ACI Research Paper #02-2021

Does Working From Home Work? A Natural Experiment From Lockdowns

Lucas SHEN

April 2021

Please cite this article as:

Lucas Shen, "Does Working From Home Work? A Natural Experiment From Lockdowns", Research Paper #2-2021, *Asia Competitiveness Institute Research Paper Series (April 2021)*

Does Working from Home Work? A Natural Experiment From Lockdowns

Lucas Shen[†]

Mar 2021

Abstract

Using open-source software records, this paper studies how closures of workplaces affect productivity as individuals move to working from home. Both the event study and microsample difference-in-differences analyses lead to the same conclusion: there are minimal increases in performance measures derived from the publicly-observed records once changing individual profiles are accounted for. The microsample estimates are precise, allowing me to rule out productivity gains as small as .28 percent in collaborative projects. In the discussion, I place my findings alongside three related studies that focus on different work contexts.

I Introduction

This study uses both temporal and geographical variation in state-imposed closure of workplaces during the COVID-19 pandemic to estimate whether working from home (WFH) changes work patterns. Specifically, I start with a census of public and timestamped activities of open-source software projects on the Github platform in Jan–Jun 2020 and map these individuals to countries. If WFH does indeed increase productivity, then one should observe higher activity rates for those individuals during periods where their stated country of location has implemented state-imposed WFH.

Why should WFH improve productivity? Flexibility and autonomy is one commonly cited reason (eg BBC 2020b; Khanna 2020). The company Fitjitsu, in particular, believes that "increased autonomy offered to its workers will help to improve the performance of teams and increase productivity" (BBC 2020a). Less time spent

[†]Preliminary draft; latest draft here. Email: lucas@lucasshen.com. Giovanni Ko, Jing Zhi Lim, Gaurav Sood, and Walter Theseira provided early and helpful feedback. Majority of the work in this paper is done WFH. Interpretations and errors are those of the author.

commuting is also a commonly cited reason, although it conflates increases via the intensive and extensive margin (Whiting 2020). Yet another reason is office distractions (eg Banbury and Berry 1998), conditional on having a conducive remote workspace. For employees in tech-related industries, the argument narrows substantially, with conventional wisdom suggesting that these jobs are well-suited for remote working to begin with (Alipour et al. 2020; Lerman and Greene 2020). The in-house report by Github in particular examines trends and finds that productivity has indeed increased (Forsgren 2020).

The seminal (and to my knowledge, only) study on the causal effects of WFH on productivity comes from Bloom et al. (2015), who conduct a randomized controlled experiment for call-center representatives of a travel agency. Over a nine-month experimental period with 249 participants, they find that WFH led to a 3 percent increase in calls per minute. During state-imposed lockdowns, however, employees WFH come from various sectors with less routine and transactional tasks, making a comparison with the participants in the Bloom et al. (2015) study difficult.

The ambiguity about WFH can be seen in the pandemic press coverage. The coverage tends to include large tech companies (Lerman and Greene 2020; BBC 2020c), with some suggesting that WFH improves productivity (BBC 2020a; BBC 2020b; Whiting 2020). The CEO of Netflix, on other hand, suggested to the media that WFH "has no positive effects and makes debating ideas harder" (BBC 2020d). The Bloom et al. (2015) study notwithstanding, the effect of WFH on productivity remains an open question, which is reflected in a contemporaneous survey by YouGov (2020) where the majority of the 1000 US respondents self-report that they are just about as productive WFH compared to going to the office (Figure A1).

Using individuals who are (mostly) software developers, and exploiting state-imposed variation of WFH across time (Figure 1), I present two pieces of evidence. First, I show in a flexible event study framework that distinct changes in work patterns occur around the first day of lockdowns. Overall productivity increases by as much as 60 percent after lockdowns, significant at the 5 percent level, but this appears to be driven by the extensive margin, as the event study framework also captures an increase in active users on the Github platform. Accounting for the changing age profile attenuates all estimates to the point where they are statistically indistinguishable from zero. Second, and motivated by the lack of pretrends in the event study analyses, I use the microsample records binned into individual-project-WFH arm cells in a pre vs post type of difference-in-differences analysis. The findings are similar, with any positive impact of WFH disappearing after accounting for changes in observables in the different WFH periods. The microsample estimates are precise, and this allows me to rule out even modest and optimistic ef-

Table 1—*Work-From-Home (WFH) Coding from OxCGRT*

OxCGRT WFH indicator	Type	Description from Oxford's Blavatnik School of Government (Petherick et al. 2020)
0	Non-binding	No measures
1	Non-binding	Recommended closing (or recommended work from home)
2	Binding	Required closing (or work from home) for some sectors or categories of workers
3	Binding	Required closing (or work from home) for all-but-essential workplaces (e.g. grocery stores, doctors)

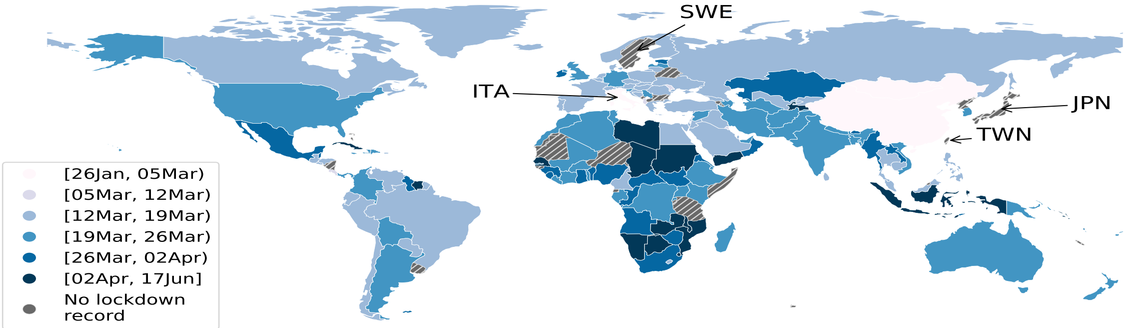


Figure 1: Staggered Timing in State-Imposed WFH

Notes—Map shows the timing of state-imposed WFH where workplaces are "required" to close. Underlying dates are for the date at which the OxCGRT WFH indicator first switches from {0, 1} to {2, 3}, that is, from no closing/recommended closing to a required closing for either some sectors or all-but-essential sectors (Table 1). Darker shades indicate later workplace closures. Countries without state-imposed WFH in the sample period (Jan–Jun 2020) are shaded in gray.

fects of WFH on productivity—the most optimistic impact is a 0.28 percent increase in productivity retrieved from the upper bound of a 95 percent confidence interval. I discuss my findings in Section V in greater detail.

This study complements the seminal study by Bloom et al. (2015) which uses a randomized controlled trial in a work context with well-defined transactional tasks. Two other contemporaneous related studies are by Choudhury et al. (2021) who estimate a substantial 4.4 percent increase in a *WFA* (work-from-anywhere) vs. WFH setting, with WFA becoming increasingly relevant, and the study by Wang et al. (2020) who find an increase in social media content generation. More generally, this study relates and pays attention to the emerging difference-in-differences literature on the variation of treatment timing across groups of units in the sample (Baker et al. 2021; Callaway and Sant'Anna 2020; Goodman-Bacon 2019).

More broadly, this paper contributes to the literature on remote working in the context of the pandemic, such as those on public goods contribution (Choudhury et al. 2020; Kummer et al. 2020; Ruprechter et al. 2021), changes in emails patterns (DeFilippis et al. 2020), uneven household costs (Stanton and Tiwari 2021), and employment and health impacts (Angelucci et al. 2020) on employees who can and cannot WFH, and how WFH potentially hurts creativity (BBC 2020d; Wang et al. 2020). Finally, a set of studies look into the share of jobs that can be done WFH (Alipour et al. 2020; Bartik et al. 2020; Bloom 2020; Brynjolfsson et al. 2020; Gottlieb et al. 2020).

The next Section II provides background to the Github platform and data construction. Section III reports the event study results. Section IV reports the microsample difference-in-differences results. Section V discusses the findings.

II Data and Background

II.A Productivity Measures

To gain empirical traction on whether WFH affects productivity, I use publicly observed activities on the Github platform as productivity measures. Github is a platform where developers (and also some researchers) host, (Git) version control, and collaborate on projects. On the official website, Github states that it is "where the world builds software", and is the "largest and most advanced development platform in the world". The platform is free and open-source, with effectively zero barriers to entry. For a sense of scale, Github states that it has over 56 million users, 100 million repositories, 3 million organizations.[1] [2]

In the Git workflow are two key milestones. First are *commits*, which are changes in the pipeline of a project. These changes for example could be in a text file in the form of code or writing. When ready, users archive these changes, in a potentially modular fashion, to the local repository (which are then eventually pushed to the corresponding remote repository).

A second and usually larger milestone is a *pull request*. In the workflow, when an individual is happy with the (set of) changes—which could be a bug fix, issue resolution, or feature addition—they submit a pull request. Once the request is submitted, other members working on the same project can review and discuss and, upon approval, merge back to the main branch which is always stable for production release. Figure A2 illustrates commits and pull request as part of a branching workflow.

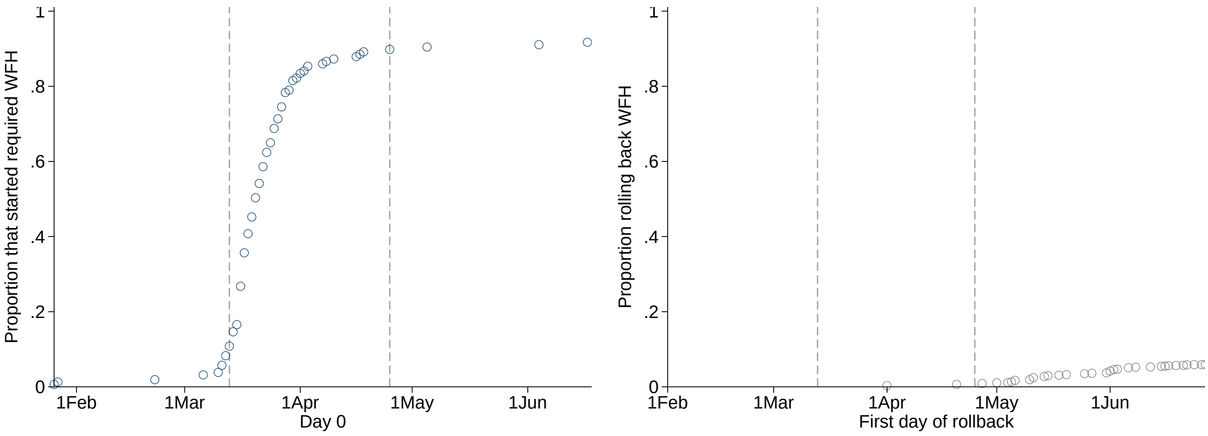
To the extent that commits and pull requests constitute milestones in a workflow, with the former being higher in frequency and the latter having greater weight, especially in non-trivial collaborative workflows, I consider them as productivity measures and as the main dependent outcomes of interest for the rest of the paper.[3]

---

[1] Github includes some of the most prominent organizations: Apple, Facebook, Github (dogfooding), Google,Microsoft, and Twitter, with repositories and users in the order of thousands.

[2] Retrieved at time of writing. `https://github.com/about`.

[3] Github is popular but not the only way to implement the Git workflow. Like Github, GitLab and BitBucket are two alternative platforms with free basic access, and some organizations implement Git using private servers.



(a) State-imposed WFH Timing (b) Rollback Timing

Figure 2: Rollout of WFH

Notes—Panel (a) plots the cumulative proportion of countries implementing state-imposed WFH across dates in the sample period. Panel (b) plots the cumulative proportion of countries rolling back state-imposed WFH (Table A1 lists these countries that rollback state-imposed WFH in the sample period). The first and second vertical gray lines mark the dates at which approximately 10 and 90 percent of countries have imposed WFH.

II.B Data

To build the panel, I first query Google BigQuery's archive of Github timestamped commits in the period Jan 2020–Jun 2020, which includes author and repository names. Only commits in public repositories are included, while those of private repositories (opted out of public view) are not. I use Github's Search API and User API to retrieve the location strings entered in the authors' user profiles and then map users to countries by querying the `OpenStreetMap` API. From this pipeline, I end up with approximately 350,000 commit records and 290,000 pull request records. For the event studies in Section III, I collapse the records to the country-date level and for the microsample difference-in-differences analyses in Section IV, I collapse the records down to the individual-project-WFH periods.[4]

To retrieve the WFH status for countries for any given date, I use the OxCGRT's repository of COVID-19 government responses trackers (Petherick et al. 2020). Table 1 lists the four types of OxCGRT WFH coding. For all the main analyses, I treat the recommended WFH from OxCGRT as non-binding while treating the two required WFH codings from OxCGRT (2 and 3) as binding and with homogeneous effect. So in the event studies for example, "Day 0" is defined as the date at which the OxCGRT WFH indicator switches from either a 0 or 1 to a 2 or 3. This assumes that the compliance of those working in ICT-related fields is the same for the two levels of state-imposed WFH. Figure 1 shows the geographical variation in the start of state-imposed WFH, while Figure 2 shows the rollout of WFH across time.[5]

[4] Appendix A.1 describes the data build in greater detail, while Section D of the Online Appendix provides randomly-sampled examples of both failed and successful geocodings.

[5] Defining "Day 0" as the OxCGRT indicator switching from 0 to either 1, 2, or 3 in the event studies does not substantially change the results (not shown in this draft).

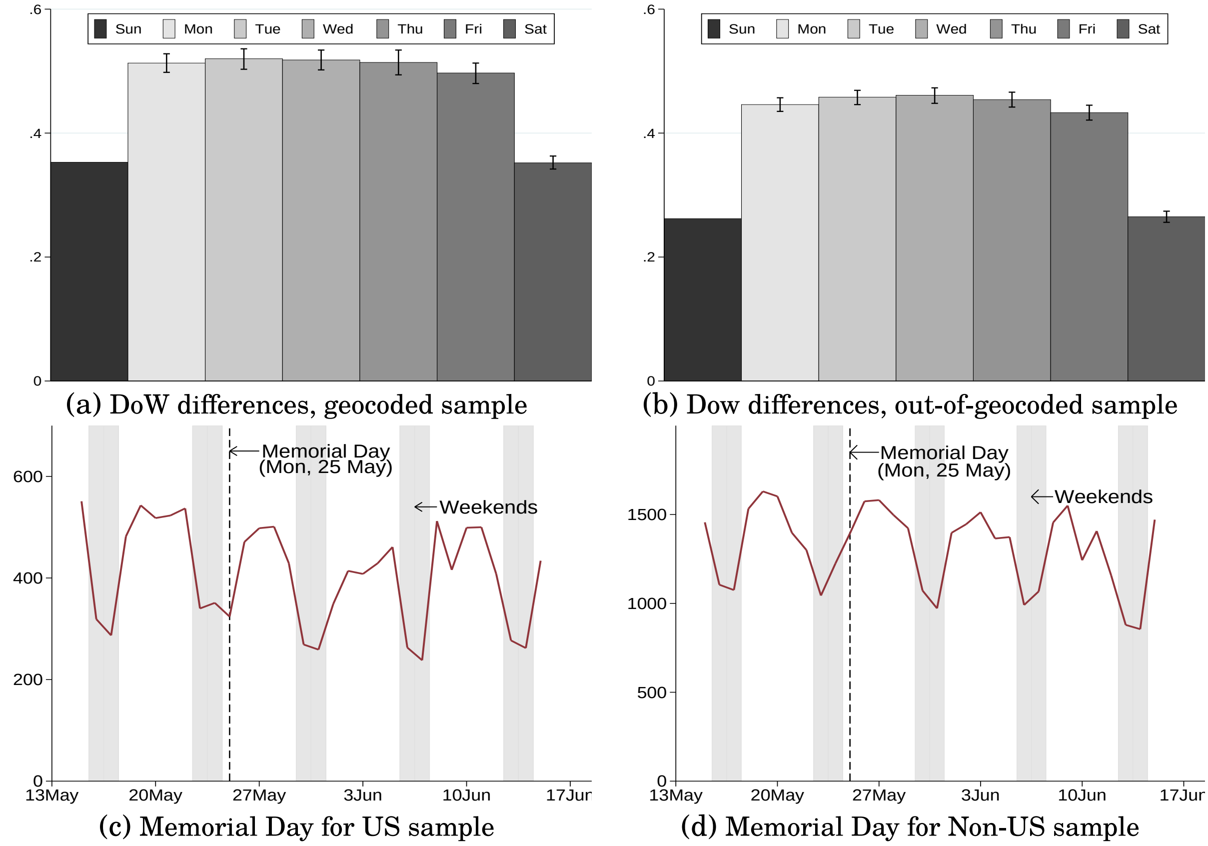


Figure 3: Sensitivity to Day-of-Week and Holidays

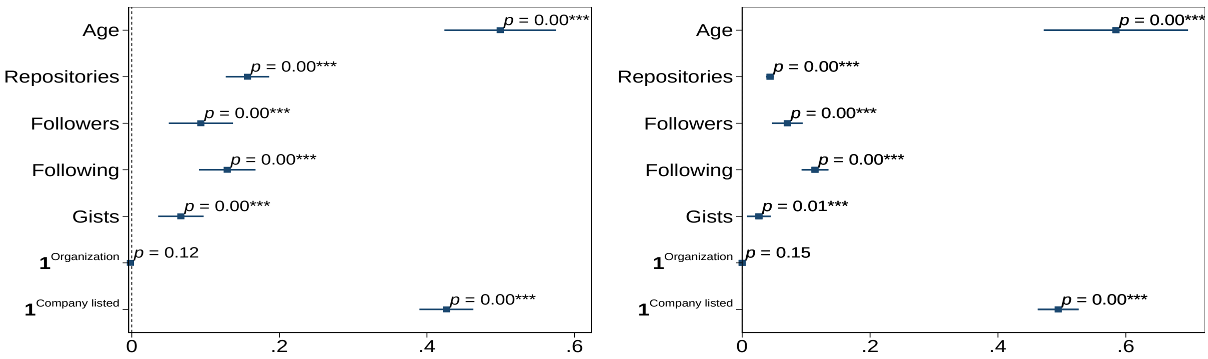
Notes— In the first row, the bar plots show the differences in log of (1+) commits by day-of-week (DoW) from regressing commits on the day-of-week dummies, plus user and repository fixed effects for the user-repository-DoW panel. The baseline day is Sunday—first bar—so that the standard errors for subsequent bars are for the estimates of the additional effects of Mon–Sat relative to Sunday. Robust standard errors are clustered at users and repositories. The second row shows the path plot (unsmoothed) of commits around May 2020 for US sample vs. the non-US sample. Black dashed line indicates Memorial Day on 25 May. Gray vertical bars indicate weekends (Sat–Sun).

II.C Benchmarking

One concern with getting only publicly recorded activity on the Github platform and geocoding them is the introduction of noise, which might affect power. For instance, with poor geocoding of users to countries, the coded WFH indicator will not properly capture the exact date at which users are mandated to WFH, creating noise. Another concern is that those Github users who are successfully geocoded have substantially different work habits than those who are not. While the concerns cannot be fully addressed, here I explore two preliminary analyses, focusing on the higher-frequency commits measure.

First, I compare the commits for the different days-of-week, finding that productivity is systematically higher during the weekends for both the geocoded and out-of-geocoded sample. One might interpret this as mitigating concerns that the open-source and public commit records captures only the "hobby"-type of projects. The first row of Figure 3 shows this.[6]

[6] To do this, I aggregate the commits log record up to the user-repository-DoW level, and then



(a) Commits sample (b) Pulls sample

Figure 4: GEOCODED VS OUT-OF-GEOCODED SAMPLE

Notes—Differences in means for geocoded records vs out-of-geocoded records (those not successfully geocoded, see Table A11 in the Online Appendix for examples). Units are in standard deviations (except the two indicator variables for scaling purposes only). Estimates derived from regressing the variables on the geocoded dummy and performing a t-test for the dummy. Number of individual observations are 44,894 and 120,614 for the commits and pulls sample, respectively. Age refers to age of the individual's user account (creation date minus 1 Jan 2020). Repositories refer to the number of public repositories listed in the account. Followers and following are the number of accounts that the individual follows and the number of accounts following the individual. Gists are the number of mini-blogs/code snippets. The last two dummies indicate whether the account type is actually an organization and whether the individual reports the company he or she works at. Tables A1–A2 of the Online Appendix tabulates the above results. Robust standard errors clustered by countries (non-geocoded counts as a "country"). ***, **, and * denotes significance at the 1, 5, and 10 percent level, respectively.

Second, I qualitatively access geocoding by comparing the productivity of the US sample vs. the rest of the world, focusing on the Memorial Day holiday— one of the largest federal paid holiday across the states in the US—which falls on a Monday (25 May) in the US. Comparing the raw path plots shows a distinct extension of the weekend productivity dip into the following Monday for the US sample but not for the non-US sample.[7]

Finally, Figure 4 reports the differences of individuals who are geocoded versus those who are not, which stresses one key point. The sample that is successfully geocoded, and included in the main analyses are more prominent on all observable aspects; they are more experienced (by account creation date), have more projects, have more followers, and follow more people. Importantly, they are more likely to list which company they work for, and this partially addresses whether the publicly-observed Github events are professional-related work.

estimate

$$\ln(1+\text{commits})_{ijd} = \alpha + \sum_{d\in\{1,\ldots,6\}} \pi_d \text{DoW}_d + \text{individual}_i + \text{repository}_j + u_{ijd\underline{t}},$$

where Sunday ($d = 7$) is the reference day.

[7] Figures A9–A12 in Section B of the Online Appendix provides additional plots to show how the raw path plots of the activities behave across days of the week and around key nationally distinct holidays.

III Event Study

III.A Specification and Results

Figure 5 shows the changes in productivity—commits and pull (requests) per user per day (all in logs)—from estimating a flexible event-study specification:[8]

$$(1) \qquad y_{ct} = \alpha_c + \alpha_t + \boldsymbol{\beta}_t \boldsymbol{X}_{ct} + \sum_{\tau=-21}^{-2} \gamma_\tau \mathbb{1}_{\{t=T_c+\tau\}} + \sum_{\tau=0}^{21} \delta_\tau \mathbb{1}_{\{t=T_c+\tau\}} + \varepsilon_{ct},$$

for a 21-day window before and after T_c, where T_c is Day 0 for country c—the date when the OxCGRT WFH indicator first switches to the binding state-imposed WFH. End points are binned and the day before T_c ($\tau = -1$) is the baseline. α_c and α_t are country and date fixed effects. $\boldsymbol{X_{ct}}$ includes group-by-week-of-year fixed effects (Goodman-Bacon 2019), where a group is defined by the Day 0 timing, two OxCGRT indices on government response,[9] and the COVID-19 epidemiology path (log of confirmed cases, recovered cases, deaths) because they potentially affect treatment anticipation in this particular context, with both sets allowed to have heterogeneous effects over time. Conditional on the observables and the group-by-week linear time trends, the γ_τ estimates constitute conditional falsification tests for pretrends, and the δ_τ estimates trace out the daily effects of WFH on the work patterns. Equation (1) also absorbs the original OxCGRT WFH indicator for each country-by-date cell to allow for rolling back of the WFH enforcement for certain countries towards the end of the sample period (see Figure 2).[10] Standard errors are clustered by country.

Figure 5 plots the estimates of γ_τ and δ_τ from Equation (1). Though measurement error is going to attenuate estimates substantially, I note distinct changes for several measures starting from Day 0. For commits, the estimated coefficient of δ_τ is .472 ($p < .047$) by day 21, which is a 60 percent increase,[11] while the number of active individuals on the other hand has an estimated δ_τ of .203 ($p < .156$), suggesting a 23 percent increase in the number of active individuals.[12] Overall, commits

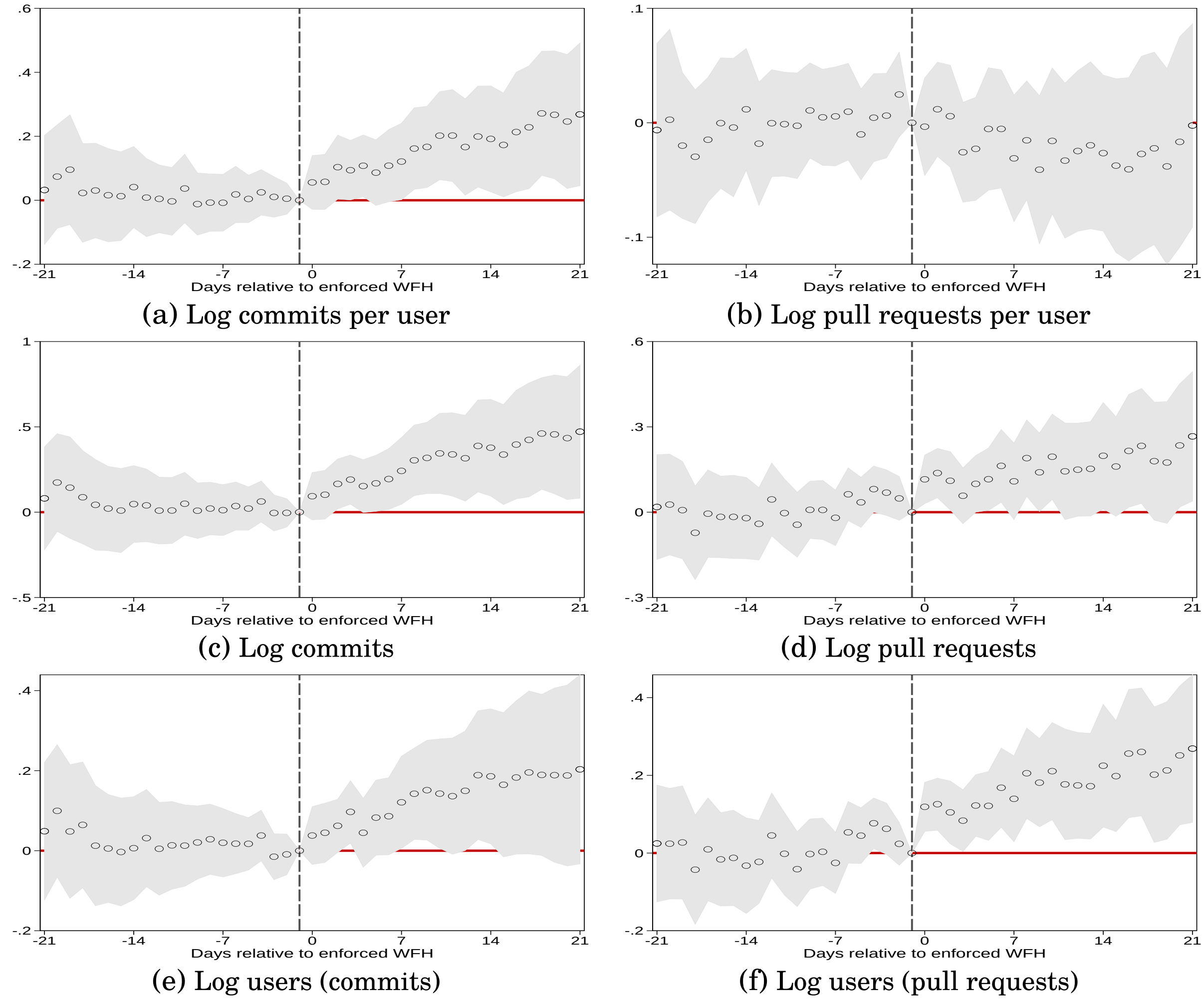
[8] To accomodate zero records for the country-date cells, the commits per user measure is $\big[$log(1 + commits) - log(1 + active commit users)$\big]$, and similarly the pull requests per user measure is $\big[$log(1 + pull requests) - log(1 + active pull users)$\big]$

[9] The "Government Response Index" and the "Economics Support Index". See `https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/index_methodology.md` for documentation.

[10] Omitting the original OxCGRT WFH indicator does not substantially change the results (not shown in the current draft).

[11] $60 \approx 100 \times [\exp(.472) - 1]$.

[12] $23 \approx 100 \times [\exp(.203) - 1]$.



(a) Log commits per user

(b) Log pull requests per user

(c) Log commits

(d) Log pull requests

(e) Log users (commits)

(f) Log users (pull requests)

Figure 5: Event Studies for Productivity

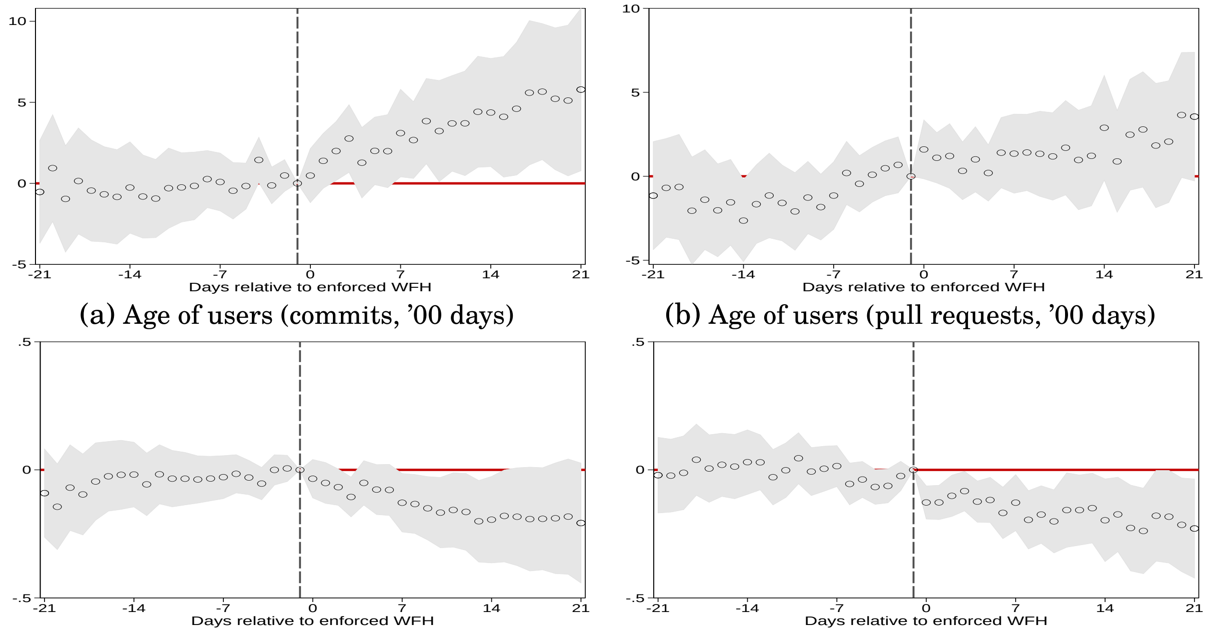
Notes—Event studies documenting changes in productivity, with figures plotting the γ_τ and δ_τ from estimating Equation (1). Day 0 is the date at which the OxCGRT WFH indicator for a country first switches from either a 0 or 1 to a 2 or 3 (Table 1). In the last row, the dependent variable is the log of (1+) the number of recorded users for the respective type of activity (column-wise). All results control for timing group-by-week, date, and country fixed effects, the OxCGRT WFH indicator, two OxCGRT government response indices, and the epidemiology records. Shaded gray areas denote 90% confidence band constructed from standard errors clustered by countries.

per user increased by 31 percent ($p < .048$).[13]

For pull requests, by day 21, I note a 31 percent increase in total pull requests per day ($p < .054$),[14] and a correspondingly large 31 increase in active individuals ($p < .02$).[15] Overall, the estimated change in pull requests per user is effectively zero, with an estimated decrease of .2 percent, and is not significant at any conventional level ($p < .978$).[16]

In sum, I take two key insights from the event study results in Figure 5. First is the validation of state-imposed WFH as an exogenous timing in treatment assignment. The γ_τ estimates, which constitute conditional falsification tests for pre-

[13] $31 \approx 100 \times [\exp(.268) - 1]$.
[14] $31 \approx 100 \times [\exp(.267) - 1]$.
[15] $31 \approx 100 \times [\exp(.269) - 1]$.
[16] $-.2 \approx 100 \times [\exp(-.002) - 1]$.



(a) Age of users (commits, '00 days)

(b) Age of users (pull requests, '00 days)

(c) New accounts (commits)

(d) New accounts (pull requests)

Figure 6: EVENT STUDIES FOR AGE PROFILE

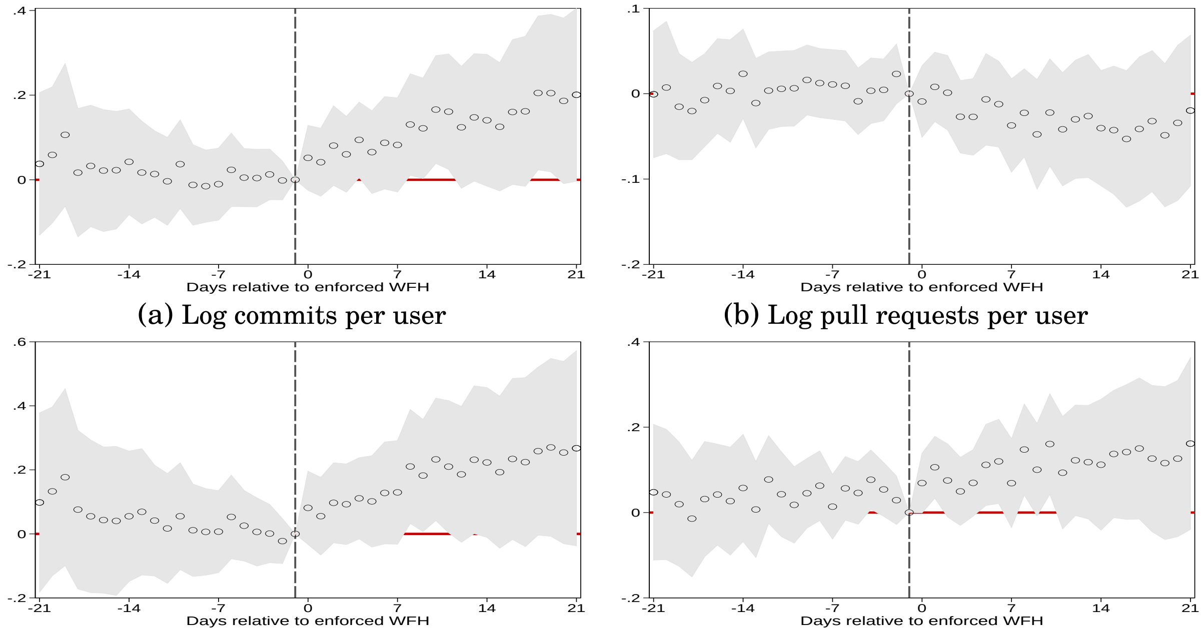
Notes—Event studies documenting changes in active individual age profile, with figures plotting the γ_τ and δ_τ from estimating Equation (1). Day 0 is the date at which the OxCGRT WFH indicator for a country first switches from either a 0 or 1 to a 2 or 3 (Table 1). In the first row, the dependent variable is the age of individuals, in days, scaled down by 100. In the second row, the dependent variable is share of "new individuals"—ln(1+new individuals) - ln(1+total individuals)—with new individuals defined as accounts created only after the last day of 2019. Shaded gray areas denote 90% confidence band constructed from standard errors clustered by countries.

trends, are all virtually zero. This implies that conditional on the covariates (which include the number of confirmed cases), the anticipatory effect of WFH is minimal. I use this finding to motivate the difference-in-differences analyses in Section IV. Second, a substantial portion of the increase in productivity, if any, comes from the extensive margin by an increase in active individuals.

III.B Changes in Age Profile

Motivated by the increase in active individuals on the Github platform following the closure of workplaces (last row of Figure 5), I repeat the event study analyses, this time with age profile as the dependent variable and, separately, controlling for age with the productivity metrics as the dependent variables.

Figure 6 plots the estimates from Equation (1), using user account age and share of new accounts as the dependent variables. The estimates imply that on average, for the commits sample, the age of active individuals increased by approximately 1.6 years by day 21 ($p < .058$). For the pull requests sample, age increases by approximately 1 year by day 21, although this is not significant ($p < .127$). The share of new user accounts also exhibit a distinct decrease after closures of workplaces. For the commits sample, the estimated δ_{21} is -.208, suggesting a -19 percent change



(a) Log commits per user

(b) Log pull requests per user

(c) Log commits

(d) Log pull requests

Figure 7: Event Studies, Controlling for Age Profile

Notes—Event studies documenting changes in productivity, with figures plotting the γ_τ and δ_τ from estimating Equation (1). Day 0 is the date at which the OxCGRT WFH indicator for a country first switches from either a 0 or 1 to a 2 or 3 (Table 1). Identical to first two rows of Figure 5, except with an inclusion of the average age of individuals as a covariate with potentially time-varying effect. Shaded gray areas denote 90% confidence band constructed from standard errors clustered by countries.

in the share of individuals from new user accounts ($p < .144$). For the pull requests sample, the estimated δ_{21} is -.229, suggesting a -20 percent change ($p < .052$). Overall, the event study results suggest, with varying statistical significance, that the average age of the user accounts is increasing, and a decrease in the share of activity from new user accounts.

Figure 7 reports the event study results for productivity, this time controlling for the age profile of active individuals (user account age). Once the average age of active individuals is included, the previous finding in Section III that productivity increases after state-imposed WFH is no longer statistically distinguishable from zero, with p-values of .106 and .149 for commits per user and total commits, respectively, by day 21. For the pull requests sample, what marginally significant result previously found also disappears, with p-values of .712 and .186 for pull requests per user and total pull requests, respectively.[17]

[17] One might consider adding the lagged value of the dependent variable in the above estimations, but this requires a different (more demanding) set of assumptions for consistency. It can be shown that if the model with a lagged dependent variable is the true model but one estimates a fixed effects model instead, then the fixed effects estimate constitutes an upper bound of the true effect (Angrist and Pischke 2009).

III.C Decomposition of DID Estimates

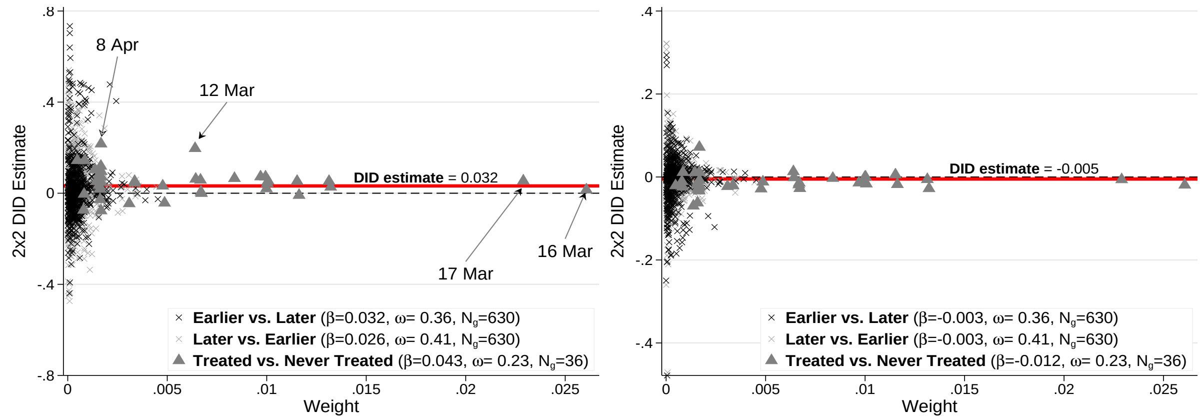
One concern with estimating difference-in-differences with variation in treatment timing is that earlier treated cohorts also end up serving as a control to later treated cohorts, and vice versa (Goodman-Bacon 2019). This potentially leads to (predictable) biases when the treatment effect exhibits variation over time, as suggested by Figure 5. Concretely, suppose that WFH improves productivity, but that individuals take up to four weeks to fully adjust to WFH. In this case, when a late treatment cohort has WFH imposed with only two weeks remaining in the sample period and is compared with an early treatment cohort, a difference-in-differences underestimates the WFH effect. This potentially explains findings of a null effect even if WFH truly improves productivity. To mitigate such concerns, I perform a decomposition of the single difference-in-differences estimate as suggested by Goodman-Bacon (2019).

Figure 8 plots the full set of "2 x 2" DID estimates, where each set corresponds to a combination of groups defined by treatment timing. The red horizontal line indicates the single-coefficient DID estimate, which in the Goodman-Bacon (2019) theorem is the variance-weighted average of the "2 x 2" DID estimates.[18] I highlight two points here. First, while heterogeneity across comparisons of timing groups is a concern with the staggered rollout of WFH, the three grouped coefficients always have the same sign, with comparable estimated magnitudes. In other words, the concern that the "Later vs. Earlier" timing group comparisons have negative estimates because the evolving trend for the late treated groups has not fully developed and thus attenuates the overall DID estimate, can be rejected.

Second, and as expected, the "2 x 2" DID estimates with the largest weights come from comparisons to "pure controls" (or the "Never Treated"—countries that never receive statewide imposed WFH in the sample period) but most of these cluster around zero. Panel (a) of Figure 8 includes annotation for "Day 0" of the treated countries for four combinations with the highest weights and estimates. France, Russia, and Norway are notable countries in this group that disproportionately contribute to the positive DID estimate (Table A2). Notably, the US is not in this group.[19]

[18] So the single-coefficient two-way fixed effects DID estimate would be $\sum_\tau \beta_\tau w_\tau$, where τ indicates one of the three timing groups. In panel (a) of Figure 8 for example, the DID estimate of $0.032 = (0.032 \times 0.36) + (0.026 \times 0.41) + (0.043 \times 0.23)$.

[19] Figure A3 shows the difference between US observations vs the rest of the world. Individuals in the US microsample tend to be more experienced (older accounts), have more followers, follow fewer people, and are more likely to list their company in their profile.



(a) Commits (b) Pull requests

Figure 8: Decomposition of 2×2 DID Estimate

Notes—Plots of the decomposition of the single-coefficient 2x2 difference-in-differences estimate for all possible "2x2" DID estimates, where the dependent variable is log productivity per user, following Goodman-Bacon (2019). In the context of this study, there are only three groups as indicated in the legend (there is no "Treated vs. Already Treated" group). The red horizontal line indicates the single-coefficient two-way fixed effects difference-in-differences estimate. Also reported are the unconditional DID estimates (β) for the three groups, the weights (ω), and group size (N_g). Dates in the plot indicate Day 0 for the treated group in the (arrowed) comparison groups. Table A2 lists notable countries that fall into the four highlighted groups.

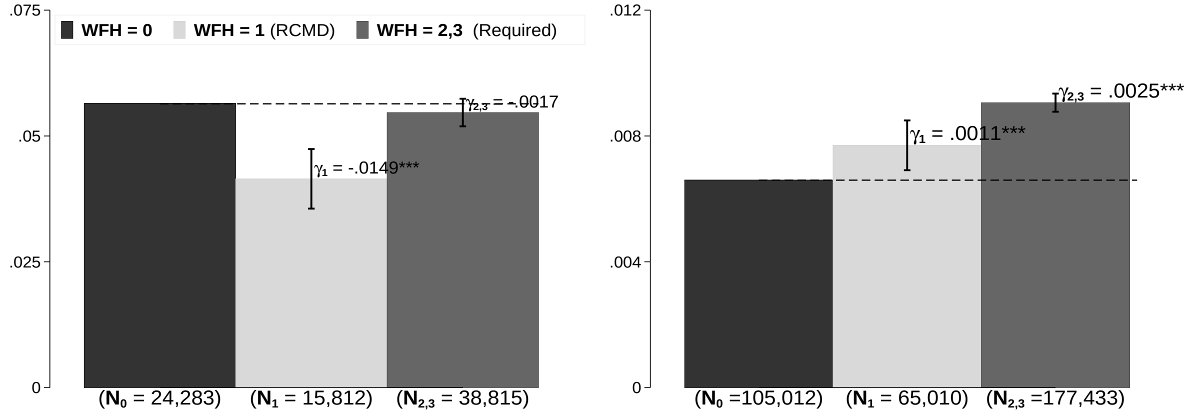
IV Microsample DID

Another way to test if WFH improves productivity is to use the microsample directly, which potentially affords greater precision at the expense of ignoring time-varying effects, and which also avoids serial correlation issues (Bertrand et al. 2004). The absence of pretrends in Section III provides confidence for a direct difference-in-differences comparison. To do this, I bin observations into individual-repository-WFH arm cells and estimate[20]

$$y_{ijk} = \alpha_i + \alpha_j + \boldsymbol{\beta_k X_{ijk}} + \sum_{k \in \{0,1,(2,3)\}} \gamma_k \mathbb{1}\{\text{WFH} = k\}_i + \varepsilon_{ijk}, \tag{2}$$

where y_{ijk} is the commits or pull requests per user i in repo j per day in the WFH arm = k period. α_i and α_j are the user and repository fixed effects, and the γ_k estimates capture the impact of state-imposed WFH on the individual work pattern of developers. Similar to Section III, I combine the OxCGRT WFH coding for 2 and 3 into a single dummy that captures binding state-imposed WFH. Finally, motivated by the above event studies where more individuals become active after lockdowns (Figure 5), and where age profile is a potential confounder (Figure 7), Equation (2) includes individual and repository level observables interacted with a dummy for whether there is *any* WFH-related state regulation—where the OxCGRT indicator

[20] In principle, one can bin observations into user-repo-day cells with indicators for WFH status of days, but most cells will end up unobserved because recorded activity for a user on a given day is likely zero.



(a) Commits

(b) Pull requests

Figure 9: MICROSAMPLE DID

Notes—Figure plots the estimated impact (estimates of γ_k from Equation (2)) of state-imposed WFH. The dependent variables are commits and pull requests per individual-repository per day in a WFH arm. The first bar in each subfigure indicates the baseline—WFH=0 (no WFH). Subsequent bars add back the estimated impacts to the baseline estimate ($\gamma_0+\gamma_\ell$, $\ell = 1$ or $2, 3$). Annotated estimates in figures are the estimates of γ_k. ***, **, and * denotes significance at the 1, 5, and 10 percent level, respectively. Parenthesized numbers ($\mathbf{N}_k$) below bars indicate size of the individual-repository observations for the corresponding WFH arm. Capped vertical bars are 95% confidence intervals from robust standard errors clustered by country.

is strictly positive (OxCGRT WFH>0).[21]

Figure 9 starts by reporting the results from estimating Equation (2) without controls ($\boldsymbol{\beta_k}$ = 0). The estimates from panel (a) suggest that state-imposed WFH has no effect on productivity, with an estimate of -.0017 for $\gamma_{2,3}$ implying that productivity in terms of commits decreases by .2 log points, but this is not significant (p<.22). For recommended WFH, the statistically significant estimate of -.0149 for γ_1 suggest that productivity dropped by approximately -1.5 percentage.

For the pull requests sample in panel (b), there appears to be a measurable increase in productivity. For recommended WFH, the estimate of .011 for γ_1 implies that those individuals in recommended WFH are 1 percent more productive than they would have been, and the estimate of .0025 for $\gamma_{2,3}$ implies a .25 percent increase in state-imposed WFH, with both estimates significant at the 1 percent level.

Figure 10 however, shows that the active individuals captured in the different WFH arms are considerably different, using the pull requests microsample.[22] This finding is similar to the finding in Section III.B where the characteristics of active individuals in the different periods are changing. To address this, I add the interaction of the individual plus repository-level characteristics with a dummy that

[21] Individual-level covariates are: number of public repositories, number of public gists, number of followers, number of following, and age. Repository-level covariates are: number of contributors, number of contributions by others, number of repo-owner contributions, number of stars, number of forks, and age.

[22] Figure A4 shows the differences for the commits microsample which is similar.

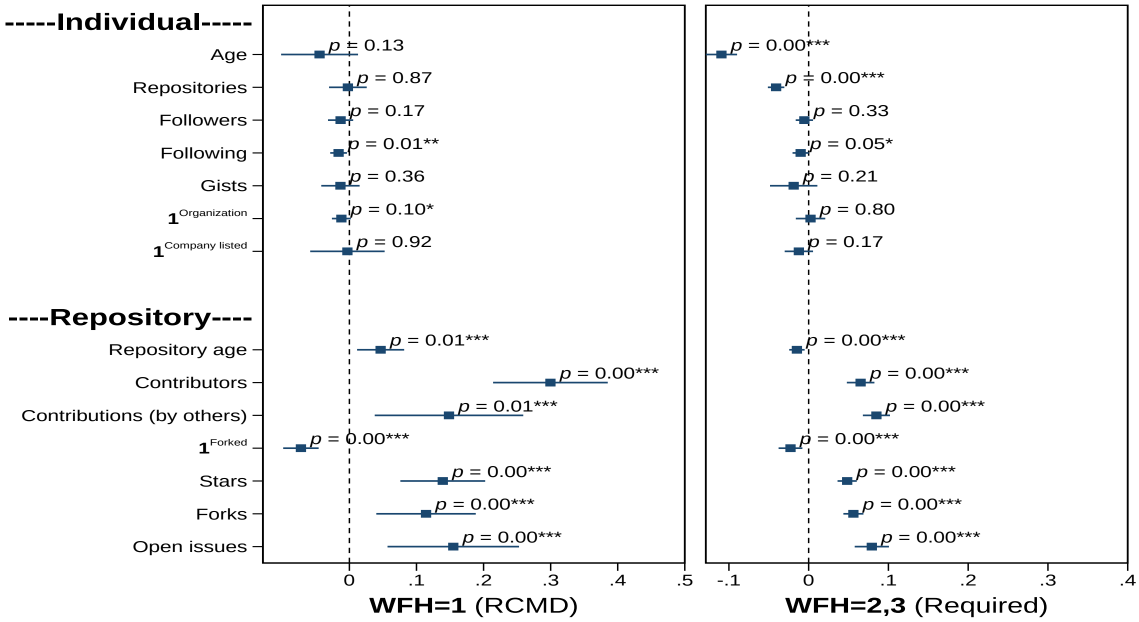


Figure 10: Differences in Observables

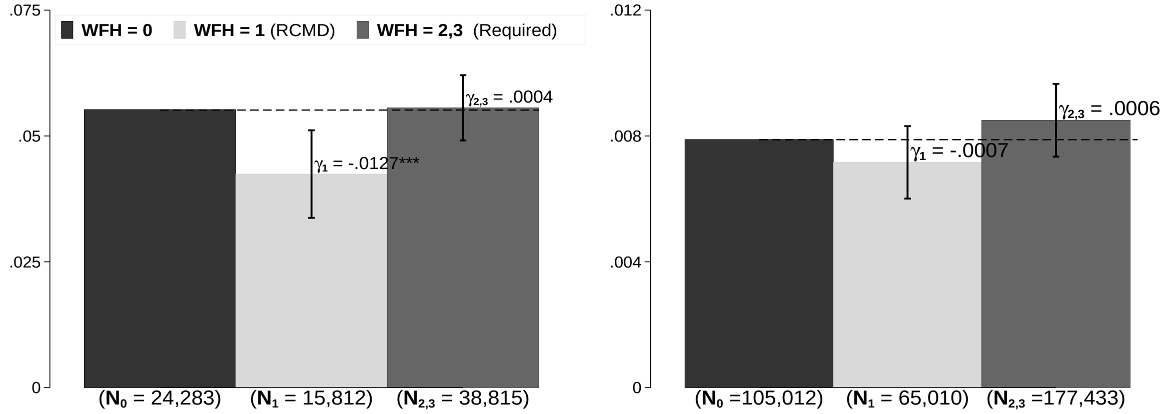
Notes—Differences in means for WFH=1 (recommended WFH) and WFH=2,3 (required WFH), compared to WFH=0 (no state regulation), using the microsample from the pull requests records. The WFH codings are the OxCGRT codings. Units in standard deviations. Repository age is defined as creation date minus 1 Jan 2020; contributions is total number of commits, pull requests, or number of issues opened; the dummy for forked indicates whether the repository was branched out from a preexisting one; stars is a measure of impact (used as a like or bookmark); forks is the number of branching out by other users; and open issues refers to the number of unresolved issues listed in the project. Tables A5–A6 in the Online Appendix tabulates the above results. Number of individuals and repositories captured are 76,830 and 72,923, respectively. The individual and repository level observations are clustered by country and programming language, respectively. ***, **, and * denotes significance at the 1, 5, and 10 percent level, respectively.

equals one for any kind of government response.[23]

Figure 11 reports the results, controlling for differences in covariates across the WFH arms, with the estimated impact of WFH mostly attenuated by an order of magnitude. The γ_k coefficients are now no longer statistically distinguishable from zero, except for γ_1 for the commits sample which still implies a decrease in commits productivity during the recommended WFH period.

Overall, based on the above microsample difference-in-differences findings, I can rule out even modest effects of WFH on productivity—based on Figure 9 without accounting for differences in characteristics across different WFH arms, the estimate implies a .25 percent increase. Given the rather precise standard errors, from the upper bound of the 95 percent confidence interval I can rule out an effect larger than even a modest .28 percent increase in productivity.

[23] $\mathbb{1}_{\{\text{OxCGRT WFH}>0\}}$, where OxCGRT WFH is as defined by OxCGRT in Table 1.



(a) Commits (b) Pull requests

Figure 11: MICROSAMPLE DID WITH CONTROLS

Notes—Figure plots the estimated impact (estimates of γ_k from Equation (2)) of state-imposed WFH. The dependent variables are commits and pull requests per individual-repository per day in a WFH arm. Similar to Figure 9 but with an additional $\mathbb{1}^{\{\text{OxCGRT WFH}>0\}}$ interaction with the individual and repository characteristics. The first bar in each subfigure indicates the baseline—WFH=0 (no WFH). Subsequent bars add back the estimated impacts to the baseline estimate ($\gamma_0+\gamma_\ell$, $\ell=1$ or $2,3$). Annotated estimates in figures are the estimates of γ_k. ***, **, and * denotes significance at the 1, 5, and 10 percent level, respectively. Parenthesized numbers ($\mathbf{N}_k$) below bars indicate size of the individual-repository observations for the corresponding WFH arm. Capped vertical bars are 95% confidence intervals from robust standard errors clustered by country.

V Discussion

V.A Interpretation and Sensitivity

All the estimates under this study fall under an intention-to-treat basis (ITT) where, geocoding notwithstanding, the assignment to state-imposed WFH status is observed but compliance is not. This framework allows one to place the estimated ITT effects in the context of a set of assumptions (Angrist et al. 1996). Under this ITT framework, the estimated ITT effect is a probability-weighted average of the ITT effects by group: ITT $= \sum_{g\in G} \text{ITT}_g \cdot \pi_g$, with $G \in \{C, D, AT, NT\}$ for compliers, defiers, always-takers, and never-takers, ITT_g defined as the compliance group-specific ITT effect $E\left[y_t(1) - y_t(0) | Z = 1, i \in g\right]$, and π_g defined as the probability that an individual is of type g.

Under state-imposed closure of workplaces, defiers are those individuals who choose to WFH before there was state-imposed WFH, and then switch to going back to the office once state-imposed WFH starts. Assuming that no such individuals exist, where they specifically oppose the WFH assignment, then π_D = 0.[24]

Further, if the exclusion restriction applies where only the actual compliance with the WFH assignment has an impact on productivity, so groups unaffected by the WFH assignment (always-takers and never-takers)[25] should have no effect on

[24] Or the monotonicity assumption, that $T_i(Z_i = 1) \geq T_i(Z_i = 0)$, where $Z = 1$ for state-imposed WFH and $T = 1$ for compliance with WFH.

[25] In this context, always-takers are Github users who are always going to WFH regardless of

productivity, then $\text{ITT}_{NT} = \text{ITT}_{AT} = 0$. The above two assumptions collapse the observed ITT effect into simply the ITT effect for compliers (ITT = $\text{ITT}_C \cdot \pi_C$) which, if given the actual compliance rate (eg 0.5) of individuals in the state-imposed WFH would retrieve the local average treatment effect (LATE) in an instrumental-variables type setup. The lower the compliance rate is, the larger the estimated impact since one needs to adjust the observed estimates the by rate of compliance to the actual WFH regulation. For example, the LATE when there is a 50 percent compliance will be the estimated ITT effect augmented by a factor of $2 = 1/(\text{compliance rate})$. Using the optimistic estimate of .28 percent increase from Figure 9, and suppose only 5 percent of the individuals in the sample complied with state-imposed WFH, then a back-of-the-envelope calculation suggests that LATE is a 5.6 percent increase in productivity. If compliance is 95 percent instead, LATE would be a .3 percent increase in productivity.

In the context of the state-imposed lockdowns, I find the first assumption where $\pi_D = 0$ to be reasonable, but the second assumption less so. There are a few plausible reasons why the exclusion restriction fails in this context. First, the state-imposed WFH is essentially a government response to the pandemic, and this likely coincides with other types of fiscal responses which might have an impact on both employee and employer behavior, including work arrangements and productivity. Second, the closure of workplaces also coincides with the closure of other places (eg parks and recreation), which also potentially alters work patterns. Third, closure of workplaces also applies to cohabitants, including kids during closures of schools, and this potentially affects productivity as well.

If the first assumption is satisfied but the second exclusion restriction assumption is not, one can still characterize the impact of WFH by focusing on the bias. With violation of exclusion restriction, the ITT becomes $(\text{ITT}_C \cdot \pi_C + \text{ITT}_{NC} \cdot \pi_{NC})$, where the second term is the effect from non-compliers (the always-takers and never-takers), implying that the bias in LATE is $\left(ITT_{NC} \cdot \frac{\pi_{NC}}{\pi_C}\right)$. So a positive ITT_{NC} implies that ITT estimates are upward biased, and would allow one to confidently rule out productivity gains from WFH given the small ITT estimates, while a negative ITT_{NC} implies that the above estimates are downward biased. Furthermore, the larger rate of noncompliance, the larger that bias. Suppose the rate of compliance is the same as the rate of noncompliance ($\pi_C = \pi_{NC}$), and that individuals who do not comply with state-imposed WFH have a .5 percent increase in productivity because of the positive impact of non-WFH-related government responses, then the LATE estimate will be upward biased by .5 percent.

state regulations on workplace closures. Never-takers are Github users who are always going work in office.

Finally, suppose that the no-interference assumption fails, where for example a team member on a project shifting to WFH affects another team member located in a place without WFH, with the disruption negatively affecting the productivity of the non-WFH team member. In such a scenario, the estimated impact of WFH would be biased upwards since the comparison group gets negatively affected as well.

V.B Related Studies

Here I place my findings in context with three related studies on WFH productivity. The finding in this study complements the Bloom et al. (2015) seminal study on the causal effect of WFH on productivity using a randomized controlled trial in Ctrip, a travel agency in China. The 249 participants (before attrition) in the study are call center representatives whose work essentially involves answering calls from customers, taking orders, and making calls to hotels and airlines to place the orders. Ensuring that employees in both the WFH and control group have the same IT equipment and internet access, Bloom et al. (2015) find that WFH improves productivity. Specifically, they find increases along both the extensive and intensive margin, where the 13 percent increase in productivity comes from a 9 percent increase in working time and a 4 percent increase from calls per minute.

Using observational data from individuals in tech-related industries complements the Bloom et al. (2015) study in at least two ways. First, the individuals captured here are not aware that they are being "monitored" to begin with, and so have no motivation to artificially inflate their work patterns to appear to be more productive when WFH. Bloom et al. (2015) are careful to note and argue against potential Hawthorne effects, but one of the authors is nonetheless a co-founder of the firm and later returns as CEO and chairman.

Second, the type of work captured in Bloom et al. (2015) is well-defined with obvious metrics for productivity. Since short-run outputs in tech-related (or even science-related) work have no obvious milestones (eg how many bugs will appear and how many features to add are seldom clear from the onset of a project), this study provides insight on how productivity is affected when work is non-transactional and non-routine, as opposed to order-taking in the context of Bloom et al. (2015).

Another related study is by Choudhury et al. (2021) in the context of patent examiner work which can be performed relatively independently. Instead of estimating productivity gains in WFH however, they focus on WFA (working-from-anywhere) and find a substantial 4.4 percent increase in productivity WFA compared to WFH. They also estimate positive effects in WFH, but since their experimental setup pivots around union negotiations providing exogenous timing in WFA but not WFH, the WFH estimates as noted by the authors do not have a causal interpretation.[26]

A third related study is by Wang et al. (2020), who have a similar approach to the one in this paper in that they take observed activities from an online platform, infer user locations from posts, and then compare productivity before and after lockdowns. Specifically, they use TikTok posts and assign users to one of ten US states in their sample using hashtags (eg #southdakota). They find a 7 percent increase in content volume but a 1 percent decrease in content novelty (via word embeddings). Using data from Github complements Wang et al. (2020) in that some creativity is needed in ad-hoc tasks and problem-solving in software development, but not to the extent required in social media content creation.

Finally, given that employees in tech-related industries have work that is particularly well-suited to remote work to begin with (BBC 2020a), a null finding suggests that productivity gains from WFH are not as obvious as some suggest (BBC 2020a,b), especially for non-routine work in other sectors that do not have obvious high-frequency milestones. The evidence in this study is also consistent with contemporaneous views that WFH mostly has no impact on productivity (YouGov 2020, Figure A1), especially in tech ["better educated and higher paid"] sectors where employers believe there has been "less productivity loss from remote working" (Bartik et al. 2020). If one is willing to accept ICT-related work as the type of work most suitable for transition to WFH (Bartik et al. 2020; BBC 2020a), then the estimates in this paper can be interpreted as upper bound estimates of WFH impact on productivity across all sectors of the economy.

V.C Other Considerations of WFH

One potential concern is that the effect of WFH on productivity is augmented by access to high-speed broadband for instance. Dingel and Neiman (2020) and Gottlieb et al. (2020) for example, find that the share of employees that can WFH is positively correlated with the wealth of a country, which in turn is likely correlated with high-speed internet penetration. Given that the selection of individuals captured in this study are mostly in tech or science-related industries, I find differences across countries in high-speed internet penetration to be less of a concern (as sug-

[26] On a related note, Kelly (2020a) reports on an unintended consequence of WFA during the pandemic—employers downward adjusting salaries because employees have moved away to cheaper locations.

gested by BBC 2020a and unlike Hjort and Poulsen 2019 for instance). A related concern is in supply-side implementations of WFH. In the context of Bloom et al. (2015), the firm ensured that employees WFH are equipped to do so. This may not necessarily be the case for ad-hoc and state-imposed WFH, where employees lack access to their usual tools and equipment.

Another consideration of WFH and productivity has to do with incentives. If it was obvious WFH would improve productivity, why are firms not already doing it and why is it not more common? As addressed in Bloom et al. (2015), a second-mover advantage type of scenario is a potential explanation. Firms want productivity gains from WFH, if any, but are not willing to first test it out themselves, choosing to wait for similar firms to experiment with WFH first and thus avoid the cost of experimenting. The COVID-19 pandemic, however, completely changes the incentive structure. Most firms are now forced to experiment, and while WFH might not be the new norm, discretionary firm-level implementation of WFH in the post-pandemic economy will reveal how effective it really is.

So what if WFH does not affect productivity? Perhaps this is good news to policy-makers during the pandemic since they can revert to considerations about health and safe-distancing without having to worry about productivity tradeoffs. Other than productivity, WFH in general have been linked to facilitation of "work-life balance" (Kelly et al. 2014), travel times and crowding on commutes to work, reduction of emissions (Choudhury et al. 2020; Bloom et al. 2015), or even just reduce office space usage, including saving on rental (Bloom 2020; Kelly 2020b). On the other hand, WFH might have adverse and unexpected effects, such as those on the psychosocial dimension, where WFH leads to reduced social interactions for example (Bloom et al. 2015; Kelly 2020b) and in the pandemic, where everyone is WFH, the adverse effects of co-telecommuting with cohabitants.[27] All these remain open questions.

I end this draft with two notes. First, the finding in this study that WFH has minimal impact on productivity does not rule out positive selection effects, where employees who know they can be (more) productive WFH choose to WFH when allowed to (Bloom et al. 2015; Choudhury et al. 2021; Kelly 2020b). Bloom et al. (2015) in particular, find the selection effect to be *twice* their estimated causal impact of WFH on productivity.

Second, one may want to consider what measures exactly would capture productivity: whether commits and pull requests are good measures of productivity in the Git workflow, and more generally what would constitute good measures of

[27] In the absence of cleaner data, surveys might shed light on how the family structure or cohabitating with others while WFH during a pandemic affects mental health and work patterns.

productivity in non-assembly line, non-routine type of work, and to what extent can one account for the quality of output. If employers cannot properly measure these to begin with, then asking how WFH affects productivity is a red herring.

References

Alipour, J.-V., O. Falck, and S. Schüller (2020). Germany's Capacities to Work from Home. *CESifo Working Paper No. 8227*.

Angelucci, M., M. Angrisani, D. Bennett, A. Kapteyn, and S. Schaner (2020). Remote Work and the Heterogeneous Impact of COVID-19 on Employment and Health. *IZA DP No. 13620*.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*(434), 444–455. doi:10.1080/01621459.1996.10476902.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton : Princeton University Press, c2009.

Baker, A. C., D. F. Larcker, and C. C. Y. Wang (2021). How Much Should We Trust Staggered Difference-In-Differences Estimates?

Banbury, S. and D. C. Berry (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology 89*(3), 499–517. doi:https://doi.org/10.1111/j.2044-8295.1998.tb02699.x.

Bartik, A. W., Z. Cullen, E. L. Glaeser, M. Luca, and C. Stanton (2020). What Jobs are Being Done at Home During the COVID-19 Crisis? Evidence from Firm-Level Surveys. *HBS No. 20-138*.

BBC (2020a). Coronavirus: Fujitsu Announces Permanent Work-From-Home Plan.

BBC (2020b). Coronavirus: Twitter Allows Staff to Work From Home 'Forever'.

BBC (2020c). Facebook and Google Extend Working From Home to End of Year.

BBC (2020d). Netflix Boss: Remote Working Has Negative Effects.

Bertrand, M., E. Duflo, and S. Mullainathan (2004, feb). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Bloom, N. (2020). How Working From Home Works Out. *Stanford Institute for Economic Policy Research*, 1–8.

Bloom, N., J. Liang, J. Roberts, and Z. J. Ying (2015). Does Working from Home Work? Evidence from a Chinese Experiment. *The Quarterly Journal of Economics 130*(1), 165–218. doi:10.1093/qje/qju032.

Brynjolfsson, E., J. J. Horton, A. Ozimek, D. Rock, G. Sharma, and H.-Y. TuYe (2020). Covid-19 and Remote Work: an Early Look At Us Data.

Callaway, B. and P. H. C. Sant'Anna (2020). Difference-in-Differences with multiple time periods. *Journal of Econometrics*. doi:10.1016/j.jeconom.2020.12.001.

Choudhury, P., C. Foroughi, and B. Z. Larson (2020). Work-From-Anywhere: The Productivity Effects of Geographic Flexibility. *Academy of Management Proceedings 2020*(1), 21199. doi:10.5465/AMBPP.2020.225.

Choudhury, P., W. W. Koo, and X. Li (2020). Working (From Home) During a Crisis: Online Social Contributions by Workers During the Coronavirus Shock. *HBS No. 20-096*.

Choudhury, P. R., C. Foroughi, and B. Larson (2021). Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal 42*(4), 655–683. doi:https://doi.org/10.1002/smj.3251.

DeFilippis, E., S. M. Impink, M. Singell, J. T. Polzer, and R. Sadun (2020). Collaborating During Coronavirus: The Impact of COVID-19 on the Nature of Work. *HBS No. 21-006*.

Dingel, J. I. and B. Neiman (2020). How Many Jobs Can be Done at Home? *Journal of Public Economics 189*(104235). doi:10.1016/j.jpubeco.2020.104235.

Forsgren, N. (2020). Octoverse Spotlight: An Analysis of Developer Productivity, Work Cadence, and Collaboration in the Early Days of COVID-19.

Goodman-Bacon, A. (2019). Difference-in-Differences With Variation in Treatment Timing. *Working paper*.

Gottlieb, C., J. Grobovsek, and M. Poschke (2020). Working From Home across Countries. *CEPR*. doi:10.4324/9781003106432-3.

Hjort, J. and J. Poulsen (2019). The Arrival of Fast Internet and Employment in Africa. *American Economic Review 109*(3), 1032–1079. doi:10.1257/aer.20161385.

Kelly, E. L., P. Moen, J. M. Oakes, W. Fan, C. Okechukwu, K. D. Davis, L. B. Hammer, E. E. Kossek, R. B. King, G. C. Hanson, F. Mierzwa, and L. M. Casper (2014, may). Changing Work and Work-Family Conflict: Evidence from the Work, Family, and Health Network. *American Sociological Review 79*(3), 485–516. doi:10.1177/0003122414531435.

Kelly, J. (2020a). Facebook CEO Mark Zuckerberg's Power Move May Be The Biggest Game Changer For The Job Market. *Forbes*.

Kelly, J. (2020b). Here Are The Companies Leading The Work-From-Home Revolution. *Forbes*.

Khanna, V. (2020). The Economic Consequences of Working From Home. *The Straits Times Covid 19 Special: Economic Affairs*.

Kummer, M., O. Slivko, and X. M. Zhang (2020, aug). Unemployment and Digital Public Goods Contribution. *Information Systems Research 31*(3), 801–819. doi:10.1287/isre.2019.0916.

Lerman, R. and J. Greene (2020). Big Tech Was First to Send Workers Home. Now It's in No Rush to Bring Them Back. *The Washington Post*.

Petherick, A., T. Hale, T. Phillips, and S. Webster (2020). Variation in Government Responses to COVID-19. *Blavatnik School Working Paper*.

Ruprechter, T., M. H. Ribeiro, T. Santos, F. Lemmerich, M. Strohmaier, R. West, and

D. Helic (2021). Volunteer contributions to Wikipedia increased during COVID-19 mobility restrictions. *arXiv*, 1–31. doi:2102.10090.

Stanton, C. T. and P. Tiwari (2021). Housing Consumption and the Cost of Remote Work. *NBER Working Paper 28483*.

Wang, X., R. Mousavi, and Y. Hong (2020). The Unintended Consequences of Stay-at-Home Policies on Work Outcomes: The Impacts of Lockdown Orders on Content Creation.

Whiting, K. (2020). Is Flexible Working Here to Stay? We Asked 6 Companies How to Make It Work. *World Economic Forum*.

YouGov (2020). Huffpost: Working From Home.

A Appendix

A.1 Data build details

To build the user-repository-WFH panels, I proceed as follows:

1. From Github's open-access public dataset on Google BigQuery, I query user commits from `Jan-Jun 2020`. This gives a user-commit log record with timestamps, author names, and repository name.

2. I curate a list of usernames in four ways: (a) using the author name of a commit through the Github search API, (b) using the username string from the repository name (e.g. `johnx/projecta` implies `projecta` belongs to user `johnx`), and later from (c) users who raise/close issues, and (d) users who submit a pull request. To retrieve user-level information (e.g. location, user type, repositories, etc.) from the list of usernames, I query Github's User API. The majority of usernames are successfully queried (620,922 of 626,488 or 99.1%), with the minority having 404 or 502 HTTP response error codes at the time of query.

3. *Geocoding*. To geocode users to a country (or U.S. state), I query the `OpenStreetMap (Nominatim) API` from the Python `geocoder` library, using the location strings that users enter into their account. Approximately half the users have a location string (309,247 or 49.4%). Each successful query returns a hierarchy of geographical information `country-state-city-county`, with a confidence score for reliability of the result (see for example Table A12). For geocoding to country-dates to get the government-enforced WFH status, I retain only country for non-U.S. countries and states for the U.S. observations (because this is the level of granularity in the OxCGRT, see below).

   Almost all users with a location string can be geocoded to a country/U.S. state (303,403 or 98.1%). Almost all location strings can be geocoded—there are 47,681 unique location strings from all the users in the sample, of which

42,552 (89.24%) can be successfully geocoded.

4. *WFH records.* From the geocoding, each activity record (commits, raise/closing of issues, pull requests) have a (`country/U.S. state - date`) tuple that I can then map to government enforced-WFH (work from home/closure of workplaces) policies in the OxCGRT. This is their `C2_Workplace` indicator. Cognizant that federal and state governments may vary in their timing, I retain the flag indicators for when a policy was targeted at sub-regions instead of a nation-wide enforcement (for subsequent robustness checks). This is their `C2_Flag`.

   In the panels, for all non-U.S. countries, the WFH indicator for each user-repo-WFH cell is the WFH based on the user's country. For those users geocoded to the U.S. states, the WFH indicator is based on the individual U.S. states, since this is the level of granularity that the OxCGRT currently offers.

   Alternatively, for county-level policies of the U.S. sample, I combine two sources of business closure records: a complete record at the state level from the COVID-19 US State Policy Database[28] and a partial record of 569 counties from crowdsourcing[29]. Where available, I use the `[business_closed_date]` record from the county-level crowdsourced records. For the remaining counties, I use their state-level `[Closed other non-essential businesses]` record. From these, 258 counties have earlier localized closures relative to the state, while 37 counties have later closures.

5. *Repository records.* Records directly from the commits archive contains 245,506 repositories. To get their repo-level information (e.g. contributors, language, open issues, etc.), I use the Github Repository API. Most of the repositories can be succesfully queried at the time of query (240,150 or 97.9%).

   A minority of commits in the sample are to multiple repositories (usually the same project under different forks). I map these multi-repository commits to the original repository (forked = False), so that language, contributors, and open issues records etc. belong to the original repository.

6. *Pull requests and issues record.* From the initial repository records, I also query their historical record of pull requests and issues (both opening and closing), retaining only records created in the year 2020. These records, with the user record, are mapped to a WFH status as described above. For closure of issues, I use the recorded assignee as the user who "resolved the issue". This step yields 39,958 closed issues that can be geocoded (included in analyses); 253,632 opened issues, and 288,175 pull requests.

[28] `http://www.tinyurl.com/statepolicysources`.

[29] `https://docs.google.com/spreadsheets/d/133Lry-k80-BfdPXhlS0VHsLEUQh5_UutqAt7czZd7ek/edit#gid=0`.

A.2 Figures and Tables

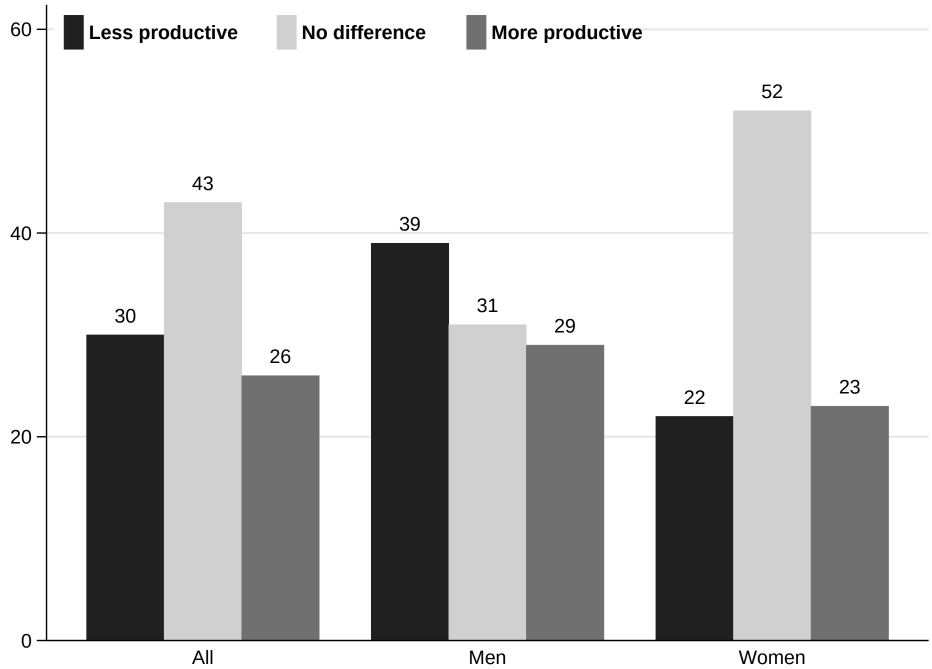


Figure A1: Self-Reported Impact of WFH on Productivity
Notes—Plot shows YouGov 2020 survey responses of 1000 US respondents on whether WFH improves productivity.
Source: YouGov (2020).

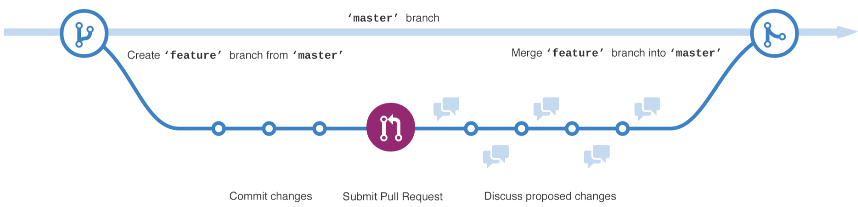


Figure A2: Git Branching Workflow
Notes—Figure shows the basic Git workflow, focusing on commits and pull requests in the context of a main deployable branch. Graphic taken directly from source.
Source: https://guides.github.com.

Table A1—*List of Countries That Rollback WFH in Sample*

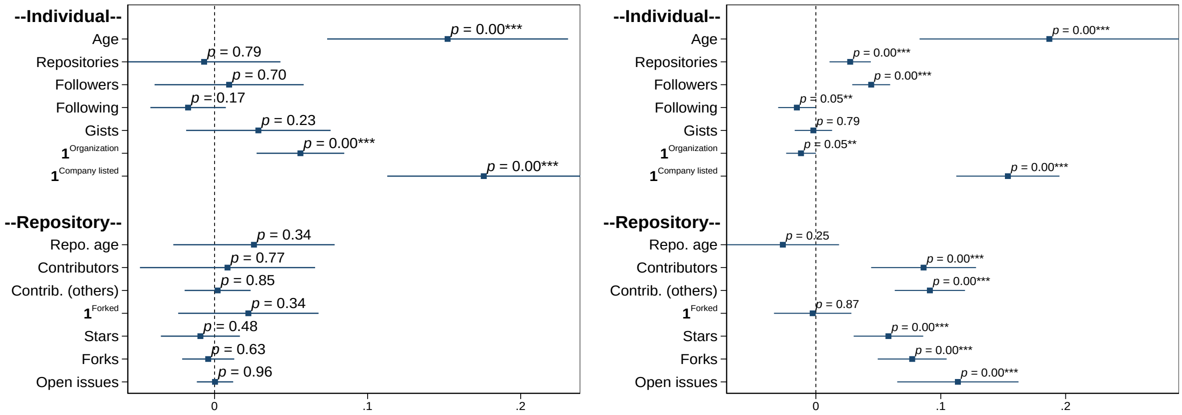
Country	WFH Day 0	Rollback Day 0	WFH rolled back to
Turkmenistan	24Mar	1Apr	0
Ghana	30Mar	20Apr	1
Greenland	18Mar	27Apr	0
Cameroon	18Mar	1May	0
Italy	22Feb	4May	1
Greece	12Mar	5May	0
Lesotho	18Mar	6May	1
Mali	25Mar	10May	1
Luxembourg	16Mar	11May	1
Burkina Faso	21Mar	14May	0
Australia	23Mar	15May	1
Slovenia	20Mar	18May	0
Botswana	2Apr	20May	1
Czech Republic	13Mar	25May	1
Timor-Leste	28Mar	27May	0
Bangladesh	19Mar	31May	1
Laos	30Mar	1Jun	0
Rwanda	21Mar	2Jun	0
Slovak Republic	13Mar	3Jun	0
Thailand	17Mar	6Jun	1
Tunisia	22Mar	8Jun	0
Mauritius	20Mar	12Jun	0
Guinea	27Mar	15Jun	1
Falkland Islands	26Mar	16Jun	0
Romania	12Mar	17Jun	1
Singapore	7Apr	19Jun	1
Saudi Arabia	16Mar	21Jun	1
Cayman Islands	23Mar	22Jun	0
Morocco	20Mar	24Jun	1
Ireland	27Mar	26Jun	1
Dominica	1Apr	27Jun	0

Notes—Table enumerates countries that rolled back state-imposed WFH during the sample period of Jan–Jun 2020, corresponding to timing of rollbacks in Figure 2. Second column shows the first day of state-imposed WFH (OxCGRT WFH $\in \{2, 3\}$). Third column shows the first day of rolling back state-impose WFH; that is, having the WFH indicating step down from either 2 or 3 to a 0 or 1.

Table A2—*Selected Treated Countries as Reference*

Country	Day 0	Commits	Users
Portugal	12Mar	2,348	676
Norway	12Mar	2,179	733
Romania	12Mar	1,064	356
Greece	12Mar	614	253
Austria	16Mar	3,359	1,059
Turkey	16Mar	2,159	676
Hungary	16Mar	1,705	562
Chile	16Mar	1,239	278
Luxembourg	16Mar	1,015	287
Sri Lanka	16Mar	868	312
Egypt	16Mar	460	187
Lithuania	16Mar	430	82
Honduras	16Mar	138	53
France	17Mar	15,457	5,508
Russia	17Mar	8,441	3,013
Switzerland	17Mar	3,661	1,546
Ukraine	17Mar	3,173	1,023
Brazil	17Mar	3,143	1,224
Thailand	17Mar	329	165
Bosnia and Herzegovina	17Mar	173	54
Seychelles	08Apr	143	54
Japan	—	20,305	6,340
Sweden	—	4,317	1,627
Taiwan	—	1,807	668
Bulgaria	—	1,477	276
Belarus	—	1,061	367

Notes—Selected treated countries for reference with Figure 8. Countries shown here are those with WFH enforcement starting on 12 March, 16 March, 17 March, and 8 April for the year 2020. Those countries with small sample share (commits < 100) are not shown. Countries sorted by WFH enforcement date and commits size in the sample period (Jan–Jun 2020).



(a) Commits sample (b) Pull requests sample

Figure A3: US vs. Rest of World

Notes—Difference in means for the micro-samples. Difference is between the US vs. the rest-of-world, estimated by regressing the baseline covariates on the US dummy and performing a t-test for the dummy. Tables A7–A10 Standard errors are robust. ***, **, and * denotes significance at the 1, 5, and 10 percent level, respectively.

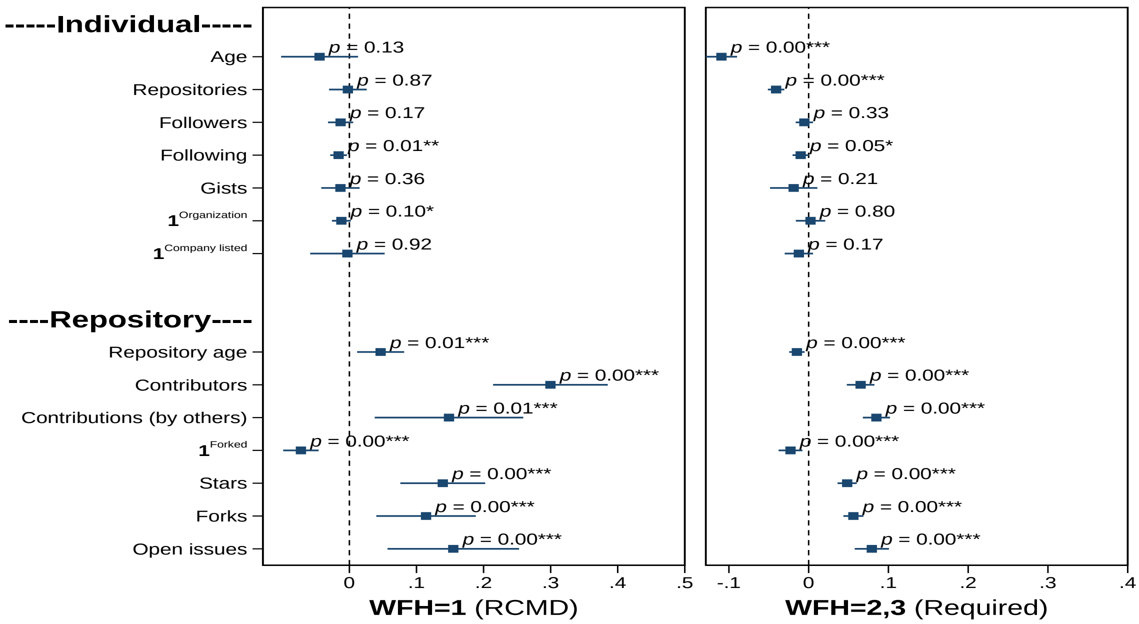


Figure A4: Differences in Observables (Commits)

Notes—Differences in means for WFH=0 (no state regulation), WFH=1 (recommended WFH), WFH=2,3 (required WFH), using the microsample from the commits records. Units in standard deviations. For the repository characteristics, repository age is defined as creation date minus 1 Jan 2020; contributions is total number of commits, pull requests, or number of issues opened; the dummy for forked indicates whether the repository was branched out from a preexisting one; stars is a measure of impact (used as a like or bookmark); forks is the number of branching out by other users; and open issues refers to the number of unresolved issues listed in the project. Tables A3–A4 of the Online Appendix tabulates the above results. Number of individuals and repositories captured are 22,183 and 25,859. The individual and repository level observations are clustered by country and programming language, respectively. ***, **, and * denotes significance at the 1, 5, and 10 percent level, respectively.