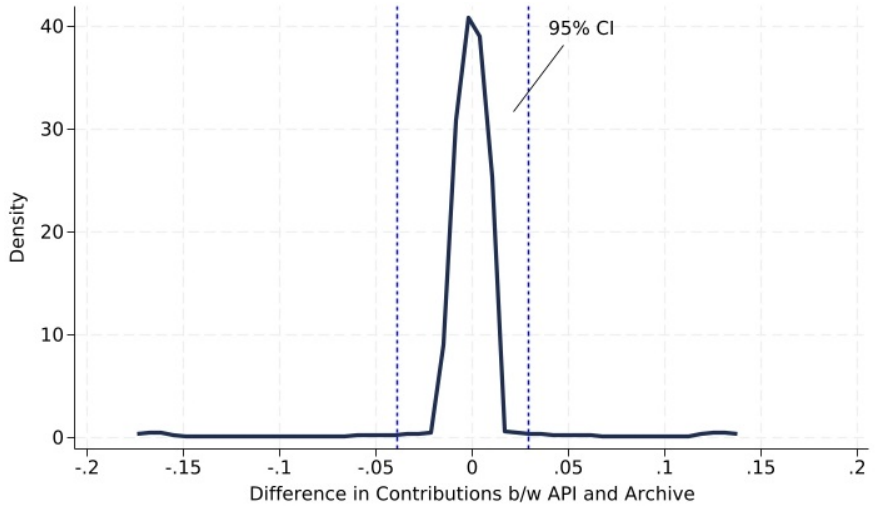
## Figure B.7: **GitHub Calendar Before and After Policy Change**

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.

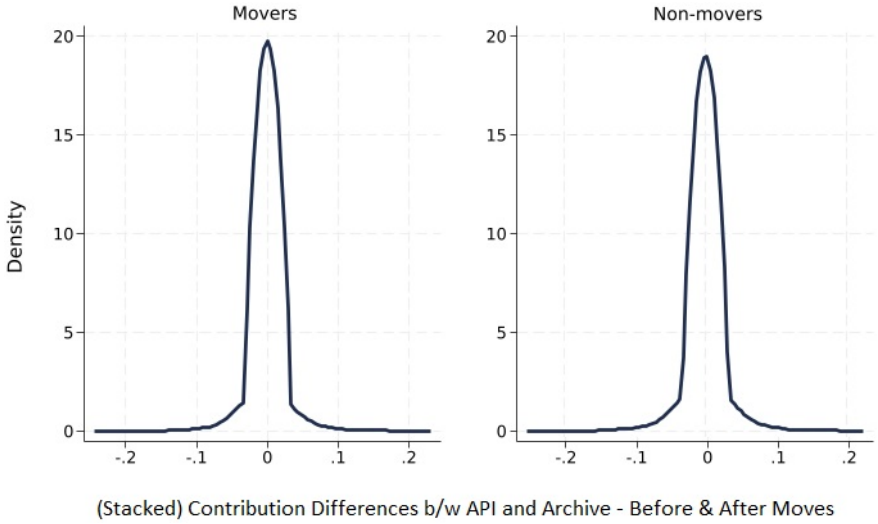


# Figure B.8: **Discrepancy in Public GitHub Contributions**

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

# Table B.1: **Association between GitHub Private Contributions and Individual/Firm Productivity**

This table presents the estimates of a correlational study between GitHub private contributions, individual career advancements and publications as well as firms' organizational capital per employee and Tobin's Q. In columns 1 and 2, the outcome variable is the employee's total number of career advancements and publications respectively during our sample period, as defined in Figure B.2. The independent variable is the percentile rank of the employee in terms of their private GitHub contributions. Both columns control for the percentile rank of the employee based on their public contributions, cohort, role and language as well as industry and location. In columns 3 and 4, the outcome variable is the firm's average organizational capital per employee and Tobin's q respectively during our sample period (years 2013-2018), as defined in Figure B.3. The independent variable is the percentile rank of firms in terms of their private GitHub contributions. Columns 3 and 4 include the percentile rank of firm's public contributions, firm size, and firm's industry as controls. Robust standard errors are reported in parentheses. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | ***Individual*** | | ***Firm*** | |
| **Outcome:** | **#(Career Advancements)** | **#(Publications)** | **Organizational Capital/Emp** | **Tobin's Q** |
| Pct(Private Contributions) | 0.0010*** | 0.0026*** | 0.0029* | 0.0040** |
| | (0.0002) | (0.0001) | (0.0016) | (0.0016) |
| Public Contribution Controls | Y | Y | Y | Y |
| Firm Size Controls | | | Y | Y |
| Cohort x Role x Language F.E. | Y | Y | | |
| Industry x Location F.E. | Y | Y | | |
| Industry F.E. | | | Y | Y |
| Observations | 143,528 | 143,528 | 1,624 | 2,016 |
| R-Squared | 0.158 | 0.256 | 0.409 | 0.382 |
| Magnitude (%) | 26.2 | 4.9 | 6.3 | 6.1 |