Discussion Paper Series – CRC TR 224

Discussion Paper No. 402
Project B 07

Air Quality, High-Skilled Worker Productivity And Adaptation: Evidence From Github

Felix Holub [1]
Beate Thies [2]

March 2023

[1] Goethe University Frankfurt, Email: holub@econ.uni-frankfurt.de
[2] University of Mannheim, Email: beate.thies@uni-mannheim.de

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Air Quality, High-Skilled Worker Productivity and Adaptation: Evidence from GitHub*

Felix Holub† Beate Thies‡

March 6, 2023

click here for most recent version

Abstract

Highly skilled knowledge workers are important drivers of innovation and long-run growth. We study how air quality affects productivity and work patterns among these workers, using data from *GitHub*, the world's largest coding platform. We combine panel data on daily output, working hours, and task choices for a sample of 25,000 software developers across four continents during the period 2014-2019 with information on concentrations of fine particulate matter ($PM_{2.5}$). An increase in air pollution reduces output, measured by the number of total actions performed on GitHub per day, and induces developers to adapt by working on easier tasks and by ending work activity earlier. To compensate, they work more on weekends following high-pollution days, which suggests adverse impacts on their work-life-balance. The decline in output arises even at concentrations in line with current regulatory standards in the EU and US and is driven by a reduction in individual coding activity, while interactive activities are unaffected. Exposure to $PM_{2.5}$ levels above the city-specific 75th percentile reduces daily output quantity by 4%, which translates into a loss in output value by approximately $11 per developer.

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

*We are grateful to Ulrich Wagner, Tatyana Deryugina, Johannes Gessner, Kathrine von Graevenitz, Nicolas Koch, Hannah Klauber, Simon Krause, Giovanni Paolo Mariani, Alessandro Palma, Nico Pestel, Linnéa Rohde, Christoph Rothe, Konrad Stahl, Kinga Tchorzewska, and Nicolas Ziebarth for helpful comments and suggestions. We received valuable feedback from audience members at the Universities of Mannheim and Frankfurt, the 15th RGS Doctoral Conference in Economics, the 2022 AERE Summer Conference, the 9th IZA Workshop on Environment, Health and Labor Markets, the 2022 ENTER Jamboree, the Workshop on Measurement and Economic Evaluation of Air Pollution at the University of Mannheim, the Heidelberg-Mannheim-ZEW Environmental Economics Brownbag Seminar, the 2022 EALE Conference, and the MCC Berlin. Support by the German Research Foundation (DFG) through CRC TR 224 (Project B07) is gratefully acknowledged. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 865181).

†Department of Economics, Goethe University Frankfurt; holub@econ.uni-frankfurt.de

‡Department of Economics, University of Mannheim; beate.thies@uni-mannheim.de

1 Introduction

Driven by technological innovation, the world of work is undergoing rapid changes. Over the last decades, computerization has been causing an increase in the demand for workers performing non-routine, analytical, and interpersonal tasks (Autor et al., 2003; Autor and Price, 2013). Skills that complement digital technologies have been growing in importance: In the US, the share of jobs intensively requiring digital skills[1] has more than quadrupled from 5% to 23% between 2002 and 2016 (Muro et al., 2017). In parallel, there have been major shifts in the organization of work, complementing the growing role of IT. Teamwork, flexible schedules, and discretion in task choice are common, replacing fixed 9-to-5 schedules and direct task assignments, especially among highly-educated workers (Bresnahan et al., 2002; Mas and Pallais, 2020; Menon et al., 2020). Because jobs characterized by these task profiles, skill requirements, and organizational features form the backbone of the modern knowledge economy and are expected to become even more important as digitization and automation proceed, it is critical to understand what determines productivity in these settings.

In this paper, we study how environmental shocks impact performance and work patterns among highly skilled knowledge workers in a flexible work environment. Vast populations are exposed to environmental conditions such as heat and poor air quality, which have been shown to reduce labor productivity in several settings. Existing research, however, has considered routine jobs and/or inflexible work contexts (e.g. Chang et al., 2019; Somanathan et al., 2021). In the settings described above, workers not only use different skills, they also have flexibility and discretion in organizing their workday, which might allow them to adapt to productivity shocks, thereby alleviating output effects. Moreover, in collaborative work settings, impacts of environmental shocks might get dampened, e.g., if co-workers can help each other to focus, or get amplified due to complementarities if co-workers depend on each others' input.

We focus on the effects of air pollution as it is a ubiquitous public health threat in urban areas across the globe.[2] 82% of the global population are exposed to levels of fine particulate matter exceeding World Health Organization (WHO) guidelines. To implement optimal air quality standards and policies to curb pollution, accurate estimates of the welfare costs of pollution are fundamental. Recent research shows that air pollution not only causes premature mortality and severe health damages, but also sub-clinical effects on labor market outcomes, student performance, or decision-making (see Aguilar-Gomez et al., 2022, for a review). The sub-clinical impacts play an important role for the total economic cost of air pollution as they affect a broad population, while morbidity and mortality effects are concentrated among vulnerable groups, like infants and the elderly.

We study the causal effect of air pollution exposure on professional software developers,

[1]Examples of digital skills are the abilities to handle information and communication technology and to conduct data analyses.

[2]In an extension, we also provide some evidence on the effects of extreme temperatures for comparison.

using data from *GitHub* to measure developer output and work patterns. Software development is a STEM (science, technology, engineering, and math) occupation that requires analytical and advanced digital skills and generates high value for consumers, other industries, governments, and the research community.[3] Adverse productivity effects of air pollution in this occupation would thus have important implications for growth, innovation, and competitiveness. GitHub is the world's largest online code hosting platform, used for storing and jointly working on coding projects. It puts great emphasis on facilitating collaboration between developers. Moreover, software developers work in highly flexible settings that usually offer discretion over working hours and the tasks a developer chooses to work on at a given point in time. With these features, software development on GitHub is representative of the settings that characterize modern knowledge work.[4]

The GitHub data allows us to address the challenge that output of knowledge workers is often difficult to observe. We collect data on 25,000 users across four continents who work on projects owned by tech companies, indicating that they are professional software developers. The data includes users' locations as well as records of all actions they conduct in public projects along with precise timestamps and some further characteristics of the underlying task. We construct a user-by-day panel including measures of work quantity and quality, working hours, and task choice for the period between January 2014 and May 2019. Based on developers' locations, we match these outcomes to city-level air quality monitor data on particulate matter smaller than 2.5 µm ($PM_{2.5}$). To account for endogeneity in air quality when estimating a model of developer activity, we follow Deryugina et al. (2019) by instrumenting $PM_{2.5}$ concentration with daily average wind direction. The 2SLS strategy exploits the effect of plausibly exogenous regional air pollution transport on local pollution levels, controlling for a wide range of other weather characteristics.

To measure daily output quantity, we count the total number of actions performed, including commits (individual code changes), opening and closing of pull requests and issues, and comments written in discussion fora.[5] As we can classify the different GitHub actions into *individual* and *interactive* activities (e.g., commits vs. comments), we can analyze heterogeneity in the effect of pollution exposure on performance in these two distinct types of work, which are both widespread in modern high-skilled jobs. To assess output quality, we compute the share of commits that get undone at a later point and the share of pull requests that get rejected as measures of error frequency. We also derive monetary estimates of the value of GitHub activities, exploiting additional data from an online marketplace where GitHub project owners offer payments for contributions to their projects. This allows us to translate the effects of

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

[4]We provide evidence that software development is representative of modern high-skilled work in Appendix Figure B.1 and Table A.1: Required skills are similar (e.g., critical thinking and complex problem solving), except for substantially stronger digital skills like programming. Both software developers and high-skilled workers in general have a lot of flexibility in organizing their work and often work in teams.

[5]Pull requests are a tool to suggest changes to the code base of a repository, for more details see Section 3.

pollution on output into monetary losses.

To study adaptation in flexible work arrangements, we investigate whether software developers exploit their discretion with respect to working hours and task choice in order to adjust to changes in air pollution. Specifically, we exploit information on the complexity of tasks addressed by developers' actions and on action timestamps to study whether they focus on easier tasks or adapt their working hours.

Our dataset covers 36 countries, including both developing and developed countries, with large variation in pollution levels, income, and pollution awareness across sample cities. We exploit this in heterogeneity analyses to explore how air pollution damages are distributed and to study the mechanisms underlying the pollution impacts.

We present three main findings. First, developers produce less output on days with higher levels of fine particulate matter. When $PM_{2.5}$ concentration reaches unusually high levels – exceeding the city-specific 75th percentile – the number of daily actions observed on GitHub falls by 4%. This effect is mainly driven by a decline in individual coding activity: The number of commits decreases by 6.2%. By contrast, collaborative or interactive work (e.g., commenting on issues) is much less affected. Compared to other occupations studied in previous research, including both physically- and cognitively-demanding jobs, the magnitude of the effects on output is relatively small. Nonetheless, the pollution-induced output declines translate into relevant monetary damages due to the high value generated by software developers. The loss in output value per developer and day amounts to $4 for a standard deviation increase in $PM_{2.5}$, i.e., common fluctuations in pollution, and $11 for days with unusually high air pollution.

Second, output quality is unaffected by changes in air pollution. We find no evidence that software developers commit more errors on days with high levels of $PM_{2.5}$. A potential explanation for the absence of quality effects, as well as the modest size of effects on quantity, is worker adaptation to pollution-induced productivity shocks in this flexible work setting.

Our third main result provides evidence for this: We find that developers exploit their flexibility in task choice by focusing on less complex tasks when air pollution increases. Among activities related to issues[6], the share that refers to issues labeled as relatively easy increases by 5% on days with $PM_{2.5}$ concentration above the city-specific 75th percentile compared to low pollution days. Similarly, code submitted or reviewed by developers changes 4% fewer files and contains 2% fewer new lines, indicating that the code addresses less complex tasks. Among developers with stronger adjustment in task choice in response to $PM_{2.5}$ exposure, effects on output quantity are attenuated.

Software developers also adapt their working hours. They reallocate work activity from high-pollution, low-productivity days to low-pollution, high-productivity weekends. In particular, developers end work activity earlier on days with unusually high $PM_{2.5}$ concentration. To compensate, they work more on weekends after a workweek with poor air quality, es-

[6] This includes creating, closing, and commenting on an issue.

pecially if pollution concentration on the weekend is moderate, i.e., below the city-specific 75th percentile. The increase in weekend work makes up for 33% of the reduction in coding output on the day of exposure. These forms of adaptation likely explain why, compared to other professions, we find moderate effects of particulate matter on output. At the same time, compensation by working more on the weekend also implies an additional welfare cost due to forgone leisure time and potential negative impacts on the work-life balance. Losses in output value are thus likely a lower bound on the overall cost of air pollution in this setting.

The adverse effects of pollution on output quantity arise at concentrations below the regulatory standards in force in the European Union and the US. Indeed, effects are strongest at low levels of $PM_{2.5}$. The effect magnitude does not vary systematically with country-level pollution awareness, indicating that the negative effects on output are not driven by avoidance behavior. This is corroborated by the fact that we find relatively small extensive margin effects. We also show that effects are substantially larger in locations with an older building stock, suggesting that differences in effective indoor pollution exposure play an important role. This points towards a physiological mechanism underlying our main results.

In a dynamic analysis, we show that pollution exposure on a given day reduces both contemporaneous output and, to a lesser extent, output produced over the following two to three days, but not thereafter. We repeat our main analysis at the monthly level, to assess the effect of pollution on output net of adjustment and accounting for dynamic impacts. We again find negative effects on the number of actions performed, driven by less individual coding activity. The impact of an increase in daily $PM_{2.5}$ concentration by one unit is roughly 60% larger according to these results compared to the results at the daily level, but still moderate in comparison to other professions. A standard deviation increase in daily $PM_{2.5}$ generates a loss in monthly output value of approximately $6. In the monthly analysis, we also consider the growth rate of a developer's followers on GitHub as a summary measure of work quantity, quality, and relevance. Air pollution exposure also reduces this outcome, indicating that it not only reduces short-run performance but also slows down the build-up of reputation, which could plausibly have long-run consequences for career paths.

Overall, our results imply that improvements in air quality generate economic benefits in terms of productivity gains among highly skilled STEM workers. While flexible work arrangements with respect to schedules and task choice allow workers to adapt to productivity shocks, even in the very flexible setting we analyze here, air pollution generates economically relevant costs. This is true even for locations with relatively low pollution levels in compliance with existing regulatory standards.

Related Literature. This paper contributes to the research on the effect of environmental factors on economic outcomes. Firstly, our paper directly links to the literature strand on air pollution and worker productivity. Several studies document a negative impact of pollution on productivity in manual and routine occupations, such as textile workers (Adhvaryu et al., 2022; 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 small number of papers also demonstrate negative effects of poor air quality on performance in more cognitively-demanding occupations. 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, 2020; Sarmiento, 2022).

While these contexts allow to create precise measures of worker performance in a specific domain, the settings analyzed do not reflect the typical features of work organization in most modern high-skill jobs. As outlined above, frequent collaboration, multi-tasking, and flexibility in work schedules are widespread in these jobs and each of these characteristics might affect the severity of pollution-induced productivity shocks. Furthermore, the rather inflexible settings studied so far do not allow to analyze worker adaptation to pollution. Related work investigates performance in cognitively-demanding tasks, but outside of standard work settings, e.g., among chess players (Künn et al., forthcoming), individual investors (Huang et al., 2020), or brain game players (La Nauze and Severnini, 2021; Krebs and Luechinger, 2021).

We contribute to this literature by expanding the analysis to a STEM profession that is representative for a large group of high-skilled workers in flexible, modern work environments. Thus, our analysis adds novel insights into the labor market cost of air pollution that will likely still be relevant after future waves of digitalization. We present first evidence on productivity effects separately for individual and collaborative activities, a distinction absent from previous work. In addition, while existing papers are based on data from a single country and often only a single site, we work with a large sample of developers across multiple countries. This allows to draw more general conclusions about the pollution-productivity relationship and to explore effect heterogeneity, e.g., with respect to local income levels or pollution awareness.[7]

Secondly, we contribute to the literature on worker adaptation to environmental shocks and connect it to research on the effect of flexible work arrangements on productivity. A number of papers study how workers adjust working hours in response to temperature shocks (Graff Zivin and Neidell, 2014; Neidell et al., 2021; LoPalo, forthcoming). With respect to air pollution shocks, Adhvaryu et al. (2022) 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 identifies a new margin of adjustment in a flexible high-skilled setting, namely task choice. Moreover, we present new evidence on temporal reallocation of work activity towards the

[7] Borgschulte et al. (forthcoming) and Fu et al. (2021) do not consider specific professions, but conduct broader analyses of air pollution and labor earnings in the US and manufacturing sector productivity in China, respectively. We add new evidence relative to these papers due to our international sample and our analysis of worker adaptation which requires high-frequency microdata.

weekend in response to pollution shocks.[8] By showing that workers exploit discretion in task choice and working hours to adapt to an environmental shock, and thereby alleviate its impact on productivity, our work links to research on the causal effects of flexible work arrangements and worker autonomy on productivity. Shepard et al. (1996), Beckmann et al. (2017), and Angelici and Profeta (2020), for instance, find across different contexts in the US, Germany, and Italy that working time autonomy increases employee productivity. Our results suggest that the ability to adapt to idiosyncratic productivity shocks might contribute to the positive relationship between flexible work arrangements and performance.

Our analysis of worker output and behavior also relates 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. Kaur et al. (2021) find that financial concerns reduce productivity via psychological channels. We complement this literature with detailed evidence on how environmental shocks affect work patterns and performance.

Lastly, another contribution of this paper is to demonstrate new ways to use publicly available data on GitHub activity. While we are not the first to use this data in economics,[9] we propose strategies to construct a sample of highly active users who are likely professional software developers, to study task difficulty, and to estimate the monetary value of the output observed on GitHub.

Outline. The remainder of the paper is organized as follows: We begin in Section 2 with a short explanation on how pollution affects the human body. Section 3 follows with a description of *Github*, the data and sample. We explain the research design and how we implement the two-stages least squares strategy of Deryugina et al. (2019) in Section 4. Our main results are presented in Section 5. Findings from heterogeneity analysis and extensions follow in Section 6. Section 7 concludes.

2 Background on Particulate Matter

In our analysis, we focus on $PM_{2.5}$ to measure air pollution, i.e., particulate matter with a diameter of less than 2.5 µm. Particulate matter refers to all solid and liquid particles suspended

[8] In parallel work, Hoffmann and Rud (2022) also show evidence that workers reallocate labor supply across days in response to changes in $PM_{2.5}$. However, they study a different setting, namely formal and informal workers in Mexico City, whereas we focus on a sample of highly-skilled STEM workers. Moreover, Hoffmann and Rud (2022) interpret the temporal substitution as a strategy to avoid pollution exposure, whereas in our setting it serves as compensation for reduced productivity.

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

in the air. In most urban areas around the world, the majority of $PM_{2.5}$ originates from anthropogenic sources, including traffic, industrial production and biomass burning (Karagulian et al., 2015). While some $PM_{2.5}$ is produced locally, for example by traffic, in most areas a significant share of local pollution arises from distant sources via long-range transport. Power plants and industrial facilities generate precursor emissions, e.g., sulphur dioxide (SO_2), which are transformed into secondary particulate matter, e.g., sulfate (SO_4^{2-}) and can get transported over long distances (Almeida et al., 2020).

A key reason to focus on $PM_{2.5}$ in our study is the fact that these small particles can penetrate indoors and are thus of major relevance for indoor office workers. Deng et al. (2017) find indoor-outdoor ratios of $PM_{2.5}$ between 0.4 and 1.2 for office and apartment buildings in Beijing, which can only be reduced to a level near zero with high-quality indoor air cleaning systems. In line with that, Xu et al. (2020) and Hoek et al. (2008) report significant and sizeable correlations between indoor and outdoor fine particulate matter for other cities in China and Europe.

Moreover, a large body of research documents that fine particulate matter plays a key role for the adverse effects that air pollution exerts on various dimensions of human health. The small particles can penetrate deeply into the lungs, causing damage to the respiratory system, including reduced lung function, asthma, and chronic obstructive pulmonary disease. Similarly, epidemiological and economic studies find evidence for cardiovascular health effects like high blood pressure and heart diseases (Lederer et al., 2021). Medical research on humans and animals points to systemic oxidative stress, inflammation and endothelial dysfunction (impaired functioning of the inner lining of blood vessels) as underlying biological mechanisms (Anderson et al., 2011; Kelly and Fussell, 2015). While severe morbidity and mortality effects are concentrated among vulnerable groups like the elderly and infants, even healthy individuals can experience mild, subclinical effects, including irritation in the nose and throat or coughing (Pope, 2000). In response to the mounting evidence on adverse health effects, several countries have introduced standards on annual ambient PM2.5 concentrations, and typically tightened them over time. Currently, standards are in force e.g. in the US (12 $\mu g/m^3$), Canada (10 $\mu g/m^3$) and the European Union (25 $\mu g/m^3$). The WHO guidelines recommend a level of no more than 5 $\mu g/m^3$.

Recent clinical and epidemiological studies imply that exposure to fine particles also exerts adverse effects on the central nervous system (Delgado-Saborit et al., 2021; Babadjouni et al., 2017). Small particles have been found to reach the brain via the olfactory pathways and the bloodstream. Animal and autopsy studies indicate that particulate matter causes neuroinflammation, which can lead to cognitive impairments and neuro-degenerative processes (Calderón-Garcidueñas et al., 2007). Associations between pollution exposure and changes in brain structure have been detected in neuroimaging studies. Consistent with this, La Nauze and Severnini (2021) find that brain game players score 0.18 standard deviations lower when

$PM_{2.5}$ concentration exceeds 25 µg/m^3 as compared to days with better air quality. Similarly, Ebenstein et al. (2016) find that short-run exposure to $PM_{2.5}$ reduces student performance on high-stake exams.

Overall, the research on particulate matter, health, and cognitive functioning implies that $PM_{2.5}$ exposure might plausibly reduce productivity both in physically- and cognitively-demanding tasks. Growing evidence in economics on negative productivity effects in manual occupations, and on reduced cognitive performance confirms this. We intend to *quantify* productivity impacts in a high-skilled work environment, and investigate potential adaptation responses that might occur in these settings.

3 Setting and Data

To analyze the effects of air pollution on productivity and work patterns in a high-skilled profession, we pair information on *GitHub* activity for a global sample of software developers with data on local air quality. This is complemented with data on meteorological conditions to construct the instrumental variables and to control for local weather. This section starts with a brief description of GitHub, followed by an overview of the GitHub data and how we use it to measure developers' productivity. After checking the validity of these outcome measures, we end with a description of the environmental data.

3.1 Setting: GitHub

GitHub is built on *Git*, an open source version control system that records who changed which part of a file at what point in time. *GitHub* is a web platform for hosting Git repositories[10] and, on top of the version control functionality, provides additional features to facilitate collaboration. For each of their repositories (or repos for short), owners can choose whether to make it public or private, i.e., whether the respective files are visible to everyone, or only to the repository members. In 2019, more than 30 million accounts were registered on GitHub, who 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 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 was conducted by the commit author. Only repository owners and team members invited by them can modify files via commits.

The primary additional collaboration features offered on GitHub are *pull requests* and *issues*. A pull request (PR) is a tool to propose code changes to a repository. To create a PR, a user generates a copy of the repository in question, implements the changes in his copy via

[10] The term repository refers to the location where all files belonging to a project are stored.

commits, and then submits these to the original repo.[11] Repo members then review the suggested changes and decide whether to accept (i.e., merge) or reject them. Each PR includes a discussion forum where users can comment directly on the proposed changes. Feedback provided there can be implemented within the same PR. Due to these features, PRs facilitate collaborative coding and are thus not only used to contribute to projects of which a user is no member but also within project teams.

Issues are text messages, typically used to suggest improvements and organize tasks in a given repo.[12] Like PRs, issues contain a discussion forum where users can comment on the problem or question at hand. Repository members can assign labels to issues in order to highlight their category (e.g., bug, feature request), priority, or difficulty. The platform provides nine default labels, and repository teams can create additional ones specific to their repo. Once an issue is resolved, it can be marked as closed.

On top of that, GitHub provides social network functions, e.g., options to follow other users and subscribe to specific repositories and issues to receive notifications about new activities.

3.2 GitHub Data on Productivity and Work Patterns

GitHub actions related to commits, PRs, and issues reflect productive work aimed at building or improving software products. Hence, we collect data on these activities to measure output generated by highly skilled developers. The GHTorrent project provides information on GitHub users and all actions they conduct in *public* repositories in the form of a relational SQL database. We use the version of the database containing data up to June 1st, 2019. The user table comprises a unique identifier, login name and registration date for all users registered on the platform at this point. In addition, location and company information as stated on the user profile on this date is reported. The projects table provides identifiers and names of all public repositories as well as a reference to the user owning the repo. Data on activities is available separately by type of action (e.g., commits, opening issues, PR comments, etc.) and includes the exact timestamps and the identifiers of the acting user and the repository where the event was conducted in. For specific actions, further information is reported, e.g., the labels attached to issues.

We complement this with data from GHArchive, which also provides a record of actions in public repositories, and contains additional information on some events, e.g., the title of a commit (called commit message), and 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 favorable features for our analysis. First, the precise records of activities conducted on GitHub enable us to quantify daily output produced by software

[11] An example for a PR can be found at https://github.com/microsoft/vscode/pull/54244.

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

developers. In this way, we address the long-standing challenge that work conducted by high-skilled workers during a given period is often difficult to measure. Second, the data cover all GitHub users which gives us a much broader geographic coverage and thus 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. Moreover, the rich information included allows us to investigate not only changes in output quantity, but also quality and work patterns, which are of major relevance in high-skill professions.

The data also have clear limitations. In order to assign local air quality to users, we rely on self-reported locations. Some users might report wrong or outdated locations, giving rise to measurement error. Under the assumption of classical measurement error which is not correlated with pollution levels, this issue leads to attenuation bias such that any adverse effects we find can be considered a lower bound on the true effect. Additionally, we have no information on work conducted in private repositories or outside the platform. Many GitHub users conduct no or only little work in public repositories such that it would be impossible to identify any 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 software developers and do a substantial part of their formal work in public GitHub repositories.

Sample Construction. 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. Among them, we keep only users 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 focus on professionals who are in some way affiliated with the companies. To identify these users, we compile a list of repositories operated by companies[13] and then use the information on the repository where a commit was made from the GHTorrent data. To drop bots, we discard a small number of users with extremely high activity levels, very regular commit patterns, or suspicious login names.[14] 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 20 commits in public repos in a given month. They enter the sample in the month after they have passed this threshold for the first time. Users stay in the sample until the end of the observation period unless they conduct less than three *unproductive* actions in a given month. In this case, we drop users from the sample for that month, assuming that they might

[13]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.

[14]Bots are computer programs typically used to automate specific tasks. On GitHub, some company-affiliated projects for instance employ bots to comment on newly opened issues and PRs to ask users to provide specific information on their issues or to sign a contributor license agreement. To make sure not to capture bots in our sample, we drop users if the number of actions or commits conducted by them is in the top 0.1 percentile of the distribution, if more than 20% of their commits occur at full hours (indicating automated commits) or if their login name indicates that they are bots.

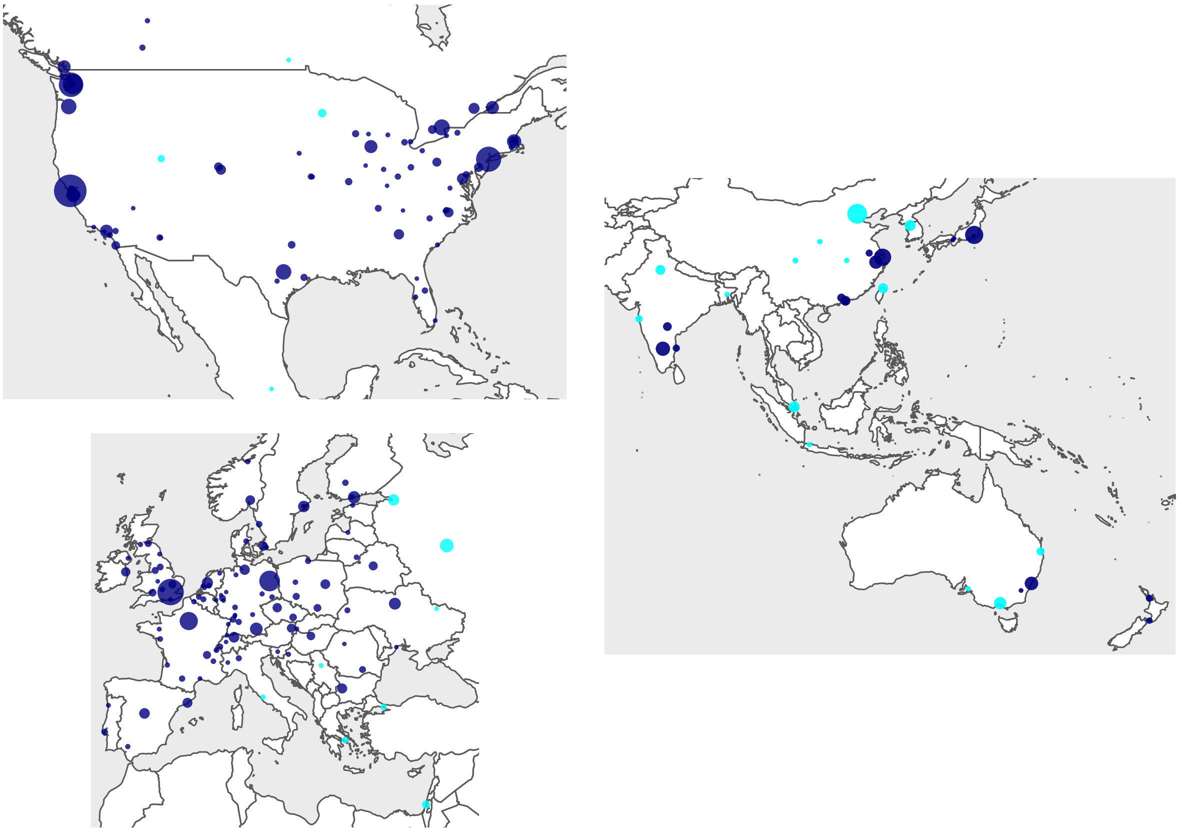


Figure 1: Sample Cities

Note: Each point represents one sample city. Circle size is based on the number of users observed in the city. Cities depicted in dark blue are used in the main analysis and cities depicted in light blue are added in extensions.

have moved to a different platform or work on projects in private repositories. Unproductive actions are activities we do not use as outcome variables, based primarily on the social network functions GitHub offers.[15] Lastly, we restrict the sample to users living in cities with at least 15 relevant users that are covered by our data on air pollution. This yields a sample of 27,701 users in 220 cities across 47 countries during the sample period from February 2014 until May 2019.[16] These locations are depicted in Figure 1. For the IV approach based on changes in wind direction, we require multiple cities in close geographic proximity, as outlined further in Section 4. Hence, in our main analysis, we focus on 193 cities across 36 countries depicted in dark blue, comprising 24,534 users. All descriptives reported in this section refer to this main analysis sample. In extensions, we also include the cities depicted in light blue.

[15] 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.

[16] 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 where they spend the biggest part of the sample period, and keep them in the sample only while they resided in this city.

Outcome Measures. For the analysis sample, we compile an unbalanced user-by-day panel including measures of output quantity and quality, as well as work patterns. To measure overall daily work *quantity*, we count the total number of productive actions conducted per user and day, after translating timestamps from UTC into local time. This is the sum of commits, comments written on PRs, issues or commits, creations of PRs and issues, closing of PRs and issues, and reopening of issues. Furthermore, we separately count the number of commits and comments because these two action categories are observed most frequently on GitHub and reflect two distinct types of work. While the number of commits provides a measure of individual coding activity, comments reflect participation in discussions about issues and code changes, i.e., collaborative work. This allows us to conduct the first analysis of the productivity impacts of air pollution in a high-skill profession that takes potential effect heterogeneity between individual and interactive work into account.[17]

To assess the *quality* of users' output, we measure the share of all PRs opened on a given day that are merged, i.e., accepted. PR rejection points to issues in the code or style. Secondly, we compute the share of commits made by a user on a given day that were reverted at a later point. Reverting a commit, i.e. removing all changes made in it, indicates severe problems that cannot easily be fixed in follow-on commits.[18]

To analyze worker *adaptation*, we build measures of task choice and working hours. Firstly, to explore whether users switch to easier tasks on high-pollution days, we leverage information on the complexity of issues and PRs. In the case of PRs, we use the number of new lines of code added, of lines of code deleted, and of code files changed as measures of their complexity. We take the average value of these variables across all PRs a user worked on a given day, either by creating, reviewing, or commenting on the respective PR. To assess the difficulty of issues, we rely on the user-assigned issue labels. We exploit the fact that there are several labels indicating that a given issue is relatively easy, e.g., the default labels *good first issue*[19] and *documentation*,[20] 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.2. We construct the share of all issue events conducted by the user (commenting, opening, closing, or reopening of issues) which refer to an *easy* issue. With this approach, we do not have

[17]The other action types occur less frequently and are thus considered as secondary outcomes. The number of PRs created also reflects individual work on code, whereas the number of issues closed, opened, or reopened provides additional measures of interactive work. The number of PRs closed reflects code review and decision-making on whether to merge or reject the proposed changes.

[18]The act of reverting a commit is itself a commit, which has a specific, auto-generated commit message. The messages are available in the GHArchive data and allow to identify both revert commits and the original commit that is reverted.

[19]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.

[20]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/).

to evaluate 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 searching for easy tasks due to an adverse productivity shock can easily identify these issues as suitable.

Secondly, we exploit the timestamps reported in the data to approximate users' working hours in order to investigate whether they try to make up for their reduced productivity by working longer hours in the evening or on the weekend. Evening work is measured by the minute of the day the last action of the day was performed and the share of actions conducted after standard working hours, i.e., after 6 pm. To measure weekend work, we use the sum of actions conducted on Saturday and Sunday of each week.

Finally, as a summary measure of the quantity, quality, and relevance of a user's work, we consider the monthly growth rate of the number of a user's followers. This allows us to investigate whether air pollution exerts only temporary effects on daily output, or whether it also generates effects on users' reputation and influence over a longer time horizon.[21]

Descriptives. Table 1 presents summary statistics on the outcome variables. On average, users perform 2.77 actions per day, of which 1.29 are commits and 0.93 are comments. The remaining productive GitHub actions—opening and closing issues and PRs—are observed less often. Hence, the sample users are indeed highly active in public repos, as casual users who work on GitHub only occasionally can hardly achieve such figures, especially given that we average across all days, including weekends and holidays.[22] On average, users are active on 37% of all days in the sample period. Conditional on being active at all, the mean number of actions per day is 7.59. A commit reversal is a very rare event, indicating severe errors. It happens among only 0.2% of all commits made in the sample. Rejection of a PR occurs more frequently, in 33% of all cases. On average, 7% of all issue events refer to an easy issue. 29% of actions are made after 6 pm. The mean time of the final action of the day is 5:45 pm.[23]

Figure 2 provides more detailed information on the distribution of activity across days of the week and hours of the day. The solid lines depict the share of all activity that is conducted during the respective hour of the day on weekdays (left) or weekends (right), respectively. We present this share for commits, comments, and total actions. Activity levels are highest during core working hours (marked in grey) and considerably lower in the evening and night hours and on weekends. Notable activity during evening hours and on weekends is not uncommon among highly educated workers (Mas and Pallais, 2020). Overall, the distribution is similar across all three variables. However, comments, i.e., interactive activities, are even more

[21] We provide a list of all outcomes and details on their construction in Appendix Table A.3.

[22] Overall, our main sample comprises only 0.076% of all GitHub users but accounts for a disproportionally large share of all activities in public repositories, e.g., 2.1% of issue creations, 7.6% of issue closings, 9.9% of comments written and 6.6% of PR actions (opening and closing).

[23] To take into account that high-skill workers often work long hours in the evening, we define a work day to last from 3 am on the calendar date to 3 am on the following day.

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

	Mean	SD	SD (within)	Min	Max	Observations
Output Quantity						
Actions	2.77	7.26	6.46	0	293	14,538,351
of which Commits	1.29	3.82	3.55	0	234	14,538,351
Comments	0.93	3.37	2.94	0	280	14,538,351
PRs opened	0.15	0.71	0.67	0	151	14,538,351
Issues opened	0.10	0.80	0.74	0	222	14,538,351
PRs closed	0.17	0.94	0.89	0	284	14,538,351
Issues closed	0.12	0.89	0.87	0	263	14,538,351
Any action	0.37	0.48	0.44	0	1	14,538,351
Actions | Actions > 0	7.59	10.39	9.22	1	293	5,310,794
Output Quality						
Share PRs merged	0.67	0.45	0.40	0	1.0	1,132,824
Share commits reverted	0.002	0.029	0.029	0	1	3,458,932
Task Complexity						
Share easy issue events	0.07	0.21	0.20	0	1.0	3,203,772
Files changed per PR	9.37	60.15	60.00	0	9660	1,738,027
Lines added per PR	347.05	1660.15	1618.98	0	64425	1,738,027
Lines deleted per PR	125.56	700.38	686.17	0	25037	1,738,027
Working Hours						
Share actions after 6 pm	0.29	0.39	0.36	0	1.0	5,296,035
Time last action	17:45	5.01 hours	4.63 hours	3:00	3:00	5,296,035

Note: This table describes the main analysis sample at the developer×date level. The first two panels provide summary statistics for outcome variables we use to measure output quantity and quality. The bottom panels describe variables measuring task complexity and working hours. The table displays the mean, standard deviation, within-developer standard deviation, minimum and maximum value of the variables as well as the number of observations.

concentrated during standard working hours as compared to commits, i.e., individual coding activities. This is plausible given that the more collaborative tasks are more productive during common working hours, when other users are working as well.

Finally, Figure 3 presents information about the work status of users in our sample. The left plot depicts the most frequent terms used in the biographies (bios) on their GitHub profiles. 36% (9,507 users) of the sample provide such a self-description. The data was 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. The right plot complements this with information on employers which users can report on their GitHub profiles. In our sample, 61% (16,385 users) provide some information in this field with Microsoft and Google being the most frequent employers, followed by Facebook and Red Hat, i.e., big US-based tech companies strongly engaged in open-source. While we are unable

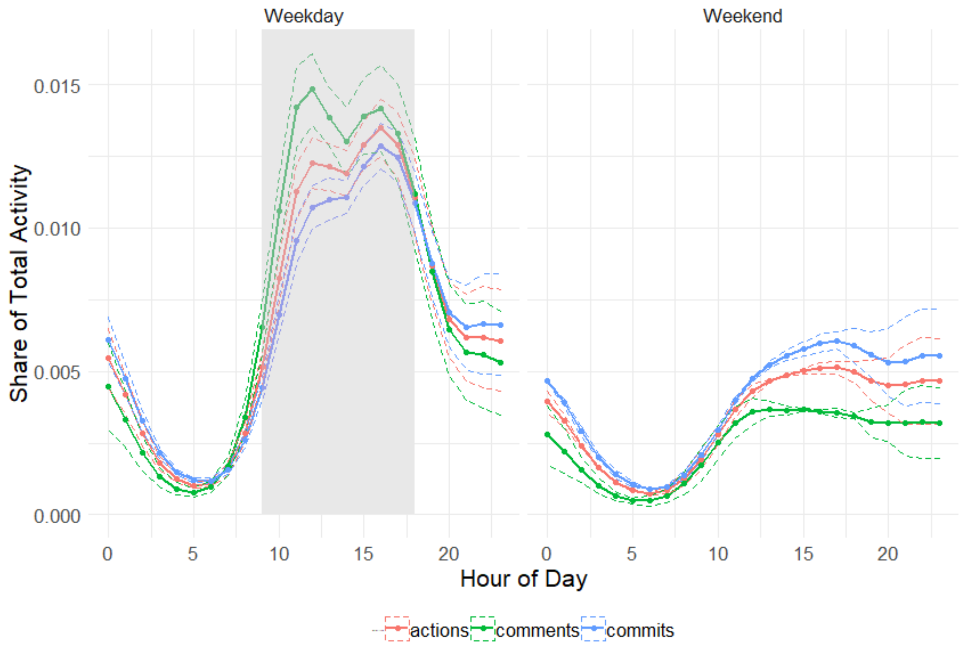


Figure 2: Distribution of Activity across Hours of the Day and Days of the Week

Note: Solid lines reflect the share of activities by sample users which is conducted during the respective hour of the day on weekdays (left) or weekends (right), separately for total actions, commits and comments. Dashed lines mark 95%-confidence intervals. Grey area reflects core working hours, 9 am to 6 pm on weekdays.

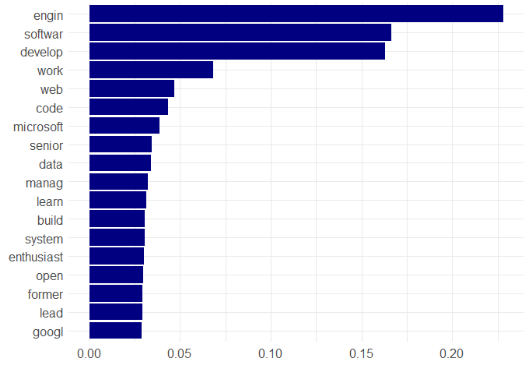
to assess whether the subsample of users who provide a bio or the company information is representative, the clear peaks in the two plots at work-related terms and well-known tech companies, together with the concentration of activity in core working hours, strongly suggest that we do capture professional software developers who use GitHub as part of their formal work. Thus, in the remainder of the paper, we use the term 'developers' when referring to the users in our sample.

3.3 Gitcoin: Monetary Value of GitHub Activity

To assess the validity of the productivity metrics constructed from the GitHub data and to translate the estimated effects of air pollution on these outcomes into monetary damages, we draw on data from a platform called *Gitcoin*.[24] Two types of agents interact on this platform: GitHub project teams aiming to incentivize external contributions to their projects post open issues from their public repos on Gitcoin, along with information on issue characteristics and a payment they offer for a solution. On the other side of the transaction, freelance developers can apply to solve these issues and earn money for their contributions.

Work on the issues is submitted in the form of a PR in the respective GitHub repo. If the

[24] Gitcoin was founded in 2019 and is complementary to GitHub. At the end of 2021, about 300,000 GitHub users were registered on the platform.



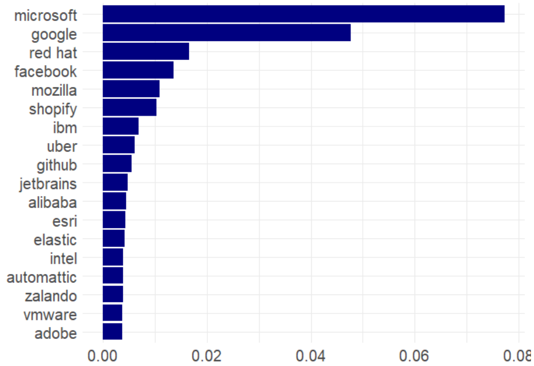


Figure 3: Most Frequent Terms from User Self-Descriptions and Company Fields

Note: The left bar plot is based on data from 9,507 user bios, accessed in 2021 via the GitHub API. Words in the bios are transferred to lowercase, stemmed, and stop words are removed. The total word count is divided by the number of bios. The plot on the right is based on data from June 2019 on 16,385 users, collected from the company column in the GHTorrent user table.

PR is accepted by the issue funder, the PR author receives the payment, typically in cryptocurrencies. We collect data on 292 issues for which PRs were submitted and payments made by March 2022 via the Gitcoin API, including the value of the payment in USD and the hours worked on the PR as reported by the submitting user. We merge this with information on the size of the respective pull request obtained via the GitHub API (number of commits, number of lines of code added and deleted, and number of files changed). A detailed description of the data is provided in Appendix C.

In the data, we find mean payments of $354 per pull request and $112 per commit, one of our primary outcome variables. On average, developers spend 1.8 hours on one commit. Hence, their implied hourly wage amounts to $62, almost coinciding with the mean hourly wage of $58 among software developers in the US in 2021 as reported by the Bureau of Labor Statistics (2021). We will use the monetary values of commits and PRs, our measures of individual coding activity, to translate the estimated effects of air pollution on these outcomes into monetary damages.

Are the outcomes we consider valid measures of productivity and task complexity? In Appendix Tables C.1 to C.3 we use the Gitcoin data to test this. We find that both the payment awarded for the PR and the time spent on creating it are consistently positively correlated with the number of commits the PR comprises. We view this as a confirmation that changes in the number of commits reflect fluctuations in developer productivity. Holding the number of commits constant, adding more lines of code and changing more files in a PR is associated with a higher payment, suggesting that these variables indeed reflect task complexity. 12% of the Gitcoin issues are labeled as *easy* according to our definition. PRs addressing these issues are on average rewarded $186, only half the amount among PRs addressing other issues.

Even holding all aforementioned PR characteristics constant, these issue labels are negatively associated with the value of a PR, supporting their validity as indicators for easy tasks.

3.4 Environmental Data

Air Quality. The pollutant of interest in our analysis is $PM_{2.5}$. We collected data on $PM_{2.5}$ concentration measured at outdoor monitors in and around the sample cities from several environmental agencies. For a few cities, we could not obtain monitor-measured data but instead used high-resolution reanalysis data from the Copernicus Atmosphere Monitoring Service (CAMS). Reanalysis datasets are constructed by combining measurements taken at ground-level monitors, satellite images, and atmospheric transport models. 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.[25] Our data on $PM_{2.5}$ covers 96% of all city×day observations.

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 the sample is 12.4 μg/m^3 (standard deviation: 14.5 μg/m^3, within-city: 11.8 μg/m^3), i.e., slightly above the annual standard of 12 μg/m^3 set by the U.S. Environmental Protection Agency (EPA) and clearly above the WHO guideline value for the annual mean concentration (5 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}.[26] Air quality exhibits substantial heterogeneity across regions: Cities in North America, Oceania, and Northern Europe have relatively clean air, with concentrations above 20 μg/m^3 rarely being observed. Locations in Southern and Eastern Europe by contrast experience this level of pollution on 28% of all days, and Asian cities even 60% 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. The u- and v-component of wind, i.e. the eastward and northward wind vectors, are reported every six hours (in UTC) on a global grid with a spatial resolution of 1.25° longitude×1.25° latitude, which corresponds to roughly 137.5km×137.5km at the equator.[27] We translate timestamps into local time and aggregate to

[25] CAMS reanalysis data is reported on a 0.1° longitude×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.

[26] We show the distribution of observations in our developer×date panel across these regions in Table A.5

[27] 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

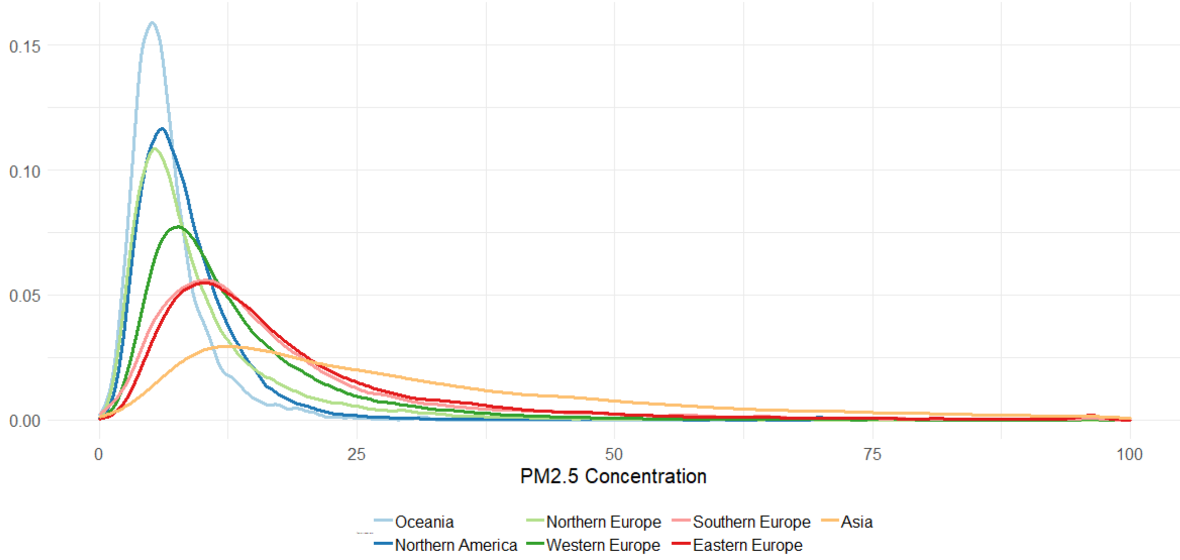


Figure 4: Distribution of Daily $PM_{2.5}$ Concentrations by Geographic Region

Note: The plot shows densities of $PM_{2.5}$ concentration based on 353,445 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, Croatia, Slovenia. Asia: China, India, Japan, Hong Kong. Northern America: US, Canada. Western Europe: Switzerland, Austria, France, Germany, Belgium, and the Netherlands. Eastern Europe: Poland, Czech Republic, Hungary, Belarus, Ukraine, Slovakia, Bulgaria, Romania.

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 = 92.3km). 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 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×0.1° latitude horizontal resolution, corresponding to roughly 11km×11km. To construct city×day level variables, we follow the 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 = 10.9km). The variables constructed are daily mean, minimum and maximum temperature, precipitation, and relative humidity.[28]

Wildfire Smoke. The North American west coast frequently experiences severe wildfires generating heavy smoke that strongly increases 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

small area. The choice of data follows a suggestion by Tatyana Deryugina which we gratefully acknowledge.

[28]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.

Bay Area and around Seattle). Given recent evidence that exposure to heavy wildfire smoke can trigger avoidance behavior, especially among high-income individuals (Burke et al., 2022), we construct control variables for heavy smoke to make sure the effects we estimate reflect physiological impacts of $PM_{2.5}$ exposure, not behavioral responses to 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. It 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 15km radius around its centroid. We aggregate the data to the daily level by summing over the intensity measure of all smoke plumes covering a city on a given day. We define a heavy smoke indicator which is one if the city was covered by a plume of the maximum intensity or if the total daily smoke intensity exceeds a value of twice the maximum intensity, and zero otherwise. This yields 0.3% of all city-by-day observations and 8.9% of all observations with any smoke exposure as heavy smoke days.

Thermal Inversions. In extensions and 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[29]. Hourly temperature at the surface level as well as several pressure levels is reported on a 0.25° longitude×0.25° latitude grid. We compute the difference between upper air temperature, at the pressure level 25 hPa above the surface, and surface air temperature. Following several recent papers, e.g. Jans et al. (2018), we then calculate the average temperature difference during local nighttime hours (midnight to 6 am). Cities are assigned the inverse distance weighted average from the four closest grid points. We use the temperature difference as a measure of inversion strength, $inv\ strength_{cd} = \Delta T_{cd}$, to instrument for pollution.

4 Research Design

The first part of this section presents our baseline regression model and discusses why endogeneity issues are likely to arise. In the second part, we describe the instrumental variable approach based on wind direction we adopt to address these issues.

Baseline Regression Model. To analyze how short-run variation in local particulate matter concentration affects output and work patterns of professional software developers, we specify

[29]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*

a model for outcome y of developers living in city c on day d.

$$y_{c,d} = \beta PM_{c,d} + \mathbf{w}'_{c,d}\gamma + \delta_{R(c)}h_{c,d} + \mu_c + \mu_{R(c),dow(d)} + \mu_{R(c),yr(d),m(d)} + \varepsilon_{c,d} \quad (1)$$

Here, $y_{c,d}$ denotes one of the measures of output quantity, quality, or work patterns described in the previous section. We obtain this variable through an auxiliary regression that includes the information available for each individual developer, i.e., her experience in using GitHub and a developer fixed effect. This way we can reduce the computational burden without losing variation in the regressor of interest which is observed at the city-day level. This procedure is common in the literature (e.g. Currie et al., 2015) and asymptotically equivalent to estimating the underlying individual-level regressions (Donald and Lang, 2007). Appendix D provides a more detailed description.

$PM_{c,d}$ is a measure of particulate pollution and varies across cities c and days d. The fixed effect μ_c controls for time-invariant unobserved factors at the city level. Region-year-month fixed effects $\mu_{R(c),yr(d),m(d)}$ capture time-varying productivity shocks common to all developers in a geographic region R. Region×day-of-week fixed effects $\mu_{R(c),dow(d)}$ and an indicator for holidays, $h_{c,d}$, control for fluctuations in work patterns and productivity across days of the week and public holidays. These fluctuations are allowed to vary in intensity across different world regions. $\mathbf{w}_{c,d}$ is a vector of weather variables that can be correlated with air quality and at the same time affect work patterns. 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. The omitted category is temperature falling between the 35th and the 65th percentile. The effects of temperature fluctuations are also allowed to differ across regions R. The vector further contains cubic polynomials of precipitation, relative humidity, and wind speed and a dummy indicating whether the city is affected by heavy wildfire smoke on day d. We weight all regressions by the number of underlying developer observations in each city–day cell and cluster standard errors at the city level.

The coefficient of interest β is estimated from day-to-day variation in city-level pollution and developer output, conditional on average developer output and after netting out other productivity determinants such as weather, seasonality, and region-wide business cycle dynamics.

Since air quality is not assigned randomly to the city-by-day observations, $PM_{c,d}$ may be endogenous in Equation (1) due to unobservable factors which co-vary with particulate matter and productivity. Variations in local economic conditions can for instance affect air pollution and developers' output at the same time. Similarly, local events like a football match or the closing of a bridge may impact both traffic and work patterns. The OLS estimate of β would thus likely suffer from omitted variable bias. A second issue is measurement error in developers' pollution exposure, which we cannot observe, but instead have to proxy for by city-level averages. This generates attenuation bias in the OLS estimate. While our regression model includes a wide range of controls to account for sorting into different cities or fluctuations in economic conditions, we still require an exogenous source of variation in local air pollution to address these two concerns.

IV estimation. We address endogeneity in Equation (1) by instrumenting local pollution levels with wind direction. This approach was 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, for instance, carries substantially lower amounts of pollution into the city than wind blowing from more densely populated or industrial areas.

It is important to note that local weather conditions can 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 the wide range of weather conditions contained in $\mathbf{w}_{c,d}$ to ensure that the instrument 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 $PM_{2.5}$ to vary. In principle, we could estimate the effect of wind direction separately for each city. In that case, however, the first stage might pick up effects of highly local transport that affects readings at local monitors due to their location relative to the pollution source, but simply redistributes particulate matter within the boundaries of a city. To ensure that the first stage only captures effects of regional pollution transport that changes $PM_{2.5}$ in the whole city, we restrict the effect of wind to vary at a geographically more aggregate level. As suggested by Deryugina et al. (2019), we use a clustering algorithm to assign cities into groups g based on their longitude and latitude. In our baseline specification, we form 50 groups, using hierarchical clustering with a complete-linkage criterion. They are illustrated in Appendix Figure B.2.[30]

We parameterize the pollution-wind relationship by a trigonometric function.[31] 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-

[30]When using all 220 cities depicted in Figure 1 in the clustering algorithm, it forms singletons for some cities which are very distant from their closest neighbor, e.g. Beijing or Salt Lake City. We drop these cities (depicted in light blue) in the main analysis because whenever a city forms its own first-stage group, the first stage might pick up local pollution transport.

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

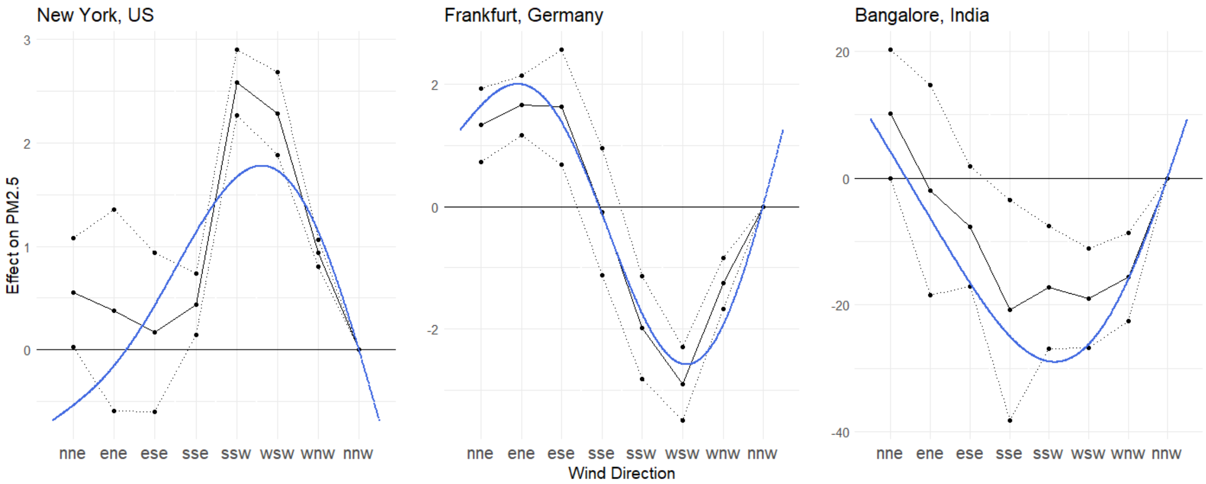


Figure 5: The effect of wind direction on $PM_{2.5}$

Notes: This figure provides a graphical illustration of the first stage. Graphs present estimated coefficients from regressions of $PM_{2.5}$ measured in $\mu g/m^3$ 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, Walldorf, Heidelberg, Bern, Basel, Strasbourg (center), Bangalore, Chennai, Hyderabad (right).

pollution relationship. The first stage of the IV estimation is as follows.

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

The coefficients ρ_1^g and ρ_2^g are allowed to vary across city-groups $g \in \{1, 2, ..., 50\}$.

Figure 5 illustrates how this trigonometric function can capture the effect of wind direction on $PM_{2.5}$ levels for the city groups represented by Frankfurt, New York City, and Bangalore. The estimated relationships are depicted in blue. They strongly resemble the results we obtain when we instead measure wind direction by eight indicators representing 45° sections of the wind rose, i.e., (0°-45°], (45°-90°], etc. In Appendix Figure B.3 we present the respective plots for all 50 city groups.

We adopt this 2SLS approach for all analyses at the city×date level. In some parts of our analysis, e.g. when exploring effect dynamics or impacts at the monthly level, we use modified versions of this framework that will be presented in the respective sections.

Measures of pollution. Our primary measure of air pollution is daily $PM_{2.5}$ concentration in $\mu g/m^3$. As an alternative, we define a binary variable that indicates whether $PM_{2.5}$ is unusually high relative to common levels in city c. More formally, it takes the value one, when the city-day level $PM_{2.5}$ exceeds the city-specific seventy-fifth percentile, $\mathbb{1}\{PM_{c,d} > Q_{.75}\left(PM \mid c\right)\}$. The proposed measure, therefore, captures non-linear effects of pollution and allows these to differ by location.

5 Main Results

In this section, we first present results on how $PM_{2.5}$ exposure affects the quantity and quality of output developers produce. Thereafter we show that they use two margins of adjustment, task choice and working hours, to adapt to increases in pollution concentration.

5.1 Work Quantity

Columns 1 to 3 of Table 2 display 2SLS estimates of the effect of $PM_{2.5}$ exposure on the three primary quantity outcomes—the number of total actions conducted, the number of commits as a measure of individual coding activity, and the number of comments written in discussion fora as a measure of collaborative work. In Panel A, we use $PM_{2.5}$ concentration as regressor and find that an increase by 1 $\mu g/m^3$ causes developers' output, measured by total actions, to fall by 0.0032 or 0.12% of the sample mean. This decline is mainly driven by a reduction in the number of commits, which decreases by 0.0026 or 0.20% of the sample mean. The number of comments is much less affected by air pollution. The point estimate is close to zero and not statistically significant. The first stage F-statistic on the excluded instruments exceeds 100, indicating that the IVs based on wind direction are sufficiently strong. For an increase in ambient $PM_{2.5}$ concentration by one within-city standard deviation (11.8 $\mu g/m^3$), the estimates imply reductions in the number of commits and total actions by 0.030 (2.3%) and 0.038, (1.4%) respectively.

In Panel B, we repeat the analysis, now using the binary variable indicating that $PM_{2.5}$ concentration exceeds the city-specific 75th percentile. The F-statistic is again well above the common threshold for a sufficiently strong first-stage relationship. The 2SLS estimates imply that on a day with relatively high pollution, the number of total actions falls by 0.11 or 4% of the mean value. The number of commits falls by 0.08 or 6.2%. Again, no significant effect on the number of comments is found.

In sum, these results imply that fine particulate matter exposure exerts a negative effect on developer output which is mostly driven by days with relatively poor air quality. The effect of a high-pollution day in Panel B corresponds to an increase in $PM_{2.5}$ concentration by more than 30 $\mu g/m^3$ based on the coefficients in Panel A. A novel finding is the strong effect heterogeneity across different types of work commonly conducted in high-skilled occupations: We observe a highly significant negative impact on individual work on code, but no effect on interactive work.

In Appendix Table A.6, we investigate the effect of $PM_{2.5}$ on further action types which occur less frequently than commits and comments—the number of issues and PRs opened and closed, respectively. 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 decision-making about whether to accept or reject the proposed changes. Consistent with the results discussed above, the number of PRs opened falls significantly in $PM_{2.5}$ concentration, and the relative effect magnitude is similar to the effect on commits. While we find marginally significant effects for closing of PRs, interactive issue events are unaffected by air pollution. Overall, these results confirm the conclusions drawn from Table 2.

Table 2: Effect of $PM_{2.5}$ on Work Quantity

	Actions (1)	*Commits* (2)	*Comments* (3)	*Any actions* (4)
Panel A.				
$PM_{2.5}$	-0.0032^{***}	-0.0026^{***}	-0.0005	-0.00013
	(0.0011)	(0.0008)	(0.0005)	(0.00009)
	[0.003]	[0.001]	[0.374]	[0.155]
	{0.006}	{0.002}	{0.374}	{0.208}
First Stage F-Stat.	102.1	102.1	102.1	102.1
% change in Y	-0.12	-0.20	-0.05	-0.04
% of full effect				11.2
Panel B.				
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	-0.1104^{***}	-0.0801^{***}	-0.0169	-0.0068^{***}
	(0.0302)	(0.0159)	(0.0170)	(0.0025)
	[0.0004]	[0.000002]	[0.323]	[0.008]
	{0.0014}	{0.00002}	{0.374}	{0.013}
First Stage F-Stat.	80.5	80.5	80.5	80.5
% change in Y	-4.0	-6.2	-1.8	-1.9
% of full effect				17.1
Observations	353,445	353,445	353,445	353,445
Mean Dep. Var.	2.77	1.29	0.93	0.37

Note: The table presents IV estimates of the parameter β in Equation (1). In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Panel B, a binary variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city-specific 75th percentile. The first stage specification is given in Equation (2). 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, and year-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. Unadjusted p-values and p-values adjusted for multiple hypothesis testing according to the Benjamini-Hochberg procedure are shown in squared and curly brackets, respectively. *p<0.1; **p<0.05; ***p<0.01

In Column 4 of Table 2 we explore the contribution of the extensive margin to the overall reduction in work quantity. The dependent variable is an indicator for a positive activity level, i.e., $\mathbb{1}\{actions_{id} > 0\}$. For both measures of pollution, we find negative point estimates whose magnitude implies that the extensive margin effect contributes approximately 11-17% to the full reduction in actions. The estimate is statistically significant only for the dummy indicating $PM_{2.5}$ concentration above the 75th percentile. 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. They are unlikely to suffer severe health damages from short-run pollution exposure which prevent them from working, especially at moderate levels of concentration, but rather subtle effects on health and cognitive function.

Since we derived our main results by testing eight hypotheses, we also report significance levels that correct for multiple hypothesis testing following the Benjamini-Hochberg procedure in Table 2.

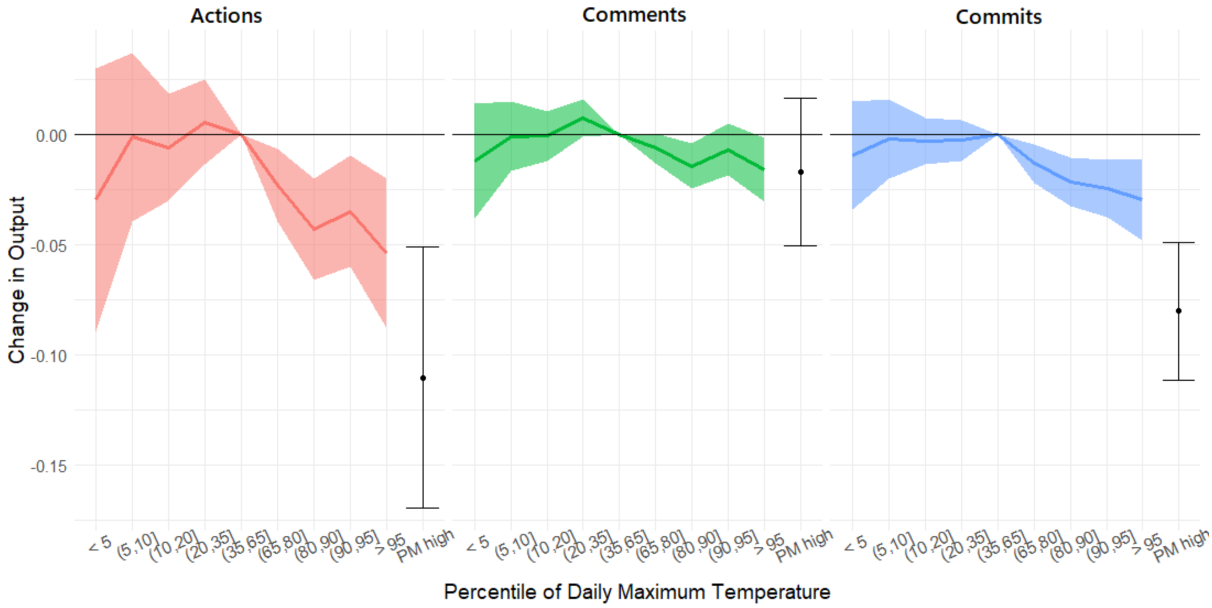
In Columns 1 to 3 of Table A.7 we present results from estimating the model in Equation (1) by OLS for the three main measures of output quantity to assess the direction and size of the bias. We obtain negative estimates for total actions and commits with both $PM_{2.5}$ in $\mu g/m^3$ and the binary indicator for unusually high pollution levels, but they are significantly different from zero only when using the binary regressor. The results replicate the pattern that effects on commits are larger than on comments. Mirroring a common finding in the literature on short-run impacts of air pollution exposure, all estimates are substantially smaller than the 2SLS results, pointing towards attenuation bias due to measurement error. The ratio of 2SLS-to-OLS estimates ranges between 7 and 13 across specifications.[32] In Table A.8 we re-do the OLS analysis on the extended sample of 220 cities. Results are almost identical for the binary indicator, but the coefficients also attain statistical significance with the continuous regressor.

Effect Magnitude. We conduct three exercises to assess the magnitude and the economic relevance of the estimates. Firstly, we compare the impact of $PM_{2.5}$ concentration above the 75th percentile on developers' output to the effect of another highly relevant environmental shock, exposure to extreme outdoor temperatures.[33] 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.

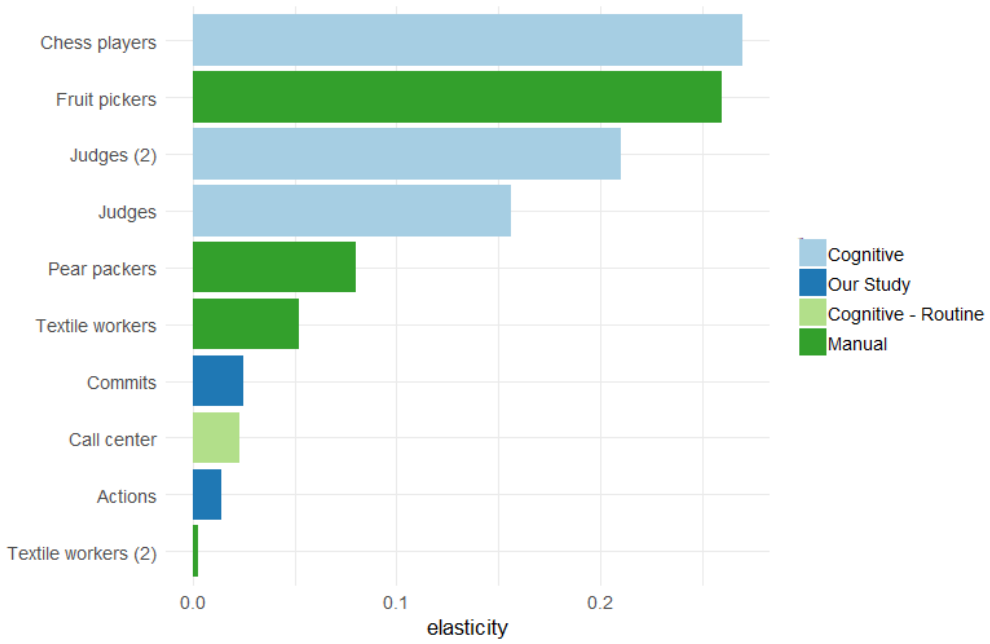
Figure 6a reproduces the estimated effects of $PM_{2.5}$ concentration exceeding the 75th percentile on actions, commits, and comments in graphical form (point estimates with 95% confidence intervals displayed in black on the right). In addition, coefficients from OLS regressions of the same outcomes on maximum daily temperature are presented. The regressors of interest

[32]This is in line with e.g. 2SLS-to-OLS ratios found by Deryugina et al. (2019).

[33]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.



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



(b) Effects of Air Pollution Across Occupations

Figure 6: Effect Magnitude

Note: Figure 6a reproduces the estimated effects of a $PM_{2.5}$ concentration exceeding the city-specific 75th percentile on actions, commits, and comments from Panel B of Table 2 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, and year-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. Figure 6b shows the elasticities of commits and actions with respect to $PM_{2.5}$, based on the estimates in Columns 1 and 2 of Table 2. Besides, it presents elasticities of performance with respect to air pollution from other studies, in particular: Künn et al. (forthcoming) (Chess players), Sarmiento (2022) (Judges), Kahn and Li (2020) (Judges (2)), Chang et al. (2019) (Call center agents), He et al. (2019) (Textile Workers (2)), Adhvaryu et al. (2022) (Textile Workers), Chang et al. (2016) (Pear Packers), and Graff Zivin and Neidell (2012) (Fruit pickers).

are eight dummy variables indicating whether maximum daily temperature falls into 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 (1). 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.[34] Even though the developers 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. Importantly, for both commits and total actions, the point estimate on the $PM_{2.5}$ dummy is more than twice as large as the point estimate for the highest temperature bin which reflects maximum daily temperature above the 95th percentile.[35] The IV estimates ($PM_{2.5}$) are less precise than the OLS estimates (temperature), but still, even the lower bounds of the 95% confidence intervals on the pollution effects are as large as or even exceeding the point estimates 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.

Next, we compute elasticities of total actions and commits w.r.t. $PM_{2.5}$ based on the estimates in Panel A of Table 2. We obtain elasticities of -.014 and -.025 for actions and commits, respectively. These values, along with elasticities of productivity or performance found in previous studies, are depicted in Figure 6b.[36] 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 for judges and chess players, who are also engaged in cognitively demanding tasks. As outlined above, a potential explanation is that chess players and judges face more inflexible settings, namely chess tournaments 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 Section 5.3. This underscores the importance of our analysis: It might be misleading to draw conclusions on the total economic cost of air pollution based on the estimates for cognitively-demanding tasks in highly inflexible settings because in many high-skilled occupations workers have at least some degree of flexibility in organizing their work day.

Even though productivity effects are small in comparison to other contexts, they might

[34] 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.

[35] Median maximum temperature in this bin is 30.5° C, while the median value in the omitted bin is 16.9° C.

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

still be economically relevant, given that software development is a high-paying occupation generating large economic value. We use the average monetary value of commits and PRs opened (derived in Section 3.3) to translate the estimated negative effects of $PM_{2.5}$ exposure on these two outcomes into changes in output value.[37] For a within-city standard deviation increase in $PM_{2.5}$ (11.8 $\mu g/m^3$) the implied reduction in daily output value amounts to $4.06 per software developer. 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. On days when $PM_{2.5}$ concentration exceeds the city-specific 75th percentile output value falls by $11.0 relative to days with better air quality. Given that we ignore losses from reductions in task complexity in the calculation (see below), these estimates can be interpreted as a lower bound.

In summary, the impact of air pollution shocks on productivity exceeds the effect of heat. In comparison to other professions, 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.

Effect Dynamics. The existing literature on air pollution and worker productivity found mixed results on the lagged impact of exposure. He et al. (2019) show 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. (forthcoming) find that chess players' performance is unaffected by pollution exposure on the previous days. To explore effect dynamics in our setting, we regress the three measures of output quantity on eleven dummies indicating whether wind was blowing to city c from the high-pollution direction on day d and each of the previous ten days, $WDir\ highPM_{c,d-k}$, where $k \in \{0, 1, \ldots, 10\}$ denotes the lag order. 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 five dummies for average daily wind direction falling into a specific 60° bin as instruments, interacted with the city-group indicators.

Appendix Table A.9 shows estimated coefficients from the "reduced form" model for total actions, commits, and comments, including just the indicator for the same day, $WDir\ highPM_{c,d}$. The signs and significance of the estimated coefficients are in line with the 2SLS results. The first stage effect, reported at the bottom of the table, implies that wind from a city's high-pollution direction raises $PM_{2.5}$ concentration on average by 3.7 $\mu g/m^3$ relative to days where wind arrives from another direction. While this approach is much less flexible than our main 2SLS model, it captures the underlying idea in a single variable and thus allows us to easily

[37] As stated in Section 3.3 we work with an average monetary value of $112 per commit. In the case of PRs, we do not use the mean value of $354 found in the Gitcoin data because PRs in that sample are larger on average than PRs created in our main analysis sample. Instead, we value PRs with $2.78 \times \$112 = \$311$ given that they comprise, on average, 2.78 commits.

analyze effect dynamics by including lags. Since we are mostly interested in qualitative results – does lagged exposure reduce productivity or does it induce developers to work more in order to catch up – we opt for this approach.

The distributed lag model we use to explore effect dynamics includes the same covariates as our main contemporaneous model plus ten lags of weather conditions and is given by:

$$y_{c,d} = \sum_{k=0}^{10} \beta_k \, WDir\ highPM_{c,d-k} + \sum_{k=0}^{10} \mathbf{w'}_{c,d-k}\gamma_k + \mu_c + \mu_{R(c),yr(d),m(d)} + \mu_{R(c),dow(d)} + \delta_R h_{c,d} + \varepsilon_{c,d} \quad (3)$$

From the corresponding estimates $\widehat{\beta_k}$, we compute the cumulative effect of exposure to wind from the high-pollution direction for s consecutive days, $\sum_{k=0}^{s} \hat{\beta}_k$ for $s = 0, 1, \ldots, 10$, which we plot in Figure 7.

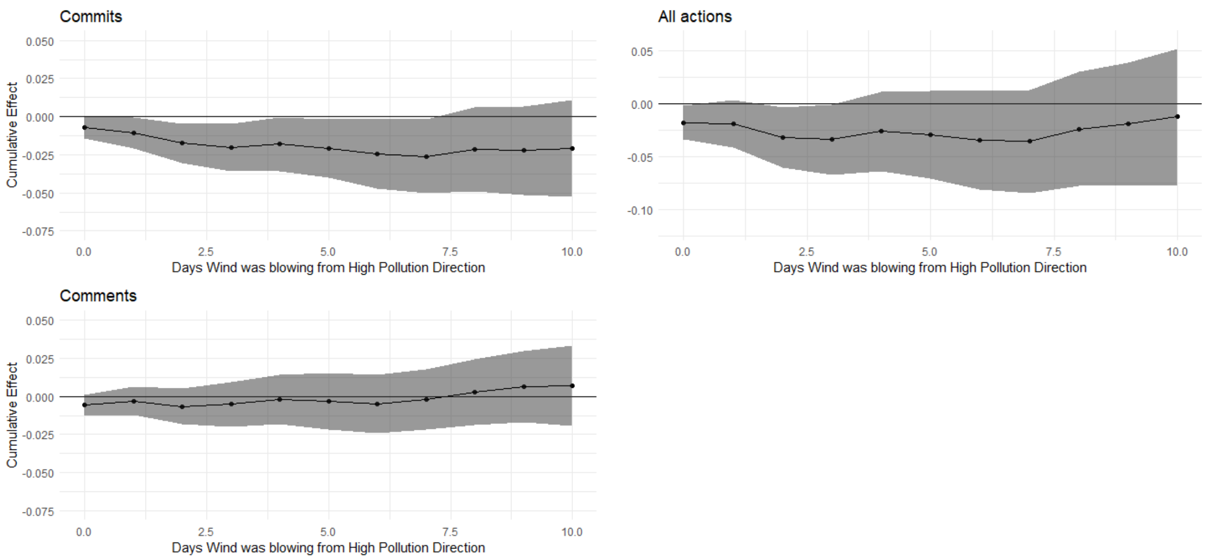


Figure 7: Effect Dynamics: Work Quantity

Note: The plots depict estimates of the cumulative effect of wind blowing from the high-pollution direction on three measures of work quantity. Effects are derived from a distributed lag model and given by $\sum_{k=0}^{s} \hat{\beta}_k$ for $s = 0, 1, \ldots, 10$. 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, Region-by-day-of-week, and Region-by-year-by-month fixed effects, a holiday indicator, and weather controls for the current day and ten 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.

For both total actions and commits, same-day exposure to pollution generates negative and significant effects. The cumulative effect magnitude grows monotonically up to the third lag. The current day effect of $WDir\ highPM_{c,d}$ is estimated to be -.018 for total actions and -.007 for commits. After four consecutive days of wind from the high-pollution direction, the cumulative effect is -.037 and -.022, respectively. For more prolonged exposure the point estimate of the cumulative effect remains rather constant but becomes noisier, likely due to serial correlation in the wind direction variable. For commits, there is no decline in the cumulative effect magnitude at higher lags, i.e., no evidence that developers completely make up for the output loss within the first ten days after exposure. After ten days, the point estimates for commits and total actions are still negative and as large as or even larger than the point estimate for same-day exposure. For comments, the cumulative effect always remains close to zero throughout the full time window.

In sum, pollution exposure generates an adverse effect on current-day output and, to a smaller extent, also reduces productivity on the following three days. Compared to health impacts, the productivity effects are rather immediate.[38]

5.2 Work Quality

So far, we have shown that exposure to $PM_{2.5}$ reduces the quantity of output developers produce per day. In high-skill jobs, output quality is of major relevance and might also be affected by pollution shocks.

Table 3 displays 2SLS estimates of the effect of $PM_{2.5}$ on two measures of work quality. The first is the share of all PRs a user opened on a given day that is later merged, i.e., accepted. PR rejections suggest issues with code quality or style, indicating low work quality. The second is the share of commits made by a user on a given day that were later reverted. Commit reversals point toward severe errors that cannot easily be corrected in follow-on commits, i.e., major issues with the work quality. Sample sizes are reduced relative to the results on output quantity, because the outcomes are only defined for city×day observations with any PRs opened and any commits, respectively. Moreover, information on commit reversals is from the GHArchive data which is only available from 2015 onward.

We find small, insignificant point estimates for both outcomes when using $PM_{2.5}$ in $\mu g/m^3$ as regressor. PRs opened on days when $PM_{2.5}$ concentration exceeds the 75th percentile, are 0.8 p.p. more likely to get accepted. This represents an increase of 1.3% relative to the mean, i.e., a small improvement in work quality. Estimates remain insignificant for the share of commits that are reverted, but the negative sign is also in line with minor reductions in error frequency.

The null effect on quality contrasts with findings by Archsmith et al. (2018) who show that baseball umpires conduct more errors when exposed to higher pollution levels. In the next section, we present evidence that developers change their work patterns when exposed to high levels of pollution. This form of adaptation might explain why effects on output quantity are relatively modest and quality is unaffected in this flexible high-skilled setting.

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

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

	Share PRs merged (1)	*Share Commits reverted* (2)
Panel A.		
$PM_{2.5}$	0.00013	−0.0000004
	(0.00019)	(0.00001)
First Stage F-Stat.	56.0	78.8
% change in Y	0.19	- 0.02
Panel B.		
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	0.0083**	−0.00015
	(0.00399)	(0.00022)
First Stage F-Stat.	42.9	61.1
% change in Y	1.3	-8.2
Observations	135,433	215,728
Mean Dep. Var.	0.665	0.002

Note: The table presents IV estimates of the parameter β in Equation (1). The outcome in Column (1) is defined as the share of all PRs opened by a developer on a given day that later gets accepted. The outcome in Column (2) is defined as the share of all commits made by a developer on a given day that later get reverted, i.e., undone (see Section 3 for details). In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Panel B, a binary variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city-specific 75th percentile. The first stage specification is given in Equation (2). 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, and year-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

5.3 Worker Adjustment

In this section, we investigate whether work patterns change in response to increases in particulate pollution. We consider two potential margins of adjustment, task choice and working hours.

Switching to Easy Tasks. We analyze whether developers switch towards easier tasks when exposed to higher levels of pollution, for both activities related to issues, i.e., interactive tasks, and activities related to pull requests, i.e., coding and review tasks. Table 4 presents 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×day observations with non-zero issue events (issue opened, closed, or reopened, or a comment written on an issue), the number of observations is 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 within-city standard deviation raises the share by 2.6%. On days when fine particulate matter

Table 4: Effect of $PM_{2.5}$ on Task Complexity

	Share Easy Issue Events (1)	*Issue Events* (2)	*Lines added per PR* (3)	*Lines deleted per PR* (4)	*Files changed per PR* (5)
Panel A.					
$PM_{2.5}$	0.00015^{**}	-0.00046	-0.0013	-0.0013	-0.0010^{**}
	(0.00007)	(0.00041)	(0.0008)	(0.0010)	(0.0004)
	[0.036]	[0.264]	[0.112]	[0.202]	[0.017]
First Stage F-Stat.	86.3	102.1	62.1	62.1	62.1
% change in Y	0.2	-0.5	-0.1	-0.1	-0.1
Panel B.					
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	0.0032^{**}	-0.0219	-0.0404^{**}	-0.0196	-0.0244^{**}
	(0.0015)	(0.0134)	(0.0181)	(0.0214)	(0.0098)
	[0.028]	[0.105]	[0.027]	[0.362]	[0.014]
First Stage F-Stat.	66.1	80.5	45.6	45.6	45.6
% change in Y	4.9	-2.4	-4.0	-2.0	-2.4
Observations	250,376	353,445	164,883	164,883	164,883
Mean Dep. Var.	0.067	0.90			

Note: The table presents IV estimates of the parameter β in Equation (1). 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 3 for details) and the total number of issue events. In Columns (3) to (5), outcomes are defined as the average number of lines of code files changed, number of new lines of code added, and number of lines of code deleted across all PRs a developer opened, closed, or commented on. Inverse hyperbolic sine transformations are applied to these outcomes. In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Panel B, a binary variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city- specific 75th percentile. The first stage specification is given in Equation (2). 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, and year-by-month fixed effects. Day-of-week and year-by-month fixed effects and 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

levels exceed the city-specific 75th percentile, the variable even increases by 4.9% of the mean, or 0.32 percentage points.

In the second column, we analyze how the denominator of the share changes. Consistent with earlier results on the impact of $PM_{2.5}$ on interactive tasks, we find no statistically significant effect on the number of issue events. 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. When hit by a pollution-induced productivity shock, developers seem to exploit the fact that certain issue labels provide a prominent signal of issue complexity to focus on easier tasks.

This finding is corroborated when we consider the complexity of PRs in Columns (3) to (5). We consider three PR characteristics: lines of code added, lines of code deleted, and number of files changed, averaged across all PRs a developer worked on a given day.[39] While there can

[39] These variables are based on GHArchive data, whereas work quantity results used GHTorrent data.

be very difficult tasks that involve a lot of thinking but require only small changes in the code, we believe that these variables provide reasonable proxies of PR 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. Results presented in Table C.3 indicate that PRs with more lines of code added and files changed are rewarded higher payments on Gitcoin, validating the use of these variables as complexity metrics. Similarly, reviewing a PR is more demanding when it contains large changes across different files. The characteristics we use to measure complexity are prominently displayed when opening a PR on GitHub, such that reviewers can easily assess them and jugde their difficulty level.

We apply the inverse hyperbolic sine transformation to the outcome variables such that coefficients approximate percentage changes. The sample size is reduced relative to previous tables because the outcomes are defined only for city×day observations with a positive number of PRs opened, closed, or commented on. Point estimates are negative across all three outcomes and the two distinct measures of air pollution. On days with unusually high $PM_{2.5}$ levels, the number of files changed falls by 2.4%, while the number of lines added drops by 4%. The effect on the number of lines deleted is also negative, but only half as large and not significantly different from zero. This pattern is plausible since 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 pattern is similar for the continuous regressor, but the effect on new lines added is not significant at conventional levels, indicating that developers move towards less complex tasks mostly in response to large productivity shocks on high-pollution days.

In sum, the results imply that, on top of the overall reduction in the number of actions completed, developers switch towards less complex tasks when exposed to high levels of $PM_{2.5}$. Thus, the estimates of monetary effects of air pollution exposure presented above provide a lower bound, given that we found in Section 3.3 that less complex pull requests and those addressing issues labeled as *easy* are rewarded lower payments, even when holding the number of commits constant.

This form of adjustment might also explain why the magnitude of effects on output quantity is relatively small in comparison to results found for other occupations, and why work quality is not adversely affected. We next present some evidence that switching to easier tasks is indeed an adaptation strategy to surges in particulate matter exposure that allows developers to prevent large declines in work quantity. Table 5 presents results from a heterogeneity analysis based on developer characteristics. Specifically, we combine tenure (time since registration on GitHub) and the number of followers at the point in time the developer enters our sample into an index that represents their experience and popularity.[40] We split the sample

GHArchive data is only available from 2015 onward. In Appendix Table A.10 we show that using data on pull requests from GHArchive we can replicate the results presented in Table A.6 for PRs opened or closed measured in the GHTorrent data.

[40]The index is computed as the average of tenure and the number of followers, after standardizing both vari-

developers into terciles based on experience and run the IV regression at the developer×day level (the second stage is given by Equation D.1). Panels A and B present results estimated separately for the bottom and the top tercile, respectively.[41]

Table 5: Effect Heterogeneity by Experience

	Actions (1)	*Commits* (2)	*Comments* (3)	*Share Easy Issue Events* (4)	*Lines added per PR* (5)	*Files changed per PR* (6)
Panel A: Bottom Tercile of Experience						
$PM_{2.5}$	-0.0033^{*}	-0.0026^{**}	-0.0008	0.00035^{**}	-0.0022	-0.0019^{**}
	(0.0017)	(0.0011)	(0.0008)	(0.00014)	(0.0015)	(0.0009)
First Stage F-Stat.	3187	3187	3187	586	211	211
% change in Y	-0.12	-0.19	-0.10	0.47	-0.22	-0.19
Observations	4,774,247	4,774,247	4,774,247	900,318	327,297	327,297
Mean Dep. Var.	2.48	1.24	0.77	0.074		
Panel B: Upper Tercile of Experience						
$PM_{2.5}$	-0.0062^{***}	-0.0044^{***}	-0.0009	0.0001	-0.0015	-0.0003
	(0.0023)	(0.0014)	(0.0010)	(0.0001)	(0.0023)	(0.0013)
First Stage F-Stat.	3010	3010	3010	756	222	222
% change in Y	-0.18	-0.30	-0.06	0.17	-0.15	-0.03
Observations	4,387,377	4,387,377	4,387,377	1,105,410	324,527	324,527
Mean Dep. Var.	3.16	1.37	1.15	0.061		

Note: The table presents IV estimates of the parameter β in equation (D.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 (2). 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 developer, day-of-week, and year-by-month fixed effects. Day-of-week and year-by-month fixed effects and 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

The first three columns display estimated effects of $PM_{2.5}$ on the three primary measures of work quantity. Point estimates for total actions and commits are negative in both samples, but larger in absolute terms as well as relative to the sample means for the most experienced developers. Effects on comments are not significantly different from zero in either sample. For the adaptation variables examined in the last three columns, the pattern is reversed: While the direction of the effects is again the same in both samples, effects are now stronger among the less experienced developers and not significantly different from zero in the upper tercile.

These results confirm that switching to easier tasks is a form of adjustment to pollution-induced productivity shocks among highly-skilled workers. A potential reason why the least

ables.

[41] Median tenure (number of followers) is 1.4 years (11) in the bottom tercile and 5.7 years (31) in the upper tercile.

Table 6: Working Hours

	Time of Last Action (minutes) (1)	*Share of Actions after 6 pm* (2)
Panel A.		
$PM_{2.5}$	-0.226^{***}	-0.0001
	(0.085)	(0.0001)
	[0.009]	[.147]
First Stage F-Stat.	96.1	96.1
Panel B.		
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	-4.162^{**}	-0.002
	(2.061)	(0.003)
	[0.045]	[.368]
First Stage F-Stat.	73.9	73.9
Observations	302,575	302,575

The table presents IV estimates of the parameter β in Equation (1). In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Panel B, a binary variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city-specific 75th percentile. The outcome variables are the timestamp of the last action conducted by a developer on a given day in minutes in Column (1), and the share of all activities conducted after 6 pm in Column (2). Estimates are based on all developer×date observations with at least one recorded action in Column (1) and at least two recorded actions in Column (2). The first stage specification is given in Equation (2). 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, and year-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. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

experienced developers show a stronger 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 peers or potential employers. Work complexity on the other hand is less easily observable.

Working Hours. A second potential adjustment margin available in flexible work environments is a change in working hours. We start by analyzing whether developers expand or reduce activity in the evening in response to increasing pollution exposure. Table 6 presents the estimated effects of $PM_{2.5}$ on the timestamp of the last action performed by a developer on a given day (in minutes) and on the share of total actions conducted after 6 pm.[42]

We find that developers on average end the work day 0.23 minutes earlier in response to an increase in $PM_{2.5}$ concentration by 1 $\mu g/m^3$. On days with $PM_{2.5}$ levels above the city-specific 75th percentile, the work day ends 4.2 minutes earlier than on days with better air quality. To

[42] These outcomes are only defined for developer×day observations with at least one action, which explains the reduction in sample sizes.

put this into perspective, we approximate the average time spent per action as

$$\frac{1}{N}\sum_{i}\sum_{d}\Big(timestamp(last\ action_{i,d}) - timestamp(first\ action_{i,d})\Big)\,/\Big(number\ of\ actions_{i,d} - 1\Big)$$

where N denotes the number of developer×date observations with at least two actions. This yields an average of 72.5 minutes spent per action.[43] Point estimates for the share of actions conducted in the evening are also negative, but small and not statistically significant. Subtle effects of pollution might make developers feel unproductive, inducing them to end their work activity earlier on high-pollution days due to reduced opportunity cost of leisure time. If $PM_{2.5}$ exposure, e.g., triggers headaches or fatigue, developers might experience this as an off day and decide to reallocate work to days when they perform better. In many jobs, knowledge workers are very flexible in when and where they want to work. Thus, shifting work intertemporally from low productivity days to the weekend, a period with relatively low activity levels and thus scope for compensation (see Figure 2), might be an important adjustment margin in these settings. To investigate this, we estimate the effect of $PM_{2.5}$ exposure during the first half of the workweek on output produced on the weekend. This analysis is conducted at the developer×week level, using the following, slightly modified, regression model.

$$\begin{aligned} y_{i,c,w}^{weekend} = \beta PM_{c,w}^{Mo\text{-}We} + \mu_i + \mathbf{x}'_{i,t}\pi + \mathbf{w}'^{weekend}_{c,w}\gamma + \mathbf{w}'^{Mo\text{-}We}_{c,w}\alpha + \\ \delta_{R(c)} h_{c,w} + \mu_{R(c),yr(w),q(w)} + \mathbf{z}'_{c,w}\varphi + \varepsilon_{i,c,w} \end{aligned} \tag{4}$$

$y_{i,c,w}^{weekend}$ denotes the sum of actions conducted by developer i living in city c on the weekend of week w. $PM_{c,w}^{Mo\text{-}We}$ is a measure of $PM_{2.5}$ concentration in city c between Monday and Wednesday of week w. Specifically, we consider average $PM_{2.5}$ concentration or the number of days with $PM_{2.5}$ concentration exceeding the city-specific 75th percentile. Due to the finding that exposure to pollution exerts negative effects on output not only on the same day, but also over the next two to three days, we focus on $PM_{2.5}$ during the first half of the workweek to make sure that we pick up developers' *behavioral* adjustment to a productivity shock during the workweek, and do not confound it with *physiological* effects.

Pollution is instrumented by the same variables as described in Equation (2), with the only difference that wind direction $\theta_{c,w}$ is averaged between Monday and Wednesday. To account for auto-correlation in the instruments, we add the vector $\mathbf{z}_{c,w}$ to the model, which includes the instrumental variables measured on the weekend and on Thursday to Friday. This ensures that we do not pick up the effects of wind direction-induced changes in pollution on the weekend itself or the days immediately before. The model further includes a developer fixed effect μ_i, a region-by-year-by-quarter fixed effect, $\mu_{R(c),yr(w),q(w)}$, the number of public holidays during the workweek, $h_{c,w}$, a vector $\mathbf{x}'_{i,t}$ of bin variables capturing the developer's tenure on GitHub,

[43] Importantly, this computation gives the time input including breaks, as we cannot disentangle time spent working on activities and breaks.

Table 7: Effect of $PM_{2.5}$ in the First Half of the Workweek on Weekend Work

	Actions		*Commits*		*Comments*	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.						
$PM_{2.5}$	0.0043	0.0133**	0.0017	0.0062**	0.0016	0.0037**
	(0.0030)	(0.0055)	(0.0017)	(0.0030)	(0.0012)	(0.0019)
	[0.158]	[0.016]	[0.308]	[0.041]	[0.158]	[0.048]
First Stage F-Stat.	1406	1147	1406	1147	1406	1147
Panel B.						
High $PM_{2.5}$ Days	0.0630***	0.1362***	0.0268*	0.0646***	0.0213**	0.0341***
	(0.0225)	(0.0321)	(0.0136)	(0.0207)	(0.0084)	(0.0107)
	[0.006]	[.00004]	[0.051]	[0.002]	[0.013]	[0.002]
First Stage F-Stat.	1455	801	1455	801	1455	801
Observations	2,011,797	1,352,191	2,011,797	1,352,191	2,011,797	1,352,191
Weeks	all	only low PM weekends	all	only low PM weekends	all	only low PM weekends

Note: The table presents IV estimates of the parameter β in equation 4. Outcomes are the sum of all actions, commits, and comments made on the weekend, respectively. In Panel A, the regressor of interest is average $PM_{2.5}$ concentration between Monday and Wednesday. In Panel B, the count of days on which the city×day $PM_{2.5}$ concentration exceeds the city- specific 75th percentile during this period is used instead. The first stage specification is given in equation 2. Regressions control for developer and region-by-year-by-quarter fixed effects, the number of public holidays during the workweek, and the leads of the instrumental variables for both the weekend and the period from Thursday to Friday. Further covariates are the number of days with heavy wildfire smoke, and third-order polynomials in average wind speed, precipitation, and relative humidity during both the weekend and the period between Monday and Wednesday. Temperature controls are included in the form of eight bin variables for the period Monday to Wednesday, and in the form of a third-order polynomial for the weekend, and are allowed to vary across regions R. 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

and two sets of weather controls, $\mathbf{w}_{c,w}^{Mo\text{-}We}$ and $\mathbf{w}_{c,w}^{weekend}$, covering the exposure period and the weekend, respectively.[44]

Table 7 presents estimates of coefficient β. Results in Columns (1), (3), and (5) indicate that developers produce significantly more output on weekends if they were exposed to unusually high levels of $PM_{2.5}$ between Monday to Wednesday of the same week (Panel B). In terms of magnitude, one additional day with $PM_{2.5}$ concentration exceeding the city-specific 75th percentile causes an increase in total actions on the weekend by 0.063 or 2.1% of the mean. Effects are positive and significant for both commits and comments, amounting to 1.6% and 2.8% of the mean values, respectively. When we consider average $PM_{2.5}$ concentration between Monday and Wednesday instead, we find positive point estimates, but these are not significantly different from zero. The compensation thus seems to be most relevant after high-

[44] $\mathbf{w}_{c,w}^{Mo\text{-}We}$ and $\mathbf{w}_{c,w}^{weekend}$ both comprise the number of days with heavy wildfire smoke exposure as well as third-order polynomials in average precipitation, relative humidity, and wind speed during the respective period. $\mathbf{w}_{c,w}^{Mo\text{-}We}$ further includes eight variables counting the number of days on which daily mean temperature is falling into the temperature bins described above. $\mathbf{w}_{c,w}^{weekend}$ includes a third-order polynomial in average temperature.

pollution days, which is consistent with the finding that unusually high pollution levels result in disproportionately large declines in output quantity on the day of exposure.

In Columns (2), (4), and (6), we repeat the same analysis, but using only developer×week observations with low pollution levels on the weekend, i.e., levels below the city-specific 75th percentile on both days. We find substantially larger coefficients, indicating that developers reallocate work from low to high productivity periods, i.e., weekends without air pollution-induced productivity shocks.

To put these effects into perspective, we can compare the effect magnitudes with the estimates in Table 2 depicting the reductions in daily output due to same-day $PM_{2.5}$ exposure. Additional work on the weekend makes up for 33% and 57% of the reduction in commits and total actions due to $PM_{2.5}$ exceeding the 75th percentile, respectively.[45]

To check that the estimates in Table 7 do not pick up any effects of unobservable confounders, but indeed reflect a behavioral response of developers to pollution-induced productivity shocks, we conduct a falsification test. We shift both the weekend and the exposure period forward by four days. The placebo weekend comprises Wednesday and Thursday and the placebo exposure period ranges from Friday to Sunday of the week before. Since activity levels are low on weekends, productivity shocks on these days should not induce compensation during the following week. Moreover, activity is already high on Wednesday and Thursday, such that there is not much scope for additional work. Hence, we expect no significant effects of $PM_{2.5}$ exposure. Appendix Table A.11 presents the results which confirm this hypothesis. Effects are neither significant in the full sample, nor when considering only placebo weekends with low pollution levels.

In summary, we find that developers work more on weekends to catch up on coding tasks not completed due to pollution-induced productivity declines during the workweek.[46] The opportunity for compensation might allow them to end work early on high-pollution work days. This reallocation option could thus also contribute to the absence of effects on work quality in this setting. If developers can end work when their health or cognitive capacity deteriorates and they face an increased risk to commit errors, this will mitigate impacts of pollution on work quality. At the same time, sacrificing leisure time on the weekend, when it is likely most valuable, implies a welfare cost and potentially adverse effects on the work-life-balance.

[45] In the case of comments, we find positive effects on weekend activity, even though there are no significant reductions in comments due to pollution exposure. When developers work on the weekend because they are behind on coding tasks, they might decide to also conduct some interactive actions given that they are active on GitHub anyways, even though there was no negative effect of pollution exposure in this domain. This might explain the positive effect for comments and the large coefficient for total actions relative to the direct effect.

[46] This result is similar to the finding by Hoffmann and Rud (2022) that workers in Mexico City reallocate labor supply across days in response to changes in $PM_{2.5}$. However, the authors find strong extensive margin effects and show that reallocation likely serves as a strategy to *avoid* pollution exposure and its adverse health impacts. In our setting focused on a global sample of high-skill tech workers reallocation is likely rather a response to low productivity during high-pollution periods.

Overall, worker adaptation likely plays an important role in explaining the modest effects of $PM_{2.5}$ on output. By focusing on easier tasks and reallocating work from high-pollution, low-productivity to low-pollution, high-productivity periods, developers alleviate the impact of the environmental shock.

6 Heterogeneity and Further Results

Our main results are based on a linear measure of $PM_{2.5}$ concentration and an indicator for unusually high levels. In this section, we exploit the large variation in air quality in our international sample to investigate whether effects on output arise across the full range of concentrations, and how they vary in intensity. Furthermore, we analyze effect heterogeneity based on location characteristics, repeat our analysis at the monthly level, and conduct several robustness checks.

Non-Linearity. To analyze the shape of the dose-response function between $PM_{2.5}$ and output quantity, we replace $PM_{c,d}$ in equation (1) with a series of dummy variables indicating whether $PM_{2.5}$ concentration falls into a specific bin.[47] We estimate the model by OLS on the extended sample of 220 cities. Since we cannot rely on exogenous variation in air quality due to wind direction in this analysis, we opt for a more conservative specification with stricter fixed effects for region×date and city×month. These absorb (i) region-wide shocks to developer output on a given date that might be correlated with $PM_{2.5}$ concentration and (ii) seasonal fluctuations in activity and air quality which are allowed to vary across cities.[48] Given the finding that the OLS results underestimate the true effects, we need to bear in mind that all results should be interpreted as reflecting lower bounds.

Figure 8 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 16 and 22 $\mu g/m^3$ to the respective bin. The x-axis measures the average concentration within the bins. The baseline bin is chosen to ensure that for each city some observations fall into this range. For perspective, it falls below the EU limit value for $PM_{2.5}$ during the sample period (25 $\mu g/m^3$), but above the EPA annual standard (12 $\mu g/m^3$). We find significant negative effects starting at a concentration of approximately 75 $\mu g/m^3$, but no significant differences in the outcomes for concentrations between the reference bin and 60 $\mu g/m^3$. Being exposed to a $PM_{2.5}$ level below 5 $\mu g/m^3$ has a significant positive impact on both total actions (point estimate = 0.038, p-value = 0.061) and commits (point estimate = 0.020, p-value = 0.031). This implies that even

[47] Bins are defined for ≤ 5, (5-6], (6-8], (8-10], (10-13], (13,16], (16,22], (22,25], (25,28], (28,31], (31,35], (35,40], (40,50], (50,60], (60, 70], (70, 85], (85,100], (100, 160] and > 160, all in $\mu g/m^3$, with (16-22] as reference bin.

[48] OLS results from this model with either $PM_{2.5}$ in $\mu g/m^3$, or the binary indicator for $PM_{2.5}$ above the city-specific 75th percentile as regressors are presented Columns 4 to 6 of Tables A.7 and A.8 for the main sample and the extended sample, respectively.

in cities with low to moderate levels of $PM_{2.5}$, further improvements in air quality will generate positive effects on worker productivity.

Figure 8: Non-linear effects of $PM_{2.5}$ on Work Quantity (OLS estimates)

Note: Plot depicts point estimates on different bins of $PM_{2.5}$ concentrations from an OLS regressions of total actions (left) and commits (right), respectively, on indicators for each bin. Covariates: Weather and holiday controls as in Equation 1, region×date and city×month fixed effects. X-axis: Average $PM_{2.5}$ concentration in each bin in $\mu g/m^3$. Shaded areas indicate 95%- and 90%-confidence intervals.

For both outcomes, the average slope of the function is larger than the estimate found in the linear OLS specification, especially at low $PM_{2.5}$ concentrations. To zoom in on the different parts of the function, we split the sample into terciles based on cities' average pollution concentration. For each subsample we define seven bin variables for $PM_{2.5}$ such that each bin includes the same number of city×date observations. The reference category is given by the lowest bin. Figure 9 presents the results, focusing on total actions for the sake of exposition.

In the subsample of cities in the bottom tercile, mean $PM_{2.5}$ concentration ranges between 5.1 $\mu g/m^3$ and 8.6 $\mu g/m^3$. Most of the distribution falls below current regulatory thresholds.[49] Moving from the lowest bin to higher concentrations generates significant negative effects on output. The implied slope is -0.0044, i.e., much steeper than the estimate from the linear specification on the full sample, and even exceeds the size of the 2SLS estimate. In the middle tercile, by contrast, the estimates imply a flat slope. Average $PM_{2.5}$ concentration in this sample ranges between 8.7 $\mu g/m^3$ and 13.0 $\mu g/m^3$, and it includes mostly cities in the European Union and the US. Among the most polluted cities in the upper tercile we again find a negative slope, but less steep than in the low pollution subsample. This subsample includes most Asian and Eastern European cities, as well as some cities in Western Europe, with mean concentration

[49] This subsample includes most cities in Australia, New Zealand, Scandinavia, and Canada and more than half the cities in the US.

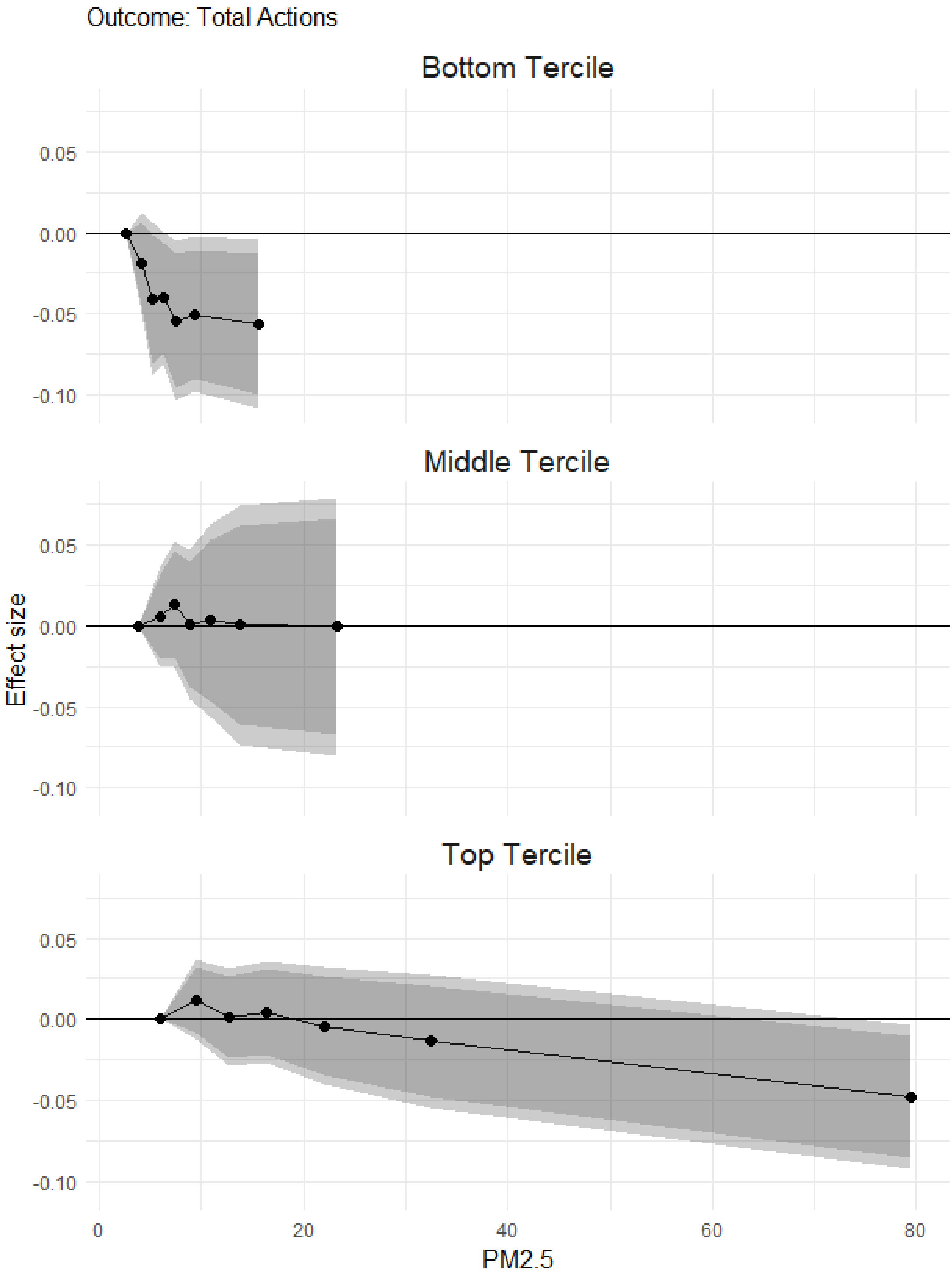


Figure 9: Non-linear effects of $PM_{2.5}$ on Work Quantity across subsamples based on average $PM_{2.5}$

Note: The Figure depicts point estimates on different bins of $PM_{2.5}$ concentrations from OLS regressions of total actions on indicators for each bin for three distinct samples. Cities are assigned into subsamples based on average $PM_{2.5}$ concentration. Covariates: Weather and holiday controls as in Equation 1, region×date and city×month fixed effects. X-axis: Average $PM_{2.5}$ concentration in each bin in $\mu g/m^3$. Shaded areas indicate 95%- and 90%-confidence intervals.

ranging between 13 μg/m^3 and 133 μg/m^3.

We are unable to pin down what drives the differences in the dose-response function across samples. They might arise because of differences in the extent of measurement error or omitted variable bias. Moreover, individuals in high-pollution cities might engage more strongly in avoidance behavior and the use of protective devices such as air purifiers.[50] An important result, however, is that $PM_{2.5}$ exposure exerts adverse effects on productivity even below relatively strict current regulatory thresholds like those by the U.S. EPA. Given that the OLS estimates likely underestimate the true effects of $PM_{2.5}$ exposure, the results imply relevant economic benefits from complying with the stricter WHO standard for $PM_{2.5}$.

Effect Heterogeneity. Next, we analyze heterogeneity in the effect of fine particulate matter on work quantity by location characteristics in order to shed light on the distribution of air pollution damages and on potential mechanisms driving the adverse productivity effects.

We start by analyzing how the effect of an increase in $PM_{2.5}$ differs between places with low vs. high average pollution levels. To this end, we compute the average $PM_{2.5}$ concentration for each first-stage city group g and form two subsamples comprising city-groups with below and above median average $PM_{2.5}$ levels. We assign city-groups to the two subsamples instead of single cities to ensure that the IV approach does not capture impacts of local pollution transport such that we cleanly identify the causal effects of interest.

Average $PM_{2.5}$ concentration is 7.9 and 18.6 μg/m^3 in the two subsamples. Panel A of Table 8 presents the estimated effects on total actions and commits. We find larger point estimates in the low pollution sample. The impact on total actions, however, is not significant at conventional levels, likely due to the reduced sample size. This confirms the result from the OLS estimation of the dose-response function which also suggests stronger impacts at lower pollution levels. As mentioned above, more frequent use of air purifiers and other protective measures in high-pollution places might explain this result. Within the US, a similar pattern has been found by Bishop et al. (forthcoming) who analyze the impact of $PM_{2.5}$ on dementia.

Secondly, we investigate differences in effect magnitude between places with relatively high vs. low income levels. We collect data on GDP per capita in 2014, the first year of our sample period from the OECD, World Bank, and national statistical offices.[51] As before, we compute average values at the city-group level and assign groups into either the above or below median subsample. Average GDP per capita amounts to $38,400 in the low-income

[50] There is anecdotal evidence that big (tech) companies equip their offices in highly polluted places with air purifiers and filters, e.g., Microsoft, Google, SAP and Coca-Cola in Delhi or Nokia in Beijing.

[51] 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., 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. Values are converted by the purchasing power parity conversion factor to adjust for differences in local price levels.

Table 8: Effect Heterogeneity

	Actions (1)	*Commits* (2)	*Actions* (3)	*Commits* (4)
Panel A.	*Below Median $PM_{2.5}$*		*Above Median $PM_{2.5}$*	
$PM_{2.5}$	-0.0055	-0.0040^{**}	-0.0027^{**}	-0.0024^{**}
	(0.0037)	(0.0016)	(0.0011)	(0.0009)
Observations	179,220	179,220	174,225	174,225
First Stage F-Stat.	107.4	107.4	102.4	102.4
Mean Dep. Var.	2.96	1.31	2.54	1.28
Mean $PM_{2.5}$	7.9	7.9	18.6	18.6
Panel B.	*Above Median GDP per capita*		*Below Median GDP per capita*	
$PM_{2.5}$	-0.0031	-0.0025^{**}	-0.0030^{***}	-0.0024^{**}
	(0.0022)	(0.0011)	(0.0011)	(0.0010)
Observations	173,371	173,371	180,074	180,074
First Stage F-Stat.	88.9	88.9	113.4	113.4
Mean Dep. Var.	2.99	1.33	2.43	1.23
GDP	71,008	71,008	38,409	38,409
Mean $PM_{2.5}$	9.4	9.4	17.9	17.9

Note: Estimated coefficients reflect 2SLS estimates of the parameter β in Equation (1) for four distinct samples. The two samples used in Panel A are constructed by comparing average $PM_{2.5}$ concentration in each first stage city group g to the median value. The two samples used in Panel B are constructed by comparing average GDP per capita in 2014 in each first-stage city-group g to the median value. Data on per capita GDP is collected from the OECD, World Bank, and national statistical offices. The first stage specification is given in Equation (2). 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, and year-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.

sample, and \$71,000 in the high-income sample. Results are reported in Panel B of Table 8. We find that the point estimates for total actions and commits are of very similar magnitude in the two subsamples. In relative terms, the effects are marginally stronger in the high-income subsample.

Since GDP and air quality are systematically correlated, the two heterogeneity analyses cannot identify distinct drivers of effect magnitude.[52] The main takeaway message from the two heterogeneity analyses is that we find no evidence that the effects of air pollution on high-skilled worker productivity are more severe in disadvantaged locations. If anything, effects are stronger in places with better air quality. Health impacts of particulate pollution, on the other hand, are typically found to be larger in lower income and high-pollution locations (Colmer et al., 2021; Hsiang et al., 2019). A potential explanation for why we do not find the same pattern could be that we consider only individuals in a high-paying occupation.

Next, we investigate whether effect magnitude differs between places with low vs. high awareness of air pollution as an important issue. We use data from the Pew Research Center International Science Survey which was conducted in early 2020 across 20 countries with several thousand interviewees per country. Survey participants were asked whether they believe that air pollution is a big, a moderate, a small, or no problem at all in their country. As a country-wide measure of awareness, we compute the share of respondents stating that air pollution is a big problem. Appendix Figure B.4 shows the distribution of this variable across the 14 countries that are included in both our data and the survey and how it varies with average $PM_{2.5}$ concentration. 152 cities in our main sample are covered by the survey data. We split these cities into three groups with low, intermediate, and high awareness.[53] Table 9 presents 2SLS estimates for the effect of $PM_{2.5}$ concentration on total actions for the total sample covered by the survey data, and the three subsamples. Effects are negative and significant across all samples, and importantly there is no clear gradient in awareness. In fact, the point estimate is identical in the high and the low awareness samples. This suggests that the reduction in output is not driven by avoidance behavior, e.g., working from home on high pollution days which reduces productivity. In this case, we would expect to see larger effects in the high awareness sample.

Lastly, we investigate effect heterogeneity based on the quality of the local building stock. Effective exposure to particulate matter is likely lower for individuals inside modern buildings with low penetration rates than for individuals in older, lower-quality buildings given the same outdoor concentration. We use data on the construction period of residential dwellings as a proxy for building stock quality. We collect data on building stock age from different

[52] In Appendix Figure B.3, we present the distribution of average $PM_{2.5}$ concentration in the rich vs. poor city groups and the distribution of GDP per capita in the clean vs. polluted city groups. Lower-income locations on average have higher pollution levels.

[53] US cities form the intermediate awareness sample. Among US respondents, 63.1% believe that air pollution is a big problem. All countries where a larger (smaller) share of respondents holds this view, are assigned to the high (low) awareness subsample.

Table 9: Heterogeneity: Awareness and Building Stock Age

	Actions			
Panel A. Awareness				
	Total	*High Awareness*	*Intermediate Awareness*	*Low Awareness*
$PM_{2.5}$	-0.0043^{***}	-0.0035^{**}	-0.0083^{*}	-0.0035^{*}
	(0.0012)	(0.0014)	(0.0042)	(0.0020)
Observations	280,297	80,963	120,521	78,813
First Stage F-Stat.	98	90	67	115
Share AP is Big Problem	67.4%	78.1%	63.1%	51.6%
Mean $PM_{2.5}$	11.0	16.2	8.7	9.5
Mean Dep. Var.	2.9	2.6	2.9	3.1
Panel B. Building Stock Age				
	Total	*Above Median Old Building Share*	*Below Median Old Building Share*	
$PM_{2.5}$	-0.0039^{**}	-0.0044^{***}	-0.0025	
	(0.0015)	(0.0016)	(0.0031)	
Observations	300,844	167,307	133,537	
First Stage F-Stat.	102	124	93	
Share modern buildings	28%	19%	38%	
Share old buildings	44%	55%	32%	
Mean $PM_{2.5}$	10.4	9.7	10.9	
Mean Dep. Var.	2.92	2.98	2.90	

Note: Estimated coefficients reflect 2SLS estimates of the parameter β in Equation (1) where the outcome variable is the number of completed actions. Each Column is estimated on a different sample. In Panel A, the sample used in Column (1) includes all 153 cities covered by the Pew Research Center International Science Survey. Results in Columns (2) to (4) are estimated on subsamples formed based on country-level awareness of air pollution, measured by the share of respondents stating that air pollution is a big problem in the Pew Survey. In Panel B, the sample used in Column (1) includes all 169 cities covered by data on building stock age. Results in Columns (2) to (3) are estimated on subsamples formed based on the city-group level share of dwellings built before 1970, which are defined as old buildings. Modern buildings are those build after 1990. The first stage specification is given in Equation (2). Covariates are as described in Table 8. 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.

national statistical offices, covering 164 out of the 193 cities in our main sample.[54] For each first-stage cluster, we compute the average share of dwellings built before 1970, i.e., the share of relatively old buildings which likely have high indoor penetration rates. As before, we then assign groups into subsamples based on whether the group level share of old buildings is above or below the median value. Panel B of Table 9 presents regression results for the full sample covered by the building stock data, as well as the two subsamples with above and below median share of old buildings. We find that the negative effect of $PM_{2.5}$ on total actions is driven by the sample with a relatively high share of old dwellings, whereas the point estimate is not statistically significant and less than 60% as large in the subsample with relatively few old buildings. The fact that effects are larger in places where effective exposure is likely higher suggests that the main results are driven by physiological effects of air pollution, rather than by behavioral changes or avoidance behavior.

Monthly Level. Next, we quantify the effect of $PM_{2.5}$ on output at a more aggregate time period, by estimating a model at the developer×month level. This is motivated by the findings that pollution exposure has not only contemporaneous, but also some lagged effects on output, and that developers partially compensate for the adverse productivity shocks by working more on weekends. We analyze effects on the same output quantity measures used before (actions, commits, and comments). In addition, we consider effects on the growth rate of the number of developers' followers. We view this as a proxy for the quantity, quality, and relevance of a developer's work on GitHub because all these dimensions likely affect the decision of other GitHub users about whether to follow the developer or not.

To explore effects at the monthly level, we need to adapt the IV strategy. We use three variables measuring the share of days in month m with wind direction falling into a specific 90° bin, each interacted with indicators for the first-stage city-groups g, as instruments.

Table 10 presents the results. An increase in monthly $PM_{2.5}$ concentration by $1\,\mu g/m^3$ reduces the number of actions performed in that month by 0.17 or 0.21% of the mean. The implied effect of an increase in $PM_{2.5}$ by one $\mu g/m^3$ on a single day is .0057, and thus slightly larger than the effect found in the analysis at the daily level (.0032). Again, this effect is mostly driven by a reduction in commits, which fall by 0.31% of the sample mean, while the reduction in comments is small and insignificant. $PM_{2.5}$ also negatively affects the growth rate of the number of followers. The estimate implies a decrease of 1.5% relative to the mean rate. In sum, exposure to air pollution has negative impacts on developers' output also over a more aggregate time period. It even slows down the process of gaining reputation in the tech community,

[54] The data is collected from the American Community Survey for metropolitan areas in the US, the EU Building Stock Observatory for EU member states (country-level), the Federal Statistical Office of Switzerland (canton-level), the Statistics Bureau of Japan (prefecture-level), Statistics Canada (province-level), and Statistics Norway (municipality-level).

which might have adverse long-run consequences for developers' career paths.[55]

Table 10: Analysis at the Monthly Level

	Actions (1)	*Commits* (2)	*Comments* (3)	*Growth Rate(Followers)* (4)
$PM_{2.5}$ (monthly)	-0.173^{**}	-0.125^{***}	-0.033	-0.00011^{***}
	(0.076)	(0.035)	(0.045)	(0.00003)
	[0.024]	[0.0005]	[0.474]	[0.0004]
F-Statistics	644	644	644	636
Observations	469,373	469,373	469,373	453,443
Mean Dep. Var.	84.3	39.3	28.3	.0072

Note: The table presents IV estimates of the effect of monthly $PM_{2.5}$ concentration on the outcomes described at the top of the Table. The excluded instruments are variables measuring the share of days in the month on which wind direction was blowing from one of three 90° angles, interacted with indicators for first-stage city groups g. Regressions control for developer and region-by-year-by-month fixed effects, third-order polynomials in average monthly temperature, precipitation, relative humidity, and wind speed, the number of holidays and days with heavy wildfire smoke at the city×month level. Temperature controls and effects of holidays are allowed to vary across regions R. 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

Robustness Checks. In Appendix Tables A.12 to A.13 we show that our main results are not sensitive to specific choices on how we set up the first and second stage models. We find evidence for a reduction in output quantity—driven by fewer commits, but no or very small changes in comments—and a switch towards easier issues and PRs in response to $PM_{2.5}$ across specifications.

First, we examine robustness to the specification of the wind direction instruments. Instead of $\sin(\theta_{c,d})$ and $\sin(\theta_{c,d}/2)$, 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 Panels A and B of Table A.12.

In Panels C and D, we report results from a specification where we used a k-means clustering algorithm, instead of hierarchical 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 and task choice are very similar to the baseline results.

Secondly, we test robustness to the functional form chosen in the second stage model. Table A.13 shows the estimated effects of $PM_{2.5}$ when work output is measured by the inverse hyperbolic sine transformation of total actions, commits, and comments, respectively (Panels A and B). Again, the direction and statistical significance of the baseline results persist, but this

[55]Inspecting GitHub pages of potential employees is a common practice in hiring decisions in the tech sector as described in several tech blogs, see e.g., https://techbeacon.com/app-dev-testing/what-do-job-seeking-developers-need-their-github or https://blog.boot.dev/jobs/build-github-profile/.

specification implies somewhat smaller effect magnitudes. Panel C displays results when $PM_{2.5}$ in logs is used as regressor. This yields a high F-Statistic and the same pattern for second-stage effects on work quantity and task complexity as the baseline model.

In Table A.14 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.

Lastly, we demonstrate that the results are overall robust to changes in the included fixed effects and weather conditions. Tables A.15 to A.18 show that across specifications with different fixed effects absorbing common time shocks at different geographic and temporal levels, and across specifications with more and less detailed weather controls, our main results hold.

Extended Sample. In Table A.19 we show that our core results also hold in the extended sample including all 220 cities in our dataset and with instruments based on temperature inversions instead of wind direction. Specifically, we estimate the model in Equation 1, but use a variable measuring inversion strength (as specified in Section 3) interacted with indicator variables for geographic regions r as instruments. We allow effects of inversions to vary geographically because the strength of the first stage effect varies based on baseline emissions (Krebs and Luechinger, 2021). We form 15 regions r to make sure that each region comprises multiple cities and forms a homogenous geographic area.[56] In Appendix Table A.19, we present results for the outcomes measuring output quantity (Columns 1 to 3) and the outcomes measuring whether developers switch to less complex tasks (Columns 4 to 6). Apart from the effect on the share of issue events referring to an easy issue, which is small and insignificant, results replicate the patterns we found in our main analysis and are of comparable magnitude.

7 Conclusion

How do environmental conditions, like fluctuations in air pollution, affect workers in jobs that form the backbone of the modern knowledge economy? These jobs are focused on interpersonal and analytic tasks, often require strong social and digital skills, and are organized in a way that gives workers flexibility in schedules and task choices. As digitization and automation continue to change the world of work, these job characteristics are expected to become even more widespread.

In this paper, we use detailed data from GitHub to study how particulate matter affects daily output and work patterns in a global sample of software developers—a high-skilled occupation that can be considered as representative of the jobs of interest.

We provide evidence that pollution exposure reduces developer output. On a day with

[56]These groups are the four US census regions, Canada, China, India, South East Asia, Japan and South Korea, Australia, New Zealand, Western, Eastern, Southern, and Northern Europe.

unusually high pollution ($PM_{2.5}$ concentration above the city-specific 75th percentile) the total number of actions conducted by software developers falls by 4% relative to days with better air quality. This effect is mostly driven by a reduction in *individual* coding activity, while the level of *collaborative* activity is unaffected. Our estimates are at the lower end of air pollution effects found in other, less flexible and less collaborative occupations studied in previous research. Moreover, we find no evidence of a deterioration in output quality. Due to the high value generated by software developers, the implied monetary loss is nevertheless economically relevant and comparable to findings for workers in manual occupations. Our estimates imply that on a day with unusually high pollution, output value falls by $11 per developer.

Our second key result is that software developers exploit the flexibility of their work setting to adapt to increases in air pollution. In particular, we find that they choose to work on less complex tasks when $PM_{2.5}$ increases. Among developers who respond with a stronger shift towards easier tasks, effects on output quantity are alleviated. In addition, developers reallocate work activity from high-pollution, low-productivity workdays to low-pollution, high-productivity weekends. One additional day with unusually high $PM_{2.5}$ concentration in the first half of the week causes an increase in weekend work by 2.1%. These forms of adaptation likely explain why the effects on output quantity and quality in our setting are small relative to previous studies. At the same time, they suggest an additional welfare cost of air pollution in this setting not captured by changes in output due to forgone leisure time on the weekend and potential negative impacts on work-life balance.

While we use data on a sample of software developers who use GitHub as part of their professional work, we believe that the findings are externally valid to workers in many other high-paying occupations which offer flexible schedules and discretion in task choice, and require similar skills, e.g., problem-solving skills, attention to detail, programming, and teamwork. This applies to many high-skill workers, including business analysts or researchers. Furthermore, the fact that our data comprises developers across more than 30 countries suggests that the effects we identify and quantify in this study are not specific to a certain firm or country context but apply more generally. Based on this, we can derive estimates of the monetary benefits in terms of productivity gains among knowledge workers from reducing $PM_{2.5}$ concentration permanently by one unit. Extrapolating to all U.S. workers in the occupation group "Computer and Mathematical Workers" and to all ICT professionals in the EU suggests annual benefits of $580m (US) and $980m (EU), respectively.[57]

Hence, our findings have important policy implications. When deciding about limit values

[57] Our 2SLS estimates for the effect of $PM_{2.5}$ concentration on commits and PRs, paired with the estimates of the monetary value of these outcomes, imply that a one unit decrease in $PM_{2.5}$ increases daily output value by $0.344 per developer. Employment in "Computer and Mathematical Workers" in the US in 2021 was 4,654,750 (Bureau of Labor Statistics, 2022). Per year the total estimated gain in output value in this group is thus $0.344/worker and day × 4,654,750 workers × 365 days = $580m, if we assume that the effect of a permanent reduction in $PM_{2.5}$ is given by the sum of the daily effects. We compute the value for the EU analogously, based on an employment figure of 7,843,000 ICT professionals in 2020 (Cedefop, 2022).

on air pollutants, regulators should take the growing evidence on the economic benefits of pollution reductions in the form of productivity gains into account. Importantly, we find that adverse effects of $PM_{2.5}$ on output are large at concentrations below the regulatory standards in force in the European Union and the US. Hence, even in areas with relatively good air quality, further improvements will likely generate additional benefits. While we find slightly smaller marginal effects in high pollution locations, the fact that $PM_{2.5}$ concentration is often an order of magnitude larger in developing countries like India and Bangladesh compared to the US might be an important barrier to growth for the software industries in these countries.

Our findings on how software developers adjust work patterns also have interesting implications for the organization of work within firms: Highlighting the difficulty of certain tasks, as done by the use of issue labels on GitHub, and granting flexibility in working hours, might help workers to better adapt to idiosyncratic productivity shocks and mitigate the total impact on team or firm performance.

References

Adhvaryu, A., N. Kala, and A. Nyshadham (2022): "Management and Shocks to Worker Productivity," *Journal of Political Economy*, 130, 1–47.

Aguilar-Gomez, S., H. Dwyer, J. S. Graff Zivin, and M. J. Neidell (2022): "This is Air: The "Non-Health" Effects of Air Pollution," Working Paper 29848, National Bureau of Economic Research.

Almeida, S., M. Manousakas, E. Diapouli, Z. Kertesz, L. Samek, E. Hristova, K. Šega, R. P. Alvarez, C. Belis, and K. Eleftheriadis (2020): "Ambient particulate matter source apportionment using receptor modelling in European and Central Asia urban areas," *Environmental Pollution*, 266, 115199.

Anderson, J. O., J. G. Thundiyil, and A. I. Stolbach (2011): "Clearing the Air: A Review of the Effects of Particulate Matter Air Pollution on Human Health," *Journal of Medical Toxicology*, 8, 166–175.

Angelici, M. and P. Profeta (2020): "Smart-working: Work flexibility without constraints," Working paper, CESifo.

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.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The skill content of recent technological change: An empirical exploration," *The Quarterly Journal of Economics*, 118, 1279–1333.

Autor, D. H. and B. Price (2013): "The Changing Task Composition of the US Labor Market: An Update of Autor, Levy, and Murnane (2003)," Working paper.

Babadjouni, R. M., D. M. Hodis, R. Radwanski, R. Durazo, A. Patel, Q. Liu, and W. J. Mack (2017): "Clinical effects of air pollution on the central nervous system; a review," *Journal of Clinical Neuroscience*, 43, 16–24.

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

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?" Working Paper 3887079, SSRN.

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

Beckmann, M., T. Cornelissen, and M. Kräkel (2017): "Self-managed working time and employee effort: Theory and evidence," *Journal of Economic Behavior & Organization*, 133, 285–302.

Bishop, K. C., J. D. Ketcham, and N. V. Kuminoff (forthcoming): "Hazed and confused: the effect of air pollution on dementia," *The Review of Economic Studies*.

Borgschulte, M., D. Molitor, and E. Zou (forthcoming): "Air Pollution and the Labor Market: Evidence from Wildfire Smoke," *The Review of Economics and Statistics*.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002): "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence," *The Quarterly Journal of Economics*, 117, 339–376.

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

——— (2022): "Occupational Employment and Wage Statistics," https://www.bls.gov/oes/current/oes_nat.htm, accessed: 2022-11-07.

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 (2022): "Exposures and Behavioral Responses to Wildfire Smoke," *Nature Human Behavior*.

Calderón-Garcidueñas, L., M. Franco-Lira, R. Torres-Jardón, C. Henriquez-Roldán, G. Barragán-Mejía, G. Valencia-Salazar, A. Gonzaléz-Maciel, R. Reynoso-Robles, R. Villarreal-Calderón, and W. Reed (2007): "Pediatric Respiratory and Systemic Effects of Chronic Air Pollution Exposure: Nose, Lung, Heart, and Brain Pathology," *Toxicologic Pathology*, 35, 154–162.

Cedefop (2022): "Employed population by occupation and sector," https://www.cedefop.europa.eu/en/tools/skills-intelligence/employed-population-occupation-and-sector?year=2020&country=EU&occupation=#1, accessed: 2022-11-07.

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.

Colmer, J., D. Lin, S. Liu, and J. Shimshack (2021): "Why are pollution damages lower in developed countries? Insights from high-Income, high-particulate matter Hong Kong," *Journal of Health Economics*, 79, 102511.

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.

Currie, J. and M. Neidell (2005): "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?" *The Quarterly Journal of Economics*, 120, 1003–1030.

Delgado-Saborit, J. M., V. Guercio, A. M. Gowers, G. Shaddick, N. C. Fox, and S. Love (2021): "A critical review of the epidemiological evidence of effects of air pollution on dementia, cognitive function and cognitive decline in adult population," *Science of The Total Environment*, 757, 143734.

Deng, G., Z. Li, Z. Wang, J. Gao, Z. Xu, J. Li, and Z. Wang (2017): "Indoor/outdoor relationship of PM2.5 concentration in typical buildings with and without air cleaning in Beijing," *Indoor and Built Environment*, 26, 60–68.

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.

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

Ebenstein, A., V. Lavy, and S. Roth (2016): "The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution," *American Economic Journal: Applied Economics*, 8, 36–65.

Fu, S., V. B. Viard, and P. Zhang (2021): "Air Pollution and Manufacturing Firm Productivity: Nationwide Estimates for China," *The Economic Journal*, 131, 3241–3273.

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.

Hoek, G., G. Kos, R. Harrison, J. de Hartog, K. Meliefste, H. ten Brink, K. Katsouyanni, A. Karakatsani, M. Lianou, A. Kotronarou, I. Kavouras, J. Pekkanen, M. Vallius, M. Kulmala, A. Puustinen, S. Thomas, C. Meddings, J. Ayres, J. van Wijnen, and K. Hameri (2008): "Indoor–outdoor relationships of particle number and mass in four European cities," *Atmospheric Environment*, 42, 156–169.

Hoffmann, B. and J. P. Rud (2022): "Exposure or Income? The Unequal Effects of Pollution on Daily Labor Supply," Working paper, Inter-American Development Bank.

Hsiang, S., P. Oliva, and R. Walker (2019): "The Distribution of Environmental Damages," *Review of Environmental Economics and Policy*, 13, 83–103.

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

Isen, A., M. Rossin-Slater, and W. R. Walker (2017): "Every Breath You Take—Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970," *Journal of Political Economy*, 125, 848–902.

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 (2020): "Air pollution lowers high skill public sector worker productivity in China," *Environmental Research Letters*, 15, 084003.

Karagulian, F., C. A. Belis, C. F. C. Dora, A. M. Prüss-Ustün, S. Bonjour, H. Adair-Rohani, and M. Amann (2015): "Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level," *Atmospheric Environment*, 120, 475–483.

Kaur, S., S. Mullainathan, S. Oh, and F. Schilbach (2021): "Do Financial Concerns Make Workers Less Productive?" Working Paper 28338, National Bureau of Economic Research.

Kelly, F. J. and J. C. Fussell (2015): "Air pollution and public health: emerging hazards and improved understanding of risk," *Environmental Geochemistry and Health*, 37, 631 – 649.

Krebs, B. and S. Luechinger (2021): "Air Pollution, Cognitive Performance, and the Role of Task Proficiency," Working Paper 3947149, SSRN.

Künn, S., J. Palacios, and N. Pestel (forthcoming): "Indoor Air Quality and Strategic Decision-Making," *Management Science*.

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.

Lederer, A. M., P. M. Fredriksen, B. N. Nkeh-Chungag, F. Everson, H. Strijdom, P. De Boever, and N. Goswami (2021): "Cardiovascular effects of air pollution: current evidence from animal and human studies," *American Journal of Physiology-Heart and Circulatory Physiology*, 320, H1417–H1439, pMID: 33513082.

LoPalo, M. (forthcoming): "Temperature, Worker Productivity, and Adaptation: Evidence from Survey Data Production," *American Economic Journal: Applied Economics*.

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," Working Paper 29598, National Bureau of Economic Research.

Menon, S., A. Salvatori, and W. Zwysen (2020): "The effect of computer use on work discretion and work intensity: Evidence from Europe," *British Journal of Industrial Relations*, 58, 1004–1038.

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.

Muro, M., S. Liu, J. Whiton, and S. Kulkarni (2017): "Digitalization and the American workforce," Working paper, Brookings India.

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.

Pope, C. A. (2000): "Epidemiology of Fine Particulate Air Pollution and Human Health: Biologic Mechanisms and Who's at Risk?" *Environmental Health Perspectives*, 108, 713–723.

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

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

Shepard, E. M. I., T. J. Clifton, and D. Kruse (1996): "Flexible work hours and productivity: Some evidence from the pharmaceutical industry," *Industrial Relations: A Journal of Economy and Society*, 35, 123–139.

Somanathan, E., R. Somanathan, A. Sudarshan, and M. Tewari (2021): "The impact of temperature on productivity and labor supply: Evidence from Indian manufacturing," *Journal of Political Economy*, 129, 1797–1827.

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.

Xu, R., X. Qi, G. Dai, H. Lin, P. Zhai, C. Zhu, L. Wang, and A. Ding (2020): "A Comparison Study of Indoor and Outdoor Air Quality in Nanjing, China," *Aerosol and Air Quality Research*, 20, 2128–2141.

Appendix

A Additional Tables

Table A.1: Characteristics of High-Skill Occupations and Software Development

	Freedom to Make Decisions		Structured versus Unstructured Work		Work With Work Group or Team	
	All high-skill occupations	Software developers	All high-skill occupations	Software develop.	All high-skill occupations	Software develop.
1	0.4	0	0.6	0	1.79	0
2	2.6	3.1	2.3	2.4	4.4	5.9
3	10.5	29.1	11.3	28.1	11.0	2.7
4	35.6	38.2	39.8	45.0	30.5	9.2
5	50.9	29.6	46.0	24.6	52.4	82.3

Note: Based on data from O*NET Database Version 25.0. Work Contexts Table. All high-skill occupations refers to occupations in Job Zones 4 and 5. Software developers refers to occupation 15-1132.00 ("Software Developers, Applications"). Categories: 1 = not important at all/no freedom; 2 = Fairly important/very little freedom; 3 = Important/Limited freedom; 4= Very Important/Some freedom; 5 = Extremely important / A lot of freedom

Table A.2: 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.3: Description of Outcome Variables

Domain	Concept	Variable	Details
Output quantity	Total output quantity	Actions	Sum of number of commits, comments on issues, PRs and commits, PRs opened, PRs closed, issues opened, closed and reopened
	Coding activity	Commits	Number of commits
	Interactive activity	Comments	Sum of number of comments written on issues, PRs and commits
Output Quality	PR Success rate	Share PRs merged	PRs opened that got merged/all PRs opened
	Deficient commits	Share commits reverted	Commits that got reverted/all commits
Task choice	Easy tasks among issue events	Share easy issue events issues	(#easy issues opened + #easy issues closed + #comments written on easy issues)/(#issues opened + #issues closed + #comments written on issues)
	Average PR complexity	Lines added per PR	Average number of lines of code added in PRs opened, closed and commented on
		Files changed per PR	Average number of code files changed in PRs opened, closed and commented on
Working hours	Evening activity	Time last action	Minute of final action of the day
		Share of actions after 6 pm	Actions made after 6 pm/Total actions

Note: The Table displays information on the outcome variables we use, how they are constructed, and what they measure.

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 Dhaka	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
Taiwan	Environmental Protection Administration
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 $PM_{2.5}$. 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 developer-by-date observations across geographic regions R

Region R	Observations	Share
Oceania	273,246	1.9
Northern America	7,244,272	50.6
Northern Europe	1,809,844	12.6
Western Europe	2,312,377	16.1
Southern Europe	333,597	2.3
Eastern Europe	725,664	5.1
Asia	1,628,930	11.4

Note: The table shows the distribution of observations in the developer $\times$ date panel described in section 3.2 across geographic regions R.

Table A.6: Effect of $PM_{2.5}$ on Quantity of Issue and Pull Request Actions

	PRs closed	*PRs opened*	*Issues closed*	*Issues opened*
	(1)	(2)	(3)	(4)
Panel A.				
$PM_{2.5}$	-0.00011	-0.00018^{**}	0.00004	0.00009
	(0.00012)	(0.00008)	(0.00009)	(0.00006)
First Stage F-Stat.	102.1	102.1	102.1	102.1
% change in Y	-0.1	-1.2	0.03	0.1
Panel B.				
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	-0.0049^{*}	-0.0064^{***}	0.0005	-0.0023
	(0.0029)	(0.0023)	(0.0029)	(0.0024)
First Stage F-Stat.	80.5	80.5	80.5	80.5
% change in Y	-2.9	-4.2	0.4	2.2
Mean Dep. Var.	0.17	0.15	0.12	0.11
Observations	353,445	353,445	353,445	353,445

Note: The table presents IV estimates of the parameter β in equation (1). In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Panel B, a binary variable is used instead, which takes a value of one if city×day $PM_{2.5}$ concentration exceeds the city-specific 75th percentile. The first stage specification is given in equation (2). 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, and year-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

Table A.7: OLS Results for Work Quantity (main sample)

	Actions (1)	*Commits* (2)	*Comments* (3)	*Actions* (4)	*Commits* (5)	*Comments* (6)
Panel A.						
$PM_{2.5}$	$-.00024$	$-.0002$	$-.00001$	$-.0006^{*}$	$-.0003$	$-.0002^{*}$
	(.0003)	(.0002)	(.0001)	(.0003)	(.0002)	(.0001)
Panel B.						
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	$-.0158^{*}$	$-.0096^{**}$	$-.0026$	$-.0264^{**}$	$-.0131^{**}$	$-.0054$
	(.0084)	(.0041)	(.0036)	(.0115)	(.0051)	(.0049)
Observations	353,445	353,445	353,445	353,445	353,445	353,445
City FE	✓	✓	✓			
Region×Day-of-Week FE	✓	✓	✓			
Region×Year-Month FE	✓	✓	✓			
Region×Date FE				✓	✓	✓
City×Month FE				✓	✓	✓

Note: The table presents OLS estimates of the parameter β in equation (1), where the dependent variables are displayed in the upper part of the table. Parameters are estimated on the main sample including 193 cities. The regressor of interest is $PM_{2.5}$ concentration in $\mu g/m^3$ in Panel A, and an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile in Panel B. 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. The temperature controls can vary across world regions R. 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.8: OLS Results for Work Quantity (extended sample)

	Actions (1)	*Commits* (2)	*Comments* (3)	*Actions* (4)	*Commits* (5)	*Comments* (6)
Panel A.						
$PM_{2.5}$	$-.0004^{***}$	$-.0002^{**}$	$-.0002^{**}$	$-.0004^{**}$	$-.0002$	$-.0001^{***}$
	(.0001)	(.0001)	(.0001)	(.0002)	(.0001)	(.00004)
Panel B.						
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	$-.0160^{**}$	$-.0087^{**}$	$-.0036$	$-.0234^{**}$	$-.0119^{**}$	$-.0049$
	(.0078)	(.0039)	(.0032)	(0.0102)	(.0046)	(.0042)
Observations	398,687	398,687	398,687	398,687	398,687	398,687
City FE	✓	✓	✓			
Region×Day-of-Week FE	✓	✓	✓			
Region×Year-Month FE	✓	✓	✓			
Region×Date FE				✓	✓	✓
City×Month FE				✓	✓	✓

Note: The table presents OLS estimates of the parameter β in equation (1), where the dependent variables are displayed in the upper part of the table. Parameters are estimated on the extended sample including 220 cities. The regressor of interest is $PM_{2.5}$ concentration in $\mu g/m^3$ in Panel A, and an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile in Panel B. 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. The temperature controls can vary across world regions *R*. 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: Reduced Form

	Actions (1)	*Commits* (2)	*Comments* (3)
High-Pollution	-0.0192^{**}	-0.0111^{***}	-0.0036
Wind Direction	(0.0087)	(0.0040)	(0.0038)
	[0.029]	[0.006]	[0.350]
Observations	367,472	367,472	367,472
First Stage Effect on $PM_{2.5}$		3.683^{***} (0.456)	

Note: The Table 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 (60° angle) that has the largest positive effect on local $PM_{2.5}$ concentration. Standard errors clustered at the city level are reported in parentheses. P-values are presented in squared brackets. All regressions include covariates as described in Table 2 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

Table A.10: Effect of $PM_{2.5}$ on PRs opened and closed with GHArchive and GHTorrent data

	PRs opened (GHA) (1)	*PRs closed (GHA)* (2)	*PRs opened (GHT)* (3)	*PRs closed (GHT)* (4)
Panel A.				
$PM_{2.5}$	-0.00029^{**}	-0.00007	-0.00016^{*}	-0.00011
	(0.00014)	(0.00013)	(0.00009)	(0.00013)
	[0.036]	[0.581]	[0.067]	[0.378]
First Stage F-Stat.	89	89	89	89
Panel B.				
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	-0.00668^{*}	-0.00458	-0.00701^{***}	-0.00586^{**}
	(0.00377)	(0.00297)	(0.00239)	(0.00297)
	[0.078]	[0.124]	[0.004]	[0.050]
First Stage F-Stat.	69	69	69	69
Observations	298,566	298,566	298,566	298,566

Note: The table presents IV estimates of the parameter β in equation (1), where the outcome is the number of pull requests (PRs) opened or closed, respectively. This is measured based on GHArchive data in columns (1) to (2) and GHTorrent data in column (3) to (4). The regressor of interest is $PM_{2.5}$ concentration in $\mu g/m^3$ in Panel A, and an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile in Panel B. The first stage specification is given in equation (2). 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, and year-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.11: Placebo Test: Effect of $PM_{2.5}$ Friday to Sunday on Work Activity Wednesday to Thursday

	Actions		*Commits*		*Comments*	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.						
$PM_{2.5}$	0.0065	0.0055	0.0045	0.0028	0.0006	0.0017
	(0.0092)	(0.0086)	(0.0039)	(0.0041)	(0.0035)	(0.0034)
Panel B.						
High $PM_{2.5}$ Days	0.0401	−0.0082	0.0293	−0.0020	0.0034	−0.0007
	(0.0888)	(0.0795)	(0.0364)	(0.0352)	(0.0346)	(0.0316)
Observations	1,997,123	1,321,642	1,997,123	1,321,642	1,997,123	1,321,642
Weeks	all	only low PM weekends	all	only low PM weekends	all	only low PM weekends

Note: The table presents IV estimates of the parameter β in a placebo version of equation 4. Outcomes are the sum of all actions, commits and comments made between Wednesday and Thursday, the placebo weekend. In Panel A, the regressor of interest is average $PM_{2.5}$ concentration between Friday and Sunday of the week before. In Panel B, the count of of days on which the city×day $PM_{2.5}$ concentration exceeds the city-specific 75th percentile during this period is used instead. The first stage specification is given in equation 2. Regressions control for developer and region-by-year-by-quarter fixed effects, the number of public holidays during the workweek, and the leads of the instrumental variables for both the placebo weekend and the period from Monday to Tuesday. Further covariates are the number of days with heavy wildfire smoke, and third-order polynomials in average wind speed, precipitation, and relative humidity during both the placebo weekend and the exposure period between Friday and Sunday. Temperature controls are included in the form of eight bin variables for the placebo exposure period, and in the form of a third order polynomial for the placebo weekend, and are allowed to vary across regions R. 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.12: Robustness: First Stage Specification

	Actions (1)	*Commits* (2)	*Comments* (3)	*Share Easy Issue Events* (4)	*Lines added per PR* (5)	*Files changed per PR* (6)
Panel A.						
$PM_{2.5}$	-0.0018^{*}	-0.0020^{***}	0.0001	0.0001^{*}	-0.0019^{**}	-0.0012^{***}
	(0.0010)	(0.0007)	(0.0005)	(0.0001)	(0.0009)	(0.0004)
First Stage F-Stat.	62	62	62	52	38	38
Panel B.						
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	-0.0629^{**}	-0.0661^{***}	0.0059	0.0031^{**}	-0.0469^{**}	-0.0285^{***}
	(0.0290)	(0.0128)	(0.0182)	(0.0016)	(0.0201)	(0.0100)
First Stage F-Stat.	49	49	49	41	41	28
IV-Specification	Three wind direction bins					
Clustering	Hierarchical Clustering Algorithm					
Observations	353,445	353,445	353,445	250,376	164,883	164,883
Panel C.						
$PM_{2.5}$	-0.0030^{***}	-0.0024^{***}	-0.0005	0.0001^{**}	-0.0012	-0.0008^{**}
	(0.0011)	(0.0006)	(0.0006)	(0.0001)	(0.0008)	(0.0004)
First Stage F-Stat.	102	102	102	86	62	62
Panel D.						
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	-0.1023^{***}	-0.0739^{***}	-0.0160	0.0029^{**}	-0.0347^{*}	-0.0201^{**}
	(0.0268)	(0.0144)	(0.0178)	(0.0014)	(0.0177)	(0.0095)
First Stage F-Stat.	81	81	81	66	46	46
IV-Specification	$\sin(\theta), \sin(\frac{\theta}{2})$					
Clustering	K-means Clustering Algorithm					
Observations	353,445	353,445	353,445	250,376	164,883	164,883

Note: The table presents IV estimates of the parameter β in Equation (1). In Panels A and C, the regressor of interest is $PM_{2.5}$ concentration in $\mu g/m^3$. In Panel B and D, an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile is used instead. Relative to specifications underlying results in Table 2, the first stage model is changed: In Panels A and B, instruments are three indicators for wind direction falling in specific bins, each covering 90° of the wind rose. In Panels C and D, the first stage specification is as in Equation (2), but we form city-groups g using k-means clustering instead of hierarchical clustering. Covariates as described in Table 2. 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.13: Robustness: Second Stage Specification

	Actions	*Commits*	*Comments*	*Share Easy Issue Events*	*Lines added per PR*	*Files changed per PR*
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Inv. Hyperbolic Sine Transformation						
$PM_{2.5}$	-0.0005^{**}	-0.0005^{**}	-0.0001			
	(0.0002)	(0.0002)	(0.0001)			
First Stage F-Stat.	102	102	102			
Panel B. Inv. Hyperbolic Sine Transformation						
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	-0.0206^{***}	-0.0159^{***}	-0.0082^{**}			
	(0.0060)	(0.0043)	(0.0039)			
First Stage F-Stat.	80	80	80			
Panel C. log(PM)						
$log(PM_{2.5})$	-0.0656^{***}	-0.0444^{***}	-0.0141	0.0018^{**}	-0.0209^{**}	-0.0143^{***}
	(0.0163)	(0.0087)	(0.0090)	(0.0009)	(0.0096)	(0.0052)
First Stage F-Stat.	197	197	197	162	117	117
Observations	353,445	353,445	353,445	250,376	164,883	164,883

Note: The table presents IV estimates of the parameter β in Equation (1). In Panel A, the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Panel B, an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile is used instead. In Panel C, the regressor is the logarithm of $PM_{2.5}$ concentration. Inverse hyperbolic sine transformations are applied to outcomes in Panels A and B. The first stage specification is given in Equation (2). Covariates as described in Table 2. 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.14: Robustness: Clustering of Standard Errors

	Actions (1)	*Commits* (2)	*Comments* (3)	*Share Easy Issue Events* (4)	*Lines added per PR* (5)	*Files changed per PR* (6)
Panel A.						
$PM_{2.5}$	-0.0032^{***}	-0.0026^{***}	-0.0005	0.0001^{**}	-0.0013	-0.0010^{**}
	(0.0012)	(0.0007)	(0.0006)	(0.0001)	(0.0008)	(0.0004)
	[0.009]	[0.0003]	[0.421]	[0.031]	[0.110]	[0.031]
Panel B.						
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	-0.1104^{***}	-0.0801^{***}	-0.0169	0.0032^{**}	-0.0403^{**}	-0.0244^{**}
	(0.0301)	(0.0154)	(0.0189)	(0.0013)	(0.0171)	(0.0092)
	[0.001]	[0.000004]	[0.379]	[0.017]	[0.023]	[0.011]
Observations	353,445	353,445	353,445	250,376	164,883	164,883

Note: The Table presents IV estimates of the parameter β in Equation (1). In Panels A, the regressor of interest is $PM_{2.5}$ concentration in $\mu g/m^3$. In Panel B, an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile is used instead. The first stage specification is given in Equation (2). 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 2 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

Table A.15: Robustness to Changes in Weather controls (Output Quantity)

	Actions (1)	*Commits* (2)	*Comments* (3)	*Actions* (4)	*Commits* (5)	*Comments* (6)
Panel A.						
	-0.0041^{***}	-0.0027^{***}	-0.0009^{**}	-0.1156^{***}	-0.0749^{***}	-0.0261^{**}
	(0.0010)	(0.0007)	(0.0004)	(0.0269)	(0.0150)	(0.0115)
First Stage F-Stat.		147			110	
Weather Controls	none					
Panel B.						
	-0.0029^{***}	-0.0022^{***}	-0.0005	-0.1162^{***}	-0.0767^{***}	-0.0243
	(0.0010)	(0.0007)	(0.0005)	(0.0316)	(0.0159)	(0.0153)
First Stage F-Stat.		108			83	
Weather Controls	Quadratic functions of precipitation, wind speed, rel. humidity and cubic, continent-specific function of mean temperature					
Panel C.						
	-0.0018	-0.0021^{**}	0.0001	-0.0539	-0.0601^{***}	0.0069
	(0.0012)	(0.0009)	(0.0006)	(0.0345)	(0.0190)	(0.0190)
First Stage F-Stat.		87			67	
Weather Controls	Continent specific bins for precipitation, wind speed, rel. humidity, minimum and maximum temperature					
Observations	353,445	353,445	353,445	353,445	353,445	353,445

Note: The table presents IV estimates of the parameter β in Equation (1). Dependent variables are denoted at the top of the table. In Columns(1) to (3), the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Columns (4) to (6), an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile is used instead. Relative to specifications underlying results in Table 2, we change the included covariates to control for weather conditions. We state the included variables at the bottom of each Panel. The first stage specification is given in Equation (2). All regressions include fixed effects as described in Table 2 and 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.16: Robustness to Changes in Fixed Effects (Output Quantity)

	Actions (1)	*Commits* (2)	*Comments* (3)	*Actions* (4)	*Commits* (5)	*Comments* (6)
Panel A.						
	-0.0044^{***}	-0.0031^{***}	-0.0009	-0.1438^{***}	-0.0954^{***}	-0.0288
	(0.0013)	(0.0009)	(0.0006)	(0.0337)	(0.0169)	(0.0189)
First Stage F-Stat.		82			66	
Fixed Effects	city, region × day-of-week, region × week					
Panel B.						
	-0.0045^{***}	-0.0034^{***}	-0.0008	-0.1323^{***}	-0.0995^{***}	-0.0170
	(0.0014)	(0.0009)	(0.0007)	(0.0399)	(0.0207)	(0.0215)
First Stage F-Stat.		84			62	
Fixed Effects	city, region × date					
Panel C.						
	-0.0032^{**}	-0.0021^{***}	-0.0007	-0.1018^{***}	-0.0603^{***}	-0.0279^{**}
	(0.0013)	(0.0008)	(0.0006)	(0.0257)	(0.0142)	(0.0127)
First Stage F-Stat.		101			76	
Fixed Effects	city × month, region × day-of-week, region × year × month					
Panel D.						
	-0.0048^{**}	-0.0028^{**}	-0.0013^{*}	-0.1311^{***}	-0.0711^{***}	-0.0403^{***}
	(0.0019)	(0.0011)	(0.0008)	(0.0319)	(0.0181)	(0.0139)
First Stage F-Stat.		78			61	
Fixed Effects	city × month, region × day-of-week, region × week					
	-0.0023^{*}	-0.0012^{*}	-0.0008	-0.0762^{***}	-0.0369^{***}	-0.0283^{*}
	(0.0013)	(0.0006)	(0.0007)	(0.0289)	(0.0138)	(0.0160)
First Stage F-Stat.		114			86	
Fixed Effects	region × day-of-week, city × year, city × month					
Observations	353,445	353,445	353,445	353,445	353,445	353,445

Note: The table presents IV estimates of the parameter β in Equation (1). Dependent variables are denoted at the top of the table. In Columns(1) to (3), the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Columns (4) to (6), an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile is used instead. Relative to specifications underlying results in Table 2, we change the included fixed effects. We state the included fixed effects at the bottom of each Panel. The first stage specification is given in Equation (2). All regressions include control variables as described in Table 2 and 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.17: Robustness to Changes in Weather Controls (Task Complexity)

	Share Easy Issue Events (1)	*Lines added per PR* (2)	*Files changed per PR* (3)	*Share Easy Issue Events* (4)	*Lines added per PR* (5)	*Files changed per PR* (6)
Panel A.						
	0.0001^{*}	-0.0017^{**}	-0.0010^{***}	0.0021^{*}	-0.0412^{**}	-0.0224^{***}
	(0.0001)	(0.0007)	(0.0003)	(0.0012)	(0.0170)	(0.0083)
First Stage F-Stat.	120	88	88	90	63	63
Weather Controls	none					
Panel B.						
	0.0001^{*}	-0.0014^{*}	-0.0010^{**}	0.0027^{*}	-0.0374^{**}	-0.0244^{**}
	(0.0001)	(0.0008)	(0.0004)	(0.0014)	(0.0184)	(0.0096)
First Stage F-Stat.	90	66	66	68	47	47
Weather Controls	Quadratic functions of precipitation, wind speed, rel. humidity and cubic, region-specific function of mean temperature					
Panel C.						
	0.0002^{**}	-0.0012	-0.0009^{*}	0.0036^{**}	-0.0416^{*}	-0.0251^{*}
	(0.0001)	(0.0010)	(0.0005)	(0.0016)	(0.0222)	(0.0131)
First Stage F-Stat.	72	52	52	55	38	38
Weather Controls	Region-specific bins for precipitation, wind speed, rel. humidity, minimum and maximum temperature					
Observations	250,376	164,883	164,883	250,376	164,883	164,883

Note: The table presents IV estimates of the parameter β in Equation (1). Dependent variables are denoted at the top of the table. In Columns(1) to (3), the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Columns (4) to (6), an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile is used instead. Relative to specifications underlying results in Table 2, we change the included covariates to control for weather conditions. We state the included variables at the bottom of each Panel. The first stage specification is given in Equation (2). All regressions include fixed effects as described in Table 2 and 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.18: Robustness to Changes in Fixed Effects (Task Complexity)

	Share Easy Issue Events (1)	*Lines added per PR* (2)	*Files changed per PR* (3)	*Share Easy Issue Events* (4)	*Lines added per PR* (5)	*Files changed per PR* (6)
Panel A.						
	0.0002^{**}	-0.0011	-0.0009^{*}	0.0039^{**}	-0.0368^{**}	-0.0231^{**}
	(0.0001)	(0.0009)	(0.0004)	(0.0016)	(0.0185)	(0.0112)
First Stage F-Stat.	70	50	50	54	38	38
Fixed Effects	city, region × day-of-week, region × week					
Panel B.						
	0.0002^{*}	-0.0007	-0.0006	0.0046^{***}	-0.0275	-0.0156
	(0.0001)	(0.0009)	(0.0004)	(0.0018)	(0.0187)	(0.0113)
First Stage F-Stat.	70	53	53	49	35	35
Fixed Effects	city, region × date					
Panel C.						
	0.0001^{**}	-0.0019^{**}	-0.0011^{**}	0.0033^{**}	-0.0376^{**}	-0.0183^{**}
	(0.0001)	(0.0008)	(0.0005)	(0.0013)	(0.0178)	(0.0091)
First Stage F-Stat.	85	62	62	64	45	45
Fixed Effects	city × month, region × day-of-week, region × year × month					
Panel D.						
	0.0002^{**}	-0.0017^{**}	-0.0010^{*}	0.0040^{***}	-0.0304^{*}	-0.0148
	(0.0001)	(0.0008)	(0.0006)	(0.0015)	(0.0177)	(0.0104)
First Stage F-Stat.	67	49	49	51	36	36
Fixed Effects	city × month, region × day-of-week, region × week					
Panel E.						
	0.0001^{**}	-0.0021^{**}	-0.0010^{*}	0.0029^{**}	-0.0434^{**}	-0.0196^{*}
	(0.0001)	(0.0009)	(0.0005)	(0.0012)	(0.0199)	(0.0103)
First Stage F-Stat.	96	70	70	71	49	49
Fixed Effects	region × day-of-week, city × year, city × month					
Observations	250,376	164,883	164,883	250,376	164,883	164,883

Note: The table presents IV estimates of the parameter β in Equation (1). In Columns(1) to (3), the regressor of interest is $PM_{2.5}$ concentration measured in $\mu g/m^3$. In Columns (4) to (6), an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile is used instead. Relative to specifications underlying results in Table 2, we change the fixed effects. We state the included fixed effects at the bottom of each Panel. The first stage specification is given in Equation (2). All regressions include control variables as described in Table 2 and 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

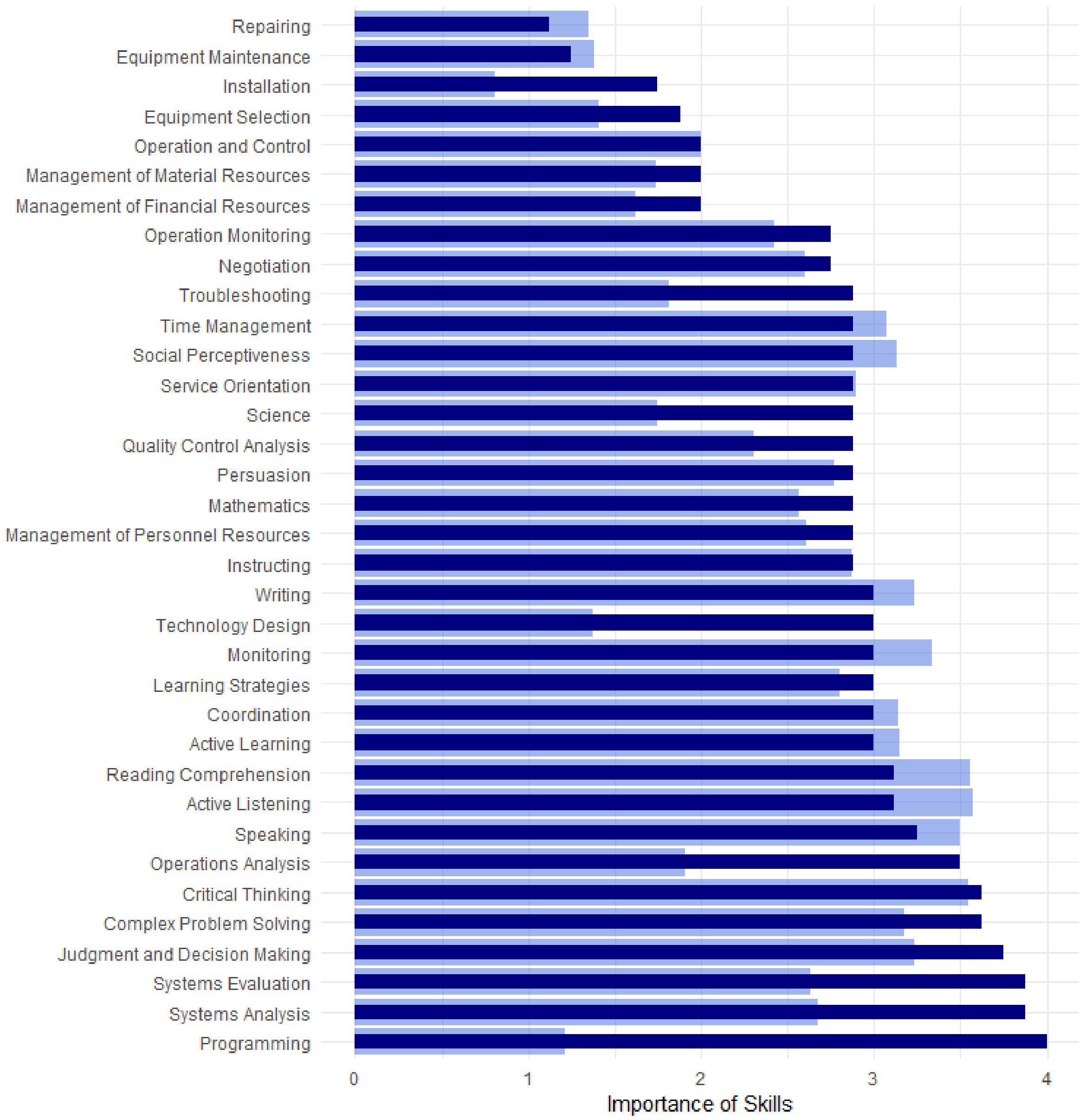
Table A.19: Effects of $PM_{2.5}$ in the Extended Sample using Inversions as IV

	Actions (1)	*Commits* (2)	*Comments* (3)	*Share Easy Issue Events* (4)	*Lines added per PR* (5)	*Files changed per PR* (6)
Panel A.						
$PM_{2.5}$	-0.0032^{***}	-0.0017^{***}	-0.0009^{***}	0.00004	-0.0023^{***}	-0.0009^{***}
	(0.0009)	(0.0006)	(0.0003)	(0.00003)	(0.0005)	(0.0003)
First Stage F-Stat.	300	300	300	301	231	231
Panel B.						
$\mathbb{1}\{PM_{2.5} > Q_{0.75}\}$	-0.1035^{**}	-0.0575^{***}	-0.0316	0.0001	-0.0383^{**}	-0.0199^{**}
	(0.0470)	(0.0196)	(0.0217)	(0.0012)	(0.0173)	(0.0092)
First Stage F-Stat.	440	440	440	385	292	292
Observations	398,687	398,687	398,687	281,985	187,935	187,935

Note: The Table presents IV estimates of the parameter β in Equation (1). In Panels A, the regressor of interest is $PM_{2.5}$ concentration in $\mu g/m^3$. In Panel B, an indicator for $PM_{2.5}$ concentration exceeding the city-specific 75th percentile is used instead. The excluded instruments in the first stage are interactions between a measure of inversion strength as specified in 3 and dummies indicating the geographic region a city is located in. Standard errors clustered at the city level are reported in parentheses. All regressions include covariates as described in Table 2 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

B Additional Figures

Figure B.1: Skill Requirements in High-Skill Occupations and Software Development



Note: Based on data from O*NET Database Version 25.0. Skills Table. Light blue bars reflect average importance of the respective skill across all high-skill occupations, i.e. occupations in Job Zones 4 and 5. Dark blue bars reflect importance of the respective skill among software developers, i.e. occupation 15-1132.00 ("Software Developers, Applications").

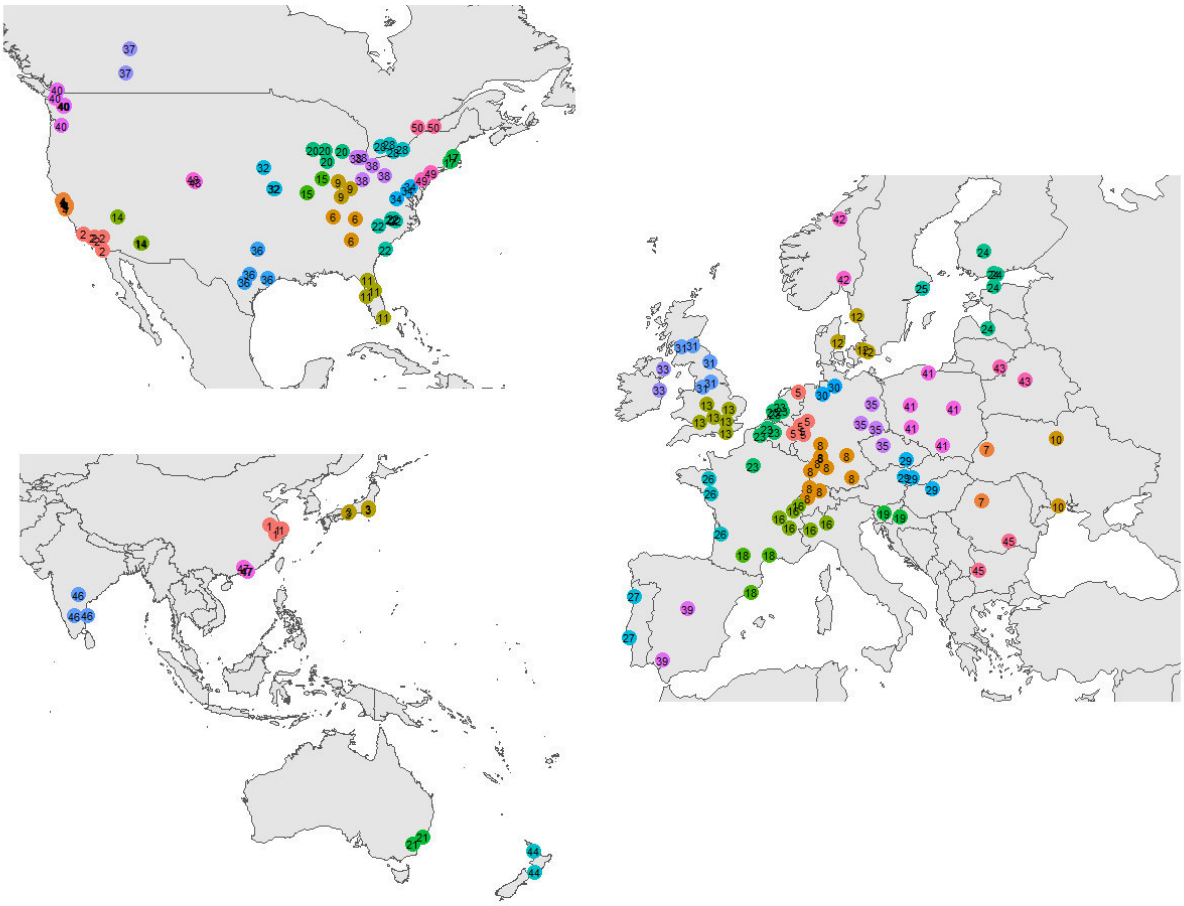
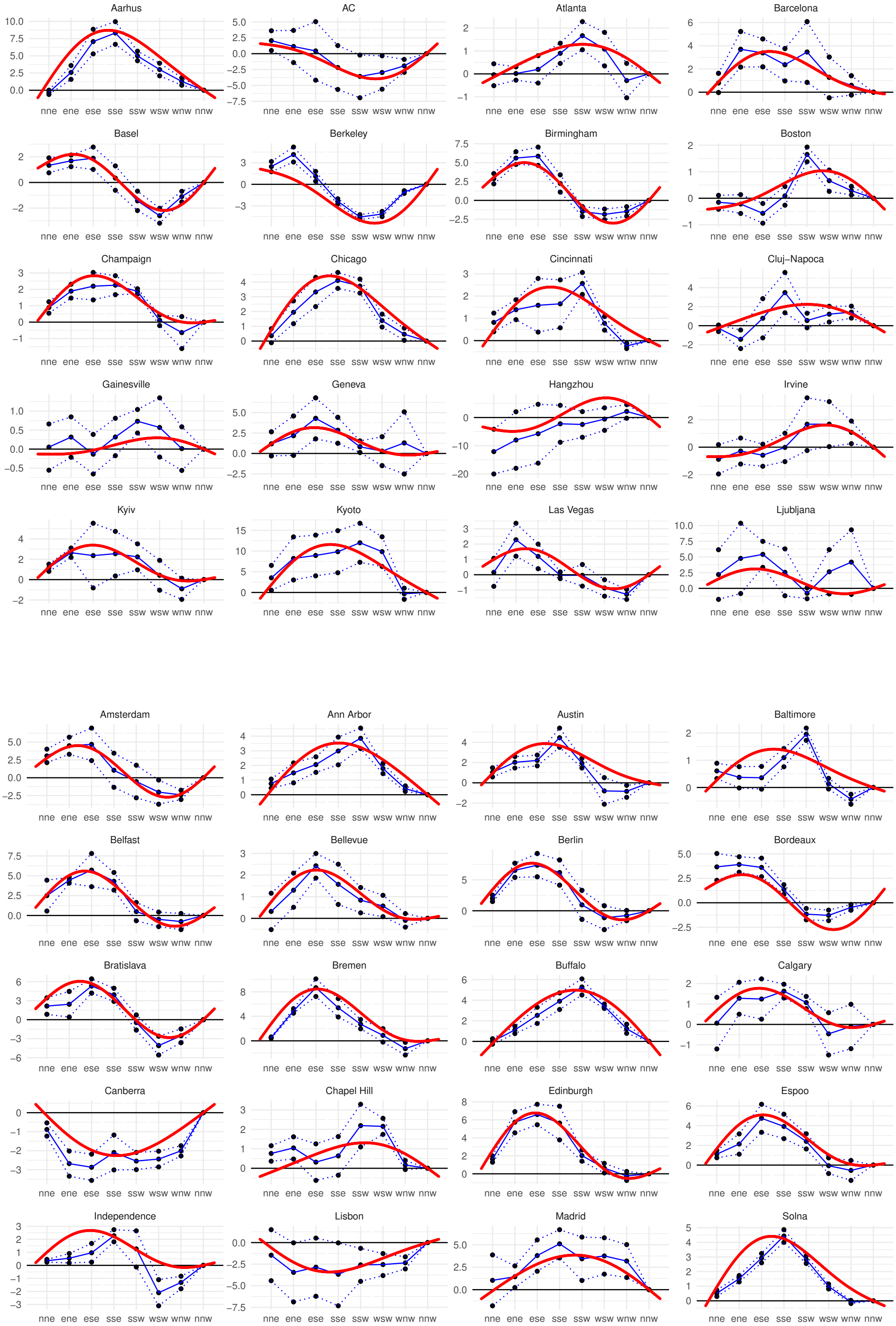
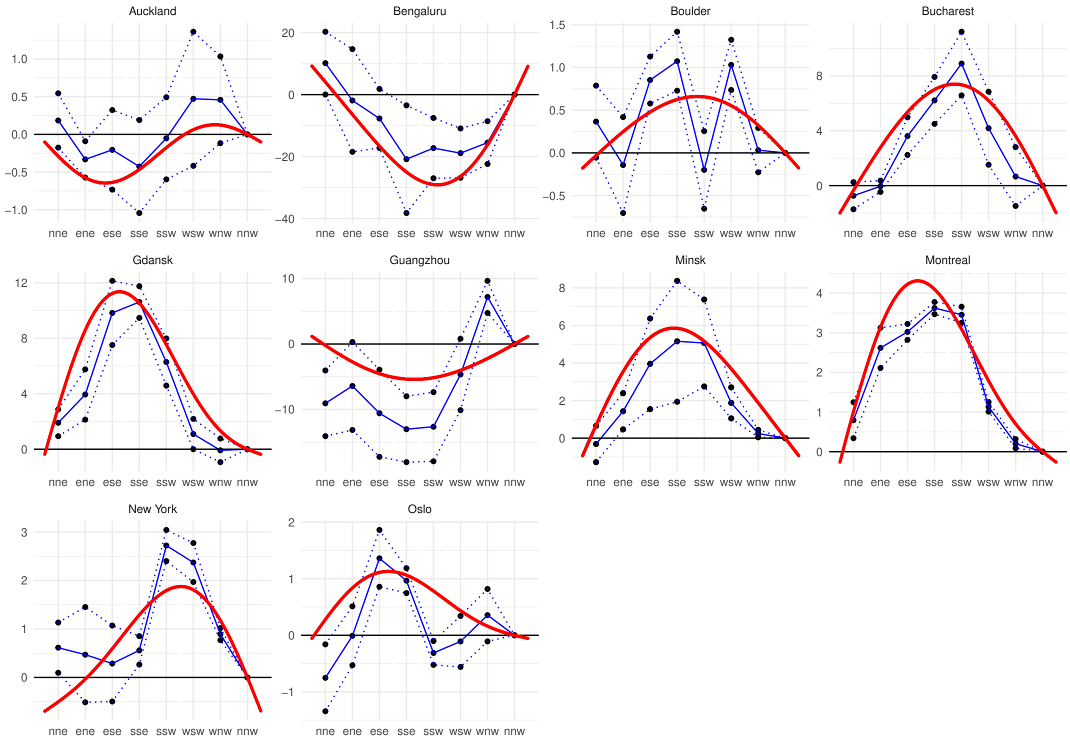


Figure B.2: Illustration of first stage city groups g

Note: Maps show our sample cities. The color of and 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 4, especially equation (2)).

Figure B.3: First stage for all 50 city groups





Notes: Plots present estimated coefficients from regressions of $PM_{2.5}$ measured in $\mu g/m^3$ on wind direction for each first stage city group as depicted in B.2. Solid blue 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. Red line: estimated relationship when wind direction is parameterized as the sine of wind direction in radians and wind direction in radians divided by two. Plot titles denote one city from the respective group.

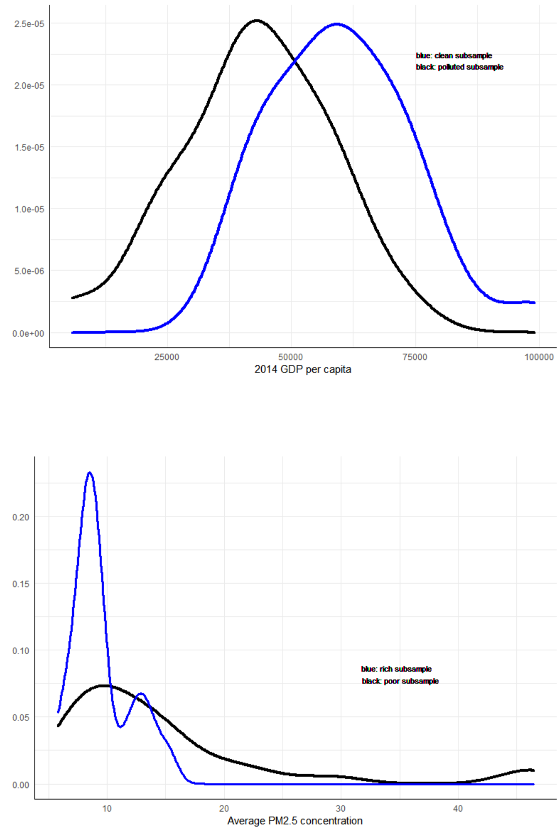


Figure B.3: GDP per capita and average $PM_{2.5}$ concentration

Note: The left plot shows the distribution of 2014 GDP per capita across city-groups g separately for the high and low pollution subsample used in the heteorgeneity analysis in Section 5. Blue: low pollution, black: high pollution.The right plot depicts the distribution of average $PM_{2.5}$ concentration separately for low and high income city-groups. Blue: high income, black: low income.

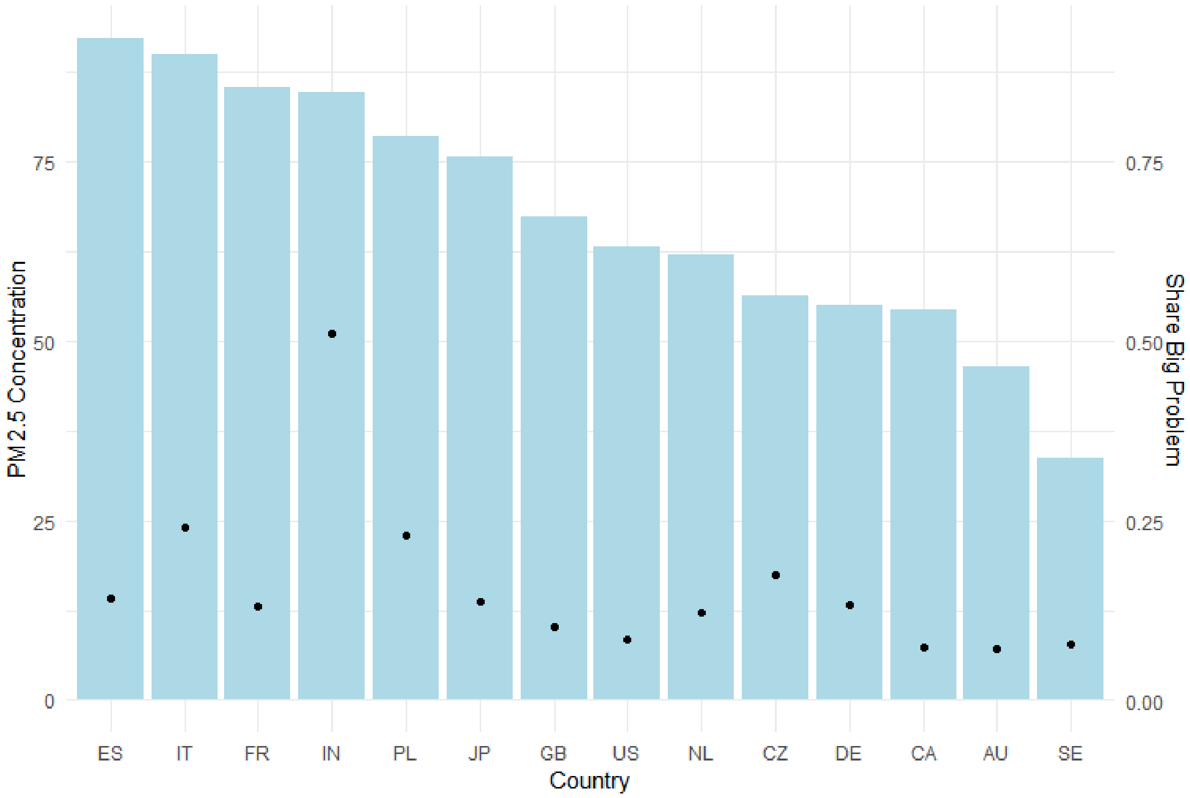


Figure B.4: $PM_{2.5}$ and Awareness by Country

Note: Bars depict the share of respondents from the respective country stating that air pollution is a big problem in their country, based on the Pew Research Center Science Survey (2020). Black dots reflect average $PM_{2.5}$ concentration in the cities within the respective country which are included in our data.

C Gitcoin

This section provides additional details regarding the data collected from Gitcoin to assess the monetary value of output produced on GitHub and to validate some of our productivity and task complexity outcomes.

We collect data on 292 Gitcoin transactions via the Gitcoin API, including the type of the posted issue (bug, documentation, improvement, feature, or other), the expected issue difficulty as assessed by the issue funders (beginner, intermediate, or advanced), the URL to the PR solving the issue and awarded the payment, the value of the payment in USD, and the number of hours worked on the PR as stated by the PR author. 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. Using the URL of the PR, we combine this with information on pull request size obtained via the GitHub API, i.e. the number of commits it comprises, the number of lines of code added and deleted, and the number of files changed. This is possible because all Gitcoin issues and PRs are created in public GitHub repos and thus visible to us. In this context, a pull request reflects the complete work on a certain issue. Commits can be interpreted as single work steps in completing this task.

Combining the data on the amount of coding work done and on the payment made we can estimate the monetary value of output produced in public GitHub repos. 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*. In the full sample, it amounts to \$112. The mean time input per commit also exhibits a steep gradient with respect to difficulty: It is 1 hour at the beginner level, but 5.3 hours at the advanced level.

To validate the use of the number of commits per day as one of our core measures of developer productivity, we analyze how the number of commits in a PR correlates with the payment awarded and the time spent on it in the Gitcoin sample.

Table C.1 depicts 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 alternatively with repository fixed effects (columns 3 and 6). The omitted difficulty category is *advanced*. 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.

In Table C.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.

To validate our proxies for PR complexity, we run the specifications from columns 4 to 6 of Table C.1 again, but add the number of files changed in the PR and the logarithm of lines of code added as additional regressors. Results are presented in Table C.3. Holding the number of commits constant, adding more lines of code and changing more files is associated with a higher payment, suggesting that these variables indeed reflect task complexity.

Table C.1: Validity Check: Number of Commits and 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		✓			✓	
Issue type		✓			✓	
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 (columns 1 to 3) or the logarithm thereof (columns 4 to 6). 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

Table C.2: Validity Check: Number of Commits and Hours Worked on a PR

	hours worked$_i$			
	(1)	(2)	(3)	(4)
$commits_i$	0.375^{***}	0.939^{***}		
	(.132)	(.341)		
$\log(commits_i)$			2.375^{***}	10.346^{***}
			(.648)	(3.535)
$\times\ \mathbb{1}\{Difficulty_i = Beginner\}$		-0.882^{**}		-9.748^{***}
		(.354)		(3.574)
$\times\ \mathbb{1}\{Difficulty_i = Intermediate\}$		-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 C.3: Validity check: PR complexity and Gitcoin payments

	$log(payment_i)$			
	(1)	(2)	(3)	(4)
$log(commits_i)$	0.143**	0.136**	0.070	0.145**
	(0.067)	(0.068)	(0.058)	(0.056)
$files\ changed_i$	0.005	0.007*	0.011***	0.004
	(0.005)	(0.004)	(0.004)	(0.004)
$log(lines\ added_i)$	0.152***	0.112***	0.091***	0.150***
	(0.036)	(0.035)	(0.028)	(0.033)
$easy\ label_i$				−0.348**
				(0.173)
Year dummies	✓	✓	✓	✓
Issue difficulty dummies		✓		
Issue type dummies		✓		
Repository fixed effects			✓	
Observations	292	274	292	270

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 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. Column 4 instead adds a dummy variable taking a value of one if the issue addressed by the PR carries a label that we classify as indicating an easy issue. 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

D Auxiliary Regressions

For estimating equation (1), measures of the output of each individual developer i are aggregated to the city-day level. Instead of forming simple averages, we take into account additional information at the developer level. This is done by estimating auxiliary regression, a common approach in this literature (e.g. Currie et al., 2015). In a first step, we estimate regressions for outcome y of developer i living in city c on day d of the following kind.

$$y_{i,c,d} = \mu_i + \mathbf{x}'_{i,d}\pi + \psi_{c,d} + \varepsilon_{i,c,d} \tag{D.1}$$

Here, $y_{i,c,d}$ denotes one of the measures of developer output, task choice, or working hours. The fixed effect μ_i captures time-invariant unobserved factors at the developer level. Including these is important as the composition of developers changes over time. A developer's experience is controlled for by $\mathbf{x}_{i,t}$, a vector of indicators for time since registration on GitHub, where each indicator represents a time span of three months. Additionally, equation (1) includes city-day fixed effects. Their estimates $\widehat{\psi}_{c,d}$ give the average outcome for a city-day after controlling for experience and composition effects. These estimates replace the dependent variable in equation (1).

This approach is computationally less costly and asymptotically equivalent to directly estimating the regressions at the individual developer level (Donald and Lang, 2007). We take into account the sample variance by weighting all regressions by the number of underlying developer observations in each city–day cell (cf. Currie and Neidell, 2005; Isen et al., 2017).