Air Quality and the Productivity of High-Skilled Workers: Evidence from GitHub

Felix Holub* Beate Thies†

March 20, 2022

DO NOT CITE OR CIRCULATE WITHOUT PERMISSION

Abstract

We study how air pollution affects productivity and work patterns among highly skilled tech workers. Using data on coding activity from *GitHub* for a sample of 26,000 software developers across 42 countries during the period 2014-2019, we estimate the causal effect of daily particulate matter ($PM_{2.5}$) concentration on activity and choice of tasks. Our findings imply that air pollution has negative effects on the number of actions performed, but these effects differ across tasks: While developers produce less code, their activity in interactive tasks is unaffected. We demonstrate that coders adjust to pollution-induced productivity declines by switching to easier tasks, which alleviates effects on output quantity. We find that a lower bound on the loss in output value due to air pollution in our sample is $15.7 million.

JEL Codes: D24, J22, J24, L86, Q52, Q53

*Department of Economics, Goethe University Frankfurt, (mail: holub@econ.uni-frankfurt.de)
†Department of Economics, University of Mannheim, (mail: bthies@mail.uni-mannheim.de)

1 Introduction

Air pollution is a key environmental threat to health. It affects mortality (Deryugina et al., 2019), hospital admissions (Schlenker and Walker, 2016), child health (Currie et al., 2015), and pharmaceutical expenditures (Deschênes et al., 2017). Health effects arise from deteriorated pulmonary and cardiovascular functioning, but also from irritation in ear, nose, throat, and lungs, or from particles entering the brain through the central nervous system, provoking inflammation and cortical stress (cf. Manisalidis et al., 2020). Through these physiological channels, air pollution can adversely affect labor market performance, both in physically- and in cognitively-demanding occupations. This is especially relevant for small particulate matter ($PM_{2.5}$), a pollutant that can penetrate into buildings.

Several studies document a negative impact of pollution on productivity in manual and routine tasks (e.g., Adhvaryu et al., 2019; Chang et al., 2016, 2019; Graff Zivin and Neidell, 2012; He et al., 2019). A small number of papers examines the relationship between air quality and performance in more cognitively-demanding tasks and occupations, e.g.baseball umpires or politicians (e.g., Archsmith et al., 2018; Heyes et al., 2019). However, none of these studies analyzes an occupation that can be considered representative of the work environment in highly-paying jobs which form the backbone of the modern knowledge economy. Such jobs are characterized by cognitively-demanding tasks, but also frequent collaboration, and flexibility in work schedules and task choice. Each of these characteristics might affect the severity of pollution-induced productivity shocks. In particular, high-skilled workers in such flexible work environments might be able to adjust to productivity shocks along several margins, e.g. working hours or task choice. In the rather inflexible settings studied so far, an analysis of this type of worker response is not feasible.

This papers examines how air pollution affects the productivity of high-skilled workers in jobs that form the backbone of the modern economy and investigates how these workers adapt their work patterns in response to pollution-induced productivity shocks. We study software development as an occupation representative for these jobs, as it is cognitively demanding, collaborative, flexible and multidimensional. Moreover, software devlopers' output generates high value for consumers, other industries, governments, and the research community.[1]

We collect data on *GitHub* users and their activity in public repositories between 2014 and May 2019. GitHub is the world's largest platform for hosting and jointly working on coding projects. We focus on 26,000 users across 42 countries who work on company-owned repositories, indicating that they are professional developers. In the data, we observe the location of a user as well as records of several actions with their respective timestamps. We construct a user$\times$day panel including measures of work quantity, quality, and task choice. The primary outcome for work quantity is the total number of actions performed. We also consider subcategories of actions, e.g. number of commits (code changes) and comments written in discussion fora because they represent distinct types of work, individual coding activity (commits) vs. interactive work (comments). We further construct variables to analyze whether workers switch to easier tasks when facing productivity shocks. Based on developers' locations, we match their output to city-level air quality monitor data on particulate matter smaller than 2.5 μm (PM2.5).

To account for the endogeneity of air quality when estimating a model of user activity, we in-

[1]Median pay of software developers in the US was $110,140 in 2020 (Bureau of Labor Statistics, 2021).

strument PM2.5 concentration with daily average wind direction, exploiting the effect of regional air pollution transport (as suggested by Deryugina et al., 2019).

Our research produces two key results. First, if a developer is exposed to unusually high levels of PM2.5, relative to the city$\times$month$\times$day of week specific average, the number of daily actions observed on GitHub falls by 3.9%. This effect is mainly driven by a decline in individual coding activity: The number of commits, i.e. submitted code changes, decreases by 6.5%. On the other hand, collaborative or interactive work (e.g., commenting on issues) is less affected.

Second, on high-pollution days users switch towards less complex tasks. Code submitted in pull requests[2] changes 4.3% fewer files and contains 7.4% fewer new lines, pointing to a reduction in the complexity of coding tasks. The finding that users focus on easy tasks is confirmed when considering actions on the platform that relate to discussing issues. The share of actions referring to easy issues increases by 7.1% relative to the mean. These findings imply that developers use their flexibility in choice of tasks to adapt to health shocks. Among users with a stronger adjustment response to PM2.5 exposure, effects on work quantity are attenuated. This form of adaptation in a flexible work environment might explain why, compared to other professions, the effect of particulate matter is relatively small in our setting.

To assess the economic relevance of our results, we leverage novel data from a platform called *Gitcoin* where GitHub users can offer payments to incentivize external contributions to their projects. We derive estimates for the value of GitHub activity in USD and find that, despite their relatively small magnitude compared to other occupations, air pollution effects on software developers translate into relevant monetary damages. A lower bound on the loss in output value due to PM2.5 exposure in our sample is estimated at $15.7 million.

Related Literature. This paper directly links to the research on the effect of environmental factors on economic outcomes. We make two contributions to this literature.

First, we extend the literature strand on air pollution and worker productivity to a profession that is representative for a large group of high-skilled workers in flexible, modern work environments. Previous research often studies low-skill, routine tasks such as textile workers (Adhvaryu et al., 2019; He et al., 2019), pear packers (Chang et al., 2016), call center agents (Chang et al., 2019), or fruit pickers (Graff Zivin and Neidell, 2012). A number of studies finds that worker productivity in high-skilled professions is also reduced by air pollution. This evidence comes from studies on error detection of baseball umpires in the US (Archsmith et al., 2018), the speech quality of Canadian politicians (Heyes et al., 2019), and case handling time by Chinese and Mexican judges (Kahn and Li, 2019; Sarmiento, 2022). While these contexts allow to create precise measures of worker performance in a specific task, these professions are rare and in general not characterized by the typical features of jobs that most high-skilled workers do (e.g. frequent collaboration, multiple different tasks). Related work investigates performance in cognitively-demanding tasks, but outside of standard work settings, e.g. among chess players (Künn et al., 2019), individual investors (Huang et al., 2020), or brain game players (La Nauze and Severnini, 2021). Our data, by contrast, measures performance in tasks from a common profession which are similar to what other knowledge workers face on their jobs. It also enables us to present first

[2]Pull requests are a tool to suggest changes to the code base of a repository, for more details see Section2.

evidence on productivity effects separately for individual and collaborative activities. We also note that our setting covers a vast number of countries, allowing to draw more general conclusions about the pollution-productivity relationship and to explore effect heterogeneity e.g. with respect to local income levels.

Second, we contribute to the literature strand on worker adaptation to environmental shocks, by highlighting a new margin of adjustment in flexible high-skilled jobs, namely task choice. The rich data we study allow us to analyze how the code written by developers changes, what kind of tasks are performed and how difficult they are. We find evidence for a switch to easier work when air pollution is high. Related papers studied e.g. how workers adjust working hours in response to temperature shocks (Graff Zivin and Neidell, 2014; Neidell et al., 2021; LoPalo, 2020). With respect to air pollution shocks, Adhvaryu et al. (2019) and Bassi et al. (2021) demonstrate an important role of managers who can mitigate productivity losses, e.g. by reallocating workers to different tasks. These studies, however, focus on rather low-skilled manufacturing workers. Our work demonstrates adaptation in a high-skilled setting where workers have greater flexibility in organizing their work day.

Another contribution of this paper is that we demonstrate new ways to use publicly available data on GitHub activity. While we are not the first to use this data in economics[3], we present a strategy to construct a sample of highly active users who are likely professional software developers and show how the data can be used, e.g. to study task difficulty. We also complement this with information from Gitcoin to estimate the monetary value of the output observed on GitHub.

With our detailed analysis of work behavior we also relate to a broader literature that studies drivers of worker productivity. For example, Lazear et al. (2015) show that productivity increased during recessions because of higher worker effort. Pencavel (2015) and Shangguan et al. (2021) examine how the output of workers is driven by their work hours. We complement this literature of detailed analyses of work patterns by demonstrating how external health shocks are an important factor to consider when studying work patterns.

Outline. We begin in Section 2 by describing how *Github* and its underlying software *Git* work, how we construct the data on developer activity, and how the different outcomes are defined. We also provide information on the environmental data that are obtained on air quality, wind, other weather conditions, wildfire smoke, and thermal inversions. In Section 3 we explain the research design and how we implement the two-stages least squares strategy of Deryugina et al. (2019). Section 4 presents result on how workers are affected by pollution in terms of work quantity, task adjustment, and quality. Section 5 concludes.

2 Setting and Data

To analyze the effect of local air pollution on productivity and work patterns in a highly-skilled profession, we pair information on GitHub activity for a global sample of software developers with data on local air quality obtained from several national environmental agencies. This is complemented with data on meteorological conditions which we use to construct the instrumental variables and to control

[3]McDermott and Hansen (2021), e.g. use the data for an analysis of the impacts of the COVID-19 pandemic on work patterns.

for local weather. This section starts with a short description of GitHub, followed by an overview of the GitHub data and how we use it to measure productivity of tech workers. After checking the validity of our primary outcome measures, we end with a description of the environmental data.

2.1 Setting: GitHub

GitHub is built on *Git*, an open source version control system, i.e., a software that facilitates keeping track of different versions of a file over time. It records who changed which part of a file at what point in time. Hence, it is particularly useful for collaborative projects where several people work on a set of files. *GitHub* is a platform for hosting Git repositories[4] and, on top of the version control functionality, provides additional features for collaboration. Although Git(Hub) can be used for various types of files, it is intended and predominantly used for storing and jointly working on coding projects. For each of their repositories, repo owners can choose whether to make it public or private, i.e., whether the respective files are visible to every user of the platform, or only to the owners themselves. In 2019, more than 30 million accounts were registered on GitHub, which together owned more than 120 million public repositories, making it the world's largest host of source code.

The core action in Git is a *commit*, which refers to the action of saving the current version of the repository after implementing a change to a file, or a set of files. As such, a commit represents that some work on code files has been conducted by the commit author.

The primary additional collaboration features offered by GitHub are *issues* and *pull requests*. An issue is a message used to suggest improvements, organize tasks or make other requests related to a given repo.[5] Each issue contains a discussion forum where users can leave comments to discuss the problem or question at hand. Pull requests (PRs) are a tool which can be used to propose code changes, including changes to repositories you do not own, i.e., whose files cannot be altered via a commit. After a PR is submitted, the repo owners can review the suggested changes and decide whether to accept (i.e., merge) or reject them. Like issues, PRs also include a discussion forum where users can comment directly on the proposed changes. Apart from allowing users to contribute to projects of which they are no member, project team members also use PRs as they facilitate collaborative coding.[6]

These distinct GitHub actions reflect productive work, mostly aimed at building or improving software products. Hence, we use these actions to measure output generated by highly-skilled tech workers. A key advantage is that the distinct actions we observe reflect different types of tasks. In particular, we consider the number of commits and the number of PRs opened as measures of individual work on code, whereas the number of comments written (the sum of comments on issues, PRs and commits) and the number of issues closed, opened, or reopened reflect collaborative work conducted in interaction with other users. Finally, the number of PRs closed reflects work on code review and making decisions on whether to merge or reject the proposed changes. This allows us to conduct the first analysis of the productivity impacts of air pollution in a high-skilled profession that takes potential effect heterogeneity between individual and team work into account. In addition to these *productive* actions, users can

[4] The term repository, or repo for short, refers to the location were all files belonging to a project are stored.

[5] For an example of an issue, see for instance `https://github.com/microsoft/vscode/issues/39526`.

[6] Collaboration is facilitated by PRs as the author can request a code review from specific team members who can write comments to provide direct feedback. Their comments can be implemented in follow-on commits within the same PR. An example for a PR can be found at `https://github.com/microsoft/vscode/pull/54244`.

also exploit the social network functions offered by GitHub, e.g., following other users or repositories.

Another important feature of GitHub allows us to study worker adjustment to the potential negative productivity shock induced by high air pollution: Repository owners can assign labels to issues in order to reflect issue category (e.g., bug, feature request), issue priority or issue difficulty. The platform provides nine default labels, and repository owners can create additional labels specific to their repo. Several labels are indicative of a relatively easy issue, e.g., the default labels *good first issue*[7] and *documentation*,[8] or individual labels such as *beginner friendly* or *low-hanging fruit*. The complete list of labels we use to identify easy tasks is depicted in Appendix Table A.1. Based on this we can measure whether users adjust the share of actions on easy tasks when air quality is poor. This approach has several advantages: We do not have to identify issue complexity ourselves but can rely on the assessment by experts who know the project in question very well. Furthermore, the label is visible to all users, i.e., workers who search for easy tasks due to an adverse productivity shock can easily identify these issues as suitable.

2.2 GitHub Data on Productivity and Work Patterns

We collect data on GitHub activity from two sources, GHArchive and GHTorrent. The GHTorrent project provides data in the form of a relational SQL database, containing information on GitHub users and all actions they conduct in *public* repositories. We use the version of the database containing data up to June 2019. The user table comprises login name, registration date, and account type (individual or organizational) for all users registered on the platform by June 1st 2019. In addition, location and company information as stated on the user profile is reported. Data on activities is available separately by type of action (e.g., commits, issue comments etc.) and includes its exact timestamp, a unique identifier for the acting user and the repository the event was conducted in. For specific actions, further information is reported, e.g., the text of pull request and commit comments.

GHArchive also provides data on all events in *public* repositories. For some event types, this source contains additional information relative to the GHTorrent data, e.g., the titles and descriptions of issues and PRs, the message attached to a commit, the number of lines of code added and deleted, as well as the number of files changed within a PR. GHArchive and GHTorrent data can be linked via users' login names.

These data have multiple advantageous features, most prominently their global coverage (all GitHub users are included) and the highly precise timestamps. This allows us, e.g., to analyze impacts of local air quality on working hours, and gives us a clear advantage in terms of external validity compared to previous studies based on data from only one country, and often even just one sampling site. However, it also has clear limitations. Firstly, in order to assign local air quality to users, we rely on self-reported locations. If some users report a wrong location or do not update the information when they move, this measurement error introduces noise into our analysis which causes attenuation bias. Hence, any adverse effects we find can be considered as a lower bound on the true effect. Secondly, the data does

[7] This issue was introduced by GitHub to encourage first time contributions, but does not imply that the issue cannot be addressed by more experienced developers.

[8] The documentation label is included because work on the documentation is typically easier than work on code to fix bugs or build new features. This follows, e.g., from Tan et al. (2020) and from the fact that GitHub also used the documentation label in their approach to construct the good first issue label (for details see `https://github.blog/2020-01-22-how-we-built-good-first-issues/`).

not provide information on work users conduct in private repositories or outside of the platform. We observe only the part of total work that is carried out in public repos. For many GitHub users this is such a marginal share that it is impossible to identify any negative productivity effects of air pollution exposure based on their activity data. Thus, when constructing our analysis sample, we aim at capturing users who are professional developers and do a substantial part of their formal work in public GitHub repositories.

In a first step, we focus on non-organizational users who report a location at the city level, which is the degree of geographic precision required to assign local air quality. Secondly, we keep only individuals who ever committed in a repository owned by a company, i.e., users with the authority to change the source code of a company-owned project. This step is intended to capture professionals who are in some way affiliated with the companies. To identify these users, we compile a list of repositories operated by companies[9] and then use the information on the repository a commit was conducted in from the GHTorrent commits table. Moreover, to focus on cases where we can observe a substantial part of an individual's total work, we only admit users into the sample once they have at least 25 commits in public repos in a given calendar month. In the following month, they enter the sample which they do not leave until the end of the sample period (May 2019), unless we do not observe any *unproductive* action in a given calendar month. In this case we drop users from the sample for that month, assuming that they might have moved to a different platform, work on projects in private repositories or be on vacation. Unproductive actions are activities we do not use as outcomes in our analysis, based primarily on the social network functions GitHub offers.[10] Lastly, we restrict the sample to users living in cities with at least 20 developers and covered by our data on air pollution. This yields a sample of 26,737 users in 180 cities across 42 countries during the sample period February 2014 until May 2019.[11] These locations are depicted in Figure 1. The median user registered in April 2012.

For these users, we compile an unbalanced user-by-day panel including measures of output and worker adjustment to productivity shocks. Both GHTorrent and GHArchive report timestamps in UTC. We translate these timestamps into local time and aggregate activity to the the daily level. To measure work quantity, we consider the total number of productive actions conducted as well as the number of commits and comments, the two most frequent action categories on GitHub. We include these actions separately because they reflect two distinct types of work: individual coding activity (commits) and collaborative work, including discussions about issues and code changes (comments). This allows us to study effect heterogeneity between individual and interactive work. We winsorize commits and total actions at the 99.99th percentile to dampen the influence of outliers. To assess effects on worker adjustment, we measure the share of commits and actions, respectively, conducted after standard working hours, i.e., after 6 pm. We hypothesize that the share of evening work might increase on high pollution days if workers try to make up for their reduced productivity by working longer hours. Secondly, we

[9]This is based on a publicly available list of firms active on GitHub, which can be accessed at `https://github.com/d2s/companies/blob/master/src/index.md` and on the lists of open-source projects operated by Google, Microsoft and Facebook published on their web pages.

[10]The unproductive actions include following another user, watching a repository, (un)subscribing to an issue, labeling an issue and (un)assigning an issue to a user.

[11]During our sample period some users changed their location. Since the GHTorrent data on users is a snapshot taken on June 1st 2019, we use earlier versions of the database (one snapshot in each year between 2015 and 2018) to check for movements. In total, 6.3% of users reported more than one distinct location during this period. We identify the city were they spend the biggest part of the sample period, and keep them in the sample only during the time they resided in this city.

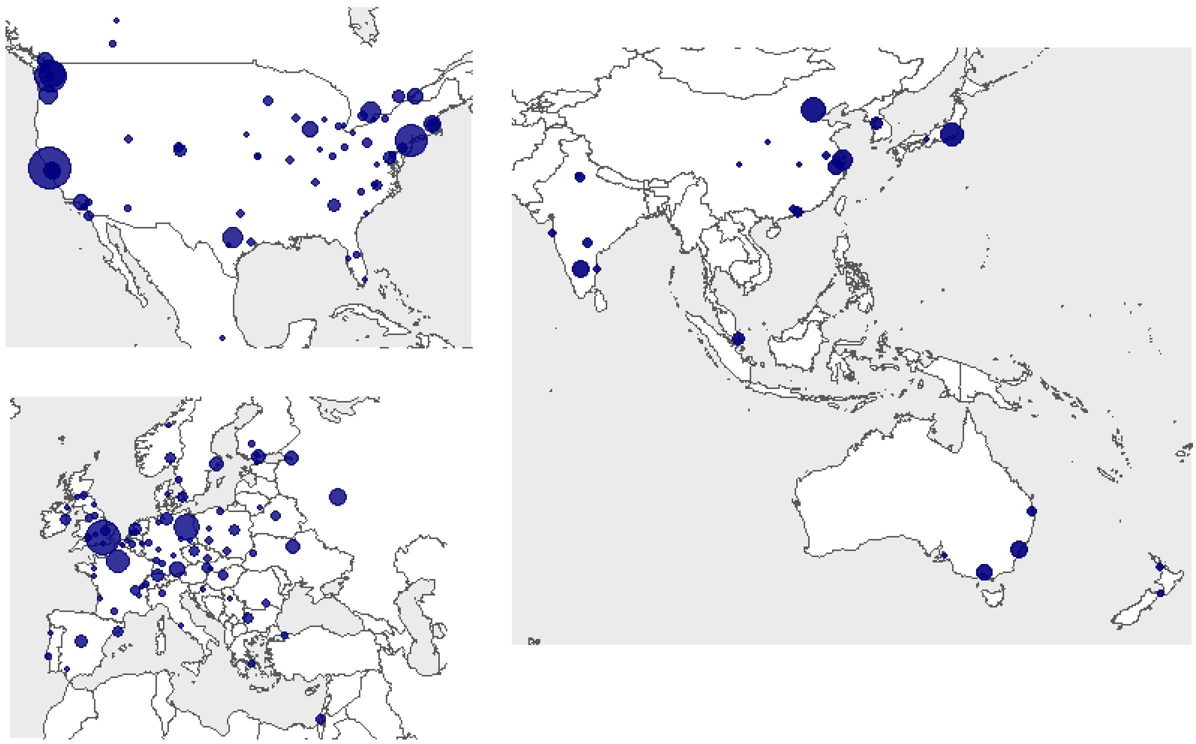


Figure 1: Sample Cities

Note: Each dot represents one sample city. Circle size refers to the number of coders observed in the city.

build outcomes indicating whether workers adjust by switching to easier tasks on high pollution days. Using the information on task difficulty derived from issue labels, we construct the share of all issue events conducted by the user which refer to an *easy* issue. With respect to individual coding tasks, we can assess the difficulty of PRs opened. In particular, we use the number of new lines of code added per PR as well as the number of code files changed per PR as proxies for complexity.

To explore whether our sample selection approach succeeds in capturing professional developers with high activity levels in public repositories, we provide a brief discussion of summary statistics, working hours and user self-descriptions. On average, users conduct 2.74 actions per day, of which 1.31 are commits and 0.90 are comments (see Table 1). This implies that users in our sample indeed conduct a lot of work in public repos, as casual users who only occasionally work on GitHub can hardly reach these values, especially given that we average across all days, including weekends, holidays and vacations. The remaining productive GitHub actions—opening and closing issues and PRs—are observed less often. Thus, we will use the sum of all actions plus the number of commits and comments (proxies for individual and interactive work) as primary outcomes when investigating effects of pollution on work quantity. On average, 7% of all issue events refer to an easy issue. Issue events comprise the actions of opening, closing and reopening an issue as well as writing an issue comment.

32% of commits are made after 6 pm. In Figure 2 we provide more detailed information on the distribution of activity across days of the week and hours of the day. Distributions are plotted for all actions (yellow) and for commits only (blue). The height of the bars refers to the share of all activity that is conducted on the respective week day and during the respective hour. For the five days of the

Table 1: Summary Statistics for the Analysis Sample of GitHub Users

	Mean	St. Dev.	Min	Max	Observations
actions	2.74	7.3	0	316	15,868,883
commits	1.31	3.95	0	256	15,868,883
comments	.90	3.32	0	280	15,868,883
PRs opened	.15	.70	0	151	15,868,883
issues opened	.10	.80	0	222	15,868,883
PRs closed	.17	.94	0	300	15,868,883
issues closed	.12	.88	0	268	15,868,883
share easy issue events	.07	.21	0	1.0	3,410,023
share commits after 6 pm	.32	.41	0	1.0	4,269,709
share actions after 6 pm	.29	.39	0	1.0	5,705,074
files changed per PR	6.12	16.37	0	1674	1,360,268
lines added per PR	178.09	501.73	0	6802	1,360,268
lines deleted per PR	65.98	200.17	0	2738	1,360,268

work week, we observe a consistent pattern: Activity is low during night hours, increases steeply in the morning, exhibits a small dip during lunchtime, before reaching a second peak in the early afternoon. Around 5 to 6 pm activity declines notably, and then continues during the evening hours, but only at about half the level that is observed during core working hours. On the weekend, the level of activity is substantially lower and it does not follow the same characteristic pattern across hours of the day that is observed on working days. In summary, activity patterns in our sample resemble standard working hours, but with notable activity during evening hours and on weekends, which is not uncommon among highly educated workers (Mas and Pallais, 2020). Interestingly, the distribution for commits and other actions seems to differ slightly. The share of commits is somewhat smaller in the morning hours and slightly larger in the evening and night hours compared to the other actions. This seems plausible given that the more interactive tasks are more productive during standard working hours, when other users are working as well.

Figure 3 presents information about the work status of users in our sample. The left panel depicts the most frequent terms individuals use in the biographies (bios) on their GitHub profiles. 36% of users in our sample (9,507 users) use the option to provide a self-description. The data is accessed via the GitHub API. For each term, we measure in what share of all bios it occurs, after stemming and removing stop words. Three terms clearly stand out: engineer/engineering, software and developer/development occur in 15% to 25% of all bios, much more often than any other words. While we cannot say whether the sub-sample of users who provide a bio is representative of the full sample, the clear peak at these work related terms strongly suggests that we do capture professional software engineers in our sample who use GitHub as part of their formal work. The plot in the right panel complements this with information on employers. Optionally, users can report the company they work at, and in our sample 61% (16,385 users) provide some information in this field. In the sub sample, Microsoft and Google are the most frequent employers, followed by Facebook and Red Hat, i.e., big US tech companies which are strongly engaged in open-source.

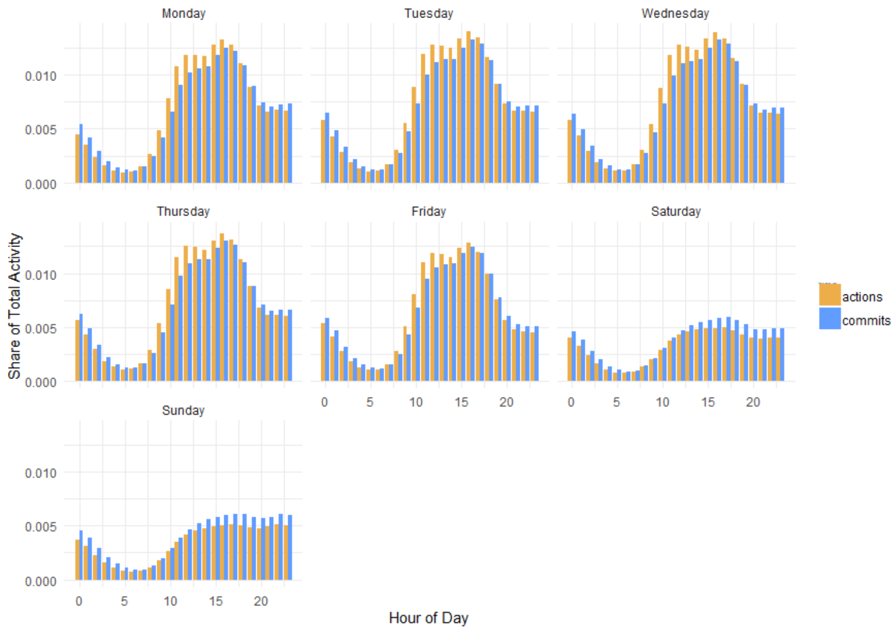
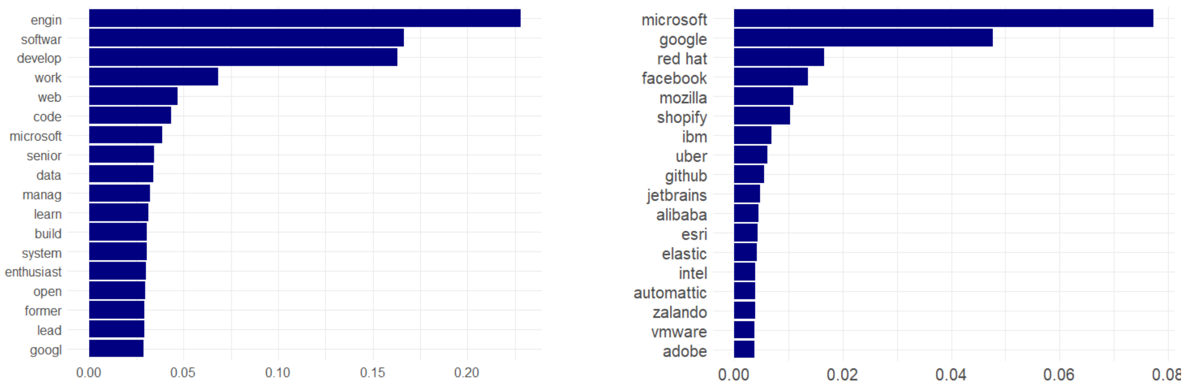


Figure 2: Distribution of Activity across Hours of the Day and Days of the Week

Note: Bar height reflects the share of total activity by sample users conducted during the respective hour on the respective day of the week. Yellow bars refer to the sum of all actions, blue bars refer to commits only.



(a) Bios

(b) Companies

Figure 3: Most Frequent Terms from User Self-Descriptions and Company Fields

Note: Panel (a) is based on data from 9,507 user bios, accessed in 2021 via the GitHub API. Words in the bios are transferred to lower case, stemmed, and stop words are removed. The total word count is divided by the number of bios. Panel (b) is based on data from the company column in the June 2019 GHTorrent user table for 16,385 users from our sample with any information in this field.

2.3 Gitcoin: Monetary Value of GitHub Activity

To assess the validity of the productivity metrics derived from GitHub data and to translate estimated effects of air pollution on these outcomes into monetary damages, we use data from a platform called *Gitcoin*. Gitcoin was founded in 2019 and is complementary to GitHub. Project teams can use Gitcoin to offer monetary payments to incentivize outside contributions to their projects. In this way, freelance developers who provide these contributions can earn money for their work on open-source. In particular, GitHub users may post open issues from their public repos along with information on issue characteristics and a payment they offer for solving the issue (called *bounty* on Gitcoin). Interested developers can apply and submit a PR in the respective repo to solve the task described in the issue. The PR is reviewed by the issue funder, and if accepted and merged into the repo, the bounty is paid to the PR author, typically in crypto currencies. All work is conducted in public GitHub repos and thus visible to us. In the end of 2021, about 300,000 GitHub users were registered on Gitcoin. We collect data on 292 issues for which PRs were submitted and payments were made by March 2022 via the Gitcoin API. These data include the URL to the PR, the type of the issue (one of bug, documentation,improvement, feature, or other), the expected issue difficulty as assessed by the issue funders (one of beginner, intermediate, or advanced), hours worked on the PR as stated by the author, and the value of the payment in USD.[12] We merge the Gitcoin data with information on pull request size obtained via the GitHub API (number of commits, number of lines of code added and deleted, and number of files changed).

In total, the work on the sample issues was rewarded with $103,313, which implies a mean payment of $354 per pull request and an average value of $112 per commit, one of our primary outcome variables. In this context, a pull request reflects complete work on a certain issue. Commits can be interpreted as single work steps in completing this task. The average monetary value per commit ranges from $32 in the subsample of issues of difficulty level *beginner* to $679 among issues marked as *advanced*. On average, developers spend 1.8 hours on one commit, again with a steep gradient with respect to difficulty: The mean time input per commit is 1 hour at the beginner level, but 5.3 hours at the advanced level.

In Table 2 we present results from regressions of the payment awarded for a PR, $log(payment_i)$, on the number of commits it comprises, $commits_i$ (columns 1-3), or the logarithm thereof (columns 4-6). We run specifications without any controls (columns 1 and 4), with controls for issue difficulty, issue type and the year of PR creation (columns 2 and 5), and lastly with repository fixed effects (columns 3 and 6).[13] Across specifications we find statistically significant positive effects, indicating that a higher number of commits is associated with higher payments. In terms of magnitude, the results from the regressions without any controls imply that one additional commit is associated with a 5.4% increase in payment (column 1), or that a 10% increase in the number of commits is correlated with a 3.5% rise in payment (column 4). When adding controls for issue difficulty and type, the magnitude of the effect is reduced. This reduction implies that part of the increase in payments in commits is driven by higher issue complexity. Even when using only variation across PRs submitted to the same repo, i.e., work on the same project, the positive relationship persists. We view this as a confirmation that changes in the number of commits observed per user and day indeed reflect fluctuations in developer productivity. In

[12] The number of issues is relatively low compared to the volume of our GitHub data because Gitcoin is much younger than GitHub and only used by a small share of GitHub users.

[13] The omitted difficulty category is *advanced*.

Table 2: Validity Check: Impact of Number of Commits on Gitcoin Payments

	Dependent variable: $log(payment_i)$					
	(1)	(2)	(3)	(4)	(5)	(6)
$commits_i$	0.054***	0.039***	0.034***			
	(.010)	(.009)	(.010)			
$log(commits_i)$				0.348***	0.264***	0.192***
				(.071)	(.068)	(.059)
$\mathbb{1}\{Difficulty_i = Beginner\}$		−2.399***			−2.412***	
		(.439)			(.419)	
$\mathbb{1}\{Difficulty_i = Intermediate\}$		−1.878***			−1.851***	
		(.415)			(.405)	
Year dummies		✓	✓		✓	✓
Issue difficulty dummies		✓			✓	
Issue type dummies		✓			✓	
Repository fixed effects			✓			✓
Observations	292	274	292	292	274	292

Note The table presents results from OLS regressions using data on the sample of Gitcoin pull requests. Observations are at the pull request level. Dependent variable is the logarithm of the payment awarded to the PR author. Explanatory variables are the number of commits (column 1) or the logarithm thereof (column 4). Columns 2 and 5 add dummies for the year the pull request was created, dummies for issue difficulty, and dummies for issue type. Column 3 and 6 instead add dummies for the year the pull request was created and fixed effects for the repository. Robust standard errors are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Appendix Table A.2 we present results from models where the dependent variable is $hoursworked_i$, the time input as reported by the PR author. We find that the time required to complete a task increases in the number of commits, and more so for issues of higher difficulty.

We also find that, holding the number of commits constant, adding more lines of code and changing more files is associated with a higher payment, confirming the use of our proxies for PR complexity. We find this by running the specifications from columns 4 to 6 of Table 2 again, but add the number of files changed in the PR and the logarithm of lines of code added as additional regressors. Detailed results are presented in Appendix Table A.3.

While the Gitcoin data allows us to assess the monetary value of commits, there are certain limitations: We only observe payments for a small subset of GitHub repositories that use Gitcoin, and these are likely not representative of the average public GitHub project the users in our sample work on. Incentivizing external contributions via Gitcoin is more attractive for small projects where the core team cannot handle all open issues by itself. Moreover, the coders who get paid via Gitcoin are mostly freelancers or casual contributors. By contrast, in our main sample we focus on professional tech workers who conduct work in repositories owned by larger tech-companies. These limitations imply that Gitcoin-derived valuations of GitHub contributions in our main sample based on Gitcoin will likely be underestimates. Gitcoin payments made by small projects to freelancers are probably lower than the payment professional tech workers at big companies would receive for the same output.

2.4 Environmental Data

Air Quality. The pollutant of primary interest in our analysis is $PM_{2.5}$ for two key reasons: It can penetrate indoors and is thus of direct relevance for indoor office workers and the smallest pollutants are known to have the largest health effects. We collect data on $PM_{2.5}$ concentration measured at outdoor monitors from several environmental agencies to cover all the cities represented in the productivity sample. For nine cities, we could not obtain monitor-measured data, but instead used high-resolution reanalysis data from the Copernicus Atmosphere Monitoring Service (CAMS). Appendix table A.4 provides a detailed list of the data sources. All data is provided at either the daily or the hourly level. Where necessary, we transfer hourly data into local time and aggregate to the daily level. Cities are assigned the simple average of all available monitor readings within a 40km radius around the city centroid.[14] Our data on $PM_{2.5}$ covers 93% of all city$\times$day observations, with best coverage in North America (97%) and lowest coverage in Asia (83%), where in some cities $PM_{2.5}$ monitors have only been installed during our sample period.

We winsorize $PM_{2.5}$ at the continent-specific 0.1^{th} percentile and the 99.9^{th} percentile to ensure that our results are not driven by extreme outliers (e.g., extremely high concentration of fine particulate matter due to heavy wildfire smoke). The population-weighted average $PM_{2.5}$ concentration in our sample is 14.9 μg/m^3 (standard deviation: 26.5 μg/m^3, within-city: 21.1 μg/m^3), i.e., close to the World Health Organization's recommended 25-hour limit value (15 μg/m^3). Figure 4 displays the distribution of daily $PM_{2.5}$ concentrations in our sample, separated by seven large geographic regions, $R \in \{$Northern Europe, Southern Europe, Western Europe, Eastern Europe, North America, Oceania, Asia$\}$.[15] Air quality exhibits substantial heterogeneity across regions: Cities in North America, Oceania and Northern Europe have relatively clean air, with concentrations above 20 rarely observed. Locations in Southern and Eastern Europe by contrast experience this level of pollution on 26% of all days, and Asian cities even 71% of the time.

Wind conditions. The instrumental variable approach is based on regional air pollution transport driven by wind direction. We collect reanalysis data on wind conditions from the Japan Meteorological Agency's JRA-55 product. Reanalysis datasets are constructed by combining measurements taken at ground-level monitors, satellite images, and atmospheric transport models. The u- and v-component of wind are reported every six hours (in UTC) on a global grid with a spatial resolution of 1.25° longitude$\times$1.25° latitude.[16] We translate timestamps into local time and aggregate to the daily level. Each city is assigned the inverse distance weighted average of u- and v-vectors at the four grid points located closest to its centroid (median distance = 92.5km). Finally, daily average wind speed and direction are computed from the city-level u- and v-vectors.

Meteorological Conditions. To construct control variables for daily weather conditions, which can be correlated with wind conditions and affect working patterns, we use the ERA5-land product from the European Centre for Medium-Range Weather Forecasts (ECMWF). It provides hourly data on air temperature two meters above the surface, precipitation and dewpoint temperature on a fine grid with 0.1° longitude$\times$0.1° latitude horizontal resolution. To construct city$\times$day level variables, we follow the

[14] CAMS reanalysis data is reported on a 0.1° longitude$\times$0.1° latitude grid. Given the large number of grid points, we only use measurement points within 25km of the centroids for the relevant cities.

[15] We show the distribution of observations in our user x date panel across these regions in table A.5. The regions are illustrated on a map in figure A.1.

[16] We deliberately use data reported on such a coarse spatial grid in order to capture broad wind patterns driving regional air pollution transport instead of very local wind conditions which only affect air quality in a small area. The choice of data follows a suggestion by Tatyana Deryugina which we gratefully acknowledge.

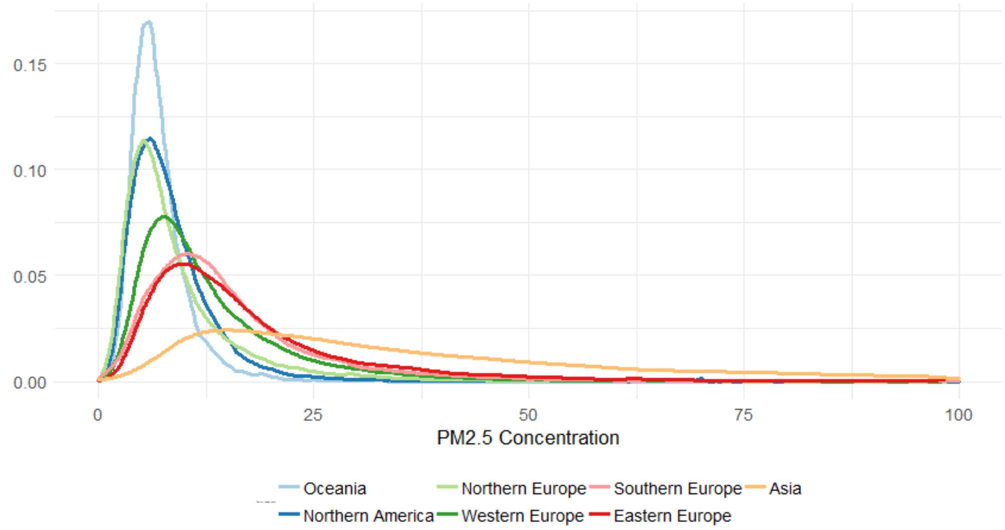


Figure 4: Distribution of daily $PM_{2.5}$ Concentrations by Geographic Region

Note: The plot shows densities of $PM_{2.5}$ concentration based on 331,025 city×day observations, separately by geographic region. Oceania: Australia, New Zealand. Northern Europe: Scandinavia, UK, Ireland, and the Baltic countries. Southern Europe: Portugal, Spain, Italy, Greece, Croatia, Serbia, Turkey, Israel. Asia: China, India, Japan, Singapore, Hong Kong, S-Korea. Northern America: US, Canada, Mexico. Western Europe: Switzerland, Austria, France, Germany, Belgium, and the Netherlands. Eastern Europe: Poland, Czech Republic, Hungary, Belarus, Ukraine, Russia, Slovakia, Bulgaria, Romania.

same approach as taken with the wind data, the only difference being that sample cities are assigned the inverse distance weighted average weather conditions from the eight, instead of four, closest grid points (median distance = 11.0km). The variables constructed are daily mean, minimum and maximum temperature, precipitation, and relative humidity.[17]

Wildfire Smoke. In recent years, especially in 2017 and 2018, the North American west coast experienced several severe wild fires generating heavy smoke that strongly increased the concentration of air pollution. Some of the largest cities within our sample are located in this area (the tech clusters in the San Francisco Bay Area and around Seattle). Given recent research by Burke et al. (2021), showing that exposure to heavy wildfire smoke can trigger avoidance behavior, especially among high income individuals, we construct control variables for heavy smoke to make sure the effects we estimate reflect physiological productivity impacts of $PM_{2.5}$ exposure, not behavioral responses to the phenomenon of wildfires. The required data is derived from satellite images and provided by the National Oceanographic and Atmospheric Administration's Office of Satellite and Product Operations in the form of shape files. The data covers the North American continent, is reported at the level of individual smoke plumes and includes a measure of smoke intensity. We define a city as being affected by a smoke event if the smoke plume overlaps with a 10km radius around its centroid. We aggregate the data to the daily level (in local time) by summing over the intensity measure of all smoke plumes covering a city on a given calendar day. Lastly, we define a heavy smoke indicator which is one if the city was covered by a plume of the highest intensity on the respective day or if the total daily smoke intensity exceeds a

[17]Relative humidity is inferred from mean daily air temperature and dewpoint temperature using the R package `weathermetrics` which uses formulas provided by the US National Weather Service.

value of 54 (which corresponds to twice the value of the highest smoke intensity), and zero otherwise. This yields 0.4% of all city-by-day observations and 8.8% of all observations with any smoke exposure as heavy smoke days.

Thermal Inversions. In robustness checks, we use temperature inversions instead of wind direction as an instrument for $PM_{2.5}$ concentration. The required data is obtained from the ECMWF's ERA5 products[18]. Hourly temperature at the surface level as well as several pressure levels is reported on a 0.25° longitude$\times$0.25° latitude grid. We collect surface temperature as well as temperature at the pressure level 25 hPa above the surface level. To construct daily inversion measures, we compute the temperature difference between upper air and surface level, averaged during local nighttime hours (midnight to 6 am), following several recent papers (e.g., Jans et al., 2018). Cities are assigned the inverse distance weighted average from the four closest grid points. An inversion indicator is defined to be one if the difference between upper air and surface temperature is positive, $inv_{cd} = \mathbb{1}\{\Delta T > 0\}$. A measure of inversion strength, $inv\ strength_{cd}$, is defined as the temperature difference whenever it is positive and as zero whenever the difference is negative.

3 Research Design

Regression Model. We want to model how short-run variation in local particulate matter concentration affects the output of professional software developers. To do so, we specify a model for the work output y of developer i living in city c on day d.

$$
\begin{aligned}
y_{i,c,d} = \beta PM_{c,d} + \mu_i + \mathbf{x}'_{i,t}\pi \\
+ \mathbf{w}'_{c,d}\gamma + \delta h_{c,d} + \mu_c + \mu_{dow(d)} + \mu_{R(c),yr(d),m(d)} + \mu_{r(c),m(d)} + \varepsilon_{i,c,d}
\end{aligned}
\tag{1}
$$

$PM_{c,d}$ is a measure of particulate pollution and varies across cities c and days d. The fixed effect μ_i captures unobserved heterogeneity at the developer level. A developer's experience is controlled for by $\mathbf{x}_{i,t}$, a vector of indicators for time since registration on GitHub falling into a specific bin, where each bin has a length of three months. The vector $\mathbf{w}_{c,d}$ contains weather variables measured for city c on day d. It includes a series of indicator variables for daily mean temperature falling into bins defined based on the 5th, 10th, 20th, 35th, 65th, 80th, 90th, and 95th percentiles of the city-specific temperature distributions. We also control for cubic polynomials of precipitation, relative humidity, and wind speed. The vector further contains a dummy indicating whether the city is affected by wildfire smoke on day d. The regression equation also accounts for public holidays $h_{c,d}$ as well as city fixed effects μ_c and day-of-week fixed effects $\mu_{dow(d)}$. Year-month fixed effects $\mu_{R(c),yr(d),m(d)}$ can vary for each of the seven large geographic regions R, which are listed in table A.5. Monthly seasonality $\mu_{r(c),m(d)}$ is allowed to vary across smaller geographic region r, which we define in a way to assure that each region contains several sample cities. Examples are Scandinavia, France, Canada, or US Census Divisions. Figure A.1 illustrates the large and small geographic regions.

The coefficient of interest β is identified from city-level day-to-day variation in pollution and de-

[18]We use the products *ERA5 hourly data on single levels from 1979 to present* and *ERA5 hourly data on pressure levels from 1979 to present*

veloper level output, conditional on average developer-output and after netting out other drivers of output such as weather, the workweek, and region-wide business cycle dynamics.

As the variable of interest only varies at the city-day level and the number of developer level observations is large, we make use of an auxiliary regression to aggregate Equation (1) to the city level (cf. Donald and Lang, 2007; Currie et al., 2015). To do so, we regress $y_{i,c,d}$ on all variables at the developer level, i.e., $\mathbf{x}_{i,t}$ and μ_i, and on a city-day fixed effect $y_{c,d}$. The estimates of the city-day fixed effects $\tilde{y}_{c,d}$ then become the dependent variable of a new regression model.

$$\tilde{y}_{c,d} = \beta PM_{c,d} + \mathbf{w}'_{c,d}\gamma + \delta h_{c,d} + \mu_c + \mu_{dow(d)} + \mu_{R(c),yr(d),m(d)} + \mu_{r(c),m(d)} + \varepsilon_{c,d} \tag{2}$$

Regression model (2) can then be estimated at the city level where the number of underlying developer-observations are used as weights.

Air quality is not assigned randomly. In Equation (2), pollution $PM_{c,d}$ may be endogenous for several reasons. Local variation in economic conditions will affect air pollution and developer level output at the same time. Similarly, local events like a football match or the closing of a bridge may affect both traffic and work patterns at the same time. A naïve estimation of the model might not only suffer from such unobserved variable bias, but also from reverse causality. For example, if coders are more productive when working at home than in the office, the reduction in commuting may affect air pollution readings. A third source of bias is measurement error in pollution. We cannot observe individual-level exposure to air pollution and instead have to proxy for it by city-level averages. While our design includes a wide range of controls to account for sorting into different cities or fluctuations in local economic conditions, we still require an exogenous source of variation in local air pollution and rely on an instrumental variable estimation.

IV estimation. We address endogeneity in Equation (2) by instrumenting local pollution levels with wind direction. This approach was first introduced by Deryugina et al. (2019) and is based on the idea that wind direction affects local particulate matter concentration because it is a key driver of pollution transport. Wind blowing from the ocean or less densely populated areas may carry substantially lower amounts of pollutants into the city than wind blowing from more densely populated or industrial areas.

It is important to note that local weather conditions, e.g., rainfall, also depend on wind. For example, wind blowing from the ocean could reduce temperatures. These local conditions could affect labor-leisure trade-offs (Graff Zivin and Neidell, 2014) and thereby the output of developers via channels other than air quality. Therefore, it is important to control for a wide range of weather conditions contained in $\mathbf{w}_{c,d}$ to ensure that the instrument wind direction does not violate the exclusion restriction.

The effect of wind direction is certainly not uniform across all cities in our global sample due to differences in geography. In some cases, more pollution might be transported into the city by wind blowing from the east, in other cases west wind might carry in most pollution. To account for this, we allow the impact of wind on PM2.5 to vary. In principle, we could estimate the effect of wind direction for each city separately. To ensure that identification comes from regional pollution transport that affects the whole city instead of highly local transport that simply redistributes particulate matter within the boundaries of a city, we resort from doing so and restrict the effect of wind to vary at a geographically more aggregate level. As suggested by Deryugina et al. (2019), we use a k-means cluster

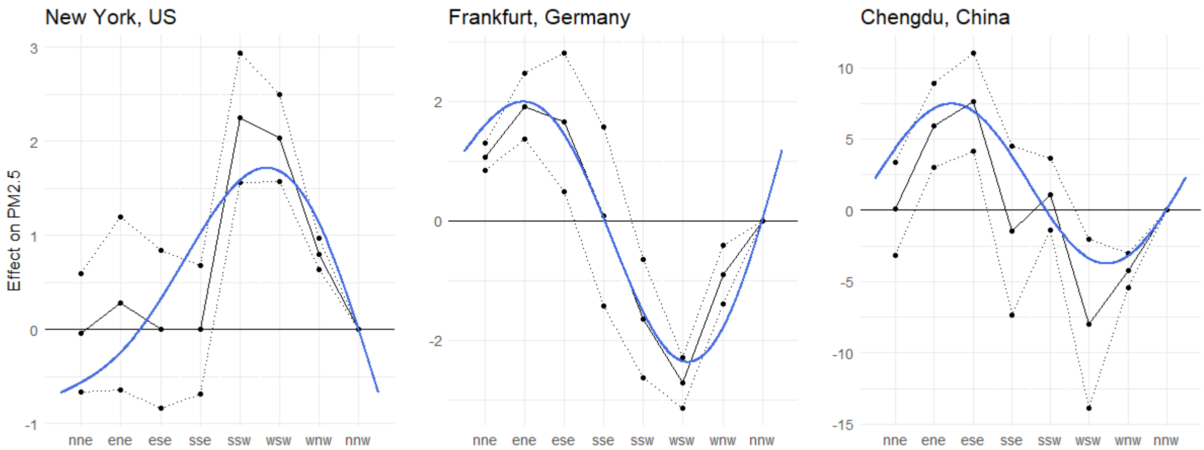


Figure 5: The effect of wind direction on $PM_{2.5}$

Notes: This figure provides a graphical illustration of the first stage. Graphs present estimated coefficients from regressions of $PM_{2.5}$ measured in μg/m^3 on wind direction. Solid black line: connects estimated coefficients on seven dummies for seven 45° bins of wind direction. The omitted direction is north-north-west, (315°, 360°]. Dashed lines: 95% confidence intervals. Blue line: estimated relationship when wind direction is parameterized as the sine of wind direction in radians and wind direction in radians divided by two. City groups comprise New York and Philadelphia (left), Frankfurt, Nuremberg, Munich, Stuttgart, Karlsruhe (center), Chengdu and Xi'an (right).

algorithm to assign cities into groups based on their longitude and latitude. In our baseline specification we form 65 groups.

We parametrize the pollution-wind relationship by a trigonometric function.[19] By specifying wind direction $\theta_{c,d}$ in radians instead of using many indicators for wind direction bins we can substantially reduce the required number of variables to appropriately model the wind-pollution relationship. The first stage of the IV estimation is as follows.

$$\begin{aligned} PM_{c,d} = \rho_1^g \sin(\theta_{c,d}) + \rho_2^g \sin\left(\frac{\theta_{c,d}}{2}\right) + \mathbf{w}'_{c,d}\gamma + \delta h_{c,d} \\ + \mu_c + \mu_{dow(d)} + \mu_{R(c),yr(d),m(d)} + \mu_{r(c),m(d)} + \varepsilon_{c,d} \end{aligned} \tag{3}$$

The coefficients ρ_1^g and ρ_2^g are determined at the city-group level.

Figure 5 illustrates for the city groups represented by Frankfurt, New York City and Chengdu how this trigonometric function can capture the effect of wind direction (on the horizontal axis) on $PM_{2.5}$ levels (on the vertical axis). This estimate (in blue) is close to an alternative specification (in black, 95% confidence interval dashed) where wind direction is measured by eight indicators representing 45° sections of the wind rose, e.g. (0°-45°], (45°-90°], etc.

Measures of pollution. Given the large variation in pollution levels across cities in our sample (see figure 4), we define PM_{ct} in Equations (2) and (3) as PM2.5 concentration measured in city-specific standard deviations. This ensures that we estimate the effect of usual fluctuations in pollution levels

[19] We are grateful to Tatyana Deryugina for this suggestion.

across all cities. As an alternative, we define what we will refer to as a pollution *shock*:

$$\text{PM2.5 shock}_{cd} = \mathbb{1}\left\{PM_{c,d} > \sqrt{\widehat{\text{Var}}\left[PM \mid c\right]} + \widehat{\text{E}}\left[PM \mid c, m(d), dow(c)\right]\right\} \tag{4}$$

This shock is an indicator that takes the value one if the city-day level $PM_{2.5}$ is more than one city-specific standard deviation above the level expected for the given city, month of year, and day of week. The proposed measure, therefore, captures non-linear effects of pollution and allows these to differ by location and time. On average, 10.5% of all days experience such a $PM_{2.5}$ shock.

4 Results

4.1 Work Quantity

Columns 1 to 3 of Table 3 display 2SLS estimates of the effect of $PM_{2.5}$ exposure on the three primary quantity outcomes—the number of total actions, commits, and comments. In panel A, we use $PM_{2.5}$ measured in city-specific standard deviations as regressor. The first stage F-statistic on the excluded instruments exceeds 100, indicating that the IVs based on wind direction are sufficiently strong. Our first main finding is that developers' output, measured by total actions, falls if the concentration of fine particulate matter increases. In terms of magnitude, an increase in $PM_{2.5}$ concentration by one local standard deviation reduces the number of actions by 0.026, which corresponds to 0.9% of the mean. This effect is mainly driven by a reduction in individual coding activity, measured by the number of commits. This outcome falls by 0.018 or 1.4% of the mean, if $PM_{2.5}$ concentration rises by one standard deviation. The number of comments—our proxy for interactive work conducted in collaboration with other users—is much less affected by air pollution. The point estimate is close to zero and not statistically significant.

In panel B we repeat the analysis, now using the binary $PM_{2.5}$ shock variable as regressor. The F-statistic drops to a value of 34.6, indicating that wind direction is a better predictor of usual fluctuations in $PM_{2.5}$ concentration than of the dummy measuring unusually poor air quality. Still, the F-statistic is well above the common threshold for a sufficiently strong first stage relationship. The 2SLS estimates imply that on a day with unusually high pollution relative to the city$\times$calendar month$\times$day of week specific average, the number of total actions (commits) falls by 0.1 or 3.9% (0.08 or 6.5%) of the mean value. Again, no effect on the number of comments is found.

In sum, the results in columns 1 to 3 of Table 3 suggest that the effect of fine particulate matter on output is mostly driven by days with unusually poor air quality given that the effect of a high pollution day in Panel B corresponds to an increase in $PM_{2.5}$ concentration by roughly four local standard deviations based on the coefficients in Panel A. Secondly, a novel finding is the strong effect heterogeneity across different types of work: We observe a highly significant negative impact on individual work on code, but no effect on interactive work.

In column 4 we explore the contribution of the extensive margin to the overall reduction in work quantity. To that end, we replace the dependent variables with an indicator for a positive activity level, i.e., $\mathbb{1}\{actions_{id} > 0\}$. We find that the probability to conduct any action falls when $PM_{2.5}$ increases. For both measures of pollution, extensive margin effects contribute approximately 20% to the full effect

Table 3: Effect of $PM_{2.5}$ on Work Quantity

	Actions (1)	*Commits* (2)	*Comments* (3)	*Any actions* (4)
Panel A.				
$PM_{2.5}$ (st. dev.)	-0.0258^{**}	-0.0177^{***}	-0.0059	-0.0020^{**}
	(0.0104)	(0.0053)	(0.0052)	(0.0009)
F-statistic	112.3	112.3	112.3	112.3
% change in Y	-0.9	-1.4	-0.7	-0.6
% of full effect				21.2
Panel B.				
$PM_{2.5}$ shock	-0.1075^{**}	-0.0845^{***}	-0.0164	-0.0080^{*}
	(0.0503)	(0.0262)	(0.0233)	(0.0042)
F-statistic	34.6	34.6	34.6	34.6
% change in Y	-3.9	-6.5	-1.8	-2.2
% of full effect				20.3
Observations	331,025	331,025	331,025	331,025
Dep. Variable Mean	2.74	1.31	0.90	.36

Note: The table presents IV estimates of the parameter β in Equation (2). In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in city-specific standard deviations. In Panel B, a binary $PM_{2.5}$ shock variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city×month×day-of-week specific average by at least one city-specific standard deviation. The first stage specification is given in Equation (3). Covariates include eight bins for mean daily temperature, third-order polynomials in wind speed, precipitation, and relative humidity, indicators for heavy wildfire smoke and holidays, as well as city, day-of-week, year-by-month, and region-by-month fixed effects. Day-of-week and year-by-month fixed effects and the temperature controls can vary across world regions R. Regressions are weighted by the number of active workers in a city during the current month. Standard errors clustered at the city level are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

on actions. Hence, the extensive margin does explain part of the effect on output, but the intensive margin response is quantitatively much more important. This result is plausible given that our sample of GitHub users likely comprises mostly young and middle-aged adults who generally do not suffer severe health damages from short-run pollution exposure. The results are consistent with subtle effects on health and cognitive function slowing down users progress while working.

In Appendix Table A.6, we investigate the effect of $PM_{2.5}$ on further action types—the number of issues and PRs opened and closed, respectively. These actions occur less frequently than commits and comments, with mean values between 0.11 and 0.17. Like a commit, opening a PR reflects individual coding work whereas opening/closing issues generally starts/ends a discussion with other users and thus constitutes interactive work. Closing a PR implies a code review and decision-making about whether to accept or reject the proposed changes. Consistent with the results in Table 3, the number of PRs opened falls significantly in $PM_{2.5}$ concentration, and the relative effect magnitude is similar to the effect on commits. Issue events are unaffected by air pollution. Overall, these results confirm the conclusions drawn from Table 3.

Effect Magnitude. We conduct several exercises to assess the magnitude and the economic relevance of the estimates. Firstly, we compare the impact of a $PM_{2.5}$ shock on users' output to the effect of another highly relevant environmental shock, exposure to extreme outdoor temperatures.[20] Secondly, we compute elasticities based on the estimated effect of $PM_{2.5}$ on commits and total actions and compare these to elasticities found in previous studies on other occupations. Finally, we leverage the information from Gitcoin to translate the effects into monetary damages.

The upper panel of Figure 6 reproduces the estimated effects of a $PM_{2.5}$ shock on actions, commits, and comments in graphical form (point estimates with 95% confidence interval displayed in black on the right). In addition, coefficients from an OLS regression of the same outcomes on maximum daily temperature are presented. The quantity measures are regressed on eight dummy variables indicating whether maximum daily temperature falls in a specific percentile range, as displayed on the x-axis. The reference category is a maximum temperature value between the 35th and the 65th percentile. In addition, regressions control for minimum daily temperature, measured in the same way, further weather controls and fixed effects as in Equation (2).

For all three outcomes the effects of temperature follow the familiar inverse u-shape: Both unusually cold and unusually hot temperatures have adverse effects, but only the impact of heat is statistically significant.[21] Even though the users in our samples might work in climate-controlled office buildings, exposure to heat during commuting times or while running other errands might plausibly generate these negative effects. The impacts of heat are negative for both categories of actions considered, with slightly larger magnitude and higher level of statistical significance for commits than for comments, but the differences are not as pronounced as for the air pollution shock. Importantly, for commits and total actions, the point estimate on the $PM_{2.5}$ shock is substantially larger than the point estimate for the highest temperature bin which reflects maximum daily temperature above the 95th percentile.[22] For total actions the effect of the pollution shock is more than twice, for commits more than four times as large. The IV estimates ($PM_{2.5}$ shock) are less precise than the OLS estimates (temperature), but still, at least for commits also the lower bound of the 95% confidence interval in the pollution effect is larger than the point estimate for heat. Hence, the adverse productivity effects of poor air quality exceed those of extreme temperatures, an environmental shock of high relevance given climate change.

The height of the bars displayed in the bottom panel reflects the size of the elasticities of productivity or performance with respect to air pollution found across different studies.[23] We obtained the elasticities of total actions and commits with respect to $PM_{2.5}$ from a version of our IV model where the regressor is the logarithm of $PM_{2.5}$ (see Appendix Table A.10).[24] Given that these estimates are derived from very different settings and rely on different approaches (IV vs. OLS estimation, measurements of indoor vs. outdoor pollution), we need to proceed with caution when drawing comparisons between them. However, it stands out very clearly that our estimates are at the lower end of the range of effect sizes found so far. In particular, the effect on developers' output is much smaller than the estimates

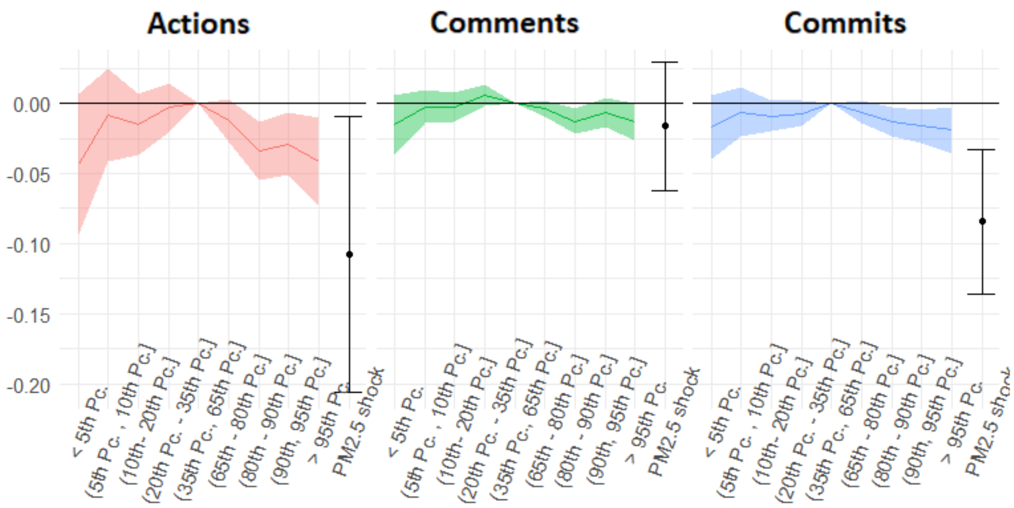
[20] This is motivated by recent findings that, in the U.S., heat exposure exerts adverse effects, e.g., on student performance on high stake exams (Park, 2020), on sentiment among twitter users (Baylis, 2020), and on mental health (Mullins and White, 2019). Please refer to these papers for more complete overviews of this literature and potential mechanisms.

[21] This is unsurprising given that by analyzing the effects of maximum daily temperature, we can better capture the impact of heat than the effect of cold, and, especially in Europe, not all office buildings are equipped with air conditioning, while heating devices are omnipresent.

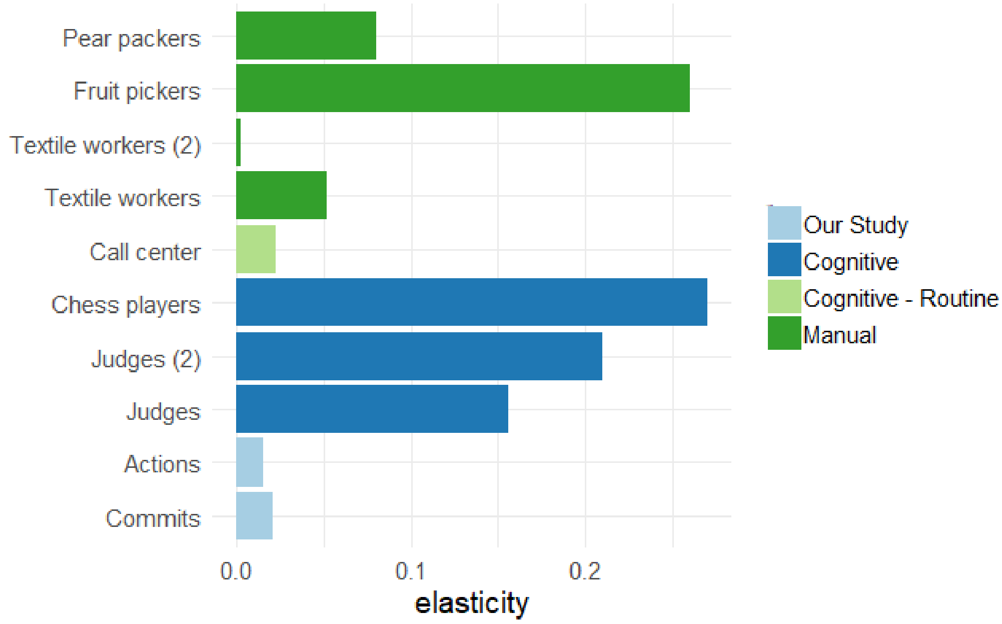
[22] Median maximum temperature in this bin is 30.9° C, while the median value in the omitted bin is 16.7° C.

[23] Air pollution is measured by $PM_{2.5}$ in all cases except for call center agents and fruit pickers.

[24] When computing elasticities based on the results in panel A of Table 3, we obtain similar values, -0.012 and -0.018 for actions and commits, respectively.



(a) Effects of $PM_{2.5}$ and Heat on Work Quantity



(b) Effects of $PM_{2.5}$ Across Occupations

Figure 6: Effect Magnitude

Note: Panel (a) reproduces the estimated effects of a $PM_{2.5}$ shock on actions, commits, and comments from Panel B of Table 3 in graphical form (point estimates with 95% confidence interval displayed in black on the right). The colored lines represent estimates from an OLS regression of the same outcomes on maximum daily temperature measured by eight dummy variables indicating whether maximum daily temperature falls in a specific percentile range, as displayed on the x axis. The reference category is a maximum temperature value between the 35th and the 65th percentile. The shaded areas are 95% confidence bands. Control variables are eight corresponding dummies for minimum daily temperature, third order polynomials for precipitation, wind speed, and relative humidity, indicators for heavy wildfire smoke and holidays, as well as city, day-of-week, year-by-month and region-by-month fixed effects. Day-of-week and year-by-month fixed effects can vary across world regions R. Standard errors are clustered at the city level and regressions are weighted by the number of active workers in a city in the current month. Panel (b) shows the elasticities of commits and actions with respect to $PM_{2.5}$, based on the estimates in column (4) and (5) of table A.10. Besides, it presents elasticities of performance with respect to air pollution from other studies, in particular: Künn et al. (2019) (Chess players), Sarmiento (2022) (Judges), Kahn and Li (2019) (Judges (2)), Chang et al. (2019) (Call center agents), He et al. (2019) (Textile Workers (2)), Adhvaryu et al. (2019) (Textile Workers), Chang et al. (2016) (Pear packers), and Graff Zivin and Neidell (2012) (Fruit pickers).

for judges and chess players, who are also engaged in cognitively demanding tasks. As outlined further above, a potential explanation for this is that the chess players and judges considered act in quite inflexible settings, namely a chess tournament and court hearings. These circumstances offer no possibility to adapt working hours or the choice of tasks to productivity shocks. This is very different in our setting, and we provide evidence on worker adjustment to an increase in $PM_{2.5}$ in the next section. This underscores the importance of our analysis: drawing conclusions on the total economic cost of air pollution based on the estimates for cognitively-demanding tasks in highly inflexible settings might be misleading because many high-skilled occupations are characterized by some degree of flexibility of workers in organizing their working day.

Even though productivity effects are small in comparison to other contexts, they might still be economically relevant, given that software development is a high-paying occupation generating large economic value. Combining our estimates with the average monetary value per commit of 112$ (see Section 2.3) implies that a standard deviation increase in $PM_{2.5}$ reduces the value of daily output by \$1.98. This is of the same order of magnitude as effects reported by Chang et al. (2016) who find that a 10 $\mu g/m^3$ increase in $PM_{2.5}$ reduces hourly output among pear packers by \$0.41, which would imply a damage of \$3.28 for a working day of eight hours.[25] During a $PM_{2.5}$ shock, which occurs on 11% of all days, output value falls by \$9.5. As discussed in Section 2.3, a value of 112$ per commit is likely a very conservative estimate for our setting. Thus, these estimates can be interpreted as a lower bound. When we base calculations of monetary damages on the average value per commit derived from Gitcoin issues of advanced difficulty level, we find a reduction in daily output by \$12 for a standard deviation increase in $PM_{2.5}$ and by \$57 for a $PM_{2.5}$ shock.

In summary, in our setting, the impact of air pollution on productivity exceeds the effect of heat, which can be informative for individuals and firms, showing that the benefits of an investments in air purifiers or HEPA filters might be as large as investments into air conditioning. In comparison to other professions, however, the effect of particulate matter is relatively small, pointing towards an important role of worker adaption in flexible work environments. Economically, the productivity effects are nevertheless relevant, given the high monetary value of software.

4.2 Worker Adjustment

Switching to Easy Tasks. Exploiting the information on issue labels, we analyze whether workers switch to easier tasks on days with worse air quality, which might dampen the adverse effect on output quantity. In Table 4 we present estimates of the impact of pollution on the share of issue events completed that refer to an easy issue (Column (1)). As this outcome is only defined for city$\times$day observations if any issue event was conducted (issue opened, closed or reopened, or a comment written on an issue), the number of observations is somewhat reduced. We find that the share of events referring to easy issues increases if $PM_{2.5}$ concentration rises. In terms of magnitude, an increase in pollution by one local standard deviation raises the share by 1.2%. On days hit by a fine particulate matter shock, the variable even increases by 7.1% of the mean, or 0.5 percentage points.

In the second column we analyze how the denominator of the share changes. Consistent with prior results on the effect of $PM_{2.5}$ on interactive tasks, we find no statistically or economically significant

[25] The average city-specific standard deviation of $PM_{2.5}$ in our sample is 11.38 $\mu g/m^3$.

Table 4: Effect of $PM_{2.5}$ on Activity on Easy Issues

	Share Easy Issue Events	*Issue Events*
	(1)	(2)
Panel A.		
$PM_{2.5}$ (st. dev.)	0.0008*	−0.0037
	(0.0005)	(0.0044)
	[0.092]	[0.399]
F-Statistic	94.3	112.3
% change in Y	1.2	-0.4
Panel B.		
$PM_{2.5}$ shock	0.0047**	−0.0103
	(0.0022)	(0.0197)
	[0.038]	[0.600]
F-Statistic	30.0	34.6
% change in Y	7.1	-1.2
Observations	258,031	331,025
Mean Dep. Var	0.066	0.87

Note: The table presents IV estimates of the parameter β in Equation (2). The outcome in Column (2) is defined as the sum of actions referring to issues, i.e., the number of issues opened, closed, reopened, and the number of issue comments written. The outcome in Column (1) is defined as the ratio of the number of these activities which refer to an issue classified as *easy* based on issue labels (see Section 2 for details) and the total number of issue events. In Panel A, the regressor of interestis $PM_{2.5}$ concentration measured in city-specific standard deviations. In Panel B, a binary $PM_{2.5}$ shock variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city×month×day-of-week specific average by at least one city-specific standard deviation. The first stage specification is given in Equation (3). Covariates include eight bins for mean daily temperature, third-order polynomials in wind speed, precipitation and relative humidity, indicators for heavy wildfire smoke and holidays, as well as city, day-of-week, year-by-month and region-by-month fixed effects. Day-of-week and year-by-month fixed effects and the temperature controls can vary across world regions R. Regressions are weighted by the number of active workers in a city during the current month. Standard errors clustered at the city level are reported in parentheses. P-values are reported in brackets. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

effects. Thus, changes in the share of easy issue events are not driven by changes in the denominator, but by a switch towards more easy issues for a relatively constant activity level with respect to issue events.

We find similar evidence for switching towards easier tasks, when we consider the complexity of coding tasks. In particular, we analyze effects of $PM_{2.5}$ on three measures for the scope of the task addressed in the pull request: number of files changed, lines of code added, and lines of code deleted, all measured per PR opened.[26] While there can be highly complex tasks that involve a lot of thinking but require only a small change in the code, we believe that these measures provide reasonable proxies of action complexity. Fixing a severe bug for instance likely requires changes in different parts of the source code, which implies a larger number of files changed. While our primary measure of coding activity is the number of commits, information on the number of lines of code written or files changed in a commit is not included in the data. Given that both, commits and PRs, reflect individual coding activity and that we found very similar effects in terms of relative magnitude for both types of actions,

[26]These variables are based on GHArchive data, whereas work quantity results used GHTorrent data. GHArchive data is only available from 2015 onward. In Appendix Table A.7 we show that using data on pull requests opened from GHArchive we can replicate the result presented in Table A.6 for PRs opened measured in the GHTorrent data, i.e., we find a negative and statistically significant effect on the variable when air quality deteriorates.

results on PRs likely carry over to commits as well.

Table 5 presents the results. The sample size is reduced relative to the previous tables, because the outcomes are defined only for city×day observations were any PR was opened. Consequently, the F-statistics also drop, but still reach a level above 60 for $PM_{2.5}$ measured in standard deviations and a value above 20 for the binary shock treatment. Given the fact that the mean values of the three outcomes are considerably larger than for the output measures analyzed before (6.1, 178.1, and 66.0), we apply the inverse hyperbolic sine transformation to the variables. Hence, coefficients can be interpreted as percentage changes. We find negative point estimates across all three outcomes and the two distinct ways to measure air pollution. The number of files changed falls by 0.9% in response to an increase in $PM_{2.5}$ concentration by one local standard deviation, while the number of lines added drops by 1.7%. The effect on the number of lines deleted is also negative, but only about half as large and not significantly different from zero. This pattern is plausible, given that tasks related to deleting code, e.g., cleaning or polishing a file or dropping a deprecated or redundant part, are often easier than creating new code. The same pattern emerges for the $PM_{2.5}$ shock regressor. However, the effects are somewhat less significant, likely due to the lower first stage strength.

In sum, the results imply that, on top of the overall reduction in the number of actions completed, users switch towards less complex tasks on high pollution days. Thus, estimates of monetary effects of air pollution exposure presented in the previous section provide a lower bound, given that estimates in Table A.3 imply that more complex pull requests, measure by the number of lines of code added and files changed, are rewarded higher payments, even when holding the number of commits constant.

Working Hours. **To be completed**

4.3 Work Quality

To be completed

4.4 Effect Heterogeneity

Experience. In this section we explore heterogeneity in the effect of fine particulate matter on work quantity and adjustment. We start by investigating heterogeneity along user characteristics. While we do not observe any user demographics, we can measure experience in working with GitHub. Specifically, we combine tenure (time since registration on GitHub) and the number of followers at the point in time the user enters our sample into an index which represents a users experience and popularity.[27] We split the sample users into terciles based on experience and run the IV regression at the user×day level (the second stage is given by Equation 1). Table 6 presents results estimated separately for the users in the bottom and the top tercile, respectively.[28]

The first three columns display estimated effects of $PM_{2.5}$ measured in local standard deviations on the three primary measures of work quantity. The point estimates are negative in both samples, but larger in absolute terms as well as relative to the sample means for the most experienced users. In the inexperienced subsample, effects are not significantly different from zero. For the adaptation variables

[27] The index is computed as the average of tenure and the number of followers, after standardizing both variables.

[28] Median tenure (number of followers) is 1.4 years (11) in the bottom tercile and and 5.6 years (30) in the upper tercile.

Table 5: Effects of $PM_{2.5}$ concentration on Pull Request Size

	Files changed per PR (1)	*Lines added per PR* (2)	*Lines deleted per PR* (3)
Panel A.			
$PM_{2.5}$ (st. dev.)	-0.0094^{**}	-0.0171^{*}	-0.0086
	(0.0041)	(0.0087)	(0.0069)
	[0.023]	[0.051]	[0.210]
F-Statistic	67.6	67.6	67.6
Panel B.			
$PM_{2.5}$ shock	-0.0429^{**}	-0.0742	-0.0501
	(0.0207)	(0.0449)	(0.0322)
	[0.040]	[0.100]	[0.121]
F-Statistic	21.0	21.0	21.0
Observations	155,060	155,060	155,060

Note: The table presents IV estimates of the parameter β in equation (2).Inverse hyperbolic sine transformations are applied to all outcomes. In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in city-specific standard deviations. In Panel B, a binary $PM_{2.5}$ shock variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city×month×day-of-week specific average by at least one city-specific standard deviation. The first stage specification is given in equation (3). Covariates include eight bins for mean daily temperature, third-order polynomials in wind speed, precipitation, and relative humidity, indicators for heavy wildfire smoke and holidays, as well as city, day-of-week, year-by-month, and region-by-month fixed effects. Day-of-week and year-by-month fixed effects and the temperature controls can vary across world regions R. Regressions are weighted by the number of active workers in a city during the current month. Standard errors clustered at the city level are reported in parentheses. P-values are reported in brackets. *p<0.1; **p<0.05; ***p<0.01

examined in the last three columns, the pattern is reverse: While the direction of the effects is again the same in both samples, effects are now stronger among the less experienced users and not significantly different from zero in the upper tercile.

We view these results as a confirmation that switching to easier tasks is indeed an adaptation strategy to increases in particulate matter exposure that allows users to maintain their usual level of work quantity. A potential reason why the least experienced users show the strongest adjustment response might be that they have the largest incentive to keep up a high activity level because the number of actions performed per day in public repositories is visualized on a user's GitHub profile and might be an important signal to other users or potential employers. Work complexity on the other hand is less easily observable.

Table 6: Heterogeneity: User Experience

	Actions	*Commits*	*Comments*	*Share Easy Issue Events*	*Lines added per PR*	*Files changed per PR*
	Panel A: Bottom Tercile of Experience					
$PM_{2.5}$ (st. dev.)	-0.019	-0.0172	-0.0056	0.0033^{***}	-0.0307^{**}	-0.0177^{***}
	(.023)	(.0106)	(.0118)	(.0011)	(.0143)	(.0065)
F-Statistics	2955.2	2955.2	2955.2	568.7	239.2	239.2
% change in Y	-0.7	-1.3		4.4	-3.1	-1.8
Observations	3,837,730	3,837,730	3,837,730	772,169	331,159	331,159
	Panel B: Upper Tercile of Experience					
$PM_{2.5}$ (st. dev.)	-0.0630^{***}	-0.0355^{***}	-0.0129^{*}	0.0009	-0.0115	-0.0058
	(.0196)	(.0097)	(.0077)	(.0006)	(.0127)	(.0071)
F-Statistic	3094.5	3094.5	3094.5	783.0	309.4	309.4
% change in Y	-1.9	-2.5	-1.1	1.5	-1.2	-0.6
Observations	4,157,687	4,157,687	4,157,687	1,059,060	410,361	410,361

Note: The table presents IV estimates of the parameter β in equation (1). Inverse hyperbolic sine transformations are applied to outcomes in columns 5 and 6. The regressor of interest is $PM_{2.5}$ concentration measured in city-specific standard deviations. The first stage specification is given in equation (3). Covariates include eight bins for mean daily temperature, third-order polynomials in wind speed, precipitation, and relative humidity, indicators for heavy wildfire smoke and holidays, as well as user, day-of-week, year-by-month, and region-by-month fixed effects. Day-of-week and year-by-month fixed effects and the temperature controls can vary across world regions R. Standard errors clustered at the city level are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Local Income. Secondly, we explore differences in effect magnitude between places with relatively high vs. low income levels. We assign cities into subsamples based on GDP per capita in 2014, the first year of our sample period, converted by the purchasing power parity conversion factor to adjust for differences in local price levels. The required data is collected from the OECD, World Bank, and national statistical offices.[29] Specifically, we split the sample cities into two groups based on whether

[29]The main data source is the OECD's database on metropolitan areas, available at `stats.oecd.org/Index.aspx?DataSetCode=CITIES`. It provides GDP per capita for metropolitan areas, i.e., for some smaller cities in our sample we do not have city-specific data, but instead assign the value for the respective metro area. Small cities in Silicon Valley, e.g.,

GDP per capita in the city is above or below the *region-specific* median[30]. This classification is chosen to ensure that this heterogeneity analysis does not simply correspond to comparing effects across different regions, as these might differ in aspects other than income. Hence, the *above median GDP per capita* subsample contains the cities with relatively high income from each of the seven geographic regions, but includes some cities with lower income than the *below median GDP per capita* sample. Median GDP per capita amounts to $41,548 in the low income sample, and $63,560 in the high income sample. Overall, GDP per capita spans from less than $10,000 in, e.g., Bengalore and Hyderabad, to more than $80,000 in, e.g., Geneva, Singapore, Greater Boston, and Greater San Francisco. The full distributions by group are displayed in panel (a) of Figure A.2. Results for the work quantity outcomes are reported in Table 7, starting with the below median GDP sample, i.e., the relatively poor cities. As in the full sample, the number of total actions conducted and the number of commits fall in response to an increase in air pollution, while comments are less affected. The effect magnitudes are larger than in the main sample. Point estimates in the high-income (above median income) subsample, displayed in the last three columns, are of smaller magnitude, and only the effect on commits is marginally significant.

Given the definition of the regressor, the differences in effects between samples could be driven by higher local standard deviations in less wealthy and more polluted cities. The finding that the effect on total actions in the rich subsample is not even statistically significant alleviates this concern. In panel (b) of Figure A.2 we plot the cumulative distribution function of the local standard deviations for both samples. While the relatively poor sample indeed includes more cities with extremely large standard deviations, the curves track each other very closely up to the 80th percentile. In sum, this suggests that the differences reflect true heterogeneity and not just a mechanical effect. Locations with low income levels bear a larger burden of the negative impacts of $PM_{2.5}$ on worker productivity, even in a high-paying, skilled profession like software development.

Non-Linearity. Lastly, we explore non-linearity of the relationship between $PM_{2.5}$ and work quantity. To that end, we turn to OLS estimation of the model in equation 2, but replace $PM_{c,d}$ by a series of dummy variables indicating whether $PM_{2.5}$ concentration falls into a specific bin.[31] As outlined above, the OLS estimator is likely inconsistent due to endogeneity from omitted variables and measurement error. In Table A.8 we present OLS estimates to assess the direction and size of the bias. We find that the OLS estimator yields negative, but insignificant results and is biased towards zero. When we include region$\times$date fixed effects, which absorb all region-wide shocks to user activity on a given date which might be correlated with $PM_{2.5}$ concentration, the bias is reduced and the effect on commits gets significant. The results replicate the pattern that effects on commits are larger than on comments. Given the finding that the OLS results underestimate the true effects, we now turn to an exploration of non-linearity, bearing in mind that all results should be interpreted as reflecting lower bounds. Figure 7 displays estimated effects of the $PM_{2.5}$ bin variables on actions and commits. Estimated coefficients reflect the impact of moving from a $PM_{2.5}$ concentration between 15 and 20 µg/m^3 to the respective bin.

Cupertino, Palo Alto and Mountain View are assigned the GDP per capita reported for Greater San Francisco. Data for cities outside OECD countries is collected from national statistical agencies, the OECD regional statistics database, or the World Bank.

[30]This refers to the big regions R, i.e., North America, Northern, Western, Eastern, and Southern Europe, Asia, and Oceania.

[31]Bins are defined for (0-5], (5-10], (10-15], (20,25], (25,30], (30,40], (40,50], (50,60], (60,70], (70,80], (80,120],(120, 160] and (160,1200], all in µg/m^3, with (15-20] as reference bin.

Table 7: Heterogeneity: GDP per capita

	Below Median GDP per capita			Above Median GDP per capita		
	Actions	*Commits*	*Comments*	*Actions*	*Commits*	*Comments*
	(1)	(2)	(3)	(4)	(5)	(6)
PM2.5 (st. dev.)	-0.0384^{***}	-0.0213^{**}	-0.0110	-0.0122	-0.0117^{*}	-0.0010
	(0.0133)	(0.0083)	(0.0072)	(0.0129)	(0.0066)	(0.0059)
	[0.005]	[0.012]	[0.128]	[0.347]	[0.082]	[0.863]
Dep. Var. mean	2.59	1.27	0.82	2.81	1.33	0.93
Observations	164,754	164,754	164,754	166,271	166,271	166,271
F- Statistic	90.8	90.8	90.8	86.0	86.0	86.0

Note: To form the sub-samples of cities with above and below median GDP per capita, a city's GDP per capita in 2014 is compared to the median value in the geographic region R the city lies in. Data on per capita GDP is collected from the OECD, World Bank and national statistical offices. Estimated coefficients reflect 2SLS estimates of the parameter β in Equation (2). The first stage specification is given in Equation (3). Covariates include eight bins for mean daily temperature, third-order polynomials in wind speed, precipitation, and relative humidity, indicators for heavy wildfire smoke and holidays, as well as city, day-of-week, year-by-month and region-by-month fixed effects. Day-of-week, and year-by-month fixed effects and the temperature controls can vary across world regions R. Regressions are weighted by the number of active workers in a city during the current month. Standard errors clustered at the city level are reported in parentheses. P-values are reported in brackets. *p$<$0.1; **p$<$0.05; ***p$<$0.01.

Interestingly, being exposed to a $PM_{2.5}$ level below 5 µg$/m^3$ has a significant positive impact on both total actions (point estimate = 0.033, p-value = 0.038) and commits (point estimate = 0.013, p-value = 0.090). This implies that even in cities with low to moderate levels of $PM_{2.5}$, further improvements in air quality can generate positive effects on worker productivity. We find significant negative effects starting at a concentration of approximately 80 µg$/m^3$, but not significant differences in the outcomes for concentrations between 5 and 70 µg$/m^3$. As mentioned above, we have to bear in mind that the OLS estimates likely underestimate true effects of $PM_{2.5}$ exposure. The point estimates suggest a roughly linear relationship between air pollution and output.[32]

4.5 Robustness and Further Results

Robustness Checks. In Appendix Tables A.9 to A.11 we show that our qualitative results are not sensitive to specific choices on how we set up the first and second stage models. First, we examine robustness to the specification of the wind direction instruments. Instead of $\sin(\theta_{c,d})$ and $sin\left(\frac{\theta_{c,d}}{2}\right)$, we use three indicator variables for average daily wind direction falling into a specific 90° bin (south-west, south-east and north-east, with north-west as omitted category), following Deryugina et al. (2019). The results are reported in columns (1) to (3) of Table A.9. In the remaining three columns we report results from a specification where we used a hierarchical clustering algorithm, instead of k-means clustering, to form the city-groups g across which the effects of wind direction are allowed to differ in the first stage. In both cases, results on work quantity are very similar to the baseline results.

Secondly, we test robustness to the functional form chosen in the second stage model. Table A.10

[32]While the point estimate on the highest bin ($PM_{2.5}$ above 160) is more than twice as high as for the second highest bin, note that average concentration in that bin is 293 µg$/m^3$, also more than twice as high as the average level of 137 in the second highest bin.

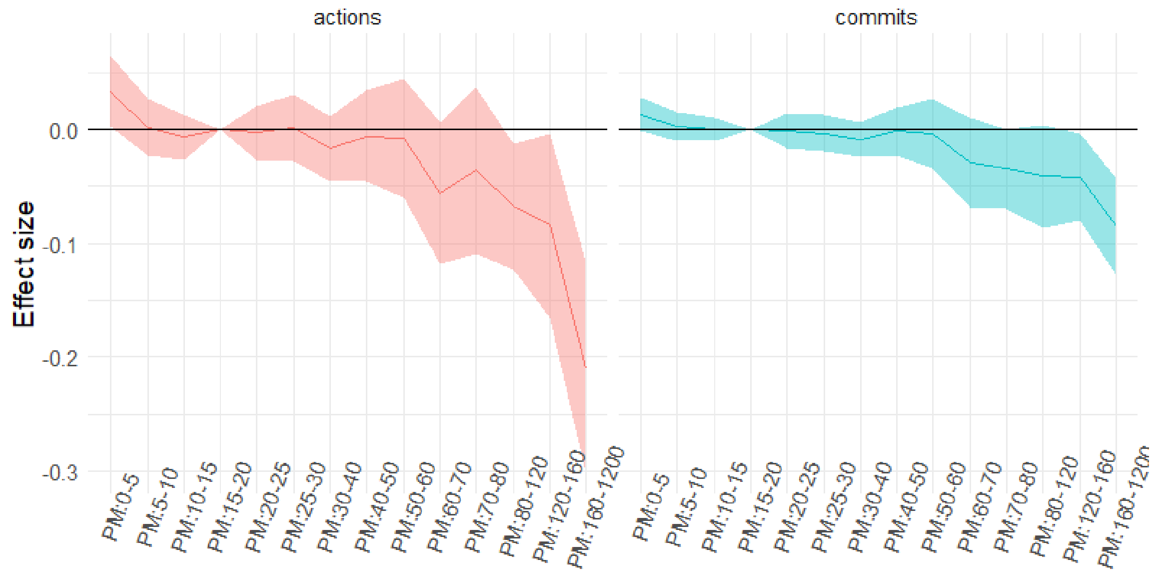


Figure 7: Non-linear effects of PM2.5 on Work Quantity (OLS estimates)

Note: Plot depicts point estimates for different bins of $PM_{2.5}$ concentrations from an OLS regressions of total actions (left) and commits (right), respectively, on indicators for each bin and covariates as in A.8, Column (2) and (4). Bin width is given at the bottom and refers to $\mu g/m^3$. Shaded areas indicate 95%-confidence intervals.

shows estimated effects of $PM_{2.5}$ when work output is measured by the inverse hyperbolic sine transformation of total actions, commits, and comments, respectively. Again, the direction and statistical significance of the baseline results persist, but the specification using inverse hyperbolic sine transformations implies somewhat smaller effect magnitudes. The last three columns of the table display results when $PM_{2.5}$ in logs is used as regressor. This yields a high F-Statistics and the same pattern for second stage effects on work quantity as the baseline model.

Finally, in panel B of Table A.11 we show that the statistical significance of our results persists if we cluster standard errors at the level of the city-groups g across which the effects of wind direction are allowed to differ in the first stage, instead of the city level. Panel A shows "reduced form" results, where we regress the outcomes on an indicator variable which is one when wind blows to the city from the direction which increases local PM2.5 levels most in the respective city-group g, $WDir\ highPM_{c,d}$.[33] For work quantity and adjustment outcomes, sign and significance of the estimated coefficients are the same as in the 2SLS estimation. The first stage effect reported at the bottom of panel A, implies that wind from a city's high pollution direction raises $PM_{2.5}$ concentration on average by 0.39 local standard deviations relative to days where wind arrives from another direction. While this approach is much less flexible than our main 2SLS model, it provides an intuitive check for the underlying idea.

[33]To identify this direction for each city-group g, we run the first stage model with the level of $PM_{2.5}$ concentration as outcome and three dummies for average daily wind direction falling into a specific 90° bin as instruments, interacted with the city-group indicators.

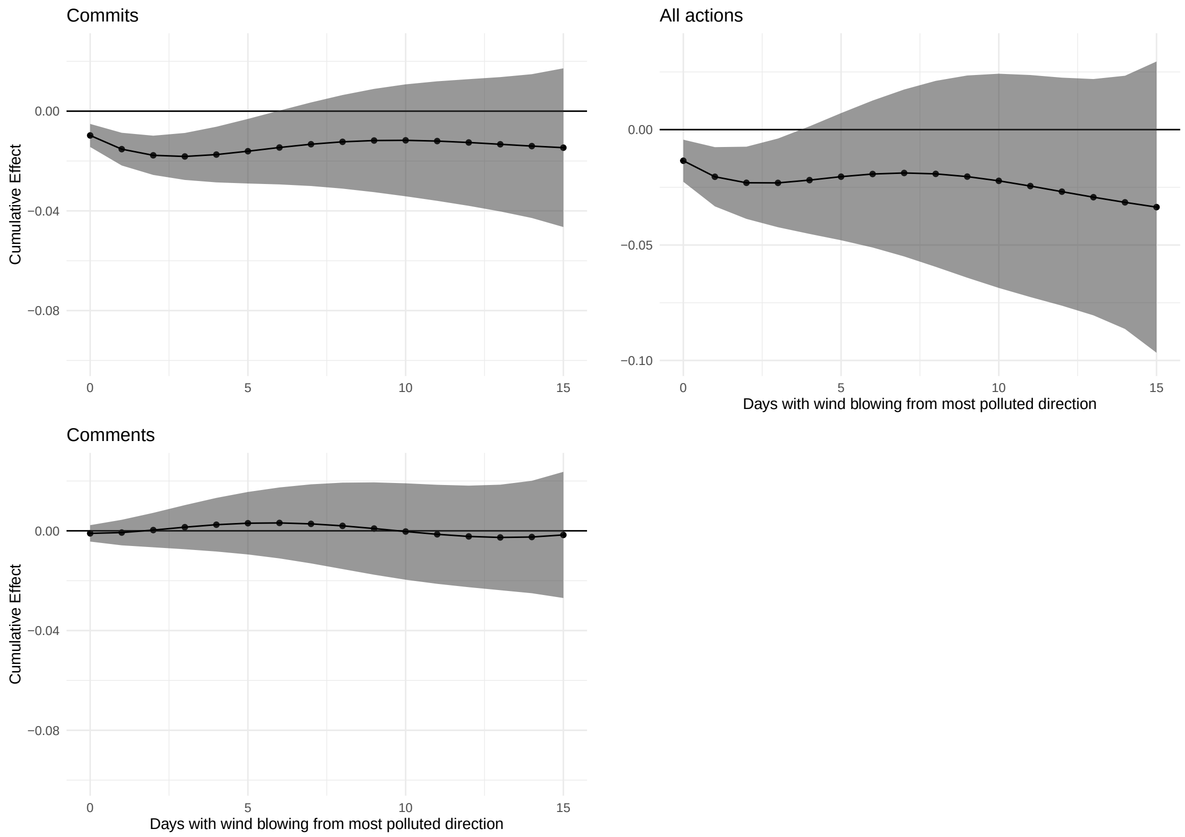


Figure 8: Effect Dynamics: Work Quantity

Note: The plots depict estimates of the cumulative effect of wind blowing from the high pollution direction on three measures of users' work quantity. Effects are derived from a fourth order polynomial distributed lag specification. The x-axis denotes the number of days over which the cumulative effect is computed. Shaded areas represent 95% confidence intervals. Regressions control for city, day-of-week, year-by-month and region-by-month fixed effects, a holiday indicator and weather controls for the current day and eight lags (third order polynomials in mean daily temperature, precipitation, relative humidity and wind speed). Regressions are weighted by the number of active workers in a city during the current month and standard errors are clustered at the city level.

Effect Dynamics. We use a version of this "reduced form" model including lags of $WDirhighPM_{c,d}$ to explore effect dynamics. The existing literature on air pollution and worker productivity found mixed results on the lagged impact of exposure. He et al. (2019) for example demonstrate evidence for lagged effects of $PM_{2.5}$ and SO_2 on the productivity of textile workers in industrial towns in China, while Künn et al. (2019) find that chess players' performance is unaffected by pollution exposure on the previous days.

We estimate the cumulative effect of elevated pollution exposure on the current day and up to 15 lags, using the indicator for wind blowing to city c from the high pollution direction, $WDirhighPM_{c,d-p}$, where $p \in \{0, 1, ..., 15\}$ denotes the lag order. Following He et al. (2019) we use a polynomial distributed lag (PDL) specification, where the coefficients on the current and lagged wind direction indicators, $\beta_0, \beta_1, ..., \beta_{15}$, are restricted to follow a smooth polynomial function. This serves to constrain the parameters as an unrestricted distributed lag model typically yields very imprecise estimates due to high serial correlation in the regressors $WDir\ highPM_{c,d-p}$. Imposing that the coefficients follow a polynomial function reduces the number of parameters that have to be estimated. Specifically we assume that $\beta_p = \alpha_0 + \alpha_1 p + \alpha_2 p^2 + \alpha_3 p^3 + \alpha_4 p^4$ for $p \in \{0, 1, ..., 15\}$, i.e., we use a fourth order polynomial. Substituting this imposed shape of the β_p into the unrestricted distributed lag model, yields five new regressors which replace the $WDir\ highPM_{c,d-p}$ and exhibit much less multicollinearity.

We estimate the PDL model for our three main quantity outcomes, recover the estimated effects $\hat{\beta}_p$, and plot the cumulative effect of exposure to wind from the high pollution direction for s consecutive days, $\sum_{p=0}^{s} \hat{\beta}_p$ for $s = 0, 1, .., 15$ in Figure 8. All models include the same covariates as our main contemporaneous model plus eight lags of weather conditions. For both total actions and commits we observe that contemporaneous exposure to particulate matter has the largest impact on activity, but exposure on the two preceding days also generates negative effects. The current day effect of $WDir\ highPM_{c,d}$ is estimated to be -.014 for total actions and -.010 for commits, very close to the estimates in the contemporaneous model (see Appendix Table A.11). After three consecutive days of wind from the high pollution direction, the cumulative effect is -.023 and -.018, respectively. For more prolonged exposure the point estimate of the cumulative effect remains rather constant, but becomes much more noisy. For comments, the cumulative effect always remains close to zero. In sum, pollution exposure generates an adverse effect on current-day output and, to a smaller extent, also reduces productivity on the following two days. Compared to health impacts, the productivity effects are rather immediate.[34] Appendix Figure A.3 shows cumulative effects estimated from PDLM for the adjustment outcomes, which confirm this result.

5 Conclusion

This paper provides evidence that air pollution adversely affects the productivity of software developers—highly-skilled workers in a work environment that can be considered as representative of many white-collar jobs that form the backbone of a modern information economy.

On a day with unusually high levels of $PM_{2.5}$, relative to the city-month-day of week specific average, total actions conducted by the developers fall by 3.9%. This effect is mostly driven by a reduction

[34]Barwick et al. (2018) for instance find that $PM_{2.5}$ exposure raises medical expenditures up to 90 days post exposure.

in *individual* coding activity by 6.5%, while the level of *collaborative* activity is unaffected.

Relative to outdoor heat, another relevant environmental shock, the impact of such an air pollution shock is sizable. However, in comparison to effects found for $PM_{2.5}$ exposure in other occupations in previous research, our estimate is at the lower end. We provide novel evidence that this might be explained by typical features of highly-paid white-collar jobs, namely a certain degree of flexibility in the choice of tasks as well as a lot of interactive work. In particular, we find that users switch towards less complex tasks on more polluted days: The share of of actions referring to easy issues increases by 7.1% relative to the mean. Besides, submitted pull requests contain 4.3% fewer files and 7.4% fewer new lines of code. Among users with a stronger adjustment response to $PM_{2.5}$ exposure, effects on work quantity are alleviated.

Even though the productivity effects are relatively small compared to other professions, they are economically relevant due to the high value generated by software developers. Using novel data from Gitcoin, we can translate our estimates into monetary effects. Given a mean value of $112 per commit, on a day with a pollution shock, output value falls by $9.50 per user. As pollution shocks occur on 10.5% of all days, the total monetary damage for our sample of 26,000 users observed between February 2014 and May 2019 amounts to $15.7 million. This is likely a conservative estimate and does not account for the lower value of work on easier tasks. Using the monetary value of commits of difficulty level *advanced*, the implied damages are as high as $95.6 million.

Our heterogeneity analysis implies that effects are larger in less wealthy cities, suggesting that high pollution levels in developing countries with large software industries, e.g., India and Bangladesh, might be an important barrier to industry growth.

In ongoing work, we explore effects of $PM_{2.5}$ on work *quality* and on worker adjustment with respect to working hours.

References

Adhvaryu, A., N. Kala, and A. Nyshadham (2019): "Management and Shocks to Worker Productivity," Working Paper 25865, National Bureau of Economic Research.

Archsmith, J., A. Heyes, and S. Saberian (2018): "Air Quality and Error Quantity: Pollution and Performance in a High-Skilled, Quality-Focused Occupation," *Journal of the Association of Environmental and Resource Economists*, 5, 827–863.

Barwick, P. J., S. Li, D. Rao, and N. Zahur (2018): "The morbidity cost of air pollution: evidence from consumer spending in China," *Available at SSRN 2999068*.

Bassi, V., M. E. Kahn, N. Lozano Gracia, T. Porzio, and J. Sorin (2021): "Pollution in Ugandan Cities: Do Managers Avoid it or Adapt in Place?" *Available at SSRN 3887079*.

Baylis, P. (2020): "Temperature and temperament: Evidence from Twitter," *Journal of Public Economics*, 184, 104161.

Bureau of Labor Statistics (2021): "Software Developers, Quality Assurance Analysts, and Testers," in *Occupational Outlook Handbook*, U.S. Department of Labor.

Burke, M., S. Heft-Neal, J. Li, A. Driscoll, P. W. Baylis, M. Stigler, J. Weill, J. Burney, J. Wen, M. Childs, and C. Gould (2021): "Exposures and Behavioral Responses to Wildfire Smoke," Working Paper 29380, National Bureau of Economic Research.

Chang, T., J. Graff Zivin, T. Gross, and M. Neidell (2016): "Particulate Pollution and the Productivity of Pear Packers," *American Economic Journal: Economic Policy*, 8, 141–169.

Chang, T. Y., J. Graff Zivin, T. Gross, and M. Neidell (2019): "The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China," *American Economic Journal: Applied Economics*, 11, 151–172.

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015): "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings," *American Economic Review*, 105, 678–709.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019): "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction," *American Economic Review*, 109, 4178–4219.

Deschênes, O., M. Greenstone, and J. S. Shapiro (2017): "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program," *American Economic Review*, 107, 2958–2989.

Donald, S. G. and K. Lang (2007): "Inference with Difference-in-Differences and Other Panel Data," *Review of Economics and Statistics*, 89, 221–233.

Graff Zivin, J. and M. Neidell (2012): "The Impact of Pollution on Worker Productivity," *American Economic Review*, 102, 3652–3673.

——— (2014): "Temperature and the Allocation of Time: Implications for Climate Change," *Journal of Labor Economics*, 32, 1–26.

He, J., H. Liu, and A. Salvo (2019): "Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China," *American Economic Journal: Applied Economics*.

Heyes, A., N. Rivers, and B. Schaufele (2019): "Pollution and Politician Productivity: The Effect of PM on MPs," *Land Economics*, 95, 157–173.

Huang, J., N. Xu, and H. Yu (2020): "Pollution and Performance: Do Investors Make Worse Trades on Hazy Days?" *Management Science*, 66, 4455–4476.

Jans, J., P. Johansson, and J. P. Nilsson (2018): "Economic status, air quality, and child health: Evidence from inversion episodes," *Journal of Health Economics*, 61, 220–232.

Kahn, M. E. and P. Li (2019): "The Effect of Pollution and Heat on High Skill Public Sector Worker Productivity in China," Working Paper 25594, National Bureau of Economic Research.

Künn, S., J. Palacios, and N. Pestel (2019): "Indoor Air Quality and Cognitive Performance," Tech. rep., IZA.

La Nauze, A. and E. R. Severnini (2021): "Air Pollution and Adult Cognition: Evidence from Brain Training," Working Paper 28785, National Bureau of Economic Research.

Lazear, E. P., K. L. Shaw, and C. Stanton (2015): "Making Do with Less: Working Harder during Recessions," *Journal of Labor Economics*, 34, S333–S360.

LoPalo, M. (2020): "Temperature, worker productivity, and adaptation: evidence from survey data production," *Available at SSRN 3711044*.

Manisalidis, I., E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou (2020): "Environmental and Health Impacts of Air Pollution: A Review," *Frontiers in Public Health*, 8, 14.

Mas, A. and A. Pallais (2020): "Alternative Work Arrangements," *Annual Review of Economics*, 12, 631–658.

McDermott, G. R. and B. Hansen (2021): "Labor Reallocation and Remote Work During COVID-19: Real-time Evidence from GitHub," Tech. rep., National Bureau of Economic Research.

Mullins, J. T. and C. White (2019): "Temperature and mental health: Evidence from the spectrum of mental health outcomes," *Journal of Health Economics*, 68, 102240.

Neidell, M., J. G. Zivin, M. Sheahan, J. Willwerth, C. Fant, M. Sarofim, and J. Martinich (2021): "Temperature and Work: Time Allocated to Work under Varying Climate and Labor Market Conditions," *PLOS ONE*, 16, e0254224.

Park, R. J. (2020): "Hot temperature and high stakes performance," *Journal of Human Resources*.

Pencavel, J. (2015): "The Productivity of Working Hours," *The Economic Journal*, 125, 2052–2076.

Sarmiento, L. (2022): "Air pollution and the productivity of high-skill labor: evidence from court hearings," *The Scandinavian Journal of Economics*, 124, 301–332.

Schlenker, W. and W. R. Walker (2016): "Airports, Air Pollution, and Contemporaneous Health," *The Review of Economic Studies*, 83, 768–809.

Shangguan, R., J. DeVaro, and O. Hideo (2021): "Enhancing Team Productivity through Shorter Working Hours: Evidence from the Great Recession," Tech. rep., The Research Institute of Economy, Trade and Industry.

Tan, X., M. Zhou, and Z. Sun (2020): "A first look at good first issues on github," in *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 398–409.

A Appendix

Table A.1: Labels Indicating *Easy* Issues

good first issues	good first bug	good-first
documentation	polish	cleanup
simple	easy	small
trivial	minor	**help wanted**
junior job	newcomer	starter
beginner	newbie	novice
low hanging	low-hanging	

*Note*If a label contains any of these terms, the issue is classified as "easy". Bolt text indicates GitHub default labels.

Table A.2: Number of Commits and Hours Worked on a PR

	Dependent variable: $hoursworked_i$			
	(1)	(2)	(3)	(4)
$commits_i$	0.375***	0.939***		
	(.132)	(.341)		
$log(commits_i)$			2.375***	10.346***
			(.648)	(3.535)
x $\mathbb{1}\{Difficulty_i = Beginner\}$		−0.882**		−9.748***
		(.354)		(3.574)
x $\mathbb{1}\{Difficulty_i = Intermediate\}$		−0.667*		−8.471**
		(.349)		(3.560)
Observations	271	267	271	267

Note The table presents results from OLS regressions using data on the sample of Gitcoin pull requests. Observations are at the pull request level. Dependent variable is the number of hours worked reported by the PR author. In colunm 1 the only explanatory variable is the number of commits in the PR. Column 2 adds dummies for issue difficulty and interactions between the number of commits and the difficulty dummies. The ommited difficulty category is *advanced*. In columns 3 and 4 report results from the same models except that the number of commits is replaced by the logarithm thereof. Robust standard errors are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table A.3: Gitcoin payments, PR size and complexity

	Dependent variable: $log(payment_i)$		
	(1)	(2)	(3)
$log(commits_i)$	0.143**	0.136**	0.070
	(0.067)	(0.068)	(0.058)
$files\ changed_i$	0.005	0.007*	0.011***
	(0.005)	(0.004)	(0.004)
$log(lines\ added_i)$	0.152***	0.112***	0.091***
	(0.036)	(0.035)	(0.028)
Year dummies	✓	✓	✓
Issue difficulty dummies		✓	
Issue type dummies		✓	
Repository fixed effects			✓
Observations	292	274	292

Note The table presents results from OLS regressions using data on the sample of Gitcoin pull requests. Observations are at the pull request level. Dependent variable isthe logarithm of the payment awarded to the PR author. Explanatory variables are the number of commits and the number of lines of code added in the PR (both in logs), the number of code files changed and dummies for the year the pull request was created. Column 2 adds dummies for issue difficulty and issue type. Column 3 instead adds fixed effects for the repository. The number of lines of code added and of files changed in the PR are winsorized at the 1st and the 99th percentile. Robust standard errors are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table A.4: Sources of Air Quality Data

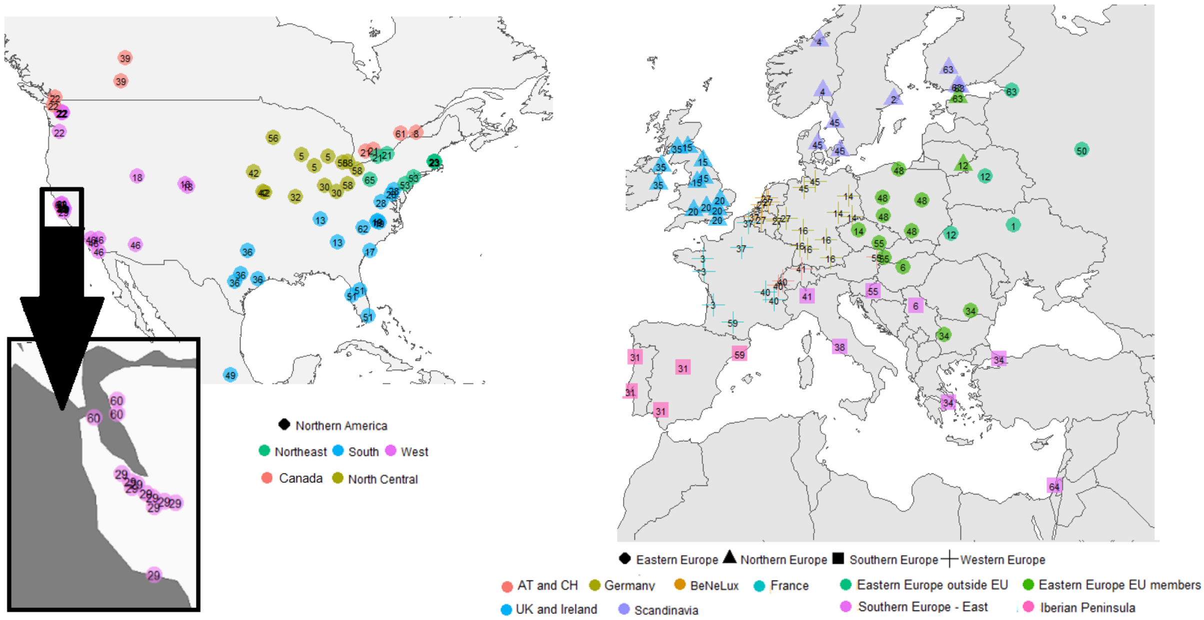
Geographic Area	**Data Source**
United States	U.S. Environmental Protection Agency (EPA)
Canada	Canadian National Air Pollution Surveillance (NAPS) Program
Mexico City	Gobierno de la Ciudad de México
Europe	European Environment Agency (EEA)
Russia, Ukraine, Belarus, Turkey, Israel	Copernicus Atmosphere Monitoring Service (CAMS)
China	National Environmental Monitoring Centre
Mumbai Hyderabad Chennai New Delhi	US Embassies (AirNow.gov)
Bengaluru	Central Pollution Control Board (CPCB)
Japan	National Institute for Environmental Studies
Hong Kong	Hong Kong Environmental Protection Department
Singapore	National Environment Agency
South Korea	Air Korea
Australia	New South Wales Department of Planning and Environment Victorian Government open data portal Queensland Government open data portal South Australian Government Data Directory
New Zealand	Stats NZ Tatauranga Aotearoa

Note: Data sources for data on PM2.5. Airbase, the EEA's database on air pollution, contains monitor data for 33 countries, including all EU members, as well as further EEA member and cooperating countries, e.g., Switzerland, Norway and Serbia.

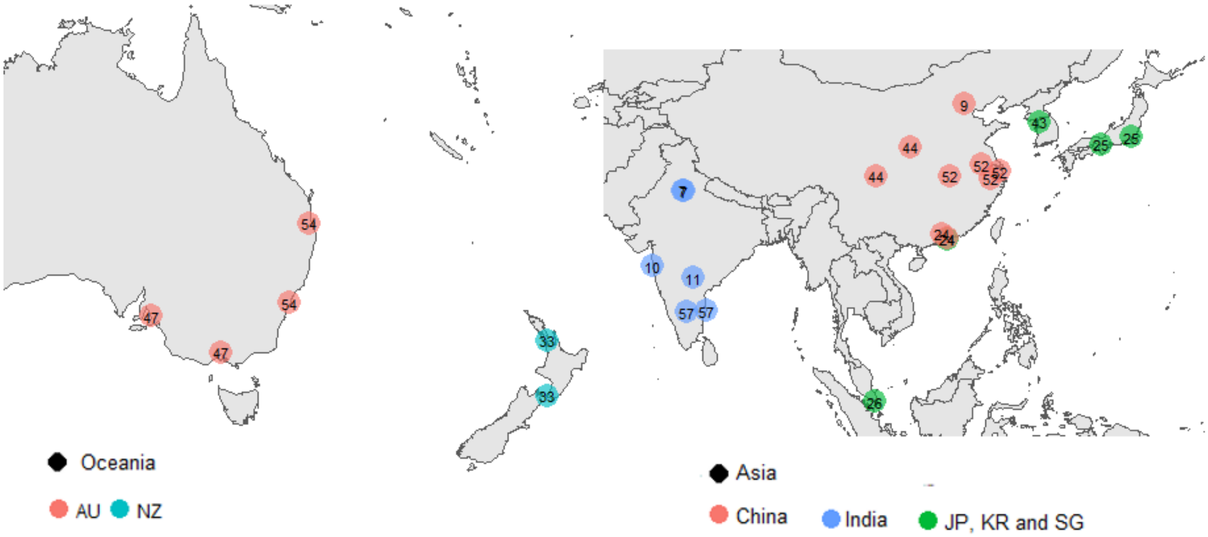
Table A.5: Distribution of user-by-date observations across large geographic regions R

region R	observations	share of total
Northern America	7,279,903	46.7
Asia	2,375,988	15.2
Western Europe	2,209,076	14.2
Northern Europe	1,780,537	11.4
Eastern Europe	1,006,935	6.5
Oceania	487,242	3.1
Southern Europe	457,136	2.9

Note: The table shows the distribution of observations in the user x date panel described in section 2.2 across large geographic regions R, displayed in figure A.1



(a)



(b)

Figure A.1: Illustration of large regions R, small regions r, and city groups g

Note: Maps show our sample cities. The number on top of the city markers refers to the group g we assign a city to for the first stage estimation of the effect of wind direction on air pollution (see section 3, especially equation (3)). Shape of the city markers refers to the large geographic regions R (see section 2 and equation (2)), and color refers to the small geographic regions r we use for region x month fixed effects (see section 3, especially equation (2)).

Table A.6: Effect of PM2.5 on Quantity of Issue and Pull Request Actions

	PRs closed (1)	*PRs opened* (2)	*Issues closed* (3)	*Issues opened* (4)
Panel A.				
$PM_{2.5}$ (st.dev.)	-0.0012	-0.0018^{**}	0.0008	0.0001
	(0.0011)	(0.0008)	(0.0009)	(0.0007)
F-statistic	112.3	112.3	112.3	112.3
% change in Y	-0.1	-1.2	0.1	0.1
Panel B.				
$PM_{2.5}$ shock	-0.0043	-0.0085^{**}	0.0058	0.0006
	(0.0048)	(0.0036)	(0.0042)	(0.0030)
F-statistic	34.6	34.6	34.6	34.6
% change in Y	-2.5	-5.7	4.8	0.5
Dependent Variable mean	0.17	0.15	0.12	0.11
Observations	331,025	331,025	331,025	331,025

Note: The table presents IV estimates of the parameter β in equation (2). In Panel A, the regressor of interest is PM2.5 concentration measured in city-specific standard deviations. In Panel B, a binary PM2.5 shock variable is used instead, which takes a value of one if city x day PM2.5 concentration exceeds the city x month x day-of-week specific average by at least one city-specific standard deviation. The first stage specification is given in equation (3). Covariates include eight bins for mean daily temperature, third-order polynomials in wind speed, precipitation and relative humidity, indicators for heavy wildfire smoke and holidays, as well as city, day-of-week, year-by-month and region-by-month fixed effects. Day-of-week and year-by-month fixed effects and the temperature controls can vary across world regions R. Regressions are weighted by the number of active workers in a city during the current month. Standard errors clustered at the city level are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

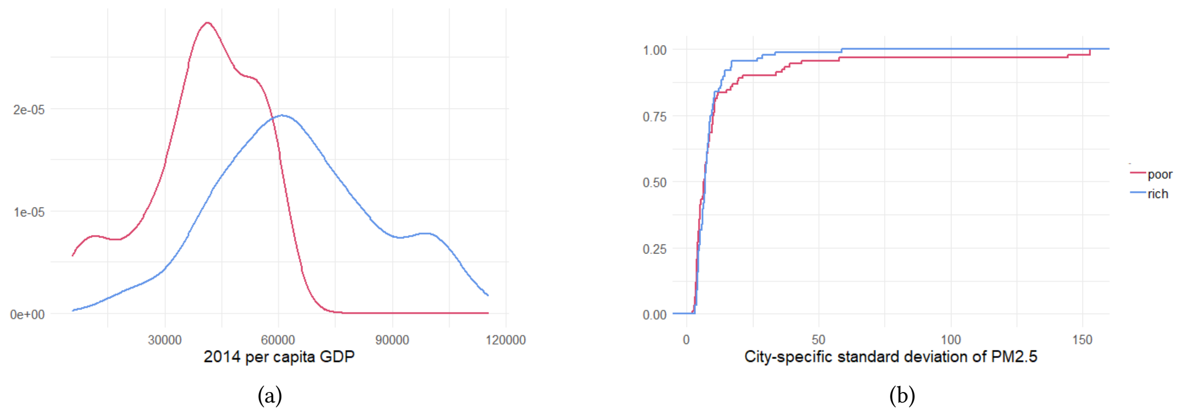


Figure A.2: Subsamples based city-level GDP per capita 2014

Note: Panel (a) shows distributions of 2014 GDP per capita across cities separately for the poor and rich subsample, as described in Section 4.4, using data from OECD, World Bank and national statistical officies. Panel (b) shows the empirical cumulative distribution function of standard deviations of PM2.5 across cities in the two subsamples. Red: below median GDP per capita subsample, blue: above median subsample.

Table A.7: Effect of PM2.5 on Quantity of PRs opened based on GHArchive and GHTorrent data

	PRs opened (GHA) (1)	*PRs opened (GHT)* (2)
Panel A.		
$PM_{2.5}$ (st. dev.)	-0.0031^{**}	-0.0019^{**}
	(0.0013)	(0.0009)
	[0.020]	[0.032]
F-statistic	98.1	98.1
Panel B.		
$PM_{2.5}$ shock	-0.0184^{***}	-0.0096^{**}
	(0.0069)	(0.0044)
	[0.008]	[0.030]
F-statistic	29.6	29.6
Observations	279,537	279,537

Note: The table presents IV estimates of the parameter β in equation (2), where the outcome is the number of pull requests (PRs) opened. This is measured based on GHArchive data in column (1) and GHTorrent data in column (2). In Panel A, the regressor of interest is PM2.5 concentration measured in city-specific standard deviations. In Panel B, a binary PM2.5 shock variable is used instead, which takes a value of one if city x day PM2.5 concentration exceeds the city x month x day-of-week specific average by at least one city-specific standard deviation. The first stage specification is given in equation (3). Covariates include eight bins for mean daily temperature, third-order polynomials in wind speed, precipitation and relative humidity, indicators for heavy wildfire smoke and holidays, as well as city, day-of-week, year-by-month and region-by-month fixed effects. Day-of-week and year-by-month fixed effects and the temperature controls can vary across world regions R. The sample period is 2015 to May 2019. Regressions are weighted by the number of active workers in a city during the current month. Standard errors clustered at the city level are reported in parentheses. P-values are reported in brackets. *p<0.1; **p<0.05; ***p<0.01

Table A.8: OLS Results for Work Quantity

	Actions		*Commits*		*Comments*	
	(1)	(2)	(3)	(4)	(5)	(6)
$PM_{2.5}$ (st. dev.)	-0.003	-0.007	-0.003	-0.005^{*}	-0.0002	-0.001
	(0.005)	(0.006)	(0.002)	(0.003)	(0.002)	(0.002)
Observations	331,025	331,025	331,025	331,025	331,025	331,025
city FE	✓	✓	✓	✓	✓	✓
region $R\times$day-of-week FE	✓		✓		✓	
region $R\times$year-month FE	✓		✓		✓	
small region $r\times$month FE	✓	✓	✓	✓	✓	✓
region $R\times$date FE		✓		✓		✓

Note: The table presents OLS estimates of the parameter β in equation (2), where the dependent variables are displayed in the upper part of the table. The regressor of interest is PM2.5 concentration measured in city-specific standard deviations. Covariates include eight bins for mean daily temperature, third-order polynomials in wind speed, precipitation and relative humidity, indicators for heavy wildfire smoke and holidays as well as city fixed effects. The temperature controls can vary across world regions R. Further included fixed effects are displayed in the bottom part of the table. Regressions are weighted by the number of active workers in a city during the current month. Standard errors clustered at the city level are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table A.9: Robustness: First Stage Specification

	Actions	*Commits*	*Comments*	*Actions*	*Commits*	*Comments*
Panel A.						
$PM_{2.5}$ (st. dev.)	-0.0188^{*}	-0.0181^{***}	-0.0006	-0.0245^{**}	-0.0170^{***}	-0.0056
	(0.0103)	(0.0047)	(0.0055)	(0.0102)	(0.0052)	(0.0053)
F-Statistic	68.6	68.6	68.6	106.1	106.1	106.1
Panel B.						
$PM_{2.5}$ (shock)	-0.0819^{*}	-0.0868^{***}	0.0029	-0.1098^{**}	-0.0824^{***}	-0.0186
	(0.0430)	(0.0203)	(0.0222)	(0.0507)	(0.0259)	(0.0240)
F-Statistic	22.3	22.3	22.3	33.3	33.3	33.3
IV-Specification	3 wind direction bins			Hierarchical clustering		
Observations	331,025	331,025	331,025	331,025	331,025	331,025

Note: The table presents IV estimates of the parameter β in Equation (2). In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in city-specific standard deviations. In Panel B, a binary $PM_{2.5}$ shock variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city×month×day-of-week specific average by at least one city-specific standard deviation. Relative to specifications underlying results in Table 3, the first stage model is changed: In columns (1) to (3) instruments are three indicators for wind direction falling in specific bins, each covering 90° of the wind rose. In columns (4) to (6), the first stage specification is as in Equation (3), but we form city-groups g using hierarchical clustering instead of k-means clustering. Covariates as described in Table 3. Regressions are weighted by the number of active workers in a city during the current month. Standard errors clustered at the city level are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table A.10: Robustness: Second Stage Specification

	Asinh(Actions)	*Asinh(Co'its)*	*Asinh(Co'ents)*	*Actions*	*Commits*	*Comments*
Panel A.						
$PM_{2.5}$ (st. dev.)	$-.0054^{***}$	$-.0035^{***}$	$-.0025^{*}$			
	(.0021)	(.0014)	(.0014)			
Panel B.						
$PM_{2.5}$ (shock)	$-.0229^{**}$	$-.0162^{**}$	$-.0089$			
	(.0101)	(.0072)	(.0062)			
Panel C.						
log($PM_{2.5}$)				$-.0446^{***}$	$-.0271^{***}$	$-.0102$
				(.0147)	(.0082)	(.0074)
F- Statistic				170.7	170.7	170.7
Observations	331,025	331,025	331,025	331,025	331,025	331,025

Note: The table presents IV estimates of the parameter β in Equation (2). In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in city-specific standard deviations. In Panel B, a binary $PM_{2.5}$ shock variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city×month×day-of-week specific average by at least one city-specific standard deviation. In Panel A, the regressor is the logarithm of $PM_{2.5}$ concentration. Inverse hyperbolic sine transformations are applied to outcomes in the first three columns. The first stage specification is given in Equation (3). Covariates as described in Table 3. Regressions are weighted by the number of active workers in a city during the current month. Standard errors clustered at the city level are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table A.11: Reduced Form and clustering level

	Actions	*Commits*	*Comments*	*Share Easy Issue Events*	*Lines added per PR*	*Files changed per PR*
Panel A: Reduced Form						
High Pollution	-0.0128^{**}	-0.0112^{***}	0.0017	0.0006^{*}	-0.0116^{**}	-0.0068^{**}
W. Direction	(0.0060)	(0.0026)	(0.0032)	(0.0004)	(0.0047)	(0.0026)
	[0.036]	[0.00003]	[0.584]	[0.096]	[0.015]	[0.011]
First Stage Effect on PM2.5 (st. dev.): 0.39^{***} (.030)						
Panel B: Clustering Level						
$PM_{2.5}$ (st. dev.)	-0.0258^{**}	-0.0177^{***}	-0.0059	0.0008^{*}	-0.0171^{**}	-0.0094^{**}
	(0.0128)	(0.0060)	(0.0060)	(0.0004)	(0.0082)	(0.0043)
	[0.049]	[0.005]	[0.329]	[0.056]	[0.042]	[0.032]
F-Statistic	112.3	112.3	112.3	94.3	67.6	67.6
Observations	331,025	331,025	331,025	258,031	155,071	155,071

Note: Panel A displays OLS estimates of the outcomes displayed in the upper part of the table on an indicator variable for wind blowing towards a city from the direction that has the largest positive effect on local PM2.5 concentration. Standard errors clustered at the city level are reported in parentheses. Panel B presents IV estimates of the parameter β in Equation (2). The regressor of interest is $PM_{2.5}$ concentration measured in city-specific standard deviations. The first stage specification is given in Equation (3). Standard errors clustered at the level of *city-groups* g across which the effect of instruments in the first stage are allowed to differ are reported in parentheses. P-values are presented in squared brackets. All regressions include covariates as described in Table 3 and are weighted by the number of active workers in a city during the current month. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

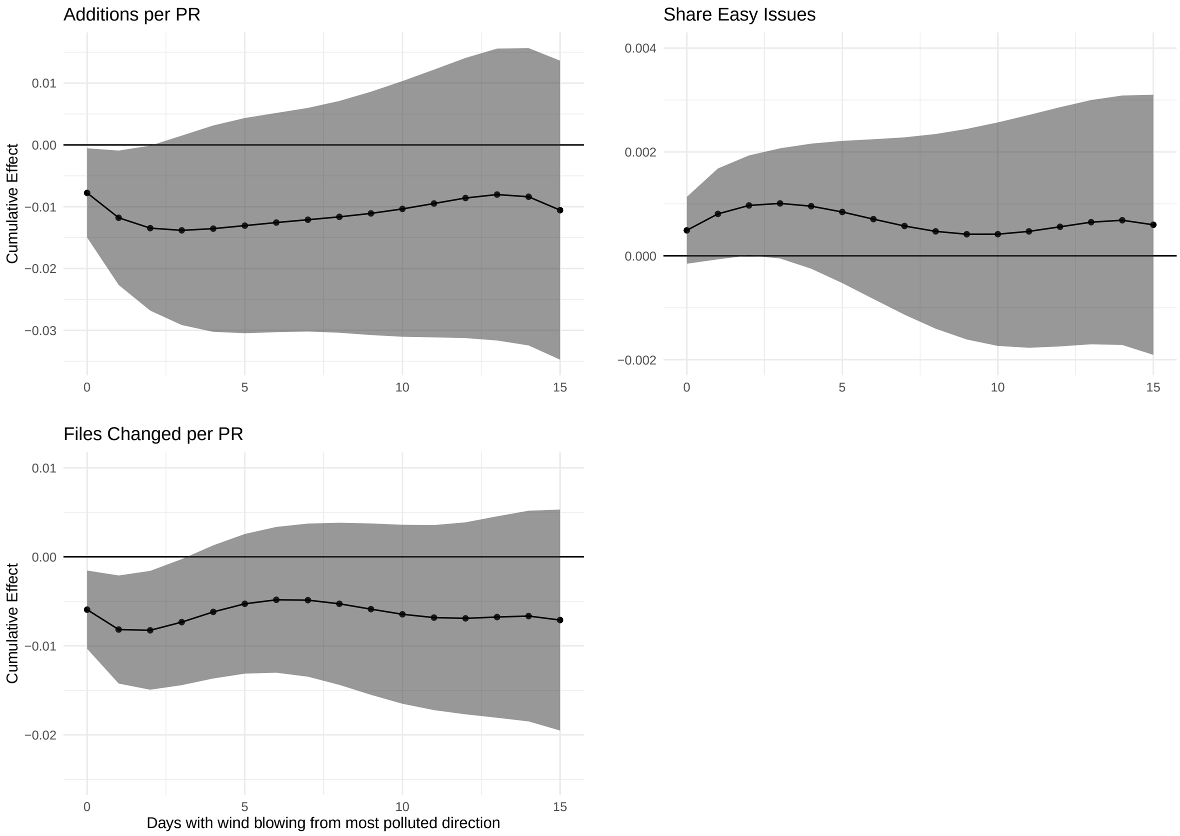


Figure A.3: Effect Dynamics: Adjustment

Note: The plots depict estimates of the cumulative effect of wind blowing from the high pollution direction on three measures of user adjustment by switching to easier tasks, the inverse hyperbolic sine transformations of the number of new lines added per PR opened, the inverse hyperbolic sine transformations of the number of files changed per PR opened, and the share of issue events that refer to an easy issue. Effects are derived from a fourth order polynomial distributed lag specification. The x-axis denotes the number of days over which the cumulative effect is computed. Shaded areas represent 95% confidence intervals. Regressions control for city, day-of-week, year-by-month and region-by-month fixed effects, a holiday indicator and weather controls for the current day and eight lags (third order polynomials in mean daily temperature, precipitation, relative humidity and wind speed). Regressions are weighted by the number of active workers in a city during the current month and standard errors are clustered at the city level.