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

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

March 24, 2026

**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.

JEL Codes: D82, J24, J62, L11

*We are grateful to seminar participants at the American Finance Association 2025, Western Finance Association 2025, SFS Cavalcade 2025, Midwest Finance Association 2025, Labor and Finance Group 2025, FSU Truist Beach Conference 2025, ISB Summer Research 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 has 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 traditionally held 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 qualifications and output, mitigating informational disparities. This paper investigates how transparency in worker productivity influences labor market reallocation, examining whether it furthers 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.

Ex-ante predictions about labor dynamics hinge on whether worker sorting under informational frictions is driven more by adverse selection or by signaling. In adverse-selection models, current employers observe worker quality more accurately and thus retain their best workers. Outside firms, lacking this information, offer low wages to job changers, so only low-ability workers switch jobs in equilibrium (Greenwald, 1986; Gibbons and Katz, 1991). These models imply little reallocation of high-ability workers toward more productive firms. Given that large firms are often viewed as the most productive (Idson and Oi, 1999), reducing informational frictions should shift top workers toward larger employers, an effect potentially amplified by their scale economies in data-driven recruitment (Brynjolfsson and McElheran, 2016). By contrast, signaling models emphasize that job assignments reveal worker ability (Waldman, 1984). High-ability workers may therefore seek employment at large firms to generate stronger productivity signals. Greater transparency would then enable smaller firms to compete more effectively for talent. Consequently, the predicted impact of productivity disclosure on labor reallocation is theoretically ambiguous.

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. These contributions contain meaningful information on 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[2] 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 that these private contributions are typical of employer-managed projects, the quantity of individuals' private contributions captures their productivity in their current firm. We validate that private contributions predict individual- and firm-level outcomes, even after controlling for public contributions, indicating that 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 display their total contributions on their profiles.

We construct a novel database that integrates employee work history with individual output to implement our research design. Employee work history is obtained from Revelio Labs, which compiles online résumés from platforms such as LinkedIn to construct a universal HR database. We

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

[2] 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.

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 GitHub's private contributions. This is achieved by regressing total private contributions implemented prior to the GitHub policy change 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. While our baseline relies on these AKM estimates, our results are robust to using the raw number or share of private contributions as measures of individual productivity.

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 increase their mobility to larger firms (those employing over a thousand individuals), by 5.2% over the sample mean in our most robust specification, after the GitHub policy change. This effect is driven by 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. 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).

A key concern is that employees with more private contributions may be inherently different, so the results could reflect underlying productivity trends. We address this by showing that the effects are specific to the visibility of GitHub private contributions. Placebo tests using public GitHub contributions, employee salaries, and publication records, which were always visible, show no significant effects. This supports the interpretation that the results reflect the revelation of new information. Consistent with this, we also find an immediate decline in stock prices for small firms with high exposure to private contributions on the day of the policy change.

Moreover, we address potential biases from self-selection, where only individuals seeking new

employment might choose to reveal private contributions. Our results persist when restricted to users who reveal their private contributions, reducing the likelihood of such biases. Also, an analysis *within only the set of job movers*, 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.

After the shock, talent reallocates from smaller toward larger firms. Workers with one standard deviation higher productivity become less likely to move to medium or small firms and more likely to move to large firms, with flows rising almost monotonically from -6% at firms with eleven to fifty employees to +6.6 % at firms with more than ten thousand employees. These reallocations come largely at the expense of the smallest firms (those employing fifty employees or less), with high-productivity workers becoming 4.8% more likely to move from small to large firms after the shock. Such moves typically constitute career upgrades, with talented employees transitioning into more senior or higher-paying roles in large firms. Large firms simultaneously increase within-firm promotions to retain productive workers, consistent with improved outside options.

We show that reducing information asymmetry about worker quality via GitHub diminishes the relevance of traditional metrics of employee quality. We find that high-productivity individuals with less experience, less schooling, or non-elite degrees are more likely to move to large firms after the policy change. These results support the hypothesis that web-based platforms have led to a 'democratization' of screening in the labor markets.

Talent flows to larger firms reflect both firm screening ability and worker preferences. High productivity workers disproportionately sort into large firms whose HR departments use GitHub, relative to otherwise similar large firms, consistent with greater investment in screening technologies (Eckel and Yeaple, 2017). The effects are stronger for large firms with above-median pre-treatment salary growth. More broadly, high-productivity workers move toward firms with higher ex ante employment and sales growth, consistent with sorting toward firms with stronger demand for talent.

The broad scope of our shock allows us to examine firm-level consequences of labor reallocation. Among small treated firms (those with less than fifty employees, and with above-median average employee productivity), the share of productive employees falls by 32% of the sample mean following the shock. Consequently, these firms experience 17% slower quarter-on-quarter employment growth relative to the sample mean and face a higher probability of market exit. Small treated firms also see

a 6% decline in productivity, measured by public GitHub contributions, not mechanically affected by the policy change.

The labor market reallocation towards large firms also has implications for aggregate industry dynamics. We find that industries with higher pre-shock GitHub activity experience larger increases in national labor market concentration after the shock, consistent with recent U.S. trends (Autor et al., 2023). Though correlational, this pattern underscores a key trade-off: transparency can improve talent allocation, but may also accelerate consolidation and weaken competition.

Our paper contributes to the literature on the growing concentration of talent in large firms. Over recent decades, employment has shifted toward large, mature firms and away from small, young firms (Haltiwanger, 2015), alongside weaker employment responses to productivity shocks (Decker et al., 2020) and a declining labor share among superstar firms (Autor et al., 2020). This shift is especially pronounced among high-skilled workers, with more than 80% of inventors employed by large firms (Giroud et al., 2024). While prior work links the rise of large firms to big data–driven productivity and quality improvements (Bessen, 2020; Aghion et al., 2023; Brynjolfsson et al., 2023; Tambe and Hitt, 2012), its implications for labor flows remain unsettled. Large firms may attract talent through wage premia, R&D-driven overhiring, or acquisitions (Akcigit and Goldschlag, 2023; Acemoglu et al., 2018; Tate and Yang, 2023), yet top workers also exit to found startups or leave after IPOs (Gompers et al., 2005; Bernstein, 2015). We study a setting in which information frictions about worker productivity are reduced to test whether greater transparency causes top talent to flow toward large firms or away from them.

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 settings, such as increased turnover of 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, whether improved metrics shift talent toward larger or smaller firms, or generate broader market effects remains unclear. We address this question using a shock that revealed individual-level productivity for a large share of U.S. tech workers.

Finally, our results provide rare empirical validation of central theories in organizational economics, which predict that in the absence of frictions, productive workers should concentrate in the largest firms. Lucas (1978) models the largest firms as being run by the most talented

individuals. Kremer (1993)'s O-Ring theory implies that complementarities lead high-skill workers to cluster together, making such firms disproportionately productive and large. Consistent with these predictions, we show that as information frictions decline, productive employees increasingly sort into larger firms.

## 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 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 within projects known as *repositories*. By providing centralized online repositories for code storage and management, GitHub enables developers to work collaboratively on projects. It also helps stakeholders track modifications to repositories over time. Whenever a repository is modified, a snapshot of the repository 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. Any GitHub user can view all details of a public repository. In contrast, private repositories are restricted to specified users or teams (by repository owners or designated managers), providing a secure environment for proprietary projects. While today GitHub users can create an unlimited number of public or private repositories for free, in the past only paid GitHub accounts could maintain private repositories.[3] Subsequently, until recently, private repositories were mostly held by organizations

[3]GitHub initially charged per repository, with plans ranging from $7 per month for up to 5 private repositories to

seeking to maintain confidentiality while facilitating within-team collaboration on projects.

### 2.2 GitHub Contributions: A Measure of Productivity

In addition to facilitating version control and collaboration on software projects, GitHub provides users with individual profile pages. These profiles include the individual's name, bio, location, affiliation with firms, and other relevant information.[4] GitHub's user profile prominently features the *contribution calendar*, offering a representation of a user's GitHub activity over time. This calendar provides a real-time high-frequency update on a user's GitHub contributions. The granularity and high frequency of the contributions and the large-scale coverage of GitHub make it an appealing measure of individual output and productivity of high-skilled professionals.

We validate this measure in Internet Appendix B.2 by relating employees' contribution ranks to career advancements and publications. We find strong positive and linear correlations, indicating that total contributions effectively capture individual employee quality and productivity. We also provide firm-level validation by showing that a firm's GitHub contributions are positively correlated with its organizational capital (Eisfeldt and Papanikolaou, 2013) and Tobin's Q. Together, these patterns suggest that GitHub contributions are a meaningful proxy for productivity, consistent with prior work that similarly leverages GitHub contributions to measure software developer output (Holub and Thies, 2023; El-Komboz and Fackler, 2023; Conti et al., 2023).

In addition, given our focus on changes in the visibility of private contributions, we assess whether they capture employee productivity beyond what is measured by public contributions. We find that while public and private contributions are positively correlated, public contributions explain less than 5% of the variation in private contributions. Internet Appendix B.2 further shows that private contributions predict individual and firm-level outcomes even after controlling for public contributions, indicating that they contain substantial incremental information.

### 2.3 GitHub and Recruitment

GitHub has become an important tool for recruiters to identify and screen talent. Montandon et al. (2021) report that most recruiters hiring for technical roles use GitHub in recruiting. Internet

$200 per month for up to 125. In October 2015, GitHub introduced unlimited private repositories for $7 per month, and made them 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–April 2014, ensuring that pricing changes do not drive our findings.

[4] See Internet Appendix B Figure B.1 for a sample GitHub profile.

Appendix B.3. provides a collage of third-party web platforms, discussion forums, and hiring resources to help organizations recruit technical talent contemporaneous to our period of study. Internet Appendix B.3 also plots the time-varying presence of HR professionals on GitHub using our matched GitHub–LinkedIn dataset. The share of firms with at least one HR professional on GitHub rose steadily through the 2010s, especially among large firms, where the share went up from 2% to 8% between 2009 and 2019. These patterns suggest that GitHub was increasingly embedded in recruitment practices, particularly at larger employers.

### 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 contributions made to public GitHub repositories on their profiles. On May 19, 2016, GitHub introduced the option for users to include the number of (anonymized) private contributions to their existing contribution calendar.[5] With a click of a button, users could reveal their previously hidden private contributions at no additional cost. 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.[6] 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.

Importantly, the policy was not preannounced, and GitHub released it without a beta period or prior communication. As a result, users had no reason to anticipate the change. We exploit users' pre-policy private contributions as an exogenous measure of their work output, since these contributions were made before users could anticipate the policy change and are unlikely to reflect strategic adjustments 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, prior to this policy change, users had no credible way to disclose their number of private contributions. The immediate first-order effect of the policy was therefore a change in the visibility

[5] `https://github.blog/news-insights/product-news/more-contributions-on-your-profile/`

[6] Internet Appendix Figure B.7 illustrates how the policy changed users' GitHub contribution calendar.

of past output rather than the output itself: turning on private contributions was costless and did not modify any historical activity. This distinction underlies our empirical strategy, which relies exclusively on pre-policy private contributions that were unaffected by visibility incentives.

### 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. Revelio data provides comprehensive information regarding users' educational background and employment history (listing firms, positions held, and tenure duration). We access data from Revelio as of November 2022, while our analytical sample covers the period from April–June 2014 to April–June 2019, providing ample time for users to update their profiles.

We obtain data on publicly sourced profiles of U.S.-based GitHub users from Humanpredictions[7], a 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, and their LinkedIn profile links. We further append this dataset with users' GitHub public and private contributions from GitHub's official API (Application Programming Interface) as described in the Internet Appendix, Section B.4. Finally, we merge GitHub users with their employment details from Revelio. We are able to successfully merge over 1 million GitHub users with employment data from Revelio. 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 at the quarterly level. When an employee lists multiple concurrent jobs on LinkedIn, we retain a single primary job by excluding volunteer/part-time roles and otherwise selecting the job listed higher on the profile and/or with higher seniority, as classified by Revelio.

As a next step, we append users' employment records with firm-level identifiers and attributes. LinkedIn provides firm-page URLs with information on its size, year of incorporation, industry, and location. We merge these using a 2017 snapshot of LinkedIn firm pages to measure firm size

[7]Humanpredictions was acquired in 2024 and no longer exists as a private firm

and age close to the GitHub policy change. Additionally, we also merge the firm URLs with an auxiliary FactSet dataset (also provided by Revelio Labs) that includes a firm's unique FactSet ID, its parent firm information, and GVKEY for public firms. For a small set of unmatched firms, we supplement identifiers and attributes using Crunchbase.

Because our setting utilizes GitHub's 2016 policy change, we filter out a relevant set of GitHub users for our empirical analysis from the merged sample in two steps. First, we keep users who had joined GitHub and had made at least one GitHub contribution before the policy change on May 19, 2016. And second, we restrict to users who had completed their highest educational degree before the policy change (excluding users still studying at any point after May 2016). These filters yield an initial sample of approximately 380,000 US-based employees across 103,000 firms.

We then apply additional filters to create a consistent sample of employees for whom we can estimate individual components of productivity using the Abowd et al. (1999) (AKM) methodology. Panel A of Table 1 summarizes these steps. Starting from roughly 380,000 individuals with employment history from 2011 to 2016, the period used to measure productivity, 93% belong to the connected set – the largest group of firms where employees transition between the same organizations. Requiring at least two employees observed per firm-month allows us to estimate AKM regression coefficients for 89% of the sample. Further refinement involves focusing on employees employed at for-profit firms at the time of the policy change, which yields a final sample size of around 300,000 employees (80% of the initial sample). Within this final sample, we identify approximately 36,000 firms (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

Panels B and C of Table 1 provide summary statistics for our final sample. Panel B presents cross-sectional distributions, showing that 79% of employees are male, and 69% are white. We show further breakdowns by industry, roles, coding language, and location in the Internet Appendix Table A.1: the Software, Internet, and Information Technology sectors account for over 45% of the sample, with financial services and marketing also employing substantial GitHub talent. Classifying employees by two-digit O*NET occupation categories, we find more than 75% hold roles in computer science and mathematics. The remaining workers are concentrated in non-technology industries

but in coding-intensive positions, such as design (e.g., computer-aided designers), management (e.g., Chief Technology Officers), business and operations (e.g., analysts), and architecture and engineering. We also confirm that the results are robust within both technology and non-technology sectors in Internet Appendix Table A.2.

Internet Appendix Table A.1 Panel E and Figure A.1 present the geographic distribution of our sample: most counties in the U.S. with significant technology clusters are represented, with New York, San Francisco, and Seattle comprising 25% of the sample. On average, employees have nine years of experience and a salary of nearly $87,000. Comparatively, the Bureau of Labor Statistics (BLS) data indicates that software engineers are 78% male, 60% white, with over 15% living in New York, Seattle, or San Francisco, and earning an average salary close to $98,000.

We next study time-varying employee outcomes. The average probability of an employee moving is 7% per quarter, implying an average tenure of 3.6 years (BLS median tenure in technical services is 3.9 years). Of movers, 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 involve a rank or salary increase, while 2% of employees are promoted within the same firm each quarter.

Panel C of Table 1 reports firm-level characteristics. In our sample, 15% of firms are large, 37% are medium-sized, and 48% are small. Internet Appendix Table A.3 shows that large firms employ over 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 firms are active on GitHub, with the average firm seeing ten public contributions per employee per quarter.

### 2.7 Extracting Employee Productivity: AKM Model

We construct our measure of individual productivity from employees' 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 productivity relative to peers, factoring in the influence of the firm on employee performance. To disentangle the individual component of productivity, independent of the firm, 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 is the percentile rank of private contributions made by employee $i$ at firm $f$ during month $t$. We decompose 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 to account for differences attributable to seniority. Our regression analysis spans May 2011 to April 2016, five years before the GitHub policy change, allowing 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 employee's productivity information.

Private contributions, our 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 $log(1 + x)$ transformation. We also assess robustness to alternative AKM specifications that use categorical variables, raw contribution counts, and $log(1 + contribution)$ as the dependent variable. Further, we check robustness to productivity measure defined outside the AKM framework directly through private contributions or the share of private to total contributions. Across all approaches, our results remain consistent. Section 3 discusses these tests in the context of employee-level results.

One concern with using an AKM framework is whether the estimated worker effect captures inherent worker quality rather than firm–worker sorting. We address this with standard diagnostics and bias corrections and show that the individual component explains almost ten times as much variation in contributions as firm effects, while the firm–worker sorting component is minimal (Internet Appendix Section C.1). We also examine persistence of employee productivity across periods and find that employees remain in similar productivity quintiles when comparing pre-shock (2011–2016) to post-shock (2016–2021) estimates, consistent with the measure capturing stable differences in individual quality (Internet Appendix Figure A.2).

Beyond sorting, AKM estimates can be affected by limited mobility bias when employees switch between firms too infrequently, or noise for employees with shorter work histories and fewer

co-workers, or potential worker–firm complementarities that violate AKM's additivity assumption. We address these concerns in the Internet Appendix Section C.2 using recent approaches in the literature, including bias-aware estimators and alternative specifications that allow for worker–firm complementarities. Across these exercises, the resulting productivity estimates change only little, and our main results are not sensitive to these modeling choices.

# 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 equal to 1 if employee $i$ switches jobs in quarter $t$, and 0 otherwise. $Productivity_i$ represents the normalized individual component of private GitHub contributions ($\vartheta_i$) from Equation 1 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, $\beta$, captures the interaction between $Productivity_i$ and $Post_t$, an indicator that switches on after GitHub's policy shift in May 2016 (i.e., 2016Q2). We include individual ($\alpha_i$) and quarter ($\delta_t$) fixed effects to control for any time-invariant individual differences and time trends. We sequentially add several controls to our model ($X_{i,t}$). 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 time to account for time-varying job trends. Additionally, we control for employee's first job industry (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 compare employee transitions within the same market-quarter. The sample spans 2013Q3 to 2019Q2 (three years before and after the

GitHub policy shock in 2016Q2). Since $Productivity_i$ is itself estimated from an AKM regression, we compute standard errors via bootstrapping within employee clusters, which is the treatment level in our analysis, using 100 iterations. Each iteration re-estimates AKM productivity on a random 80% of employees in the connected set and then runs the mobility regressions using the resulting estimates. We report the average standard errors across iterations. This approach is standard in the literature (Bennedsen et al., 2019; Bhaskarabhatla et al., 2021; Best et al., 2023), and our results are insensitive to the number of iterations, changing by less than one percent even with 1,000 draws (Appendix Table A.4 Panel D).

Table 2 presents our main results. Individuals with productivity estimates higher by 1 standard deviation increase their likelihood of transitioning to larger firms (employing over one thousand employees) by 4–6% relative to the sample mean following the shock, depending on the specification. The estimates are statistically significant and consistent in magnitude across columns 1 to 5, as we progressively introduce controls from solely employee and time fixed effects to incorporating time-varying employee-level controls (salary and seniority), dynamic job-level (cohort-programming language-role-time), and market-level (industry-M.S.A-time) fixed effects.

We also 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 2016Q1 (January to March 2016), one quarter before the shock.

The event study corresponding to Table 2 Column 5 is presented in Panel A of Figure 1. While each pre-period coefficient is statistically indistinguishable from zero, there is a sudden surge in employee mobility among more productive employees after the GitHub announcement that persists for up to three years. If these employees are already likely to join larger firms, and the GitHub policy simply shifts the timing, we would expect the impact to reverse quickly. Instead, the sustained increase suggests the policy shift alters longer-term career trajectories for employees.

In the Internet Appendix, we further examine whether moves to large firms are accompanied by changes in industry and location (Table A.5). We find similar effects for within-industry and

across-industry moves; however, effects are stronger for moves that involve a location change than for within-location moves, consistent with disclosure helping workers and firms overcome weaker localized networks across geographies. We also test whether the effect reflects increased mobility among more productive employees versus reduced mobility among less productive employees by dividing the sample into binary treatment and control groups based on employee productivity (Internet Appendix Figures A.4). The results show both an increase in mobility among more productive employees and a decline among less productive ones, but the increase for more productive employees is roughly four times larger and drives the overall effect.

### 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. We address these concerns below.

#### Omitted Variable Bias

First, we check whether our findings simply reflect time-varying trends among more productive employees by conducting falsification tests using *placebo* measures of employee productivity available before the shock. If such trends were driving the results, they should show up for productive employees even under pre-shock observable productivity measures, and not just private contributions that were unavailable before the shock. We test this hypothesis in Panel A of Table 3 using four *placebo* measures of observable employee productivity.

Column 1 presents our baseline results with private contributions for comparison. Column 2 uses employees' pre-shock public contributions (which were visible before the shock) as the placebo measure. We exclude contributions to employees' own repositories to separate professional activity from personal projects. As in our baseline, an AKM decomposition isolates the individual component of public contributions. Column 3 employs the percentile rank of the total stars received on employees' public repositories, Column 4 incorporates the logarithm of employees' salaries[8], and Column 5 uses the percentile rank of the number of articles authored by employees (posted on

[8]Revelio estimates salaries using a model trained on over 200 million observations based on role, level, firm, and location, adjusted for inflation. This aligns with our setting, as external employers similarly observe these characteristics but do not know workers' exact salaries.

Google Scholar, DevTo, and ORCiD) – all during the pre-shock period. Across Columns 2–5, we find no economically or statistically significant changes in post-shock mobility for employees with one standard deviation higher placebo productivity, in sharp contrast to Column 1.

Additionally, we examine whether broader trends correlated with the 2016 presidential elections or local economic conditions could explain our findings. If that were the case, we would expect to find different impacts in areas based on their political leanings or economic characteristics. However, Internet Appendix Table A.6 rules this out by showing a similar effect for states irrespective of their political leaning, and no differential impact on counties that were more white, richer, or unequal.

As a final validation against concurrent trends, we study stock-price reactions around the GitHub announcement using 1,675 publicly listed firms in our sample.[9] We categorize firms along two dimensions: average employee productivity (top or bottom quartile) and firm size (with indicators for the top and bottom quartiles), and find a 71-basis-point decline in returns for small, high-productivity firms (Internet Appendix Table A.7), suggesting that the market anticipates potential talent loss in these firms.[10] 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. Placebo tests using 100 random dates show similar or more negative coefficients in fewer than 5% of cases (Internet Appendix Figure A.5), reinforcing that the observed effects are likely driven by the GitHub announcement rather than random fluctuations.

**Selection**

One may worry about selection concerns if movers differentially choose to display private contributions, or if some firms restrict employees from displaying them. To address these issues, we re-estimate our analysis in sub-samples designed to shut down these concerns: 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 are presented in Table 3, Panel B.

First, we repeat our test on the sub-sample of users with non-zero private contributions (Columns 1–2), restricting to users with at least one or ten pre-period private contributions. This within-sample analysis reduces concerns that the baseline results are driven by selective disclosure

[9] Bloomberg does not report any major tech-related news announcements between 17 May 2016 and 20 May 2016.
[10] Firms in the bottom quartile have fewer than 1,200 employees, close to the 1,000 cutoff used elsewhere to distinguish medium from large firms, while firms in the top quartile have more than 18,000 employees.

among prospective movers. Although these restrictions substantially reduce the sample size, our results remain robust in both magnitude and statistical significance.

Next, in Columns 3–4, we focus on a sub-sample where employees, as of the shock date, work at firms that have at least one or ten pre-shock private contributions. We find even stronger results in this subsample, negating concerns of bias stemming from firms not utilizing private contributions before the shock or restricting employees from disclosing private contributions after the shock.

In Column 5, we restrict our sample to moves and study the differential likelihood of more productive employees transitioning to large firms after the shock. If our baseline results were driven solely by movers selectively revealing private contributions, restricting the sample to moves should attenuate the effect. Column 5 confirms that following the shock, movers with higher private contributions were significantly more likely to transition to larger firms.

A final critique is that transitory shocks could simultaneously affect both individual productivity and mobility. To address this, we estimate employee productivity separately before and after the GitHub shock. Internet Appendix Figure A.2 shows that individual productivity remains highly persistent, with most employees remaining in the same productivity quintile across both periods. Internet Appendix Table A.8 further narrows the analysis to users with consistent productivity levels across the pre- and post-shock periods, and the results remain robust, suggesting the effects are not driven by transitory fluctuations in productivity.

**Robustness of AKM estimates**

Beyond selection concerns, we evaluate the robustness of our results to alternative ways of measuring employee productivity, particularly our choice of the AKM framework. First, we replicate our baseline tests using productivity measures that do not rely on AKM (Internet Appendix Table A.9). We use each employee's raw percentile rank in private-contribution number, and the share of private contributions. The latter also mitigates concerns that the results reflect differences in working style (e.g., many small versus few large contributions). In both cases, the estimates remain economically and statistically significant.

Next, in Internet Appendix C, we implement a battery of robustness checks around the AKM decomposition to address common critiques of the standard AKM model and our choice of dependent-variable transformation. As discussed in Section 2.7, we consider issues emphasized

in the AKM literature—including limited mobility bias, noise for workers with shorter histories or fewer co-workers, and potential worker–firm complementarities—and show that our results are not sensitive to these modeling choices. We also address concerns about continuous-treatment difference-in-difference designs raised by Callaway et al. (2024) by showing similar results along with parallel trends assumption continuing to hold when using a binary indicator.

Internet Appendix Table A.4 presents additional tests to specification choices in our analysis. We consider alternative samples, including matched samples based on demographics, role, and seniority (with and without public contributions) and one excluding Microsoft employees, given its 2018 acquisition of GitHub. We also examine sensitivity to the estimation window for AKM and incorporate more granular fixed effects (interacting all baseline controls) and add controls for public productivity information, such as public contributions, repository stars, and publications. In addition, we report results with both 100 and 1000 bootstrap iterations to calculate standard errors. Across all specifications, the main findings are robust.

Internet Appendix Table A.2 breaks down the sample by firm industry and location. The results are similar in both technology and non-technology sectors, as well as in technology hubs and other locations. This pattern is consistent with the GitHub shock affecting all industries that employ tech talent, rather than being specific to technology firms or sectors.

### 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 4, reveal that individuals with productivity estimates higher by 1 standard deviation reduce 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 Panels B and C of Figure 1 validate the absence of pre-trends in both those outcomes pre-treatment.

In the Internet Appendix (Figure A.3 and Table A.10), we further show that the effect increases almost monotonically with destination firm size, from -6% of the sample mean for firms with 11–50 employees to +6.8% for firms with more than 10,000 employees. The exception is firms with fewer

than 10 employees, where the effect is small (0.4% of the mean) and statistically insignificant, as many of these firms likely consist of founders, rather than regular employees.

Building further on this size gradient, Internet Appendix Table A.11 shows that the effect is also strongly non-linear in worker productivity: mobility to large firms rises with employee productivity and is largest in the upper tail (e.g., a 32% increase relative to the mean for the top 1% employees versus 20% for the top quartile). Further tests in Internet Appendix Table A.12 reveal that the results are consistent with the gains accruing to "superstar" employers – the top 1% or 10% firms by market capitalization or revenue (as defined in Tambe et al. (2020)). Finally, decompositions by firm age indicate reduced mobility into younger firms (less than 10 years since incorporation) and increased mobility into older ones. Even within startups, gains are concentrated among the largest, highly funded startups (Internet Appendix Table A.13).

### 3.3 Heterogeneity by Initial Firm Size

We next identify the origin firms for these employees to establish patterns of labor reallocation. We analyze the post-GitHub policy change in the mobility of productive employees to large firms by the size of the employee's prior firm through a triple-difference specification:

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

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. We also estimate a dynamic version

of Equation 4 to verify the absence of pre-trends.

Column 2 of Table 5 presents the results. We find that mobility to large firms for individuals with productivity estimates higher by 1 standard deviation, who were already employed at large firms, decreases by 7% of the sample mean after the GitHub announcement. Conversely, the probability of individuals with 1 standard deviation higher productivity moving to large firms increases by 3.2% and 4.9% of sample mean after GitHub's policy change for those currently employed at medium and small firms, respectively. The absence of pre-trends, when plotting the triple-difference coefficient for both mid-sized and small firms in Figure 2, underscores that we are capturing the impact of the change in GitHub policy.

Our results indicate that increased public information on employee productivity has heterogeneous effects by firm size. Productive employees at small and medium firms migrate to larger firms, generating substantial reallocation of talent after the GitHub policy change, while inter-firm moves among productive employees already at large firms decline. These patterns can be reconciled by two forces: efficient sorting and internal labor markets.

Greater transparency improves sorting: high-quality workers at smaller firms can signal ability and move to large firms, where scale and complementarities raise their marginal product. For high-quality employees already at large firms, however, the gains from moving are smaller as their marginal product is already high. However, these workers may not typically earn their full marginal product because firms strategically delay promotions to avoid revealing worker quality (Waldman, 1984; Bernhardt and Scoones, 1993). After the GitHub policy change, reduced information asymmetry limits this withholding, letting large firms better match workers' marginal product and reducing large-to-large mobility. Within-firm promotions provide a natural test of this mechanism.

We study firm promotions to verify if efficient sorting and internal labor markets can indeed explain our results. Under this hypothesis, large firms should attract productive employees from medium and small firms with larger promotions, while also retaining their most productive incumbents through internal promotions once signaling frictions decline. We present these results in Columns 3–8 of Table 5, dissecting 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.

Column 3 demonstrates that the likelihood of transitioning to a large firm with an increase in salary and title rises by 5.8% above the mean for employees with 1 standard deviation higher productivity, following the GitHub announcement. Column 4 shows that this increase is mainly driven by employees from medium and small-sized firms. Specifically, the probability of large-to-large moves with promotion decreases by 6.9% of the sample mean, while the probability of transitioning from small and medium firms to large firms with promotion increases by 5.1% of the sample mean. These findings suggest that large firms attract productive employees from small and medium firms by offering improved salary and title, after the GitHub change.

In contrast, Column 5 reports that the moves to small or medium firms with promotion decreases by 4.5% of the sample mean for individuals with 1 standard deviation higher productivity, after the GitHub announcement. Column 6 reveals that this decrease is mainly attributed to employees from large firms. Specifically, moves from large-to-small or large-to-medium firms with promotion decreases by 11.2% of the sample mean, while there is a 2.2% change in the probability of moving from one small and medium firm to another with promotion.

Column 7 notably illustrates that current employers also react to the disclosure of new productivity information as within-firm promotions for employees with 1 standard deviation higher productivity increases by 5.8% over the mean post-shock. The corresponding dynamic event study in Figure 3 finds no pre-trends. Column 8 shows the effect is stronger in large firms, where within-firm promotion probability rises by 7.5% post-shock versus 5.1% in small firms.

### 3.4 Impact on Productivity

We also explore whether these career upgrades are associated with a change in employee productivity. We check whether productive employees experience a larger post-shock change in productivity (Internet Appendix Table A.14). We use public contributions as a measure for productivity, as these were not directly influenced by the shock. We exclude contributions made by employees to their own repositories, as these may represent personal activity. We find a 2% post-shock increase in percentile rank over the mean in public contributions for employees with 1 standard deviation higher pre-shock private-contribution productivity. These results are driven by

changes in the probability of experiencing very high (top decile) contributions.

Overall, the disclosure of productivity information on GitHub benefits more productive employees by improving both performance and career progression: large firms retain top workers through internal promotions, while small firms lose valuable talent to large firms offering better career opportunities.

# 4 Potential Mechanisms

## 4.1 Heterogeneity by Individual Characteristics

The existing literature highlights substantial information asymmetry surrounding employee quality (Greenwald, 1986), which implies that, absent direct productivity indicators, firms rely on hard metrics such as experience and education for recruitment (Spence, 1973; Waldman, 1984). We examine this mechanism by exploring the heterogeneity of our results across three individual metrics: work experience, years of schooling, and elite education in Table 6.

Column 1 shows that productive employees with below-median experience at the time of the shock (7 years) experience a 1.7× larger post-shock increase in mobility to large firms than those with above-median experience. Column 2 shows a 2× larger increase for productive employees with below-median schooling than for those with above-median schooling. Column 3 shows that productive employees without an elite degree (top 20 universities per U.S. News and World Report) experience a 2.7× larger mobility increase to large firms than those with an elite degree.

Taken together, our results support the hypothesis that web-based platforms reduce reliance on hard metrics, such as experience and education, and network-based screening, leading to a "democratization" of screening in the labor market.

## 4.2 Heterogeneity by Firm Characteristics

We next examine which firm- and job-side characteristics are associated with productive workers' transitions to large firms. Larger firms may have better IT resources (Eckel and Yeaple, 2017), helping them utilize and recruit based on GitHub information. Large firms also offer steeper career paths (Di Porto et al., 2024). A preference for such steeper career paths may drive more productive employees toward larger firms. We test these hypotheses using heterogeneity tests (Table 7).

First, we examine tech-savvy recruiting. Using our matched GitHub–LinkedIn dataset, we

identify HR professionals active on GitHub; by 2016, over 6% of large firms had at least one such recruiter (Internet Appendix B.3). We classify large firms based on whether their number of GitHub-active recruiters in 2016 is above or below the median. Columns 1–2 of Table 7 show that employees with productivity 1 standard deviation above the mean increase mobility to large firms with above-median GitHub-active recruiters by 7.7% of the sample mean post-shock, compared to 2.1% for large firms with below-median recruiters on GitHub (difference significant at 1%).

Second, we proxy for career-path steepness using average annual salary change at a firm level from LinkedIn data for all employees in the same job roles as those in our GitHub sample during the pre-shock period (2013Q3–2016Q2). 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 7, Columns 3–4 show a 6.5% increase in mobility to large firms with above-median salary growth versus 2.2% for below-median firms (difference significant at 1%), consistent with faster career growth attracting productive employees.

Finally, we explore whether productive workers sort into large firms with higher growth potential. In Column 5–6 of Table 7, we split large firm destinations by ex-ante employment growth, measured using pre-period LinkedIn data. We find individuals with 1 standard deviation higher productivity increase mobility to above-median growth large firms by 7% of the sample mean after the GitHub policy change, compared to 4.8% for below-median growth large firms (difference significant at the 5% level). In Columns 7–8, we repeat the split using Compustat-based pre-period sales growth: mobility to above-median sales-growth large firms increases by 8% of the sample mean versus 3.9% for below-median sales-growth firms (difference significant at the 1% level).

# 5 Firm Level Impact

Section 3 shows that high-productivity employees reallocate from small and medium-sized firms to large firms following the GitHub productivity disclosure. We next assess whether this labor reallocation has an aggregate impact on firm outcomes. If smaller firms can readily replace departing workers or reallocate work internally, the disclosure should have little impact on firm growth or productivity; if not, we should observe an aggregate effect.

We examine the impact of the GitHub policy change on firm growth and productivity using firm-level aggregates and a triple-differences design. The basic difference-in-differences setup

compares 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. The third difference interacts this effect with firm size, allowing us to compare the impact of the policy on treated small and medium firms relative to treated large firms. We estimate the following regression equation:

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

The dependent variable is an outcome for firm $j$ in quarter $t$. The coefficient $\beta_1$ captures the interaction of $\mathbb{1}(Productivity_j > Median)$, an indicator that takes the value 1 if the average AKM-based employee productivity at the time of GitHub's policy change is above-median for firm $j$, and $\mathbb{1}(Post_t)$, an indicator that switches on after the policy (i.e., after 2016Q2). Coefficients $\beta_2$ and $\beta_3$ further interact these terms with indicators for firm size. Our sample covers a period of six years: 2013Q3 to 2019Q2, three years before and after the shock. We include firm and industry×state×time fixed effects, controlling for constant firm-level differences and market-level time trends in our results. Additionally, we add outcome-specific controls: lagged employment for employment outcomes, and lagged total contribution stock (depreciated at 15% annually) for contribution-based productivity metrics. We cluster our standard errors at the firm level.

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

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

The model is similar to Equation 5, except it estimates a separate $\beta_\tau$ for each quarter $\tau$ relative to the shock. The omitted quarter is 2016Q1 – 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 8. Columns 1 reports firm employment growth: small firms with above-median individual productivity grow 17% slower than the mean after the GitHub announcement.[11] Column 2 examines a proxy for firm exit, defined as

[11] A maximum of 1.5% of our sample is acqui-hires (where more than 80%) of the team moves to a new firm.

a headcount reduction of more than half relative to the previous quarter. We find that small firms with above-median productivity witness a substantial 56% increase over the mean in the probability of exit after the GitHub announcement.[12] Panels A and B in Figure 4 demonstrate that our results are not driven by differential pre-trends across firms.

Columns 3 to 6 use productivity outcomes, measured using public contributions for firms' employees (excluding contributions to employees' own repositories), as these were not directly influenced by the shock. We find declines in public contributions among small firms with above-median productivity after the shock: a 6% reduction in percentile rank and a 20% decline in the probability of top-decile contributions (relative to the mean). Public contributions per employee similarly fall (Column 5–6), with a 6% decline in percentile rank and a 23% reduction in the probability of being in the top decile. Panels C and D in Figure 4 display the corresponding event studies, with no discernible pre-trends.

The absence of strong positive effects for large firms with higher GitHub contributions is consistent with all large firms broadly benefiting from the influx of talented labor, regardless of their own GitHub usage. Instead, small firms reliant on GitHub lose critical talent and struggle to offset these losses, leading to declines in both growth and productivity.

A final question concerns the magnitudes of these effects. At the employee level, our results show that the shock increases the reallocation of productive employees from small to large firms by about 5% of the mean mobility. 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.15. Following the shock, small firms with above-median productive employees see the share of productive employees fall by about 1 percentage point per quarter–roughly one-third of the pre-shock mean–driven by employee exits from the top productivity quintile. These results underscore the critical role of the most productive employees within firms. A one-third decline in their share coincides with a 17% decline in firm growth and about a 6% decline in productivity, both relative to 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 firm growth and productivity by size. We next examine

[12]The magnitude is large because the pre-period mean of exits is only 0.47%, particularly because large firms rarely exit. The average exit among small firms is 0.82%, implying that the shock increases exit probability by 32%.

industry dynamics, asking whether employment concentration changes more in the industries most exposed to GitHub's policy. To do so, we aggregate our data to the industry level and relate changes in concentration to industry-level private contributions. Despite debate over local labor market concentration, LinkedIn data for workers in our sample show that the software developer labor market operates largely at the national level. Among job switchers in the three years before the shock (2013–2016), 35% relocated across states. We measure change in concentration by the difference in industry's Herfindahl-Hirschman Index (HHI) over two periods: five years before the shock (2011–2016) and five years after (2016–2021). We use the average number of private contributions in each industry as our independent variable. While these correlations do not establish causality, they are informative given the broader rise in industry concentration over the past decade (Autor et al., 2023).

The results, in Internet Appendix Table A.16, show that industries with greater exposure to GitHub private contributions experience larger post-shock increases in labor market concentration. A one standard deviation increase in exposure is associated with roughly a 6–7% higher rise in HHI under LinkedIn industry definitions and about a 14% higher rise under SIC (3-digit) definitions when comparing 2016–2021 to 2011–2016. These magnitudes suggest that productivity disclosure, and the resulting talent reallocation, may be linked to broader shifts toward employment concentration in the industries where such information is most salient.

## 6 Conclusion

This paper represents one of the first examinations of how individual-productivity-related big data affects labor distribution within the economy. We study 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 shift: improved productivity disclosure triggers the reallocation of talented individuals toward larger firms. Smaller and medium-sized firms face challenges in both recruiting and retaining talent. In contrast, large firms retain existing talent through internal promotions and attract new talent.

However, this redistribution has adverse effects on the growth and productivity of smaller

firms. Correlational evidence also suggests that this reallocation increases market concentration in industries with greater informational disclosure facilitated by GitHub.

Our findings unearth a novel labor channel for big data: increased employee productivity information reallocates labor from small to large firms, amplifying the significance of large firms in the economy. The overall welfare consequences 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., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, reallocation, and growth. *American Economic Review*, 108(11):3450–91.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2023). A theory of falling growth and rising rents. *Review of Economic Studies*, 90(6):2675–2702.

Akcigit, U. and Goldschlag, N. (2023). Where have all the "creative talents" gone? employment dynamics of us inventors. Technical report, 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(3):488–500.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709.

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

Bennedsen, M., Tsoutsoura, M., and Wolfenzon, D. (2019). Drivers of effort: Evidence from employee absenteeism. *Journal of Financial Economics*, 133(3):658–684.

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

Bernstein, S. (2015). Does going public affect innovation? *The Journal of Finance*, 70(4):1365–1403.

Bessen, J. (2020). Industry concentration and information technology. *The Journal of Law and Economics*, 63(3):531–555.

Best, M. C., Hjort, J., and Szakonyi, D. (2023). Individuals and organizations as sources of state effectiveness. *American Economic Review*, 113(8):2121–2167.

Bhaskarabhatla, A., Cabral, L., Hegde, D., and Peeters, T. (2021). Are inventors or firms the engines of innovation? *Management Science*, 67(6):3899–3920.

Bradshaw, R. (2024). Apollo technical research, surprising social media recruiting statistics. `https://www.apollotechnical.com/social-media-recruiting-statistics/`.

Brynjolfsson, E., Jin, W., and Wang, X. (2023). Information technology, firm size, and industrial concentration. Technical report, National Bureau of Economic Research.

Brynjolfsson, E. and McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5):133–139.

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

Chen, J. and Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, 139(2):891–936.

Conti, A., Gupta, V., Guzman, J., and Roche, M. P. (2023). Incentivizing innovation in open source: Evidence from the github sponsors program. Technical report, National Bureau of Economic Research.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2020). Changing business dynamism and productivity: Shocks versus responsiveness. *American Economic Review*, 110(12):3952–90.

Di Porto, E., Pagano, M., Pezone, V., Saggio, R., and Schivardi, F. (2024). Careers and wages in family firms: Evidence from matched employer-employee data. *Unpublished Manuscript*.

Eckel, C. and Yeaple, S. R. (2017). Too much of a good thing? labor market imperfections as a source of exceptional exporter performance. Technical report, National Bureau of Economic Research.

Eisfeldt, A. L. and Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns. *The Journal of Finance*, 68(4):1365–1406.

El-Komboz, L. A. and Fackler, T. (2023). Productivity spillovers among knowledge workers in agglomerations: Evidence from github. *Available at SSRN 4600283*.

Floyd, E., Tomar, S., and Lee, D. J. (2024). Making the grade (but not disclosing it): How withholding grades affects student behavior and employment. *Management Science*, 70(4):2497–2517.

Gibbons, R. and Katz, L. F. (1991). Layoffs and lemons. *Journal of labor Economics*, 9(4):351–380.

Giroud, X., Liu, E., and Mueller, H. (2024). Innovation spillovers across u.s. tech clusters. Working Paper 32677, National Bureau of Economic Research.

Goldstein, I., Spatt, C. S., and Ye, M. (2021). Big data in finance. *The Review of Financial Studies*, 34(7):3213–3225.

Gompers, P., Lerner, J., and Scharfstein, D. (2005). Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999. *The Journal of Finance*, 60(2):577–614.

Greenwald, B. C. (1986). Adverse Selection in the Labour Market. *The Review of Economic Studies*, 53(3):325–347.

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

Idson, T. L. and Oi, W. Y. (1999). Workers are more productive in large firms. *American Economic Review*, 89(2):104–108.

Jiang, W. and Li, T. (2024). Corporate governance meets data and technology. *Foundations and Trends in Finance*, 14(2):61–136.

Kahn, L. B. (2013). Asymmetric information between employers. *American Economic Journal: Applied Economics*, 5(4):165–205.

Kremer, M. (1993). The o-ring theory of economic development. *The Quarterly Journal of Economics*, 108(3):551–575.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 9(2):508–523.

Montandon, J. E., Valente, M. T., and Silva, L. L. (2021). Mining the technical roles of github users. *Information and Software Technology*, 131:106485.

Rockoff, J. E., Staiger, D. O., Kane, T. J., and Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *American Economic Review*, 102(7):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(3):355–374.

Tambe, P., Hitt, L., Rock, D., and Brynjolfsson, E. (2020). Digital capital and superstar firms. Technical report, National Bureau of Economic Research.

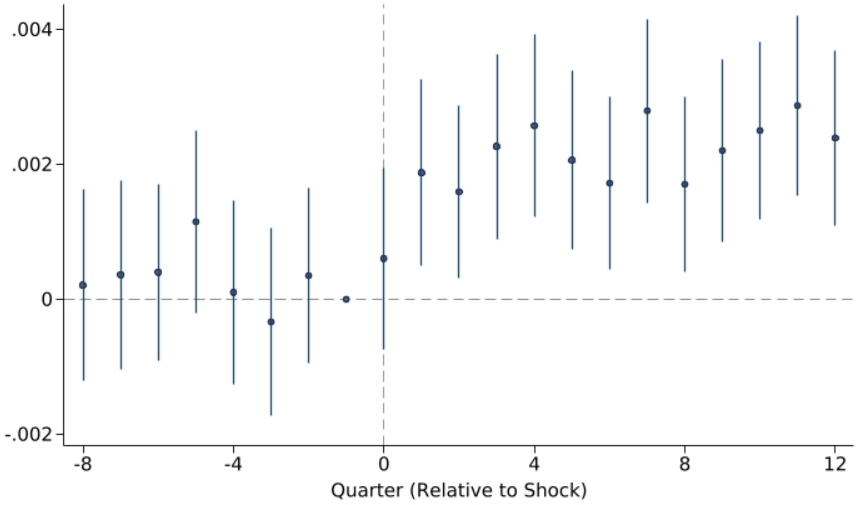
Tambe, P. and Hitt, L. M. (2012). The productivity of information technology investments: New evidence from it labor data. *Information systems research*, 23(3-part-1):599–617.

Tate, G. and Yang, L. (2023). The human factor in acquisitions: Cross-industry labor mobility and corporate diversification. *The Review of Financial Studies*, 37(1):45–88.

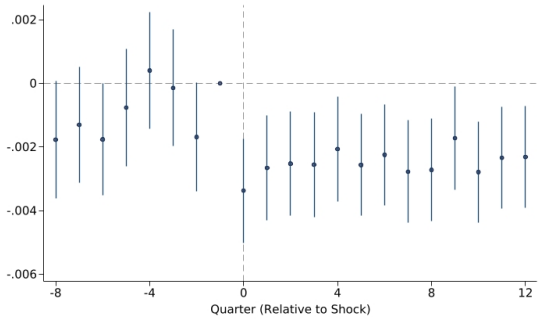
Waldman, M. (1984). Job assignments, signalling, and efficiency. *The RAND Journal of Economics*, 15(2):255–267.

Figure 1: **GitHub Shock & Employee Mobility to Large, Medium, & Small Firms**

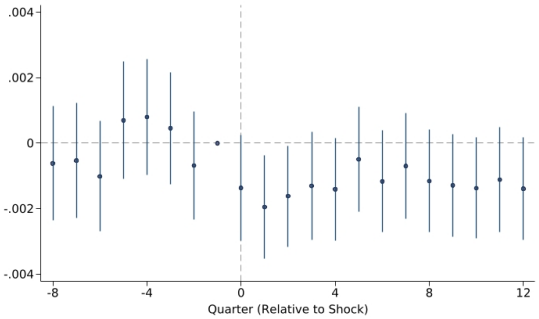
This figure presents an event study on employee mobility to large, medium, and small firms around the disclosure of employee productivity information on GitHub. The x-axis plots the quarter around the shock. The dashed line indicates the quarter 2 (April-June) of 2016, during which GitHub introduced 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 from Equation 3. The figure plots the effect corresponding to job switch to a large firm (>1000 employees) in Panel (a), a medium-sized firm (50 < employees <= 1000) in Panel (b), and a small firm (<= 50 employee) in Panel (c). We control for employee fixed effects, time-varying employee salary and seniority, as well as fixed effects for the interaction of employee cohort, role, programming language, and time, and a three-way interaction of industry, location, and time. Standard errors are bootstrapped within employee clusters using 100 repetitions. Light blued lines show 95% confidence intervals for estimates.




(a) Move to Large Firms




(b) Move to Medium-Sized Firms




(c) Move to Small Firms

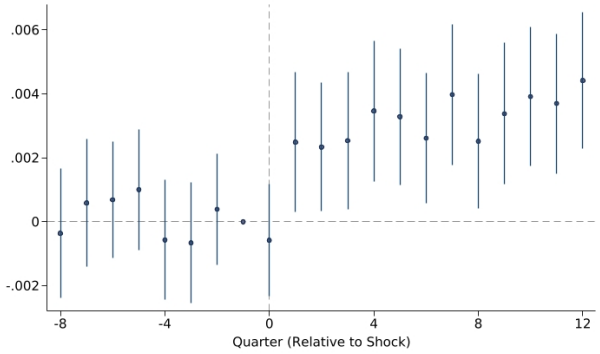
Figure 2: **GitHub Shock & Employee Mobility to Large Firms: Heterogeneity by Initial Firm Size**

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

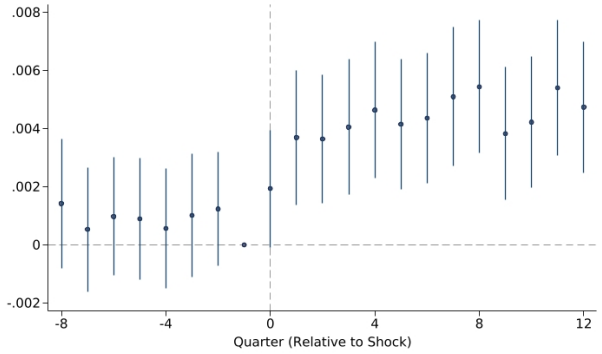


Figure 3: **GitHub Shock & Employee Promotions at Large Firms**

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.

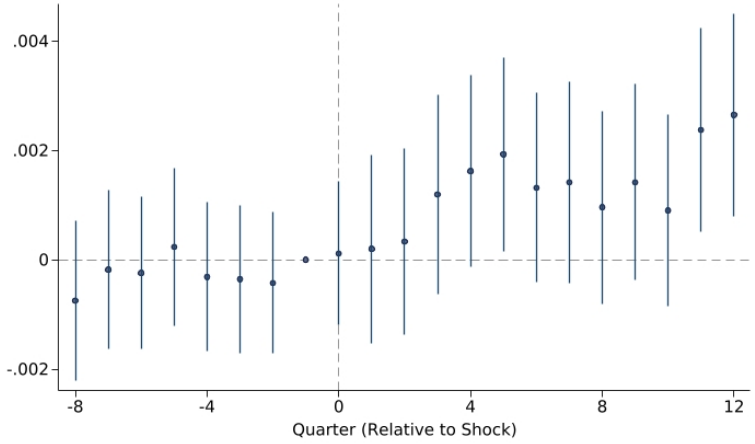
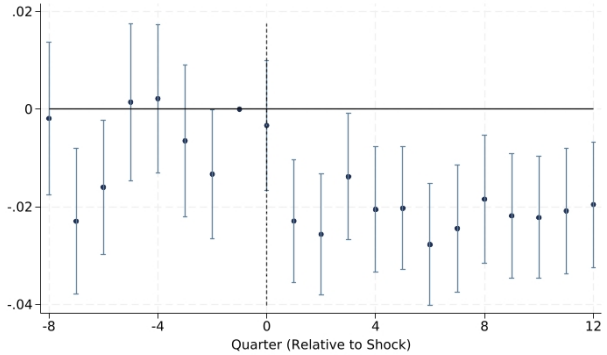
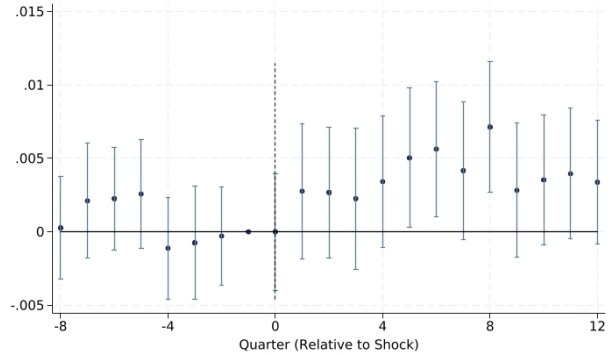


Figure 4: **GitHub Shock & Employment and Productivity at Small Firms**

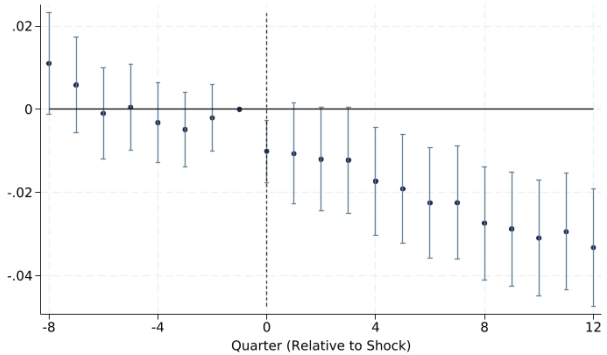
This figure presents an event study on firm-level employment growth and productivity for small firms (<= 50 employees) around the disclosure of employee productivity information on GitHub. The x-axis plots the quarter around GitHub's policy change. The dashed line indicates the quarter 2 (April-June) 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 on small firms with above-median employee productivity estimated by the $\beta$ coefficients from Equation 5. We plot the effect on firms' QoQ employment change (in %) in Panel (a) and the likelihood of firms' total employment dropping by more than 50% in Panel (b). In Panel (c) (Panel (d)), we plot the likelihood of firms' total public contributions (total public contributions per employee) being in the top decile. Firm contributions only include public contributions made by incumbent employees towards repositories not owned by the employees themselves. We control for firm and industry × state interacted with time fixed effects in all panels. Additionally, we control for firms' time-varying lagged employment in Panels (a) and (b), and firms' time-varying lagged stock of total contributions (depreciated at 15% annually) in Panels (c) and (d). We cluster the standard errors at the firm level. Light blued lines show 95% confidence intervals for estimates.



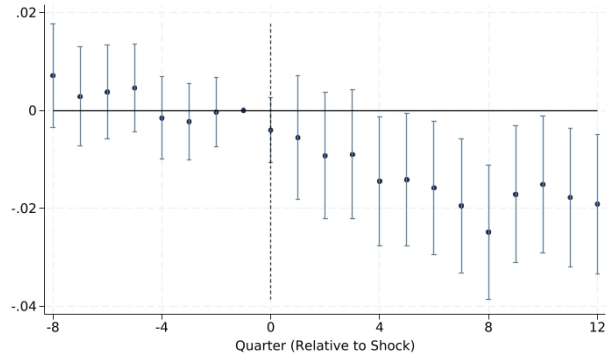

(a) Employment Growth




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




(c) Public Contributions




(d) Public Contributions per Employee

Table 1: **Sample Breakdown & Descriptive Statistics**

This table presents the details of our sample construction and the descriptive statistics of our key variables. Panel A shows the number of employees 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 using the AKM (Abowd et al., 1999) decomposition (Equation 1), and the final sample remaining after excluding users working in non-profit industries at the time of GitHub's policy change. Panel B (Panel C) shows descriptive statistics for employee-level (firm-level) variables.

**Panel A: Sample Breakdown for AKM Test**

| | 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: 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 C: 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: **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 3: **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 2, 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 4: **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 5: **GitHub Shock, Employee Mobility & Promotions : 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 and employee career advances 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:

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

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. In columns 1 & 2, 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. The dependent variable is an indicator variable that is 1 if employee $i$ switches job to a large firm (>1000 employees) in columns 1 & 2. In columns 3-8, the dependent variable captures a career promotion indicator (salary or job-seniority increasing compared to the previous quarter) along with the employee switching job to a large firm (> 1000 employees) in columns 3 and 4, switching job to a small or mid-sized firm (<= 1000 employees) in columns 5 and 6, and remaining at the same firm in columns 7 and 8. 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 graduation cohort, job role, programming language, and time, as well as a three-way interaction of employee's industry, location, and time. In columns 7 and 8, 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) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| **Outcome:** | **𝟙(Job Change to Large Firms)** | | **𝟙(Promotion)** | | | | | |
| | | | *Move to Large* | | *Move to Small & Med.* | | *Within-Firm* | |
| Productivity × 𝟙(Post) | 0.0017*** | -0.0023*** | 0.0014*** | -0.0016*** | -0.0016*** | -0.0039*** | 0.0010*** | 0.0014** |
| | (0.0002) | (0.0005) | (0.0001) | (0.0004) | (0.0003) | (0.0006) | (0.0003) | (0.0006) |
| Productivity × 𝟙(Post) × 𝟙(Mid) | | 0.0033*** | | 0.0026*** | | 0.0039*** | | 0.0001 |
| | | (0.0006) | | (0.0005) | | (0.0007) | | (0.0007) |
| Productivity × 𝟙(Post) × 𝟙(Small) | | 0.0038*** | | 0.0028*** | | 0.0047*** | | -0.0005 |
| | | (0.0006) | | (0.0005) | | (0.0008) | | (0.0007) |
| Employee F.E. | Y | Y | Y | Y | Y | Y | Y | Y |
| Time F.E. | Y | Y | Y | Y | Y | Y | Y | Y |
| Salary & Seniority Controls | Y | Y | Y | Y | Y | Y | Y | Y |
| Cohort × Language × Role × Time F.E. | Y | Y | Y | Y | Y | Y | Y | Y |
| Industry × Location × Time F.E. | Y | Y | Y | Y | Y | Y | Y | Y |
| Firm Cohort | | | | | | | Y | Y |
| Observations | 3,199,173 | 3,022,682 | 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 | 143,446 | 142,754 | 143,022 | 142,797 |
| R-squared | 0.212 | 0.227 | 0.204 | 0.216 | 0.212 | 0.22 | 0.274 | 0.275 |
| Y-Mean | 0.033 | 0.032 | 0.024 | 0.023 | 0.036 | 0.035 | 0.019 | 0.019 |
| Magnitude (Main) (%) | 5.23 | -7.05 | 5.76 | -6.91 | -4.52 | -11.16 | 5.84 | 7.5 |
| Magnitude (Small) (%) | | 4.86 | | 5.14 | | 2.21 | | 5.11 |

Table 6: **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 7: **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 (7-8), the subset is based on whether the average employment (sales) growth at the firm in the pre-period is high, i.e., above median, or low. $\alpha_i$ 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 Large Firms with)** | | | | | | | |
| | **Tech-savvy Recruiters** | | **Steep Career Growth** | | **High Employment Growth** | | **High Sales Growth** | |
| | Yes | No | Yes | No | Yes | No | Yes | No |
| Productivity × 𝟙(Post) | 0.0014*** | 0.0003*** | 0.0016*** | 0.0001 | 0.0008*** | 0.0010*** | 0.0011*** | 0.0002*** |
| | (0.0002) | (0.0001) | (0.0002) | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| Employee F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Salary & Seniority Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort × Language × Role × Time F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry × Location × Time F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 3,199,173 | 3,199,173 | 3,192,168 | 3,192,168 | 3,191,534 | 3,191,534 | 3,152,320 | 3,152,320 |
| # Unique Employees | 143,446 | 143,446 | 143,445 | 143,445 | 143,445 | 143,445 | 143,441 | 143,441 |
| R-Squared | 0.189 | 0.23 | 0.2 | 0.221 | 0.18 | 0.221 | 0.19 | 0.217 |
| Y-Mean | 0.019 | 0.015 | 0.024 | 0.007 | 0.011 | 0.02 | 0.014 | 0.005 |
| Magnitude (%) | 7.69 | 2.09 | 6.48 | 2.02 | 7.03 | 4.85 | 8.02 | 3.86 |
| T-test ($H_0 : \beta_{Yes} = \beta_{No}$) | (0.0000)*** | | (0.0000)*** | | (0.0129)** | | (0.0003)*** | |

## Table 8: **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 |

# Big Data and Bigger Firms: A Labor Market Channel

## Internet Appendix

**Abhinav Gupta** **Naman Nishesh** **Elena Simintzi**

March 24, 2026

# Internet Appendix A

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

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

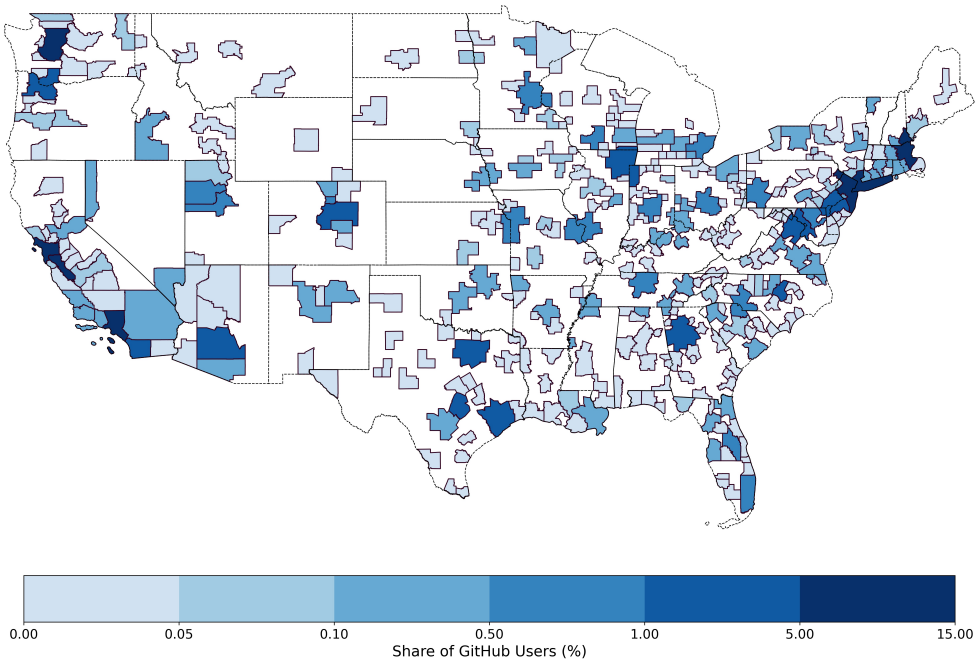
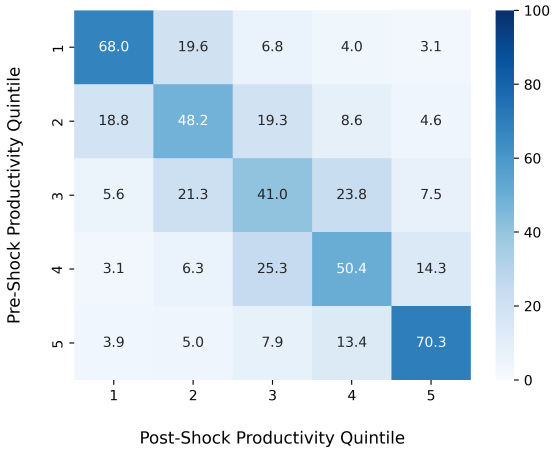
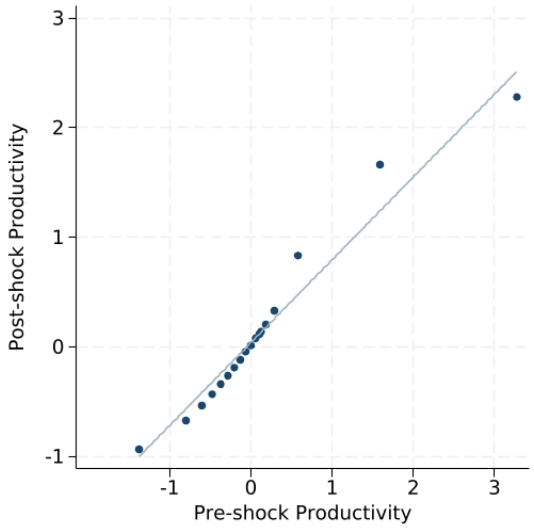


Figure A.2: **Persistence of Employee Productivity: Evidence from AKM Decomposition Before and After Shock**

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




Panel A




Panel B

Figure A.3: **GitHub Shock & Employee Mobility to 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.



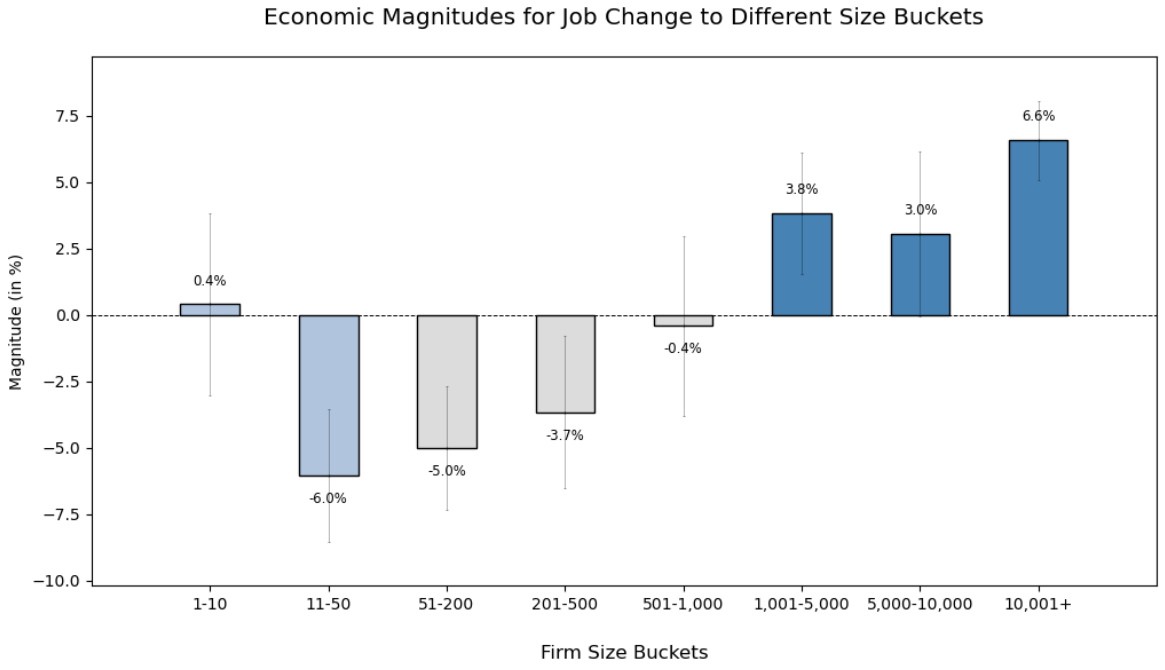


Figure A.4: **GitHub Shock & Employee Mobility to Large Firms: Breakdown by Treated and Control Group**

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. 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.

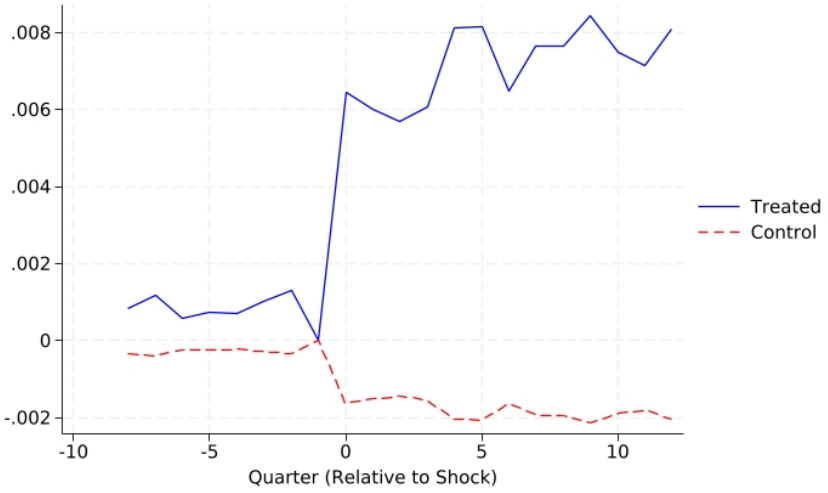


Figure A.5: **Firm Market Reaction: Placebo Test**

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.7. 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.7.

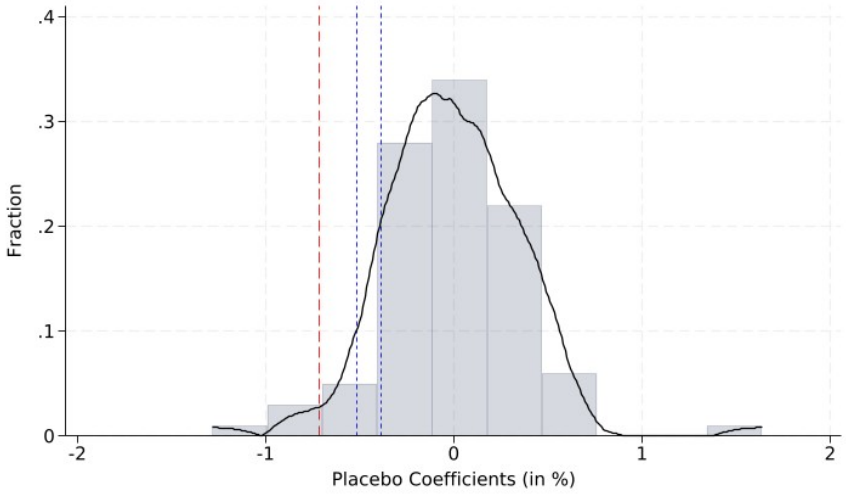


Table A.1: **Sample Characteristics**

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**

| Occupation | 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**

| Occupation | 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: **GitHub Shock & Employee Mobility in Tech vs Non-Tech Sectors**

This table presents estimates on the impact of GitHub's disclosure of employee productivity information on employee mobility towards large firms in the technology sector versus the non-technology sector. The specification is similar to the one reported in column 5 of Table 2. In column 1 (2), we estimate the effects on a restricted sample of employees who work in the technology (non-technology) industry at the time of GitHub's policy change. We categorize employees into the technology industry if their listed industry on LinkedIn is one of the following: Computer Software, Internet, Information Technology & Services, Telecommunications, Semiconductors, Computer Networking, Computer Games, Computer & Network Security, Computer Hardware, Information Services, and Nanotechnology. In column 3 (4), we restrict the sample to employees who live (do not live) in technology hubs at the time of GitHub's policy change. We categorize employees as living in a technology hub if their listed MSA on LinkedIn is either in California or Massachusetts. $\alpha_i$ 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) |
|---|---|---|---|---|
| **Output:** | **𝟙(Job Change to Large Firm)** | | | |
| | **Tech Industry** | | **Tech Hub** | |
| | **Yes** | **No** | **Yes** | **No** |
| Productivity × 𝟙(Post) | 0.0023*** | 0.0013*** | 0.0021*** | 0.0016*** |
| | (0.0003) | (0.0004) | (0.0004) | (0.0003) |
| Employee F.E. | Yes | Yes | Yes | Yes |
| Salary & Seniority Controls | Yes | Yes | Yes | Yes |
| Cohort × Language × Role × Time F.E. | Yes | Yes | Yes | Yes |
| Industry × Location × Time F.E. | Yes | Yes | Yes | Yes |
| Observations | 1,640,903 | 1,272,996 | 846,149 | 2,125,324 |
| # Unique Employees | 73,376 | 57,637 | 38,263 | 95,240 |
| R-Squared | 0.222 | 0.285 | 0.233 | 0.24 |
| Y-Mean | 0.033 | 0.034 | 0.034 | 0.033 |
| Magnitude (%) | 7.01 | 3.69 | 6.28 | 4.75 |

## Table A.3: **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.4: **GitHub Shock & Employee Mobility to Large Firms: Robustness**

This table presents robustness tests for estimates reported in column 5 of Table 2. In Panel A, we run the test on alternative samples. In A1 and A2, we 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 A1 and demographic characteristics along with public contributions in A2. In A3, we exclude employees who have ever worked at Microsoft, the firm that acquired GitHub in 2018. In Panel B, 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 C presents results with more granular controls. In C1, we control for employee's cohort, role, language, location, and industry interacted with time fixed effects. In C2, 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 D, we present our results with an alternative number of bootstrap iterations performed to compute the standard errors and report the errors up to six digits after the decimal. All panels run tests corresponding to our baseline specification, Column 5 in Table 2. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Outcome:** | **Coefficient** | **Std. Error** | **Observations** | **R-squared** | **Y-Mean** | **Magnitude (%)** |
| **Panel A: Alternate Employee Sample** | | | | | | |
| *A1: Sample Matched on Individual Demographics* | | | | | | |
| Productivity × 𝟙(Post) | 0.0019*** | (0.0003) | 1,241,915 | 0.254 | 0.03 | 6.18 |
| *A2: Sample Matched on Individual Demographics and Public Contributions* | | | | | | |
| Productivity × 𝟙(Post) | 0.0017*** | (0.0003) | 1,405,593 | 0.25 | 0.031 | 6.37 |
| *A3: Sample Excluding Microsoft Employees* | | | | | | |
| Productivity × 𝟙(Post) | 0.0017*** | (0.0002) | 3,035,617 | 0.217 | 0.032 | 5.13 |
| **Panel B: Alternate AKM Sample Period** | | | | | | |
| *B1: Pre Sample AKM (2011-14)* | | | | | | |
| Productivity × 𝟙(Post) | 0.0015*** | (0.0003) | 2,014,005 | 0.219 | 0.031 | 4.68 |
| **Panel C: More Controls** | | | | | | |
| *C1: Control for Cohort × Role × Language × Location × Industry × Time* | | | | | | |
| Productivity × 𝟙(Post) | 0.0023*** | (0.0005) | 745,659 | 0.37 | 0.036 | 6.35 |
| *C2: Control for Public Contributions, Stars, and Publications* | | | | | | |
| Productivity × 𝟙(Post) | 0.0020*** | (0.0002) | 3,180,434 | 0.211 | 0.033 | 5.92 |
| **Panel D: Robustness with Number of Bootstrap Iterations** | | | | | | |
| *D1: Baseline - 100 Iterations* | | | | | | |
| Productivity × 𝟙(Post) | 0.0017*** | (0.000198) | 3,199,173 | 0.212 | 0.033 | 5.23 |
| *D2: Robustness - 1,000 Iterations* | | | | | | |
| Productivity × 𝟙(Post) | 0.0017*** | (0.000197) | 3,199,173 | 0.212 | 0.033 | 5.23 |

## Table A.5: **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.6: **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.7: **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.8: **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 2 by excluding users with sizeable changes in productivity estimates derived from the AKM decomposition in the pre- and post-shock period. Change in productivity estimates is defined as the change in productivity quintile in Panel A of Figure A.2. Column 1 (2) (3) presents robustness by limiting the sample to employees whose absolute productivity quintile change, i.e. $\Delta Prod$, is not more than three (two) (one). Column 4 only retains employees who fall in the same productivity quintile in the pre- and post-period. Standard errors are bootstrapped within employee clusters using 100 repetitions. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| **Outcome:** | **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 × 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.9: **GitHub Shock & Employee Mobility to Large Firms: Alternative Productivity Definition**

This table presents robustness tests for estimates reported in Table 2 using alternative productivity definitions. Instead of decomposing employee productivity using the AKM, we define productivity directly based on employees' raw private contribution numbers. In Panel A, we define employee productivity as the percentile rank of employees' pre-period private contributions. In Panel B, productivity is defined as the ratio of employees' pre-period private contributions to their pre-period total contributions. Both panels run tests corresponding to our baseline specification in Table 2. Standard errors are clustered at the employee level. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

| **Panel A: Defining Productivity as Pct(Private Contributions)** | | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| **Outcome:** | **𝟙(Job Change to Large Firms)** | | | | |
| Pct(Pvt Contributions) × 𝟙(Post) | 0.0020*** | 0.0019*** | 0.0021*** | 0.0020*** | 0.0020*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0002) | (0.0002) |
| 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 (%) | 6.32 | 6.23 | 6.83 | 6.33 | 6.08 |
| **Panel B: Defining Productivity as Private Contributions/Total Contributions** | | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| **Outcome:** | **𝟙(Job Change to Large Firms)** | | | | |
| Share(Pvt Contributions) × 𝟙(Post) | 0.0017* | 0.0017* | 0.0018* | 0.0019** | 0.0042*** |
| | (0.0010) | (0.0010) | (0.0010) | (0.0009) | (0.0008) |
| Employee F.E. | Y | Y | Y | Y | Y |
| Time F.E. | Y | Y | Y | Y | Y |
| Firm Cohort | | Y | Y | Y | Y |
| Cohort × Time F.E. | | | Y | | |
| Cohort × Language × Role × Time F.E. | | | | Y | Y |
| Industry × Location × Time F.E. | | | | | Y |
| Observations | 6,804,386 | 6,804,386 | 6,804,166 | 3,811,484 | 3,199,173 |
| # Unique Employees | 302,781 | 302,781 | 302,772 | 170,880 | 143,446 |
| R-squared | 0.054 | 0.054 | 0.06 | 0.168 | 0.212 |
| Y-Mean | 0.031 | 0.031 | 0.031 | 0.032 | 0.033 |
| Magnitude (%) | 3.11 | 3.12 | 3.35 | 4.27 | 5.79 |

Table A.10: **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.11: **GitHub Shock & Mobility of High Productivity Employees**

This table presents estimates on the impact of GitHub's disclosure of employee productivity information on the mobility of high productivity employees. We report the differential impact of GitHub's policy change on productive employees estimated by the $\beta$ coefficient from the equation:

$$\mathbb{1}(\text{Job Change to Large Firm})_{i,t} = \beta\, \mathbb{1}(\text{Superstar Employee}_i) \times \mathbb{1}(\text{Post}_t) + \alpha_i + X_{i,t} + \varepsilon_{i,t}$$

The independent variable is the indicator variable $\mathbb{1}(\text{High Productivity Employee}_i)$ interacted with $\mathbb{1}(Post)$, another indicator that takes the value of 1 for quarters after GitHub's policy change (i.e., after quarter 2 of 2016). $\mathbb{1}(\text{High Productivity Employee}_i)$ equals 1 if employee $i$'s productivity estimate from the AKM decomposition lies in the top percentile of the productivity distribution. Columns (1)–(5) use alternative percentile cutoffs: top 1%, 5%, 10%, 20%, and 25%. The dependent variable equals 1 when employee $i$ switches job to a large firm ($> 1,000$ employees) in quarter $t$. $\alpha_i$ 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 Firm)** | | | | |
| | *Top 1%* | *Top 5%* | *Top 10%* | *Top 20%* | *Top 25%* |
| 𝟙(High Productivity Employee) × 𝟙(Post) | 0.0108*** | 0.0083*** | 0.0057*** | 0.0071*** | 0.0068*** |
| | (0.0014) | (0.0007) | (0.0006) | (0.0005) | (0.0005) |
| Employee F.E. | Y | Y | Y | Y | Y |
| Time F.E. | Y | Y | Y | Y | Y |
| Salary & Seniority Controls | Y | Y | Y | Y | Y |
| Cohort × Language × Role × Time F.E. | Y | Y | Y | Y | Y |
| Industry × Location × Time F.E. | Y | Y | Y | Y | Y |
| Observations | 3,199,173 | 3,199,173 | 3,199,173 | 3,199,173 | 3,199,173 |
| # Unique Employees | 143,446 | 143,446 | 143,446 | 143,446 | 143,446 |
| R-Squared | 0.211 | 0.212 | 0.212 | 0.212 | 0.212 |
| Y-Mean | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 |
| Magnitude (%) | 32.41 | 25.02 | 17.16 | 21.36 | 20.4 |

## Table A.12: **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.13: **GitHub Shock & Employee Mobility by Firm Age**

This table presents estimates on the heterogeneous impact of GitHub's disclosure of employee productivity information on employee mobility to firms based on their age, size, and funding. 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). Panel A classifies destination firms by age and size. Column (1) captures moves to firms with age $\geq$ 10 years, column (2) to firms aged $<$ 10 years, column (3) to firms with age $\geq$ 10 years and over 1,000 employees, column (4) to firms with age $\geq$ 10 years and 50–1,000 employees, and column (5) to firms with age $\geq$ 10 years and less than 50 employees. Panel B restricts the sample to startups (firms with age $<$ 10) and focuses on moves to star startups. Across the five columns, star startups are defined based on different thresholds of venture funding prior to the 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).

**Panel A: Heterogeneity by Firm Age**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| **Outcome:** | **𝟙(Job Change to to Firm)** | | | | |
| | *Age ≥ 10* | *Age < 10* | *Age ≥ 10 & Size > 1,000* | *Age ≥ 10 & 50 < Size ≤ 1,000* | *Age ≥ 10 & Size ≤ 50* |
| Productivity × 𝟙(Post) | 0.0015*** | -0.0017*** | 0.0016*** | 0.0000 | -0.0001* |
| | (0.0002) | (0.0003) | (0.0002) | (0.0001) | (0.0001) |
| Observations | 3,180,055 | 3,180,055 | 3,178,043 | 3,178,043 | 3,178,043 |
| # Unique Employees | 143,444 | 143,444 | 143,444 | 143,444 | 143,444 |
| R-Squared | 0.224 | 0.208 | 0.213 | 0.225 | 0.238 |
| Y-Mean | 0.045 | 0.035 | 0.028 | 0.014 | 0.004 |
| Magnitude | 3.29 | -4.83 | 5.91 | -0.05 | -3.61 |

**Panel B: Heterogeneity by Funding among Startups**

| **Outcome:** | **𝟙(Job Change to Startup)** | | | | |
|---|---|---|---|---|---|
| | *Funding > Series B* | *Funding > Series D* | *Funding > $ 100 Mn* | *Funding > $ 500 Mn* | *Funding > $ 1 Bn* |
| Productivity × 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 |
| Time F.E. | Y | Y | Y | Y | Y |
| Salary & Seniority Controls | Y | Y | Y | Y | Y |
| Cohort × Language × Role × Time F.E. | Y | Y | Y | Y | Y |
| Industry × Location × Time F.E. | Y | Y | Y | Y | Y |
| Observations | 3,145,396 | 3,145,396 | 3,145,396 | 3,145,396 | 3,145,396 |
| # Unique Employees | 143,437 | 143,437 | 143,437 | 143,437 | 143,437 |
| R-Squared | 0.173 | 0.168 | 0.166 | 0.155 | 0.151 |
| Y-Mean | 0.005 | 0.003 | 0.003 | 0.001 | 0.001 |
| Magnitude (%) | -5.04 | 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.14: **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 2. 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.15: **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 8. 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.16: **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).

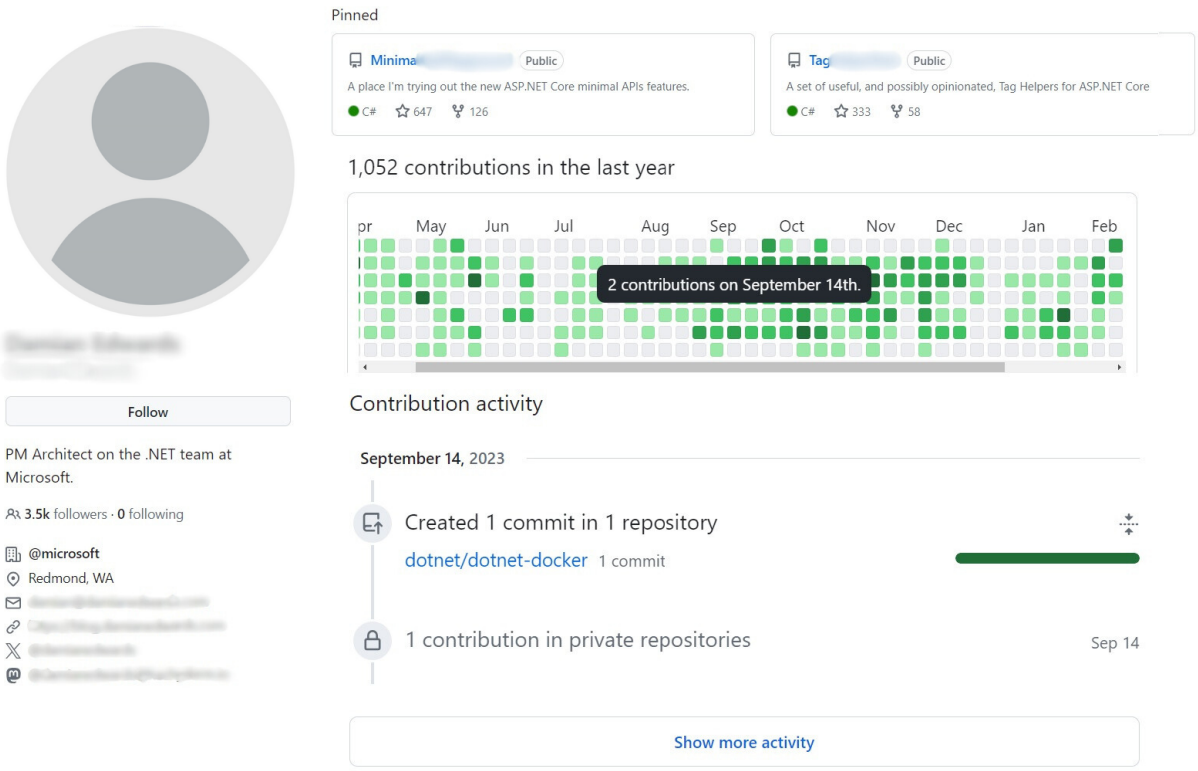
| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Outcome: | Change in Industry Labor Market Concentration (ΔHHI) | | | |
| | All Firms (LinkedIn) | | Compustat Firms (SIC 3-Digit) | |
| Share Emp with Private Contr > 0 | 0.064** | | 0.140*** | |
| | (0.029) | | (0.002) | |
| Average Private Contr | | 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: GitHub Contributions

## B.1 GitHub Contribution Calendar

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

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



## B.2 Correlation between GitHub Contributions and Other Productivity Measures

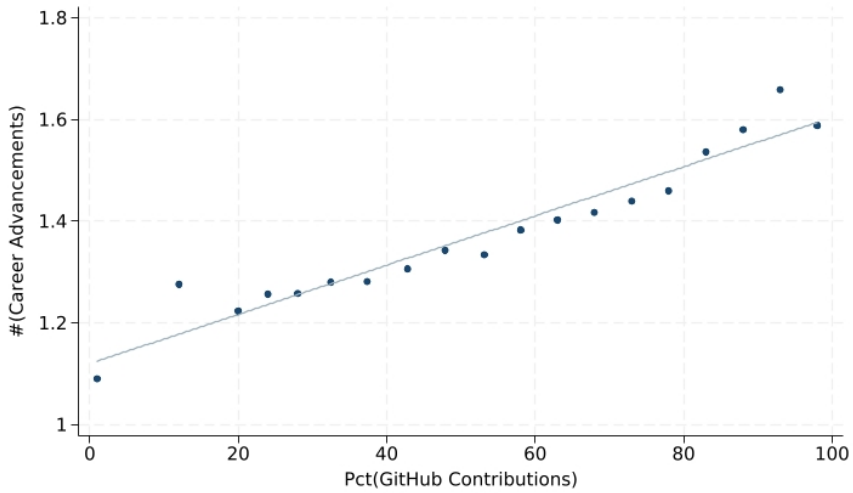
This section provides evidence supporting our use of GitHub contributions as a proxy for employee productivity, and motivates our focus on private contributions in the main analysis. Two measurement issues are central. First, GitHub contributions are a non-traditional output measure, so it is important to validate that they are meaningfully related to worker quality and firm performance. Second, because our empirical design relies on changes in the *visibility* of private contributions, we need to establish that private contributions contain incremental information beyond what could have been inferred from public contributions that were always visible on users' profiles.

**Validation of GitHub contributions as a productivity proxy:** In Figure B.2, we first relate employees' contribution ranks to independent proxies for individual quality and productivity observed in our linked career-history data, including career advancements and publications. We find strong positive and approximately linear relationships between contributions and these outcomes, consistent with contributions capturing economically meaningful differences in worker productivity. We also provide a firm-level validation by relating a firm's GitHub contributions to standard measures of firm value and intangible capital. In particular, Figure B.3 provides the complementary firm-level validation by showing that firms with higher GitHub contributions have higher organizational capital (Eisfeldt and Papanikolaou, 2013) and higher Tobin's Q. Together, these patterns support interpreting GitHub contributions as a meaningful proxy for productivity, consistent with prior work that uses GitHub activity to measure software developer output (Holub and Thies, 2023; El-Komboz and Fackler, 2023; Conti et al., 2023).

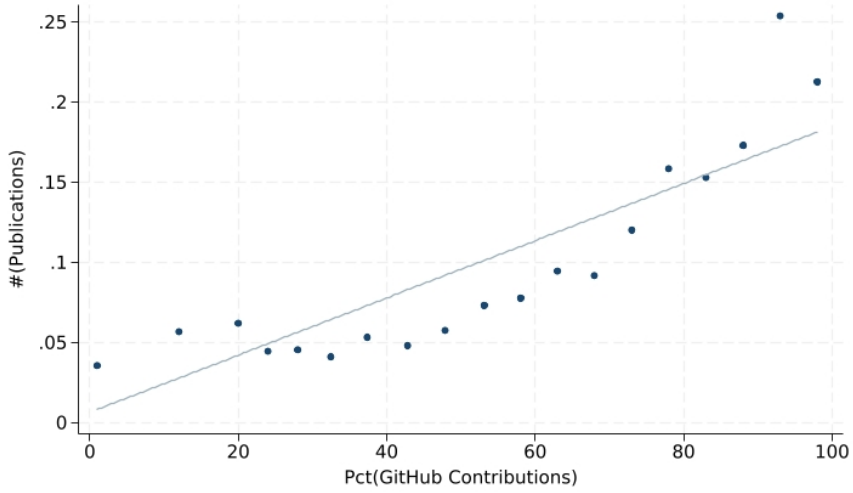
**Why private contributions matter for our setting:** While overall GitHub activity is informative, our setting hinges on private contributions because the 2016 policy change allows users to publicly display (anonymized) private activity that was previously hidden. The key question is whether private contributions simply mirror public contributions or whether they capture additional productivity information that is not otherwise visible to outside employers. In Figure Figure B.4, we therefore compare public and private contributions and show that, although they are positively correlated, public contributions explain less than 5% of the variation in private contributions. We then test whether private contributions contain incremental information by relating them to individual and firm outcomes while controlling for public contributions. Private contributions remain strongly predictive of individual career outcomes and firm-level measures such as organizational capital and Tobin's Q, indicating that they capture meaningful dimensions of productivity and firm value beyond what is reflected in public GitHub activity.

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.



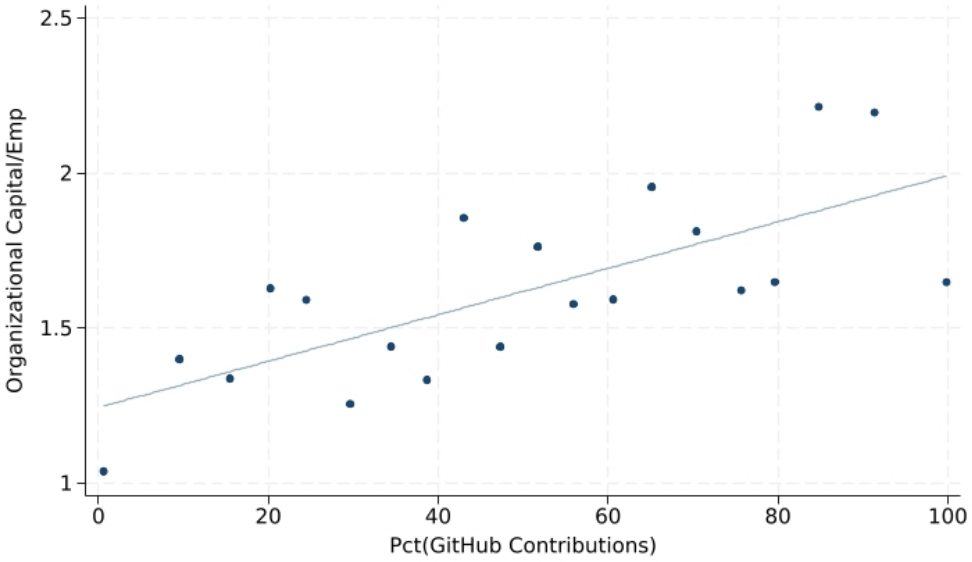

Panel A



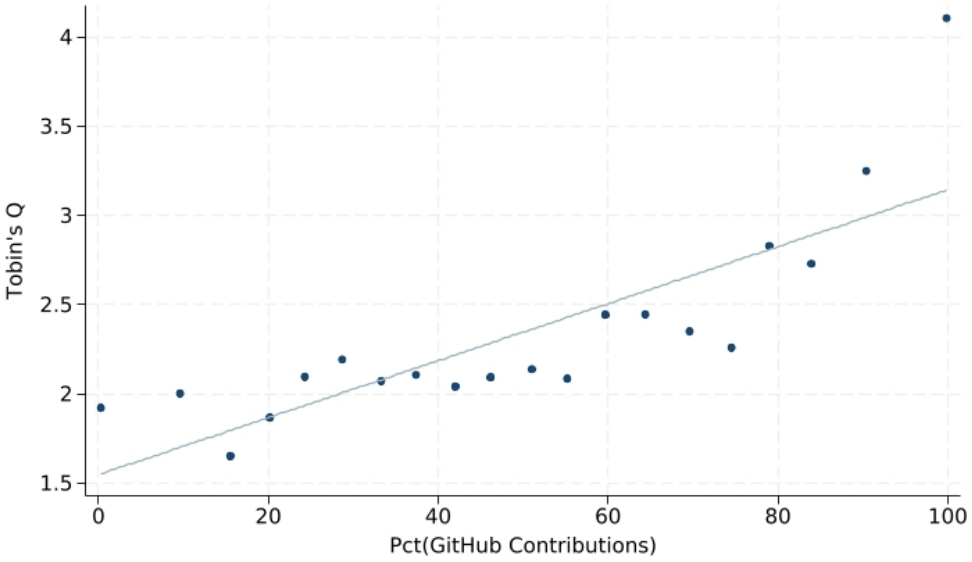

Panel B

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.

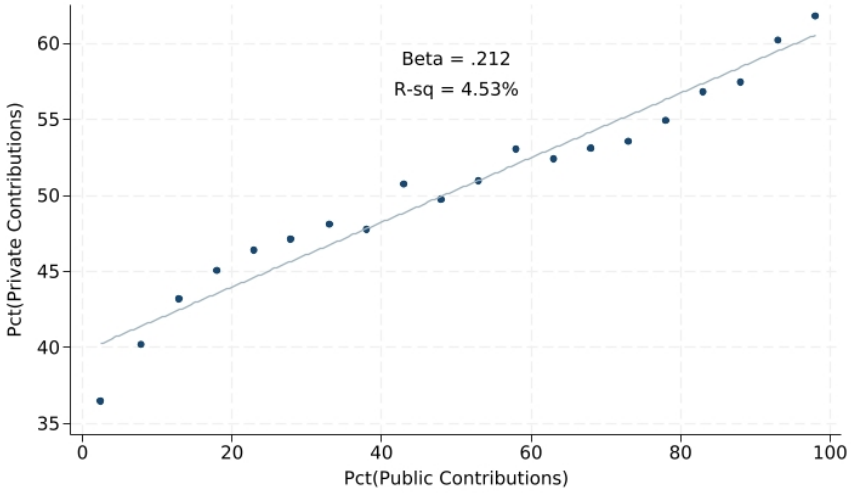


Table B.1: **Association between GitHub Private Contributions and Individual/Firm Productivity**

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 |

## B.3 Use of GitHub by HR professionals

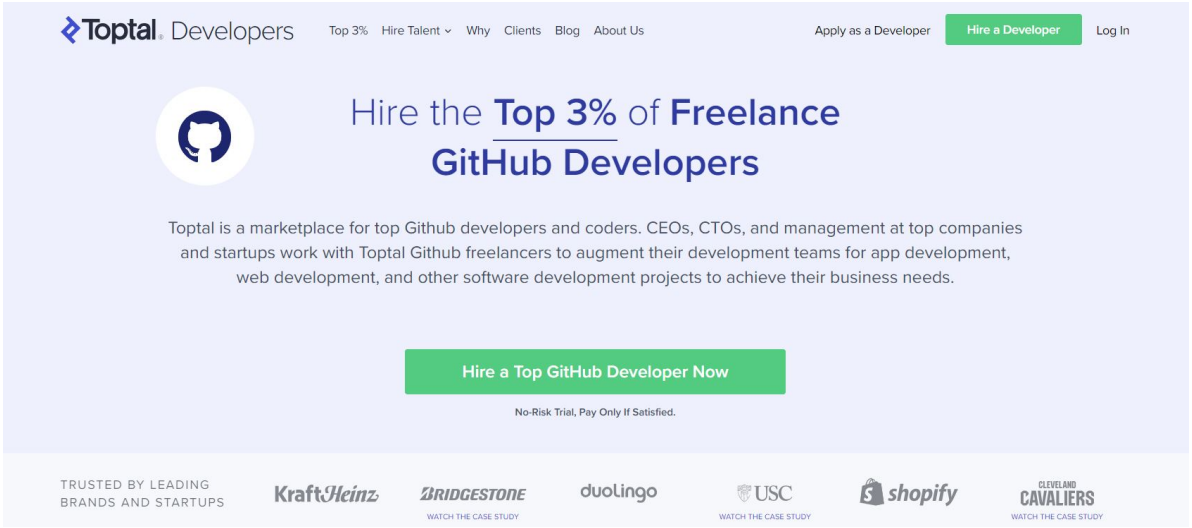
In this section, we provide additional evidence that GitHub had become an input into recruiting for technical talent well before the 2016 policy change. Because GitHub profiles summarize a developer's activity and output at high frequency, recruiters can use this information to screen candidates on dimensions that are otherwise difficult to observe in standard resumes. Consistent with this view, Montandon et al. (2021) report that most recruiters hiring for technical roles use GitHub in their recruiting process, often to assess technical ability and the technologies candidates have worked with. Complementing this evidence, a growing body of industry hiring guides and recruiting resources—emerging in the early 2010s—explicitly discuss how employers can evaluate developers using GitHub profiles, frequently emphasizing contribution histories as a proxy for quality and effort. For instance, `https://philipwalton.com/articles/how-to-find-qualified-developers/` outlines heuristics for identifying qualified developers via GitHub, while `https://beamery.com/resources/blogs/recruiting-and-sourcing-developers-on-github-the-complete-guide` offers step-by-step guides for recruiters to find talent on the platform. Notably, these sources highlight GitHub activity as a proxy for an individual's quality and effort, while also discussing practical frictions in sourcing and screening (e.g., identifying relevant profiles and interpreting activity), which helped motivate the rise of intermediaries.

Figure B.5 provides snapshots of third-party web platforms and recruiting resources that explicitly facilitate hiring through GitHub. These platforms illustrate that a broader ecosystem developed around GitHub-based recruiting during our sample period, further supporting the view that GitHub activity was salient to recruiters and embedded in screening practices.

To complement this qualitative evidence on GitHub-based recruiting, Figure B.6 uses our matched GitHub–LinkedIn dataset to document the presence of HR professionals on GitHub over time. We identify HR professionals based on their LinkedIn job titles and mark their GitHub activity when they join GitHub. The figure plots, by firm size, the fraction of firms with at least one HR professional on GitHub. The share rises steadily through the 2010s and is especially pronounced among large firms, consistent with larger employers adopting more formalized and technology-enabled recruiting processes. This evidence should be interpreted as conservative: our counts rely on public GitHub–LinkedIn links, and recruiters may be less likely than developers to link GitHub to professional profiles.

Figure B.5: **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.



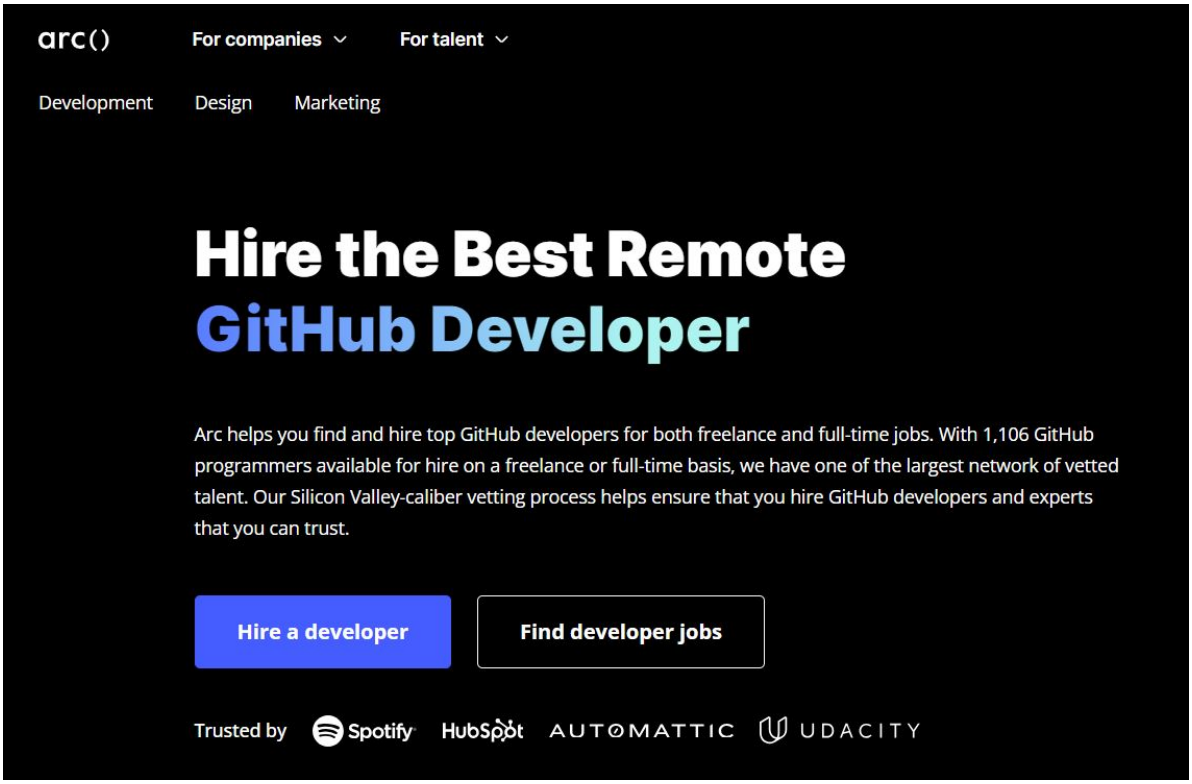



Figure B.6: **Presence of Recruiters (HR Professionals) on GitHub**

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.

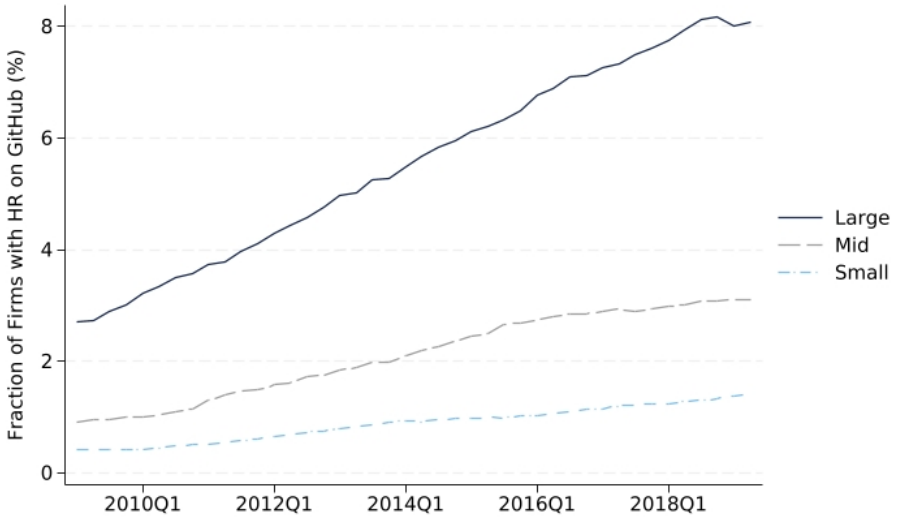
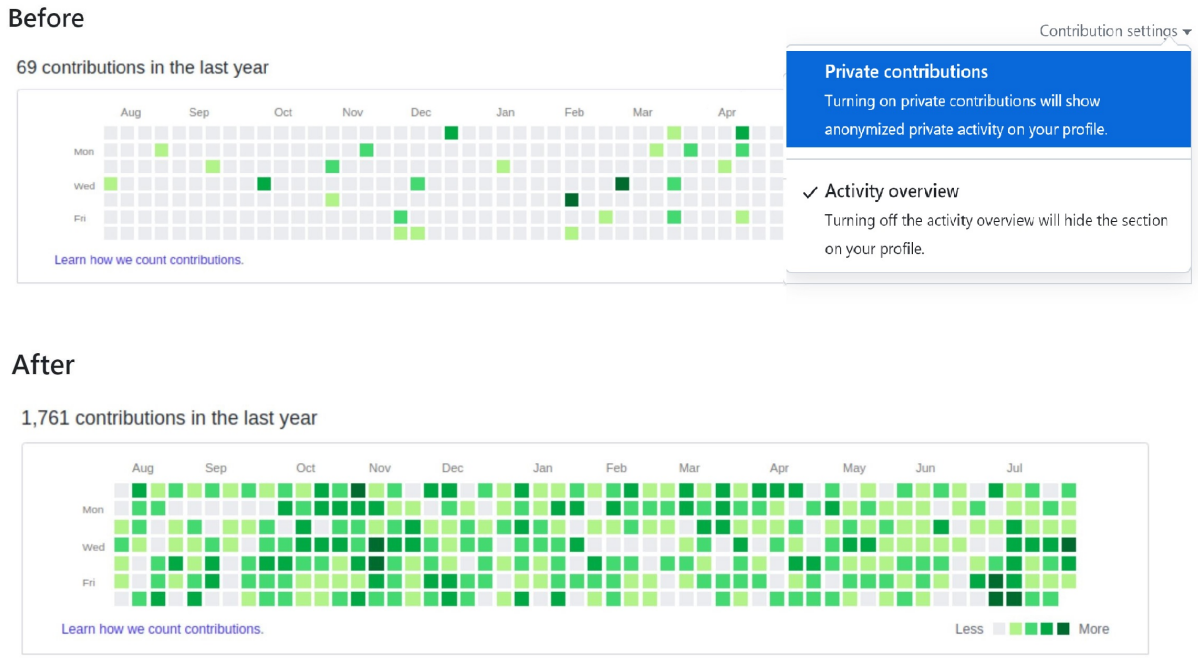


Figure B.7: **GitHub Calendar Before and After May-2016 Policy Change**

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.



## B.4 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.

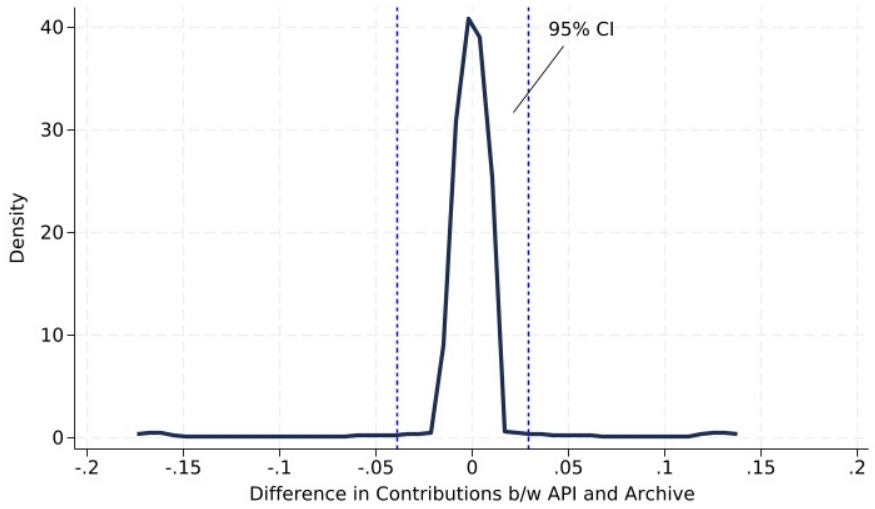
The GitHub Archive (`https://www.gharchive.org/`), a third-party repository, records and stores all public events on GitHub, making daily snapshots of public contributions available for download. This dataset offers comprehensive historical coverage of public contributions that cannot be altered retroactively,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

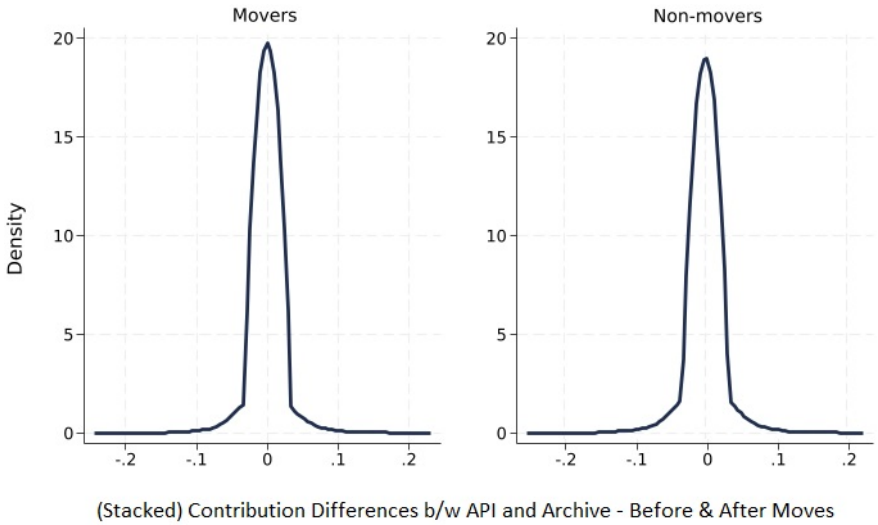
A concern is that the API reflects the current state of user contributions, which may have been amended after the original event, and may not fully align with historical records. To assess the extent of this potential discrepancy, we compare public contributions reported by the GitHub Archive (at the 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. Moreover, 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.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

# Internet Appendix C: Robustness of AKM Estimates

This appendix details our AKM estimation procedure and presents robustness checks that validate both the modeling choice and our productivity estimates. As described in the main text, we measure individual productivity using a two–way fixed effects AKM model with controls for time and role:

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

where $\vartheta_i$ and $\vartheta_f$ capture employee and firm heterogeneity, $\vartheta_t$ and $\vartheta_k$ absorb common shocks at the month and role level, and $X_{i,f,t}$ includes experience and tenure. We take $\widehat{\vartheta_i}$, the estimated employee fixed effect, as our proxy for individual productivity in the subsequent analysis. The AKM is estimated on the five years preceding the GitHub policy change.

## C.1: Role of the Individual Component

A natural concern with AKM is whether firm–worker sorting, rather than worker heterogeneity, drives most of the observed productivity variation. To investigate this, we follow Card et al. (2013) and estimate an event study of job transitions during the pre-shock period (2011–2016). Figure C.1 presents an event study analysis that examines the impact of job transitions on productivity during the pre-shock period (2011-2016), which is used to estimate productivity. The outcome of interest is the 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 results show strong firm effects: workers moving from quartile 1 to quartile 4 firms experience large productivity gains, while moves within quartile 1 or quartile 4 generate little change. Conversely, transitions from quartile 4 to quartile 1 lead to sharp productivity declines. These patterns are consistent with additive worker and firm effects, but not with models where improved firm–worker matching alone boosts productivity.

We further assess the importance of sorting by decomposing the total variance in contributions into three components: individual effects, firm effects, and worker–firm sorting. Column 1 of Table C.1 shows that employee heterogeneity explains about 68% of the variation, firm effects about 10%, and the sorting term about −7%. Thus, individual heterogeneity accounts for nearly seven times as much variation as firm heterogeneity. 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).

One potential concern is *limited mobility bias*, which arises when few workers switch firms. In such cases, sampling noise inflates the firm component and biases the worker–firm covariance downward. To address this, we apply two bias-aware estimators (Table C.1, Panel B). Following Andrews et al. (2008), we implement a homoskedastic "de-noising" correction, and following Kline

et al. (2020), we apply a heteroskedasticity-robust leave–one–out estimator. Both these estimators are implemented based on the computer package provided by Bonhomme et al. (2023). Both adjustments shift variance components in the expected direction: the firm share falls, the negative sorting term attenuates, and the residual rises, while the worker share remains dominant. The magnitude of these changes is far smaller than in wage decompositions studied by Andrews et al. (2008) and Kline et al. (2020), suggesting limited mobility bias is not first-order in our context.

## C.2: Alternative Specifications

We next test robustness to alternative AKM specifications. Our baseline outcome, the within-month percentile rank of contributions, follows Chen and Roth (2024), who show that log transformations of skewed, sparse data may be problematic. Panel A of Table C.2 demonstrates that results are unchanged when using contribution bins, raw counts, $\log(1 + \text{count})$, or a zero-inflated log specification that treats zeroes explicitly.

Another concern is our use of a continuous treatment in a difference-in-differences framework. Callaway et al. (2024) caution that parallel trends may not fully guard against selection when treatments are continuous. To test this, we replace the continuous treatment with a binary indicator. Figure C.2 and Panel B of Table C.2 show that our parallel trends assumption and main results remain intact.

We also verify that implementation details of the AKM estimation do not drive our findings. Specifically, we re-estimate worker productivity $\widehat{\vartheta}_i$ using the algorithm and connected-set construction of Bonhomme et al. (2023), then map these estimates back into our baseline sample. As reported in Panels A1–A2 of Table C.3, the resulting treatment effect (5.18) is nearly identical to the baseline (5.23).

We then test robustness to the additivity assumption in AKM, which rules out employee–firm complementarities. If such complementarities were first-order, AKM could misattribute match-specific gains to either the worker or the firm. To relax additivity, we estimate the BLM model of Bonhomme et al. (2019), which assigns employees to latent types and firms to latent classes, allowing outcomes to depend on their interaction. Using the algorithm of Bonhomme et al. (2023), we estimate a 5×5 specification with Expectation–Maximization. We then rank employee types by their average outcome across firm classes and define productivity based on membership in the top quintiles. Panels A3–A4 of Table C.3 show that our main results are unchanged, confirming that our findings are not an artifact of AKM's additivity restriction.

Finally, we address the concern that AKM employee fixed effects may be noisy for employees with shorter work histories or with fewer coworker comparisons. To quantify estimation noise directly, we follow the Empirical-Bayes shrinkage approach used in personnel and value-added settings (for instance, see Chetty et al. (2014); Walters (2024)). We estimate the AKM model on 1,000 bootstrap subsamples (80% of employees per run) and compute, for each employee, the variance of their estimated employee fixed effect across these runs. In Panels B1–B2 of Table C.3,

we shrink the baseline AKM productivity measure toward the sample mean in proportion to this employee-specific noise (B1), and using a calibrated average reliability of 0.70, following conventions in the value-added literature (B2). In Panels B3–B4, we instead retain our original AKM productivity measure but exclude the top 10% and top 25% most noisy employees identified from these bootstrap draws. Across all versions, the treatment effect remains unchanged, indicating that our results are not driven by noisily estimated employee productivity.

Figure C.1: **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 C.2: **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 Panel A of Figure 1 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.

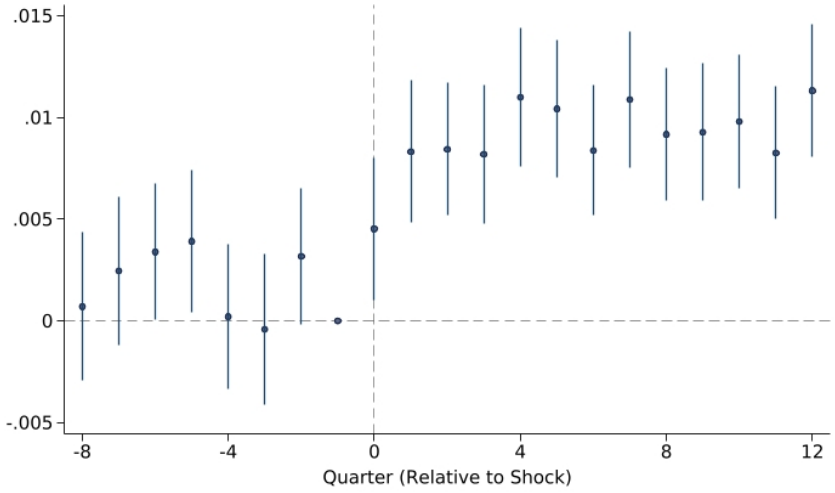


Table C.1: **Variation Breakdown for AKM Tests**

This table presents the details of variance decomposition 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 ($\vartheta_k$). The table summarizes the breakdown of the variance in the dependent variable explained by the employee and firm components as well as the covariance of the employee and firm components (i.e., the sorting term). We also compare the variance decomposition in our baseline model with bias-corrected AKM variances using the homoskedastic correction methodology by Andrews et al. (2008) and heteroskedastic leave-out correction methodology by Kline et al. (2020). We use the computer program provided by Bonhomme et al. (2023) to estimate the bias-corrected AKM and their variance decompositions.

| | **Variation Explained** | | |
|---|---|---|---|
| | (1) **Baseline** | (2) **Andrews et al. (2008)** | (3) **KSS (2020)** |
| Employee | 67.9% | 64.6% | 61.9% |
| Firm | 9.6% | 7.5% | 7.4% |
| Sorting Term | -7.4% | -3.5% | -3.4% |
| Residual | 29.9% | 31.4% | 34.1% |

Table C.2: **Robustness with Dependent Variable in the AKM and Employee Productivity Definition**

This table presents robustness tests for estimates reported in column 5 of Table 2 using alternative dependent variables in the AKM and alternative employee productivity definitions. In Panel A, we use alternative measures of productivity from 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. Panel B presents results from estimates using binary dummy indicators for employees' productivity obtained from the baseline AKM. In B1, $\mathbb{1}(Productivity > 0)$ 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 B2, B3, and B4, the dummy variable takes the value of 1 if the employee's productivity is in the top decile, top quintile, and top quartile, respectively. Standard errors are bootstrapped within employee clusters using 100 repetitions in Panels A and are clustered at the employee level in Panels B. 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 Dependent Variable in AKM** | | | | | | |
| *A1: AKM with Categorical Variables*<br>Productivity × 𝟙(Post) | 0.0016*** | (0.0002) | 3,199,173 | 0.212 | 0.033 | 5.02 |
| *A2: AKM with Raw Contribution Numbers*<br>Productivity × 𝟙(Post) | 0.0008*** | (0.0001) | 3,199,173 | 0.211 | 0.033 | 2.73 |
| *A3: AKM with Logarithm of (1+Y)*<br>Productivity × 𝟙(Post) | 0.0017*** | (0.0002) | 3,199,173 | 0.212 | 0.033 | 4.98 |
| *A4: AKM with Logarithm where Y_min = 0*<br>Productivity × 𝟙(Post) | 0.0016*** | (0.0002) | 3,199,173 | 0.212 | 0.033 | 4.89 |
| **Panel B: Dummy Productivity Definition** | | | | | | |
| *B1: Normalized Productivity Estimate > 0*<br>𝟙(Productivity > 0) × 𝟙(Post) | 0.0071*** | (0.0005) | 3,199,173 | 0.212 | 0.033 | 21.35 |
| *B2: Productivity in Top Decile*<br>𝟙(Top Decile Productivity) × 𝟙(Post) | 0.0057*** | (0.0006) | 3,199,173 | 0.212 | 0.033 | 17.16 |
| *B3: Productivity in Top Quintile*<br>𝟙(Top Quartile Productivity) × 𝟙(Post) | 0.0068*** | (0.0005) | 3,199,173 | 0.212 | 0.033 | 20.4 |
| *B4: Productivity in Top Quartile*<br>𝟙(Top Tercile Productivity) × 𝟙(Post) | 0.0068*** | (0.0005) | 3,199,173 | 0.212 | 0.033 | 20.39 |

Table C.3: **Robustness with Alternative AKM Specifications**

This table presents robustness tests for estimates reported in column 5 of Table 2 using alternative AKM specifications. In Panel A, we run the AKM model using the computer program and framework provided by Bonhomme et al. (2023). In A1 and A2, we run the two-way fixed effects models that include homoskedastic and heteroskedastic bias correction methodologies by Andrews et al. (2008) and Kline et al. (2020), respectively. In A3 and A4, we estimate productivity using the static BLM framework (Bonhomme et al., 2019), which models latent worker types and firm classes and allows non-additive (interaction) effects between them. We estimate five worker types and five firm classes and define a productive employee as a binary dummy indicator representing whether they belong to the top quintile in terms of productivity (in A3) or in the top two quintiles (in A4). Panel B performs robustness tests based on Empirical-Bayes shrinkage of the employees' productivity estimate in our baseline AKM specification. In B1, we shrink employee productivity in proportion to the variance of each employee's AKM estimate across 1,000 bootstrap runs (employee-specific reliability). In B2, we apply shrinkage calibrated to an average reliability of 0.70 for all employees. In B3 and B4, we retain the original AKM productivity measure but exclude the top 10% and top 25% of employees with the noisiest productivity estimates across bootstrap runs. Standard errors are bootstrapped within employee clusters using 100 repetitions in columns 1 and 2 of Panel A, and columns 3 and 4 of Panel B, and are clustered at the employee level otherwise. Significance levels: *(p<0.10), **(p < 0.05), ***(p < 0.01).

| **Panel A: Alternative AKM Specifications** | | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Output:** | **𝟙(Job Change to Large Firm)** | | | |
| | **KSS (2020)** | **Andrews et al. (2008)** | **BLM (2019)** | **BLM (2019)** |
| Productivity × 𝟙(Post) | 0.0018*** | 0.0018*** | 0.0046*** | 0.0034*** |
| | (0.0002) | (0.0002) | (0.0005) | (0.0005) |
| Employee F.E. | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| Cohort × Language × Role × Time F.E. | Yes | Yes | Yes | Yes |
| Industry × Location × Time F.E. | Yes | Yes | Yes | Yes |
| Observations | 3,069,222 | 3,069,222 | 3,108,102 | 3,108,102 |
| # Unique Employees | 137,549 | 137,549 | 139,493 | 139,493 |
| R-Squared | 0.212 | 0.212 | 0.211 | 0.211 |
| Y-Mean | 0.034 | 0.034 | 0.032 | 0.032 |
| Magnitude (%) | 5.18 | 5.18 | 14.39 | 10.74 |
| **Panel B: Baseline AKM with Empirical-Bayes Shrinkage** | | | | |
| **Output:** | **1(Job Change to Large Firm)** | | | |
| | **Employee-Specific Reliability** | **Calibrated Reliability = 0.70** | **Excluding Top 10% Noisiest Employees** | **Excluding Top 25% Noisiest Employees** |
| Productivity × 𝟙(Post) | 0.0017*** | 0.0016*** | 0.0018*** | 0.0020*** |
| | (0.0002) | (0.0002) | (0.0002) | (0.0003) |
| Employee F.E. | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| Cohort × Language × Role × Time F.E. | Yes | Yes | Yes | Yes |
| Industry × Location × Time F.E. | Yes | Yes | Yes | Yes |
| Observations | 3,199,173 | 3,199,173 | 2,903,282 | 2,448,646 |
| # Unique Employees | 143,446 | 143,446 | 130,257 | 109,269 |
| R-Squared | 0.212 | 0.212 | 0.211 | 0.211 |
| Y-Mean | 0.033 | 0.033 | 0.035 | 0.037 |
| Magnitude (%) | 5.08 | 4.71 | 5.24 | 5.29 |

# References


Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Andrews, M. J., Gill, L., Schank, T., and Upward, R. (2008). High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A*, 171(3):673–697.

Bhaskarabhatla, A., Cabral, L., Hegde, D., and Peeters, T. (2021). Are inventors or firms the engines of innovation? *Management Science*, 67(6):3899–3920.

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., and Setzler, B. (2023). How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics*, 41(2):291–322.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739.

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

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality*. *The Quarterly Journal of Economics*, 128(3):967–1015.

Chen, J. and Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, 139(2):891–936.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review*, 104(9):2633–2679.

Conti, A., Gupta, V., Guzman, J., and Roche, M. P. (2023). Incentivizing innovation in open source: Evidence from the github sponsors program. Technical report, National Bureau of Economic Research.

Eisfeldt, A. L. and Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns. *The Journal of Finance*, 68(4):1365–1406.

El-Komboz, L. A. and Fackler, T. (2023). Productivity spillovers among knowledge workers in agglomerations: Evidence from github. *Available at SSRN 4600283*.

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

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898.

Montandon, J. E., Valente, M. T., and Silva, L. L. (2021). Mining the technical roles of github users. *Information and Software Technology*, 131:106485.

Walters, C. (2024). Empirical bayes methods in labor economics. In *Handbook of Labor Economics*, volume 5, pages 183–260. Elsevier.