The London School of Economics and Political Science

Essays in Environmental Economics and Development Economics

Azhar Hussain

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy.

London, June 2025

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of approximately 25,000 words.

Statement of conjoint work

I confirm that Chapter 2 is jointly coauthored with Tim Besley, and I have contributed 50% to this work. I also confirm that Chapter 3 is jointly coauthored with Javad Shamsi and Ranjana Sinha, and I have contributed 34% to this work.

Acknowledgements

I first and foremost owe sincere gratitude to my supervisors Tim Besley, Robin Burgess, Gharad Bryan, and Daniel Sturm for their invaluable guidance and unwavering support throughout my PhD. They were extremely generous with their time and consistently offered feedback at every stge of this journey. They will continue to be my inspiration not only for doing research but also for supporting and mentoring junior colleagues. I am also grateful to Clare Balboni for her continuous advice and insightful comments.

This thesis further benefited from conversations with Sam Asher, Michael Callen, Arun Chandrasekhar, Don Davis, Matthias Doepke, Dave Donaldson, James Fenske, Ludovica Gazze, Michael Greenstone, Vernon Henderson, Allan Hsiao, Namrata Kala, Guy Michaels, Torsten Persson, James Rising, and Javad Shamsi. Seminar participants at Barcelona School of Economics, British Academy, Massachusetts Institute of Technology, Monash University, ISEC - Bengaluru, KAPSARC School of Public Policy, London School of Economics, NYU Abu Dhabi, Paris Dauphine University, Stockholm School of Economics, University of Belfast, University College London, University of Essex, University of Glasgow, University of Houston, University of Nottingham, University of Oxford, University of St Andrews, University of Warwick, and Wageningen University further improved this work.

I am thankful for the excellent research assistance of Viva Awasthi, Tanvi Bansal, Jake Fazzio, Joseph Marshall, and Ushmita Pareek. Financial support from British Academy, Private Enterprise Development in Low-Income Countries, and STICERD is gratefully acknowledged. I am indebted to my colleagues at LSE for making my PhD journey such a wonderful experience that I would always cherish. I thank especially Philip Barteska, Shania Bhalotia, Alix Bonargent, Tim Dobermann, Arnaud Dyèvre, Andres Fajardo Ramirez, Ananya Kotia, Rishabh Malhotra, Virginia Minni, Sidharth Moktan, Michelle

Rao, Nilesh Raut, Yannick Schindler, Javad Shamsi, Caterina Soto Vieira, Ameek Singh, Hamza Syed, Abhijit Tagade, and Anshuman Tiwari.

I extend my sincere thanks to Sadia Ali, Lubala Chibwe, Jide Gbadamosi, Tosin Lamidi, Annie-Rose Nicholas, Hitesh Patel, Lakmini Staskus, Emma Taverner, Nick Warner, Andy Wilson, and Irina Zaraisky for their support with administration and IT tasks.

The unwavering support of my wonderful family has been invaluable in allowing me to pursue a career path that may not always be easily understood from the outside. I am truly grateful to my parents, Fatima and Arshad, and my parents-in-law, Sajid and Farha, for their constant encouragement and support. I also wish to thank my brothers, Asfar, Anwar, and Asdar, my sisters, Shaheen and Mehar, my brother-in-law, Danish, my sister-in-law, Sara, and my nieces, Aabia and Faiza, for being a source of comfort throughout this journey. My cousin, Faiz and my friend, Govind were always available to hear my thoughts and offer their perspective. My deepest gratitude goes to Zoya, my greatest source of strength throughout this journey and every aspect of life, and to Azhaan, whose arrival added a touch of chaos and a far greater measure of joy to the final stage of this thesis.

Abstract

This thesis examines how climate-related risks and responses ranging from natural disasters to green transition and public mobilization, shape economic behavior, perceptions, and political outcomes. It highlights how climate risks and social responses shape both market outcomes and politics, with implications for designing effective climate policy.

The first chapter examines the economic effects of flooding in Indonesia, focusing on its impact on firm behavior and regional economic variables. Using granular firm-level data and spatial data on historical floods, I estimate the short-run effects of flood events and find that more severe floods significantly reduce aggregate output, hinder business formation, and erode firm-level capital. A model of firm entry under climate risk reveals that these effects are largely driven by perceived flood risk, thereby emphasizing the anticipatory behavior of firms and the importance of flood mitigation infrastructure, such as flood defenses.

The second chapter addresses the environmental implications of coal-fired power plants. Combining geocoded survey data from 51 countries with plant-level information, we document that individuals living within 40 km of coal plants are systematically more dissatisfied with local air quality. Employing equivalent variation measure, we show that replacing coal plants with renewable technology is feasible even if the gains from improved air quality are only considered in the benefits calculations.

The third chapter explores the political consequences of climate-related protests. Analyzing large-scale protests, such as the Fridays for Future, we find that protests significantly increase climate awareness in the short run, as evidenced by spikes in Google search trends and media coverage. These shifts also translate into increased electoral support for Green parties in Europe and influence political discourse within the UK Parliament.

Contents

List of Figures

List of Tables

Chapter 1

Economic Consequences of Flooding in Indonesia

1.1 Introduction

This chapter examines the effects of flooding on firms in a low- and middle-income country, Indonesia. Given that the Global South is expected to experience heightened direct and indirect consequences of climate change (Cruz and Rossi-Hansberg 2021), the manifestation of such changes through extreme weather events is increasingly evident (IPCC 2023). Therefore, it is crucial to investigate how such events affect firms in different regions to better understand the response of production activities amid climate change. The insights gained through this investigation would be instrumental in designing appropriate adaptation strategies, such as reforming industrial zoning policies, for the changing world.

There is a negative contemporaneous relationship between flooding and measures of regional economic activity. When a region is hit by a flood, characterized by its spatial coverage and temporal extent, it shows an immediate reduction in aggregate value-added, a measure of economic output, and capital stock and labor employment. This negative relationship is partly driven by the reduced entry of firms in flood-affected regions. At

the firm level, flooding has a more pronounced negative effect on capital stock and positive effect on temporary labor hiring. Although this analysis sheds light on the costs associated with floods, it is unclear how floods interact with firm-level decision-making. In particular, due to the persistent nature of flood shocks in areas with high economic activity within Indonesia, the flood measure partially captures the evolving flood risk that firms perceive.

To address this, I introduce a quantitative framework where firms' input choices are influenced by different aspects of flooding. Specifically, firms choose their capital levels before flood shocks, taking flood risk into account, whereas labor is chosen afterward, once the actual flood shock has realized. The framework offers a novel microfoundation for understanding how perceived flood risk and actual flood shocks interact with firm behavior. The main findings indicate that perceived flood risk, rather than the occurrence of actual flooding, plays a more significant role in influencing firms' input decisions. I also conduct a counterfactual analysis in the spirit of building flood defenses to secure flood-prone areas. Building flood defenses has a direct positive impact on aggregate output, as protected areas become less vulnerable to flooding. However, these gains are partially offset by the entry of less productive firms into the now-safer areas, along with an upward pressure on equilibrium wages due to increased competition for scarce labor inputs.

The case in point is Indonesia, the world's tenth largest economy, which has maintained a high disaster risk profile mainly due to catastrophic flooding and accelerated sea-level rise affecting its major economic centers (World Risk Report 2023). As Figure 1.1 shows, flood events frequently affect areas in the southern islands of Java and Sumatra, where a disproportionately large share ($> 90\%$) of manufacturing firms are located. This persistence of flood shocks potentially drives firms to update their perceived flood risk over time. Furthermore, as Figure 1.2 indicates, flooding is not just a recurring challenge but an escalating threat over time in Indonesia. Floods are the single most catastrophic natural disaster in terms of economic damage and human loss that Indone-

sia faces today (Government of Republic of Indonesia 2007). This confluence of factors makes Indonesia a prime case for examining the impacts of flooding.

In the first part of the chapter, I provide reduced-form evidence on the contemporaneous effects of flooding on economic variables at both aggregate and firm levels using a static difference-in-differences research design. I propose and develop a regional flood index based on the spatial and temporal expanse of flooding. Since the index is continuous, I can analyze the effects by varying the intensity of flooding. I find that a 90th percentile flood leads to 20%, 25%, and 15% declines in aggregate value-added, capital stock, and labor employment, respectively, at the sector-region level. Estimating the same relationship at the firm level suggests that a 90th percentile flood is associated with a 5.7% reduction in the value of capital stock for a typical firm, with significant heterogeneity across sectors, more capital-intensive sectors, such as Iron and Steel, show more significant decline. Notably, the association of floods with firm exit is limited, while a firm's decision to enter a sector within a region is affected by floods in that region. Specifically, a 90th percentile flood is associated with 20% less firm entry at the sector-region level in the year of flooding. Given the spatial concentration of firms in areas prone to persistent flood shocks, these effects are partly driven by firms' evolving perceptions of flood risk.

The reduced-form findings yield estimates that capture both the actual damages from flooding and the adjustments that firms make in response to their evolving perception of flood risk over time. However, these findings do not provide a framework for understanding the mechanisms through which flooding influences firm behavior. In particular, capital installation decisions are typically made well in advance of the realization of flood shocks, and firms would anticipate these shocks and choose capital accordingly. In this context, the second part of the chapter introduces a quantitative framework with flood risk and endogenous entry decisions to study the anticipatory effects of flooding. The time-varying parameters governing the regional flood exposure are estimated using the empirical distribution of firm-level production capacity utilization, which is negatively

affected by flooding. The model builds on the seminal work by Lucas (1978) on understanding the impact of managerial talent on the distribution of firms. The foundational elements of the model find roots in more recent *misallocation* research, such as Hsieh and Klenow (2009) and Besley, Roland, and Reenen (2020), particularly using a general equilibrium framework to analyze firm behavior. The model transcends further by integrating firm entry and exit along the lines of Hopenhayn (1992).

Firms use a production technology that combines capital and labor inputs with firms' idiosyncratic productivity to produce output. Decisions regarding the amount of capital to install are made before the realization of flood shocks and take into account the uncertainty surrounding its utilization based on the flood risks associated with the firms' locations. Labor adjusts flexibly after the realization of flood shocks; however, it is indirectly influenced by flood risk through prior capital investment decisions. Risk-neutral firms maximize expected profits, where the expectations are based on the share of capital that can be utilized in a given year. This results in time-varying flood risk that differs across regions and sectors, acting as aggregate misallocation force that impacts capital allocation and ultimately output in equilibrium. Additionally, firms exhibit variations in their idiosyncratic productivity levels, which remain constant over time and are drawn from a common regional distribution. To enter a market, firms must incur a one-time fixed cost, making their entry decision contingent upon expected profits net of this fixed cost. This creates a productivity cutoff below which firms opt not to enter certain sectors within a region. This cutoff productivity, combined with labor market clearing, determines the mass of firms, their allocations, and the equilibrium wages in the market.

I estimate key model parameters to conduct quantitative analysis of the equilibrium. The novel regional shape parameters of the distribution of the share of capital utilized in a given year are estimated using the empirical distribution of production capacity utilization across firms located in a region. Production capacity utilization is the percentage of actual production over the planned production by a firm in a given year. Firms in flood-affected regions report lower production capacity utilization. Based on this finding, the

share of capital that can be utilized in a given year is proxied by its production capacity utilization. Using such an objective measure that remains unaffected by prices and other short-run equilibrium adjustments ensures a more accurate assessment of the impact of flooding on firms. Additionally, the parameters governing the regional component of flood risk are strongly correlated with the empirical flood index used in the reduced-form analysis. I further estimate sector-specific production function parameters and the parameters that govern the distribution of firm productivity using standard methods from the literature.

In the analysis section, I start by disentangling the effects of flood risk and flood shocks on firm behavior. I employ firm-level equilibrium conditions for optimal capital and labor allocation to create a linear specification that can be estimated using ordinary least squares. The results indicate that firms reduce their capital investment and increase labor hiring in response to flood risk. In contrast, the impact of flood shocks, which affect equilibrium input allocations directly, is found to be limited. As an experimental counterfactual exercise in the spirit of flood defense systems used across the world, I compare observed outcomes with those generated after bringing the top 20th percentile of flood-prone regencies to the median value of the distribution by constructing flood defenses there. This intervention benefits all sectors and regions, but the benefits are larger for more capital-intensive sectors. The direct impact of the intervention increases annual aggregate output across sectors (regions) by 7% (16%). However, allowing for the entry of new firms reduces the gains by almost half, as less productive firms are now able to enter into these safer areas. The influx of new firms consumes scarce production resources, intensifying competition and ultimately driving wages upward. This underscores the potential downsides of such costly protective investments that are prevalent worldwide, particularly in low- and middle-income countries.

The chapter relates to a strand of literature that studies the impact of climate change and natural disasters on the distribution of economic activity within and across regions (see, for example, Castro-Vincenzi 2024; Balboni 2024; Hsiao 2024; Nath 2024; Bilal

and Rossi-Hansberg 2023; Desmet et al. 2021; Jia, Ma, and Xie 2022; Kocornik-Mina et al. 2020; Cruz and Rossi-Hansberg 2021; Balboni, Boehm, and Waseem 2023). I contribute to this literature in various ways. First, I employ a continuous measure of regional flooding that allows me to establish the relationship between flooding and economic variables at different flood intensities. Second, I provide a microfoundation for understanding how flooding affects firm decision-making by integrating the former within the firms' production function. I demonstrate how firms in low- and middle-income countries adjust their production inputs in response to threats posed by flooding. Specifically, they reduce investment in capital stock and increase temporary labor hiring, a more readily available resource in these settings, thus highlighting both production resilience and a form of adaptation to deal with such disruptions. Third, I develop a time-varying measure of flood risk, which is *perceived* by the firms. The literature has mostly employed time-invariant measures of flood risk, which is informed through atmospheric and hydrology models.[1] However, firms' decision-making is responsive to the actual flood events, and thus the perceived flood risk should evolve over time. Fourth, I examine the sectoral heterogeneity in the impact of flooding on both the intensive and extensive margins. On the methodological side, the chapter is related to the literature on firm dynamics, misallocation, and their aggregate productivity effects (see, for example, Hsieh and Klenow 2009; Besley, Roland, and Reenen 2020; Midrigan and Xu 2014; Bento and Restuccia 2017; Bartelsman, Haltiwanger, and Scarpetta 2013; Hopenhayn 2014; Gopinath et al. 2017; Restuccia and Rogerson 2008; Banerjee and Duflo 2005)). I contribute to this literature by applying the methodology in the context of flooding, which generates distortions for firm-level capital decisions. I also integrate firm entry and exit dynamics along the lines of Hopenhayn (1992) and solve the model equilibrium analytically.

The remainder of the chapter is organized as follows. The next section provides details on the data used. Section 1.3 presents the reduced-form findings on the contemporaneous effects of flooding. Section 1.4 develops the theoretical model for studying the

1. See, for example, Fathom Global Flood Map

effects of flood shocks and flood risk. Section 1.5 discusses the estimation of key model parameters, and Section 1.6 presents the analysis. Section 1.7 contains some concluding remarks.

1.2 Data

There are two main datasets used in the chapter. The first is an extract of historical floods with various pieces of information to facilitate the construction of a flood index. The data on outcomes is derived from the census of medium and large manufacturing establishments located in Indonesia.

1.2.1 Large Flood Events

The data on floods is obtained from the Dartmouth Flood Observatory (DFO), which is a global, dynamic archive of large flood events starting in the year 1985 (Kocornik-Mina et al. 2020). The data provides start and end dates along with the extent of *affected area* for each flood event. Polygons representing the areas affected by flooding are drawn in a GIS program based upon information acquired from governmental, instrumental, news, and remote-sensing sources. Considering a longer time frame and the reliance on media reports, there could be concerns around plausible spatial and temporal bias in the reporting of flood events. For example, media reporting has improved over the years due to the development of technology and transportation infrastructure, and floods are more likely to be reported in areas with large population settlements and economic activity. In an extreme case, Figure 1.1 could merely reflect population settlement patterns across regencies in the Indonesian archipelago, rather than the actual number of flood events hitting them. To rule out this possibility, I redraw the map using only flood events confirmed through satellite observations, which are not subject to such biases. Specifically, I use inundation maps from 41 individual flood events across Indonesia during the period 2002–2018, as identified by Tellman et al. (2021). These maps are based on satellite

imagery captured by the Moderate-Resolution Imaging Spectroradiometer (MODIS) on NASA's Terra and Aqua satellites, which image the globe daily at a spatial resolution of 250 meters.[2] The spatial pattern of flooding derived from these objective satellite-based measures aligns closely with the pattern depicted in Figure 1.1. This consistency is illustrated in Figure 1.3, which confirms the robustness of the observed spatial distribution of floods. Although 30% of the reported flood events in Indonesia last three days or less, media reporting could still miss some small flood events, particularly due to the low- and middle-income country setting (see, for example, Patel 2024, for estimates of such biases in Bangladesh). Following variables are then constructed from this data:

- $FloodAreaShare_{rt}$: It is the share of the flood-affected area of regency[3] r in year t. For a few cases where a regency witnessed multiple flooding episodes in a year, this is the average of all those flood-affected area shares. As discussed earlier, this variable captures the extent of geographic regions affected by a flood event, rather than just the extent of inundation. Flood-affected area is usually larger than the inundated area, and is a more relevant measure for studying the effect of flooding on economic activities. Using inundation maps of individual flood events from Tellman et al. (2021), I study the relationship between inundated area share and flood-affected area share at the regency level for these specific events using both non-parametric and parametric methods.[4] Table 1.4 in the Appendix report mea-

2. MODIS is well-suited for detecting large, slow-moving flood events but has limited capacity to resolve urban floods.

3. A regency is an administrative level-2 unit located within a province in Indonesia. As of 2020, there were 34 provinces containing a total of 522 regencies. However, many of these provinces and regencies were born out of administrative divisions among the existing ones through the years 1990-2010. Since the analysis starts in 1990, I merge some of these divisions to be representative of the 1990 administrative boundaries. Moreover, I drop four provinces on the Eastern islands namely, Papua, Papua Barat, Maluku, and Maluku Utara, as these provinces are sparsely populated by mainly indigenous tribes, and are predominantly engaged in activities such as forestry and fishing. After implementing all these changes, the analysis is representative of approximately 270 regencies that cover the entire economic map of the Indonesian archipelago.

4. Several areas within a regency could not be observed due to cloud cover on some days. In addition, to prevent misclassification of terrain shadows as water, areas with gradient larger than 5° were masked out in the maps. This means that the flooded area share is calculated out of the observed pixels, which do not necessarily cover the whole regency. Flooded areas would also be missed if they happen to lie within the masked areas.

sures of association between the two variables using two non-parametric methods and a regression analysis. Column 1 reports the Kendall's Tau-b coefficient and Column 2 reports the Somers' D coefficient, both of which range from -1 (perfect inversion) to +1 (perfect agreement), with 0 indicating no association. Clearly, the two variables are positively related. Using estimates of regression analysis reported in Column 3, a unit increase in flooded area increases the flood-affected area by 1.42 units.

- $FloodDaysShare_{rt}$: It is the share of days in year t that a regency r remains affected by flooding. It is calculated using the total duration (end date - start date + 1) of all flood events in a year. In most cases, these dates are derived from news reports. In a few cases for which beginning dates could not be determined, the starting dates are assumed to be the 15th day of the respective months.[5] Ending dates can either be exact, based on dates on which flood water starts to recede as per the news reports or estimated, based on a qualitative judgment concerning the flood event. Najibi and Devineni (2018) analyzed the issue of misreporting of duration using all flood events in the DFO catalogue for 1985–2015 time period. Their comparative analysis using the *in situ* streamflow observations obtained from the gauge stations suggests that the flood duration data from DFO is reliable and does not suffer from misreporting issues.

Figure 1.19 in the Appendix shows the distribution of above variables by pooling all regency-year observations used in the analysis. Flood index is then generated by taking a simple product of the above two variables and rescaling the product by its maximum value so that it lies in the interval [0,1]. Thus, the index provides a measure of the intensity of floods by capturing both the spatial and temporal extents of each flood episode. Considering the spatial and temporal extent of flooding helps capture aspects of flood risk that both affected and unaffected firms may internalize in their decision-making.

5. Only around 2.5% of the flood events that occurred in Indonesia during the 1985-2012 period have a starting date as the 15th day of a month.

This approach acknowledges that firms, regardless of direct impact, could adjust their decisions based on the potential threat posed by flooding in their locations over time. As shown in Figure 1.20 in the Appendix that includes all regency-year pairs (even those without floods), the flood index has a Pareto-like distribution with a long tail of extreme values. This is intuitive since extreme flood events are rare. Table 1.5 reports the summary statistics on the flood index by utilizing only those regency-year pairs for which the index takes non-zero values. In the reduced-form analysis, the 25th, 50th, 75th, and 90th percentiles are used to capture and report the differential effects of flooding across various intensities. For example, at the 50th percentile, the estimated coefficients would represent the relationship between an outcome variable and a flood event at the median intensity level, an event more severe than 50% of all observed flood events in Indonesia from 1990 to 2012.

1.2.2 Information on Manufacturing Establishments

Data on the manufacturing establishments is obtained from the Annual Census of Medium and Large Manufacturing Establishments in Indonesia, also known as *Statistik Industri*. This data collection exercise was initiated by the Government of Indonesia in 1975 to survey all the manufacturing establishments with twenty or more workers annually. The central statistical agency, Statistics Indonesia, manages the collection and distribution of this data across different public departments and research organizations. Statistics Indonesia sends a questionnaire (asking details about previous year's operations), containing 150+ questions in a typical year, annually to all registered manufacturing establishments. In case of no response, field agents attempt to visit these establishments to either encourage compliance or confirm that the establishment has ceased operations (Blalock and Gertler 2008).

The establishment-level data include information on industrial classification (5-digit ISIC), first year of commercial production, ownership structure, assets, income, output, value-added, expenses, capital stock, and other specialized information specific to a year

for each establishment.[6] The main variables used in the analysis include measures on value-added, capital stock, labor employment, age (based on reported birth year), and location (regency where plant is located). All monetary variables are reported in nominal terms and are deflated using the wholesale price index at the 5-digit ISIC level to obtain their real values. Establishments are expected to report both market and book values of their capital stock, broken down by categories such as land, buildings, and equipment. However, book values are missing for most observations, and not all categories of capital are consistently reported over time. Therefore, I use the market value of capital stock in all cases, unless it is unavailable and the book value is reported for those observations. All variables are winsorized at the 1% level on both the lower and upper tails each year to help mitigate potential measurement error concerns. To prevent compositional changes across variables from influencing the estimates, I include only plant-year observations with complete data on all three variables viz., value added, total capital stock, and labor employment after the final data cleaning step. Finally, I exclude all state-owned establishments, which represent less than 3% of the total establishments in any given year, to avoid potential bias related to the implicit government insurance available to this group.

In a typical year, around 21,000 establishments are surveyed, with locations identified up to the regency where they are located. This establishment-level data is representative of firm-level analysis, as more than 95% of surveyed establishments are single-branch entities. Therefore, hereafter, I use the term "firm" instead of establishment to refer to these manufacturing enterprises. Although these firms represent only about 2% of the total number of manufacturing firms operating in Indonesia in any given year, they contribute approximately 80% of the total value-added in the country's manufacturing sector. In terms of spatial distribution, these firms are primarily concentrated on the

6. Statistics Indonesia checks the reported values for inconsistencies and missing values and tries to make in-house corrections and imputations using the previous rounds of data before releasing it to the users. I do some additional data cleaning to match location identifiers consistently across years, impute some variables to correct for non-reporting in just one or two years, fix outliers identified at the firm and industry levels by interpolating between years, and fix a few obvious mistakes made in the data entry process.

southern islands of Java and Sumatra, where flooding episodes are also more frequent and intense (Asian Development Bank 2019).

1.3 Reduced-form Evidence

This section presents reduced-form results on the immediate impact of floods on aggregate and firm-level value-added, capital stock, and labor employment. It then examines the extensive margin by estimating the impact of flooding on firm entry and exit. Given the unique setting in this study where manufacturing hubs, primarily located in the flood-prone regions, contend with recurrent large flood events, it becomes challenging to isolate and interpret the long-run impact of individual flood occurrences without introducing potential biases.[7] Therefore, to examine the long-run consequences of flooding in such a context, one could use some kind of cumulative measure of flooding that *integrates* floods over an extended period of time. However, the focus of this chapter is on estimating the contemporaneous effects of flooding and learning about firms' reaction towards regional flood risk. Therefore, one such analysis is included in the Appendix Section 1.8 where I estimate the effects of cumulative flood innovations on economic variables.

1.3.1 Effect of Flooding on Economic Variables

Econometric Model

I estimate the contemporaneous effects of flooding on aggregate (sector-regency) outcomes i.e., logarithm of total (labor share-weighted) value-added, capital stock, and labor

7. Figure 1.21 in the Appendix shows that most of the regencies located on the islands of Java and Sumatra, which account for more than 95% of the manufacturing value-added, are affected by a large flood every alternate year. Figure 1.22 in the Appendix shows the effect of the *first* flood on the logarithm of aggregate value-added, capital stock, and labor employed using the imputation-based difference-in-differences estimator proposed by Borusyak, Jaravel, and Spiess (2024). The estimated coefficients are not only less representative and far from what I would *want* to estimate but are also difficult to interpret. This is because the focal manufacturing regencies are not represented in the estimation of most of the dynamic effects coefficients as they get dropped too early because of the immediate second flooding episode.

employed at sector-regency level using the following specification:

$$y_{srt} = \upsilon + \beta^J Flood^J_{rt} + \zeta_r + \nu_{st} + \varepsilon_{srt} \tag{1.1}$$

where, y_{srt} is the logarithm of total (labor-share weighted) value-added, capital stock, or labor employment in sector s located in regency r in year t. $Flood^J_{rt}$ is a dummy variable that is assigned value 1 when the flood index in a regency-year exceeds the J^{th} percentile value for each $J \in \{25, 50, 75, 90\}$. Therefore, β^J captures the effect of J^{th} percentile flood on y_{srt}. ζ_r controls for time-invariant regency level characteristics, such as pre-existing differences in flood exposure and industrial settlements across regencies. ν_{st} controls for sectoral growth over time, where s denotes 2-digit ISIC sector. Where applicable, the outcome variables have been deflated by using 5-digit ISIC industry wholesale price index, and trimmed by 1% on both the tails for each year before being collapsed at the aggregate level.

Next, I conduct the firm-level estimation using the following specification:

$$y_{isrt} = \upsilon + \beta^J Flood^J_{rt} + \iota X_{isrt} + \zeta_i + \nu_{st} + \psi_{pt} + \varepsilon_{isrt} \tag{1.2}$$

where, y_{isrt} is the logarithm of value-added, capital stock, and (permanent and temporary) labor employed for firm i, belonging to sector s, located in regency r, in year t. $Flood^J_{rt}$ has the same definition as earlier. X_{isrt} includes time-varying, firm-level controls. Given the extensive impact of floods on firms, there are few suitable candidates for valid control variables. Consequently, only the logarithm of firm age and its squared term are included as firm-level controls. Firm fixed effects are controlled for as represented by parameter vector ζ_i. Similar to Equation (1.1), sector $\times$ year fixed effects are included. Since spatial margin is a key component in the flood index, with more statistical power available, province $\times$ year fixed effects denoted by ψ_{pt} are also included to control for removing differential geographic trends in flooding and outcome variables. In both estimations, control observations are defined by regency-year pairs that are not affected by flooding,

meaning that the flood index is zero for control observations.

Different manufacturing sectors within the economy may be affected differently by flooding, depending on their production characteristics, such as input mix. To examine sectoral heterogeneity in the impact of flooding, I estimate an interaction version of the firm-level specification. This involves interacting the 2-digit ISIC sector dummies with flood dummies for different percentiles. The resulting specification is as follows:

$$y_{isrt} = \upsilon + \beta^J Flood^J_{rt} + \gamma^J_s \mathbb{1}\{Sector = s\} \times Flood^J_{rt} + \iota X_{isrt} + \zeta_i + \psi_{pt} + \varepsilon_{isrt} \quad (1.3)$$

Here, γ^J_s captures the estimated effect of the flood dummy J on firm-level economic variables, while all other symbols retain the same definitions as in Equation (1.2). When presenting the results, I report the combined main and interaction effects, that is, $\beta^J + \gamma^J_s$.

Results and Discussion

Figure 1.4 reports the results from estimating Equation (1.1). The contemporaneous effects of flooding on aggregate value-added (left), capital stock (centre), and labor employed (right) are negative, and the effects become stronger as the flood intensity increases. A 90th percentile flood is associated with 20%, 25%, and 15% decline in the sector-regency value-added, capital stock, and labor employment respectively.[8]

Figure 1.5 reports the results of estimating the firm-level specification outlined in Equation (1.2). Keeping everything else constant, a 90th percentile flood leads to a 5.7% decrease in the firm-level capital stock. While permanent labor employment does not respond to flood shocks, firms tend to increase their hiring of temporary workers during flood events. In addition, the effects on capital are driven by a few fixed capital categories, as reported in Figure 1.6.[9] The estimates indicate that the negative effects on the total

8. Though most of the regencies are observed for the entire 23 years of study period, to avoid compositional changes driving the estimates, I conduct a robustness check by using only those regencies for which at least 20 years of data are available. Results are robust to using this more balanced sample, and is reported in Figure 1.23 in the Appendix.

9. As mentioned in the data section, the reporting on different capital categories is not consistent over time, so the number of observations used in the estimation of coefficients are different across the four

capital stock are primarily driven by structures and land, which are more susceptible to lose their values due to the anticipatory effects of flooding, as firms preemptively reduce investments in response to the expected flooding. In contrast, machinery and other equipment, which are more vulnerable to the direct destruction effects of flooding, do not respond significantly to floods. This suggests that the perceived risk of flooding instead of actual flood shocks could be more important for capital investment decisions.[10] The temporal resolution of the analysis may also influence these results. Firms may be able to quickly *rebuild* some of their destroyed capital within a year, meaning such effects may not be captured in the annual analysis. There is evidence suggesting that firms often return to their initial production levels relatively quickly after floods (see, for example, Balboni, Boehm, and Waseem 2023, which examines firm responses to flooding in Pakistan). However, the impact on the capital stock in these cases remains unclear. Even if firms can quickly replace some of their damaged capital, the remaining capital that remains affected over a longer time horizon would influence the long-term development trajectory.[11]

Figure 1.7 reports the results of the heterogeneity analysis across sectors by estimating Equation (1.3) for the 90th percentile flood dummy.[12] The findings reveal two key patterns. First, the sector-specific results are qualitatively consistent with the overall trends shown in Figure 1.5, indicating that a typical firm experiences a decline in capital stock and increases temporary labor hiring after floods. Second, capital-intensive sectors—such as food processing, iron and steel, and ceramics, glass, and clay products—exhibit a more significant reduction in capital stock. This suggests that the impact of flooding on the capital margin is particularly pronounced in sectors that depend heavily

categories.

10. Similar to the aggregate results, firm-level results are robust to using a more balanced sample. Results using only those firms for which data is available for at least 20 years are reported in Figure 1.24 in the Appendix.

11. In 2023, 2,955 firms across the manufacturing, services, and retail sectors were surveyed in Indonesia between December 2022 and September 2023. Approximately 16% of these firms reported being impacted by a natural disaster during that period, but only about one-fourth of the affected firms experienced damage to their physical assets. *Source:* World Bank Enterprise Surveys, www.enterprisesurveys.org.

12. Results for flood dummies at other percentiles are provided in the Appendix.

on capital for their production processes.

1.3.2 Effect of Flooding on Firm Exit and Entry

Changes in the number of operational firms across years would contribute to the effects of flooding on aggregate economic variables. For example, reduced entries and/or increased exits could explain the negative effects obtained. Similarly, sample attrition owing to exiting firms could potentially bias the firm-level results. Firms that survive negative shocks, such as floods, are not only more adapted to deal with these shocks but are also more adept in their operations.[13] Therefore, if some small and less efficient firms are shutting down after floods, then the firm-level estimates are the lower bounds of total effects.

Econometric Model

To estimate the contemporaneous effect of flooding on firm exit, I employ the following econometric specification on the firm-level data:

$$y_{isrt} = \upsilon + \beta^{J} Flood^{J}_{rt} + \iota X_{isrt} + \zeta_{r} + \nu_{st} + \psi_{pt} + \varepsilon_{isrt} \tag{1.4}$$

where y_{isrt} is an exit dummy for firm i, belonging to 2-digit ISIC sector s, located in regency r, in year t and other terms have the same interpretation as the previous specifications.[14]

Results and Discussion

The left plot in Figure 1.8 reports the results of estimating Equation (1.4) on all the firms in the data. The results suggest that floods are not associated with firm exits. This might

13. Figure 1.28 in the Appendix suggests that exiting firms, on average, have lower value-added, capital stock, and labor employment, compared to both new entrants and survivors.

14. The exit dummy is an "implied" variable in the sense that they are backed out from the longitudinal observation of a firm in the data. A firm's last year in the data is taken as its exit year.

not be a surprising result considering that the sample comprises of medium and large manufacturing firms that are likely more adept at dealing with such shocks. Null results on the exit margin also indicate that the biases introduced in the previous analysis on firm-level variables are small.

The right plot in Figure 1.8 reports the results from estimating Equation (1.1) on the logarithm of number of new firms entering into a sector-regency in a given year. A 90th percentile flood leads to a 20% reduction in the number of new firms entering in a sector within a regency. This evidence suggests that firms avoid flood-affected regencies when setting up their operations.[15] Entry is typically a costly decision requiring investments, so firms would typically enter if they expect to make some profits net of entry costs. However, the results on entry indicate lack of good foresight on floods, as firm entry is reduced in the year of flooding. This very fact motivates the modeling assumptions in the firm entry part of the theoretical framework, in particular, an entrant's decision to enter depends on its expected profits in the current period only. Additionally, similar to the findings related to capital, these extensive margin results point towards the role of perceived flood risk at the firm level due to the evolving distribution of flood shocks.

1.3.3 Scope of the Reduced-form Approach and Way Forward

The reduced-form findings capture both the elements of destruction by flooding and anticipation due to the evolving flood risk across regencies. Firms located in regencies where floods are more persistent will take more anticipatory actions aimed at mitigating the effects of flooding, thereby reducing the actual destruction resulting from floods. Therefore, the reduced-form approach fails to offer a framework for understanding the mechanisms leading to the impact, particularly in linking to the literature on misalloca-

15. I validate this result using a more direct measure of flood risk made available for year 2013 by *Indeks Risiko Bencana Indonesia* (IRBI). I use regency-level average flood index (over the period 1990-2012) and the flood risk data available for the year 2013 to estimate their impact on average firm entry rate and firm entry rate in 2012 respectively by employing the specification: $y_r = \upsilon + \beta Flood_r + \varepsilon_r$. Entry rate is defined as the count of new firms in the current year over the total count of firms in the previous year for a given regency. Table 1.6 in the Appendix suggests that regencies with high flood index and high flood risk profile witness lower firm entry rates.

tion. The capital installation decisions are usually taken well in advance of such shocks, and firms can use some available signals on these shocks to inform their decision on the current level of capital to install. In addition, the approach becomes unreliable for investigating some sources of heterogeneity across firms that could drive the impact of flooding. Since floods are evolving shocks that are heterogeneous across space, any policy analysis would be considered incomplete without the evaluation of meaningful counterfactual scenarios beyond the existing flood experiences. A model-based approach could potentially address some of these issues.

To get to the anticipatory element of flood shocks, I exploit the uncertainty that firms face due to regional flooding when choosing the optimal capital to install in a given period. Capital installation decisions are taken in anticipation of floods experienced by the regencies where firms are located. This generates a time-varying flood risk variable, which vary across regencies and also across industries located within a regency. Modeling floods in this manner links them closely with the firm's production enterprise, thereby offering a microfoundation for understanding how perceived flood risk interacts with firm behavior.

The share of installed capital that can be utilized in a regency in a given year is proxied by the firm-level production capacity utilization (PCU) for that year. PCU is the percentage of available production capacity that a firm is able to utilize in a given year.[16] Figure 1.9 reports the results of estimating Equation (1.2) with PCU as dependent variable; a 90th percentile flood is associated with a 3.6% decline in the firm-level PCU.[17] Therefore, the empirical distribution of firm-level PCU could be used to calibrate a measure of flood risk, which captures the anticipatory effects of flooding on firms. Unlike other measures of flood risk that are generated using climate and atmospheric models and might

16. Figure 1.29 shows that the share of firms reporting lower PCU levels tend to be higher in the flood-prone regencies, thereby providing a suggestive evidence for the impact of flooding on PCU.

17. There are some firm-year observations that report zero capacity utilization even when the output is non-zero. Since the data primarily comprises of manufacturing firms, it is less likely that they engage in other businesses to generate output, therefore, these observations could be excluded from the estimation. The reported estimates use all available observations, but the results remain qualitatively unchanged if those observations are excluded.

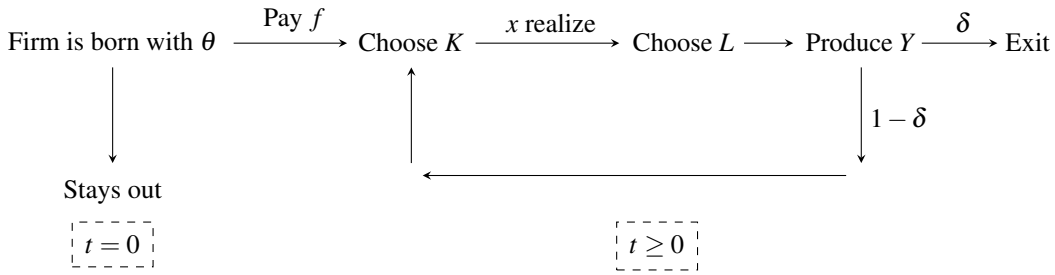
not capture local conditions, this measure is informed from the firm-level decisions itself, and is therefore more relevant for studying the effects of flooding on firms. Given that capital installation decisions are made in anticipation of current flooding, firms account for flood risk when choosing the optimal level of capital. Firms located in flood-affected regencies are able to utilize a lower share of their installed capital, which in turn impacts the level of capital that they install to start with. The potential entrants in these markets face similar constraints due to the stochastic nature of flooding and their entry decisions are driven by expected profits. Below, I propose a model with heterogeneous firms to study the effects of flood risk and flood shocks on the production side of the economy.

1.4 Theoretical Framework

The multi-sector, multi-region general equilibrium model captures the interaction between regency-level flooding and firm-level economic variables. Central to the model are parameters that account for flood exposure at the regency level and firm selection based on the idiosyncratic productivity that each firm is endowed with. The investigation of the impact of flooding on firms employs a *within* production function approach in which anticipation of floods affect firms' optimal capital installation decision. Specifically, incumbent firms internalize the constraint that flooding at their location will render only a fraction of their installed capital usable in each period. Firms are modeled as risk-neutral agents, so only the expected values of these shares matter for their decision-making. The flood risk arising from the anticipatory element of flooding also varies by industries, in particular, the capital intensity of their respective production technologies. Firms are price-takers, taking wages and rental price of capital as given. The fixed differences in productivity across firms, combined with decreasing returns to scale, generate rents for individual firms. Consequently, even though all firms within a sector are price-takers and produce a homogeneous product, the equilibrium features a distribution of firms differentiated by their idiosyncratic productivity levels.

On the extensive margin, firm entry is affected by flood risk. Each firm is born with an idiosyncratic productivity that remains constant over time. To enter a market and start production, each firm pays a one-time fixed cost. Entrants use their expected profits for the current year to decide if they should enter the market. This structure on entry decision generates firm selection, where only firms with large enough productivity decide to enter into a sector within a region.

Below is a schematic life-cycle diagram of a typical firm in the model that is born at $t = 0$ and has options to either stay out of market or pay a one-time fixed cost f to be able to choose the capital stock for period $t = 0$. The flood shock is then realized, and the firm chooses the flexible input, labor, to produce output in period $t = 0$. This three-step cycle of capital installation, flood shock realization, and flexible labor choice continues in this order with probability $(1-\delta)$, as the firm can exit the market with probability δ in any period $t \geq 0$ for some exogenous reasons.



1.4.1 Technology

The production technology of a risk-neutral firm i located in regency r at time t is Cobb-Douglas in labor L and capital K, with the sector-specific output elasticity of capital and returns to scale parameters denoted by α_s and η_s respectively. The firm can only utilize a stochastic share $x_{it} \in [1,\infty)$ of the installed capital because of floods at time t, so it forms expectations on this random variable. Idiosyncratic productivity θ remains constant over time and has an ex-ante regency-specific distribution with c.d.f. (p.d.f.) denoted by H_r (h_r), which has full support in the domain $[1,\infty)$. The production function is defined as

below:

$$Y_{it}(\theta, x, K, L) = \theta_i \left\{ \left(\frac{K_{it}}{x_{it}} \right)^{\alpha_s} L_{it}^{1-\alpha_s} \right\}^{\eta_s} \tag{1.5}$$

Labor hiring decision is made after the realization of x_{it}, so the labor demand adjusts flexibly based on the capital input demand and the realized value of x_{it}. Firms take prices as given, so with wage rate w_t, the optimal choice of labor after maximizing the profit function can be written as follows:

$$L_{it}(\theta, x, K, w) = \left\{ \frac{w_t}{(1-\alpha_s)\eta_s \theta_i (K_{it}/x_{it})^{\alpha_s \eta_s}} \right\}^{-\frac{1}{1-(1-\alpha_s)\eta_s}} \tag{1.6}$$

Putting the optimal labor choice above back into the production function, the firm-level conditional (on capital) profit function can be written as follows:

$$\pi_{it}(\theta, x, K, z) = \Gamma_{it} \left(\frac{K_{it}}{x_{it}} \right)^{\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} \tag{1.7}$$

where $\Gamma_{it}(\theta, w) \equiv [1-(1-\alpha_s)\eta_s] \, \theta_i^{\frac{1}{1-(1-\alpha_s)\eta_s}} \left\{ \frac{w_t}{(1-\alpha_s)\eta_s} \right\}^{-\frac{(1-\alpha_s)\eta_s}{1-(1-\alpha_s)\eta_s}}$ is the product of aggregate, sector-, and firm-level parameters. z_t denotes the input price vector (w_t, ρ). Firms make their decision on the optimal capital to install before x_{it} is realized, so being risk-neutral, they maximize expected profits when making this choice under uncertainty.

1.4.2 Flood Shocks and Flood Risk

Firms anticipate flood shocks in a given year and choose capital accordingly. Firms internalize the fact that flooding can reduce the share of installed capital that they could utilize in a given year. Firms do not know this share precisely but have knowledge of the distribution from which the share is drawn. By maximizing the expected profits, firms choose the optimal capital to be installed in each period.

The share variables, $x \in [1, \infty)$, follow Pareto distributions with time-varying, regency-

specific shape parameters ϕ_{rt}. The general form of the distribution is as follows:

$$G_{rt}(x) = \begin{cases} 1 - \left(\frac{1}{x}\right)^{\phi_{rt}} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

The above assumption on the distribution of flood impact on firms is natural and is motivated by the fact that some regencies are more flood-prone than others. Firm's expected share of capital that they can utilize in a given year would be lower for those regencies that experience more extreme flood events i.e., regencies with heavier Pareto tails.

Capital is perfectly mobile, so the rental price of capital ρ, does not vary across regencies. This assumption is made because the central bank of Indonesia, *Bank Indonesia*, sets a national interest rate that influences lending rates across the country. Integrated financial markets within the country ensure that the no-arbitrage condition holds, leading to convergence of interest rates across regions.[18] Going back to the profit function defined in Equation (1.7), the net expected profit function can be written as follows:

$$\Pi_{it}(\theta, x, K, z) = \Gamma_{it} \mathbb{E}\left[\left(\frac{K_{it}}{x_{it}}\right)^{\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}}\right] - \rho K_{it} \tag{1.8}$$

Using the distribution function of share variable defined above, the objective function for the optimal capital can then be written as follows:[19]

$$K_{it} = \text{argmax}\left\{\Gamma_{it} \tau_{srt} K_{it}^{\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} - \rho K_{it}\right\} \tag{1.9}$$

where $\tau_{srt}(\phi) \equiv \dfrac{\phi_{rt}}{\phi_{rt} + \alpha_s \eta_s / (1-(1-\alpha_s)\eta_s)}$ is a measure of flood risk, which captures distortions introduced in capital decisions due to flooding. This measure at the sector-

18. There could still be some regional differences in the price of capital due to a multitude of factors, including regional flood risk. Ridhwan et al. (2012) show the map of these differences for rural and regional bank interests rates for the period 2000-08. First, the total assets held by these two categories of non-central banks combined is less than 10% of the total assets in the Indonesian banking industry (Financial Services Authority of Indonesia 2024). Second, the spatial pattern of deviations in interest rates reported in Figure 1 of Ridhwan et al. (2012) does not correspond to the pattern of flooding shown in Figure 1.1.

19. Detailed derivations are shown in the Appendix Section 1.10

regency level account for both the spatial differences in flood exposure and the differences arising due to sectoral characteristics, in particular, heterogeneity in capital intensity across industries. Note that τ_{srt} is an increasing function of ϕ_{rt}, but decreasing in η_s and also α_s in case of a decreasing return to scale technology.[20] Also, by the properties of Pareto distribution, increasing ϕ_{rt} decreases the probability of realization of larger values of x_{it}, that is, increases the probability of observing higher shares. Therefore, regencies experiencing more extreme floods on average, would have lower values of respective tail parameters. Finally, solving for the optimal capital choice using Equation (1.9) delivers the equilibrium value of capital installed as outlined below:

$$K_{it}(\theta,\phi,z) = \frac{\alpha_s\eta_s}{\rho}\tau_{srt}^{\frac{1-\eta_s+\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\theta_i^{\frac{1}{1-\eta_s}} \tag{1.10}$$

where $\Lambda_{st}(z) \equiv \left\{\frac{w_t}{(1-\alpha_s)\eta_s}\right\}^{-\frac{(1-\alpha_s)\eta_s}{1-\eta_s}}\left\{\frac{\rho}{\alpha_s\eta_s}\right\}^{-\frac{\alpha_s\eta_s}{1-\eta_s}}$ is the product of aggregate and sector-level parameters. To be precise, this is the stock of capital *installed* by a firm, but the amount capital that is eventually *utilized* will depend on the realization of x_{it}. More precisely, the part of capital stock that is utilized in production will be (K_{it}/x_{it}), as the remaining capital gets *destroyed* in flood. Using Equation (1.6), the equilibrium value of labor demanded can then be written as follows:

$$L_{it}(\theta,\phi,z,x) = \frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\theta_i^{\frac{1}{1-\eta_s}}x_{it}^{-\frac{\alpha_s\eta_s}{1-(1-\alpha_s)\eta_s}} \tag{1.11}$$

Due to the multi-stage setup that firms use for choosing inputs within a period, the *ex-ante* (before realization of flood shocks) values of output and profit would differ from the respective *ex-post* (after realization of flood shocks) values. However, the entry decision of new firms will be based on the expected value of profit. Below are the expected values

20. The setup is similar to Lucas (1978) in which the source of decreasing returns is managerial limits on the production side. Alternatively, the origin of decreasing returns could be on the demand side as shown in Hopenhayn (2014). The equivalence of results obtained from these two different approaches can be ensured by calibrating the demand elasticity, ε, in the latter to be equal to $(1/(1-\eta))$.

of output and profit in equilibrium:

$$Y_{it}(\theta,\phi,z) = \tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}} \Lambda_{st}\theta_i^{\frac{1}{1-\eta_s}} \tag{1.12}$$

$$\Pi_{it}(\theta,\phi,z) = [1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt}]\,\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}} \Lambda_{st}\theta_i^{\frac{1}{1-\eta_s}} = [1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt}]\,Y_{it} \tag{1.13}$$

1.4.3 Firm Entry and Exit

Firm exit from the market is simple in that it is assumed to be exogenous and is parameterized by a constant probability of exit δ, which is independent of sector-, regency-, and firm-level variables.[21] The entry decision of firms has more subtleties involved. The pool of potential (identical) entrants is unbounded. Each entrant is born with a (constant) idiosyncratic productivity θ, which is drawn from a common regency-specific, time-invariant distribution $H_r(\theta)$. A potential entrant decides to stay out of the market in any given year if its productivity is too low. It is because each potential entrant needs to pays a one-time fixed cost, f to enter the market. Therefore, an entrant in period t would have its expected profit in period t exceed the fixed cost of entry. This generates a time-varying sector-region cutoff productivity θ^*_{srt}, below which potential entrants do not enter into sector s within regency r in year t. Combining all of the above, the expected net profit function of a potential entrant i at time t can be written as follows:

$$\sigma_{it}(\theta) = \max\{0, \Pi_{it}(\theta) - f\} \tag{1.14}$$

where $\Pi_{it}(\theta)$ is the expected profit function defined in Equation (1.13).

21. Once a firm has entered, it cannot endogenously exit the market. Equation (1.13) supports this assumption as follows. A firm can exit if its profit becomes negative in any period i.e., $1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt} < 0$, but this is not plausible in a decreasing returns to scale technology, where $\eta < 1$.

Aggregation

Owing to the above entry and exit dynamics, the equilibrium productivity distribution, $\mu_{srt}(\theta)$ could differ from the ex-ante distribution, $h_r(\theta)$. However, the exogeneity of exit decisions ensures that it does not affect the equilibrium distribution. In addition, all entrants with $\theta < \theta^*_{srt}$ stay out of the market, so the equilibrium distribution depends only on the productivity of entrants. That is, the equilibrium productivity distribution is a truncated version of the ex-ante distribution, as outlined below:

$$\mu_{srt}(\theta) = h_r(\theta|\theta \geq \theta^*_{srt}) = \begin{cases} \dfrac{h_r(\theta)}{1-H_r(\theta^*_{srt})} & \theta \geq \theta^*_{srt} \\ 0 & \theta < \theta^*_{srt} \end{cases} \tag{1.15}$$

where the probability of successful entry into sector s in regency r is $p^e_{srt} \equiv 1 - H_r(\theta^*_{srt})$.[22]

Studies have shown that the Zipf's law seems to be an empirical regularity for the firm size distribution; the results get even tighter at the upper tail, which tends to be well-approximated by a Pareto distribution (Geerolf 2017).[23] Using this information, the distribution of productivity is assumed to have the following form:

$$H_r(\theta) = \begin{cases} 1 - (\overline{\theta}_r/\theta)^{\xi} & \theta \geq \overline{\theta}_r \\ 0 & \theta < \overline{\theta}_r \end{cases}$$

where $\overline{\theta}_r$ is a regency-specific scale parameter that reflects the spatial differences in firm productivity. For example, a firm born in Jakarta might be more or less productive due to competing agglomeration and congestion externalities existing there. Without firm entry and exit, the equilibrium productivity distribution is same as the initial distribution, $h_r(\theta)$. But when firms are allowed to enter and exit the markets, the equilibrium distribution takes the form defined in Equation (1.15). I now define the expected aggregate output

22. Hopenhayn (1992) discusses the assumptions under which law of large numbers could be applied to determine the equilibrium distribution, $\mu_{srt}(\theta)$ from the initial distribution, $h_r(\theta)$.

23. Since, the data comprises only of medium and large manufacturing firms, the Pareto assumption on firm productivity distribution is even more innocuous.

for sector s in regency r at time t, $\overline{Y}_{srt} \equiv \int_{\theta^*_{srt}}^{\infty} Y_{it}(\theta)\mu_{srt}(\theta)d\theta$, which is a total weighted output of all surviving firms, where the weights are mass of firms at each productivity level in the equilibrium distribution. Combining the definition of equilibrium distribution from Equation (1.15) with the firm-level expected output defined in Equation (1.12), the expected aggregate output can be written as follows:[24]

$$\overline{Y}_{srt} = \tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}} \frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1} \Lambda_{st} \left(\theta^*_{srt}\right)^{\frac{1}{1-\eta_s}} \tag{1.16}$$

The only parameter restriction that needs to be imposed here is that $(\xi(1-\eta_s) > 1)$, which, as would be seen later in the estimation part, is true for all the 3-digit ISIC manufacturing sectors in the economy. Also, since $(\theta^*_{srt} > 1)$ and $(\xi > 0)$, given constant wages, the expected aggregate output and thereby expected aggregate profit increase after considering the firm selection above. Similar to the above, expressions for all other expected aggregate variables could be derived using Equations (1.12) and (1.13)

One key object is the cutoff productivity θ^*_{srt}, which is the productivity of least productive firm deciding to enter into sector s in regency r at time t. This can be pinned down using the firm-level expected profit function as described next.

Zero Cutoff Net Profit Condition

Each entrant needs to pay a one-time fixed cost, f to enter into any market. This fixed cost can be thought of as a permit or license fee that each new firm needs to pay to a centralized authority. An entrant would be willing to pay this fee if it can make a non-negative expected net profit, that is, the cutoff productivity, $\theta^*_{srt} \equiv \inf\{\theta : \sigma_{it}(\theta) > 0\}$, where the net expected profit of entrant is defined in Equation (1.14). Using the firm profit function defined in Equation (1.13) delivers the expression for cutoff productivity below:

$$\theta^*_{srt} = \left\{ \frac{f}{\left[1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt}\right]\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}} \right\}^{1-\eta_s} \tag{1.17}$$

24. The intermediate steps for getting to this final expression are outlined in the Appendix Section 1.10.

Keeping wages fixed, θ^*_{srt} is a decreasing function of τ_{srt},[25] so, the cutoff productivity level is larger for more vulnerable sectors and flood-prone regencies, that is, the idiosyncratic productivity needs to be high enough for firms belonging to these sectors if they intend to establish operations in the flood-prone regencies, thereby suggesting that the firm selection effects would be stronger in these cases.

1.4.4 Equilibrium

The definition of equilibrium is standard, with the labor market clearing at the aggregate level each year. Like capital, labor is mobile across regencies, so wages are constant in space, but they do adjust over time in response to flooding. Therefore, the equilibrium is pinned down by the wages w_t and the cutoff productivity levels θ^*_{srt}. The aggregate labor endowment, $\bar{L}$ is assumed to be exogenous and constant over years. The labor market clearing condition delivers the equilibrium wage equation as follows:[26]

$$w_t = \frac{f}{\bar{L}} \sum_{r=1}^{R} \sum_{s=1}^{S} \frac{(1-\alpha_s)\eta_s \tau_{srt}}{1-(1-\alpha_s)\eta_s - \alpha_s \eta_s \tau_{srt}} \frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1} \tag{1.18}$$

It is easy to see that wages increase when the fixed cost of entry rises. It is because with increase in the fixed cost, the firm selection also gets stronger, that is, fewer and more productive firms are able to enter the markets. Since, the marginal product of labor depends on firm's productivity, when labor is reallocated to higher productivity firms, the marginal product of labor increases. Thus, wages also rise to match this increased marginal product in equilibrium.

Also, the sign of first derivative of wage equation w.r.t τ_{srt}, which quantifies the overall effect of flooding for sector s located in regency r at time t, is positive. This means that equilibrium wages go down as the impact of flooding increases. To get to the intuition behind this result, first, remember that flood shocks reduce the utilization of capital

25. For this to hold with certainty, η_s needs to be smaller than unity, that is, the production technology should have decreasing returns to scale.

26. Detailed derivation in the Appendix Section 1.10

and that reduces the returns on capital in areas affected by flooding. In the model, firms rely on both capital and labor to maximize output net of input costs. When capital is less productive, the marginal productivity of labor also declines, since firms cannot use labor as effectively without undistorted capital. This lower productivity reduces firms' demand for labor at any given wage, as they need to scale down operations in response to flooding. With a reduced demand for labor, the equilibrium wages undergo downward adjustment to clear the labor markets. Wages also decrease due to an increase in flood risk that is captured by a decline in τ_{srt}. It is because with increased risk, less firms will enter into these markets, decreasing the competition for labor, and eventually driving the wages downwards to clear the labor market.

With the equilibrium wages in hand, using Equation (1.17), the equilibrium values of cutoff productivity levels can be derived to follow the general analytical expression below:

$$\begin{aligned}\theta^*_{srt} = & \frac{f^{1-\eta_s}}{((1-\alpha_s)\eta_s)^{(1-\alpha_s)\eta_s}} \left(\frac{\rho}{\alpha_s \eta_s \tau_{srt}}\right)^{\alpha_s \eta_s} \\ & \times \left\{\frac{1}{1-(1-\alpha_s)\eta_s - \alpha_s \eta_s \tau_{srt}}\right\}^{1-\eta_s} \\ & \times \frac{f}{\overline{\overline{L}}} \left\{\sum_{r=1}^{R}\sum_{s=1}^{S} \frac{(1-\alpha_s)\eta_s \tau_{srt}}{1-(1-\alpha_s)\eta_s - \alpha_s \eta_s \tau_{srt}} \frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1}\right\}^{(1-\alpha_s)\eta_s}\end{aligned} \tag{1.19}$$

The derivative of the above expression w.r.t. τ_{srt} has a negative sign, thereby suggesting that even after accounting for equilibrium wage adjustments, the cutoff productivity level for entering into more vulnerable sectors located in flood-prone regencies needs to be higher.

The set of parameters: $\left(\{\{\phi_{rt}\}_{r=1}^{R}\}_{t=1}^{T}, \{\alpha_s\}_{s=1}^{S}, \{\eta_s\}_{s=1}^{S}, \xi, \{\overline{\theta}_r\}_{r=1}^{R}, \rho, f, \delta, \overline{L}\right)$ would need to be estimated or calibrated to compute each equilibrium object in levels. However, the subsequent analyses will require estimates for most—but not all—of these parameters. Specifically, estimates for the last four parameters, which represent the rental price of capital, fixed cost of entry, exit probability, and aggregate labor supply, will not be

needed.

1.5 Estimation

This section describes the estimation of some of the parameters in the model. Table 1.1 reports the summary on these parameters along with the estimation and calibration techniques employed.

1.5.1 Flooding Shape Parameters (ϕ_{rt})

To estimate the shape parameters of the Pareto distributions of the share of installed capital that can be utilized in a regency in a given year, I use the empirical distribution of PCU for firms located in the regency in that year. One key advantage of using an objective measure, such as PCU, is that it is immune to price changes and market adjustments due to new firm entries. Other similar measures, such as output and value of capital stock, would potentially be impacted by these equilibrium adjustments. Additionally, PCU is a relative measure based on what firms were able to produce relative to what they had *planned* in a given year, so it can be used as a proxy for flood shocks at the firm level. With this, the maximum likelihood estimator of the shape parameter can be derived and that has the following form:[27]

$$\widehat{\phi}_{rt} = \frac{N_{rt}}{\sum_{i=1}^{N_{rt}} \ln(x_{it})}$$

where x_{it} is the reciprocal of PCU for firm i in year t and N_{rt} is the number of firms located in regency r in year t. Under the assumption that firms' expectations are informed *only* by past events, one could alternatively use the past realizations of PCU to estimate this parameter. However, the current realizations would have elements from the past floods due to the auto-correlated nature of flood shocks. In the model, due to the multi-stage decision process within a period, entrants would not know the realized flooding

27. Detailed derivation of the estimator is in the Appendix Section 1.10.

before they enter nor the incumbents can choose capital under perfect information about flooding.

Figure 1.10 shows the distribution of average (over years) estimated flood exposure, ϕ_{rt} for each regency in the data. Due to the property of Pareto distribution, lower the value of shape parameter, higher is the exposure of regency to extreme flooding. Clearly, regencies in the West Java province, including Jakarta, are some of the most exposed regencies in Indonesia.

Figure 1.11 shows the correlation between the average (over years) shape parameters with the empirical flood index across regencies. The relationship is statistically significant and suggests that flood-prone regencies experience more severe flooding (thicker Pareto tails) on average. Therefore, regency-level flood index and model-based flood exposure are related, and the relationship is in the expected direction.

1.5.2 Firm Productivity Parameters ($\overline{\theta}_r, \xi$)

I calibrate the scale parameter, $\overline{\theta}_r$ to match the logarithm of aggregate regency value-added (akin to regional GDP) that is computed using the data I have on medium and large manufacturing firms in Indonesia. The shape parameter ξ, is then estimated using the following maximum likelihood estimator:[28]

$$\widehat{\xi} = \frac{N_r}{\sum_{i=1}^{N_r} \ln(\theta_i / \overline{\theta}_r)}$$

where N_r is the number of firms in regency r and θ_i is the logarithm of average value-added for firm i located in regency r. The point estimate is 4.514 with a robust standard error of 0.0237.[29]

28. I use the logarithm of average value-added over the entire period for which a firm has been operational in the data. This averaging exercise fixes each firm's productivity to a constant, but alleviates some concerns around measurement error in the reported figures.

29. One potential concern here could be that the equilibrium distribution, $\mu_{srt}(\theta)$, is *related* to the ex-ante distribution, $h_r(\theta)$, but it is not the same distribution. This issue becomes important when we are dealing with small samples within sector-region pairs because the non-applicability of the law of large numbers will not let us approximate the equilibrium distribution from the ex-ante distribution.

Given that the production function parameters, α_s and η_s are already estimated, one could compute the Solow residuals as a measure of productivity. However, this approach could be problematic for two reasons. First, flood risk can affect productivity. I follow an approach that closely mirrors Besley, Roland, and Reenen (2020) to show that this case does not arise here, owing to the aggregate nature of shocks. Second, measurement error in firm-level data, especially with respect to capital, could severely bias the productivity estimates (Collard-Wexler and De Loecker 2016). However, the impact of measurement error on the estimates of production function parameters should be minimal due to their aggregate nature.

The output share of firm i, belonging to sector s and located in regency r, $\kappa_{it} \equiv Y_{it}/\overline{Y}_{srt}$, will be equal to the *relative* productivity of firm, ω_{it} in a world without friction. However, in the presence of distortions due to flooding, the relative productivity could also change. Using Equation (1.12), the productivity terms can be written as follows:

$$\theta_i^{\frac{1}{1-\eta_s}} = \frac{Y_{it}}{\overline{Y}_{srt}} \frac{\overline{Y}_{srt}}{\tau_{srt}^{\frac{\alpha_s \eta_s}{1-\eta_s}} \Lambda_{st}}$$

Further, including the general expression for aggregate output in the above, one gets ω_{it} to be equal to κ_{it}. Therefore, productivity is not affected by flood risk.

1.5.3 Production Function Parameters (α_s, η_s)

I employ the production function estimation approach using the technique proposed in Levinsohn and Petrin (2003) for each 3-digit ISIC sector. The method uses similar identification ideas as in a control-function setup to address the issue of endogeneity of input choices. Unlike the seminal work Olley and Pakes (1996), which uses investment as a proxy to control for unobserved productivity shocks, Levinsohn and Petrin (2003) uses intermediate inputs, such as materials or energy, as proxies for unobserved productivity shocks. Intermediate inputs are often more flexible and can adjust more quickly to productivity changes than investment, which also suffers from issues such as lumpiness in

choice and a lot of zero entries in the data. In the estimation, logarithms of real value-added, real capital stock, quantity of labor employed, and real material costs are used as left-hand side, state, static input, and productivity proxy variables respectively.

Table 1.2 reports the computed production function parameters for each 3-digit ISIC sector; some of the capital-intensive sectors are Industrial Chemical Products, Basic Iron and Steel, and Machines and Repairs.[30] With the production function parameters in hand, I can compute the flood risk, represented by τ_{srt} in Equation (1.9), for each sector within a given regency. Figure 1.12 shows the distribution of average (over years and sectors) flood risk across regencies. Lower value of flood risk point to larger capital distortions caused by flooding. Additionally, the spatial distribution of flood risk has noticeable differences from the flood exposure map shown in Figure 1.10, thereby highlighting significant sectoral variations in industrial settlement patterns across regencies.

With the parameter estimates in hand, I show some analysis on flood risk and flood shocks using the equilibrium conditions derived from the model. Due to the exogenous nature of flood shocks, one could look at the effects of flooding on output both with and without equilibrium adjustments. The next section delves into the details.

1.6 Analysis

This section delineates the results of the analysis conducted using the structure of model and parameter estimates from the previous two sections. In the first part, I use the equilibrium conditions from the model to disentangle the effects of flood and flood risk. Next, I conduct a counterfactual analysis on flood defenses and disentangle effects due to different margins.

30. Table 1.7 reports the estimated output elasticities of capital and labor for each 3-digit ISIC sector.

1.6.1 Flood Shocks and Flood Risk

In the model, flood risk directly affect the optimal capital installation decision, while labor hiring decision is affected by flood shocks. However, the installed capital could suffer destruction ex-post due to realized flood shocks. So, I can use the expressions for firm-level equilibrium capital stock (accounting for destruction) and labor to understand how each gets impacted by different elements of flooding. Taking logs of Equations (1.10) and (1.11) delivers the following two equations:

$$\ln K_{it} = \underbrace{\ln\left(\frac{\alpha_s \eta_s}{\rho}\right) + \ln \Lambda_{st}}_{\text{sector} \times \text{year fixed effect}} + \frac{1-\eta_s+\alpha_s\eta_s}{1-\eta_s}\ln \tau_{srt} + \underbrace{\frac{1}{1-\eta_s}\ln \theta_i}_{\text{firm fixed effect}}$$

$$\ln L_{it} = \underbrace{\ln\left(\frac{(1-\alpha_s)\eta_s}{w_t}\right) + \ln \Lambda_{st}}_{\text{sector} \times \text{year fixed effect}} + \frac{\alpha_s\eta_s}{1-\eta_s}\ln \tau_{srt} + \underbrace{\frac{1}{1-\eta_s}\ln \theta_i}_{\text{firm fixed effect}} - \frac{\alpha_s\eta_s}{1-(1-\alpha_s)\eta_s}\ln x_{it}$$

One could compute the coefficients on τ_{srt} and x_{it} directly using the estimated parameters but it would be subject of endogeneity concerns. Therefore, to estimate the elasticities of capital and labor with respect to flood risk and actual flood shock, I employ the following econometric specifications based on the above two equations, accounting for the destruction effect of flood shocks on capital:

$$\ln K_{it} = \nu_{st} + \beta^K \ln \tau_{srt} + \zeta_i + \gamma^K \ln x_{it} + \varepsilon_{it}^K \tag{1.20}$$

$$\ln L_{it} = \nu_{st} + \beta^L \ln \tau_{srt} + \zeta_i + \gamma^L \ln x_{it} + \varepsilon_{it}^L \tag{1.21}$$

In the above specifications, τ_{srt} is a measure of flood risk. The reason why this variable captures only flood risk and not the combined effects of flood risk and actual flooding is that it reflects average disruptions in production capacity under expected flooding. Therefore, the effect of flood risk is identified by the cross-regency differences in the mean probability of flooding, while the effects of actual flood shocks uses the variation induced by annual flood events. There is a plausible endogeneity concern with this setup

if PCU is itself influenced by labor and capital levels. For example, firms with less labor or capital might experience less or more production disruptions during floods, which would impact capacity utilization. One way to address this concern is to instrument the flood risk and actual flooding with some relevant objective measures of flooding. For example, flood index could be a potential IV in this case. However, as discussed in the reduced-form section, flood index potentially captures both the effects, so it will not be a valid IV for either of the two variables. It is also a coarser variable as it only varies across regencies over time, so the analysis could potentially suffer from weak instrument problem (Angrist and Pischke 2009). To partially alleviate the concerns around endogeneity, I look at whether there is heterogeneous effect of flooding on PCU due to the firm size, where the firm size is calculated based on both the labor employment and capital stock measure. The results reported in Table 1.8 in the Appendix suggests that such a heterogeneity in the impact of flooding on PCU is non-existent.

The flood risk variable above uses the estimated parameter values of regency-level flood exposure based on the distribution of firm-level PCU from past years only. This is essential for the analysis to have enough statistical power to identify both β and γ parameters capturing effects of flood risk and flood shocks respectively. Table 1.3 report the results of estimating Equations (1.20) and (1.21) in Columns 5 and 6 respectively. Result in Column 5 suggests that an increase in τ_{srt}, that is, a decrease in flood risk leads to an increase in the capital stock at firm level. In terms of magnitude, a 1% decrease in τ_{srt} leads to a 0.25% decrease in the value of capital stock. Coefficient estimates on flood risk variable in Column 6 suggests that the labor demand increase in response to an increase in flood risk. More precisely, a 1% decrease in τ_{srt} leads to a 0.13% increase in labor employment. Firms tend to reduce investment in capital, which is less flexible and more vulnerable to the effects of flooding, and increase hiring of labor, which is a relatively more flexible input in production. The coefficients on flood shock variable suggests that its impact on both capital and labor inputs is small relative to flood risk. Flood shock affect labor and capital directly, while flood risk enters in the labor decision

only through the capital margin. The results highlight that controlling for the indirect effects of flooding operating through flood risk, the direct impact of flooding on firm-level capital stock and labor employment is limited.

1.6.2 Flood Defenses

One counterfactual exercise that is natural in this setting is the installation of flood defense systems, such as flood barriers and fences in flood-prone areas. This exercise is in the spirit of various mitigation efforts undertaken by both local and central governments in Indonesia through both in-house and international support (Islam et al. 2019). For this *experimental* exercise, I assume installation of flood defenses in the most-affected regencies of Indonesia in terms of flooding. The metric used for classifying these regencies is the average flood exposure, which is estimated using the empirical distribution of firm-level PCU, across years. Flood defense systems reduce the flood exposure of regencies where they are installed by bringing the exposure level down to a lower level.

Flooding is an exogenous event, and the parameter governing it is also independent of equilibrium effects in the model viz. wage adjustments and endogenous entry decisions. Thus, one can quantify the *direct* effects of flood risk and then examine how these effects change once equilibrium forces viz. wage adjustments and endogenous firm entry, are taken into account. For this reason, the analyses are reported in two distinct scenarios outlined below:

1. *Flood Risk Only*: This scenario quantifies the change in direct effects of flood risk before and after the installation of flood defenses. Due to the exogeneity of parameter governing the direct effects of flood risk, it can be inspected separately from the equilibrium adjustments that occur as a result of it.

2. *Equilibrium With Entry*: This scenario quantifies the total change, including both direct and indirect effects of flood risk, before and after the installation of flood defenses. Flood risk affect firm entry decision and equilibrium wage, so account-

ing for these adjustments is essential for capturing the overall benefits of flood defenses.

To quantify the effects of flood defenses on the aggregate output, I compare the aggregate output after the installation of flood defenses to the observed aggregate output. In this thought experiment, flood defenses reduce the flood exposure of the top 20th percentile of regencies by bringing them down to the median level of flood exposure distribution. The flood exposure distribution is generated using average (over years) flood exposure of all the regencies in Indonesia. The counterfactual assigns same flood exposure level to the most-affected regencies in all the years, but different sectors within a regency would still be impacted differently due to their sectoral characteristics. The motivation for using the median benchmark is firstly to be realistic that flooding cannot be completely eliminated, so all regencies should experience some level of flooding in the *constrained* best-case scenario. Secondly, flooding, being a spatial shock by design, primarily creates differences across regencies, and switching off this channel by bringing all treated regencies to the same level could provide insights on the spatial misallocation effects of flooding.

To operationalize this experimental exercise, consider an exogenous change in the flood exposure of regency r from ϕ_{rt} to $\widetilde{\phi}_r$, where $\widetilde{\phi}_r$ is the median value of regency-level flood exposure. Only the top 20th percentile regencies undergo this change in flood exposure from the start of the period, while the remaining regencies remain unaffected in all the years. Figure 1.13 shows the spatial distribution of flood exposure following the installation of flood defenses. As the most-exposed regencies are now more secure from flooding, the minimum value of average ϕ increases for 53 treated regencies out of a total of 266 regencies in the sample. Similarly, Figure 1.14 depicts the spatial distribution of flood risk, which decreases for the top 20th percentile of the most-exposed regencies. For the observed outcomes, Equation (1.16) delivers the expected equilibrium value of aggregate output. The same equation can then be used to write the counterfactual output

after the installation of flood defenses as below:

$$\widetilde{\overline{Y}}_{srt} = \widetilde{\tau}_{srt}^{\frac{\alpha_s \eta_s}{1-\eta_s}} \frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1} \widetilde{\Lambda}_{st} \left(\widetilde{\theta}^*_{srt}\right)^{\frac{1}{1-\eta_s}}$$

There are various ways in which the above output could be compared to the observed output. In the most comprehensive analysis, one can ideally calibrate or estimate all the parameters involved in both the objects and compute the objects in *levels*. However, this exercise would require imposing additional assumptions on the model structure and would also be prone to measurement errors. Therefore, I take the ratio of the two objects, which cancels all the *fixed* terms that are assumed to not change in the counterfactual world.[31] The ratio can be written as follows:

$$\frac{\widetilde{\overline{Y}}_{srt}}{\overline{Y}_{srt}} = \left(\frac{\widetilde{\tau}_{srt}}{\tau_{srt}}\right)^{\frac{\alpha_s \eta_s}{1-\eta_s}} \frac{\widetilde{\Lambda}_{st}}{\Lambda_{st}} \left(\frac{\widetilde{\theta}^*_{srt}}{\theta^*_{srt}}\right)^{\frac{1}{1-\eta_s}}$$

Taking the log of the above ratio, I derive the (log) change in aggregate output as follows:

$$\widetilde{\Omega}_{srt} = \underbrace{\frac{\alpha_s \eta_s}{1-\eta_s} \ln\left(\frac{\widetilde{\tau}_{srt}}{\tau_{srt}}\right)}_{\text{flood risk}} + \underbrace{\ln\left(\frac{\widetilde{\Lambda}_{st}}{\Lambda_{st}}\right) + \frac{1}{1-\eta_s} \ln\left(\frac{\widetilde{\theta}^*_{srt}}{\theta^*_{srt}}\right)}_{\text{wage and entry adjustments}} \tag{1.22}$$

$\widetilde{\Omega}$ captures the change in (log) aggregate output in the counterfactual with respect to the real world. So, a positive $\widetilde{\Omega}$ would mean that the aggregate counterfactual output is higher, and for small changes, the magnitude would represent the percentage increase in output relative to the observed output. The total change in aggregate output can be decomposed into two parts: (A) flood risk and (B) wage and entry adjustments. The first part captures the direct impact of flood risk, while the second part provides estimates of indirect effects due to the equilibrium forces in place. In the results that follow, *Flood Risk Only* scenario reports estimates of (A), while *Equilibrium With Entry* scenario re-

31. The assumption imposed on the equilibrium is that the exogenous change in flood exposure does not affect the production function, entry fixed cost, and productivity distribution parameters.

ports the sum of (A) and (B).

Flood Risk Only. In this scenario, the estimates of the first term in Equation (1.22) is reported. Given that there are two key margins of variation viz. sector and regency, I report results on both the margins.[32] Although flooding is inherently a regional shock, its impact can vary significantly across economic sectors, depending on each sector's specific vulnerabilities to flood risk and flood events. For example, an iron and steel firm and a furniture producer might both be located on the same floodplain in Jakarta, yet the impact of flood events—and how each firm perceives these events to inform their flood risk—could differ greatly owing to their sectoral characteristics. For the aggregation step, I take simple average across all qualifying observations. Figure 1.15 shows the distribution of change in aggregate output due to flood risk across sectors, where the sectors are ordered in increasing order of their capital intensities from left to right. First, on average, all sectors derive direct benefits in terms of aggregate output from the installation of flood defenses in the top 20th percentile of most flood-affected regencies in Indonesia. This confirms a well-known empirical fact that industries in Indonesia are primarily clustered in flood-prone areas. Such clustering in high-risk zones is due to various factors, including historical path dependence, agglomeration externalities, and higher demand due to richer population. Second, the benefits increase moving from left to right on the graph, thereby suggesting that the sectors using capital-intensive technology for production reap more direct rewards from such protective investments. This is because the sectors that rely on capital heavily face more distortions in their production decisions because of flood risk, and flood defenses help alleviate those distortions. On average, the aggregate annual sector-level output increases by 7%, but with significant heterogeneity across sectors. Figure 1.16 illustrates the spatial distribution of changes in aggregate output resulting from reduced flood risk across regencies. On average, the aggregate annual output increases by 16% in the treated regencies, though the range is considerable,

32. There is also time variation but it is not key part of the analysis as all flood defense systems everywhere are installed at the start of the period.

varying from 9% to 30%. The sectoral composition of each regency plays a significant role in shaping these outcomes; for instance, regencies with a higher concentration of capital-intensive sectors tend to experience greater gains in aggregate output compared to those with a higher proportion of less capital-intensive sectors. For example, regencies in the West Java province, including Jakarta, see disproportionately higher gains in their expected aggregate output.

Equilibrium With Entry. This scenario captures the total change in aggregate output, the sum of the two parts outlined in Equation (1.22) after the flood defenses intervention. Understanding the direct effects of protective investments, such as flood defenses, is important to justify the monetary costs involved in their construction and maintenance. However, the indirect effects could potentially increase these benefits further or decrease them depending on the margin looked at in calculating these effects. One such margin that is important to consider before such interventions are commissioned is how potential firms, which are still out of the market, would respond to the changes resulting from interventions. Installation of flood defenses potentially increase the pool of new firms entering into these safer areas, but that would also increase competition among incumbent firms for the scarce resources employed in the production of final goods.

Figure 1.17 shows the distribution of total change in aggregate output across manufacturing sectors; results of the previous scenario are also included side-by-side for comparison. The sum of direct and indirect effects is positive for all the sectors. On average, aggregate output increases by 4% from the observed outcome after the installation of flood defenses. However, accounting for the equilibrium forces of wage adjustment and firm selection on entry decreases the aggregate gains for all the manufacturing sectors relative to the previous scenario, which captures only the direct effects of flood defenses. Flood defenses, by design, make risky regencies safer for economic operations. Since the potential market entrants make their entry decision on the expected profits, which depend on the anticipation of flooding, installation of flood defenses increase these ex-

pected profits. This means that less productive firms, which were unable to enter earlier due to high flood risk, would be able to enter into these markets now. Due to the larger mass of incumbent firms in equilibrium, the competition for the scarce labor inputs also increases, thereby exerting an upward pressure on the equilibrium wages. Therefore, the increased competition driving wages upwards combined with the reduction in firm selection on entry decreases the aggregate output in equilibrium relative to the direct impact on aggregate output due to flood risk after the installation of flood defenses.[33] Figure 1.18 shows the distribution of total change in aggregate output across treated regencies. Similar to the sectoral distribution above, total benefits of flood defenses decrease for all regencies when indirect equilibrium effects are accounted in the change calculation. Overall, the yearly aggregate output increases by 9%, which is about half of the gains from considering direct effects only.

1.7 Conclusion

This chapter investigates the impact of flooding on the manufacturing sector of a low- and middle-income country. Using historical data on floods, I show that severe floods are associated with significant reductions in aggregate measures of production inputs and economic output. Though at the firm level, the value of capital stock declines and hiring of temporary labor increases, with the risk of floods also acting as a deterrent to firm entry. However, in regions with persistent flood shocks, both the actual damages from floods and the anticipatory adjustments in response to evolving perception of flood risk, could play a significant role in generating these results. To address this, I develop a model of firms with endogenous entry and flood risk affecting capital installation decisions, to assess the effects of different elements of flooding on firm behavior. I provide a micro-

33. One concern might be that the results are driven by regency characteristics where the sectors are located. To address this issue, I plot the same figures keeping regencies constant across sectors. There are only six regencies where all 25 3-digit ISIC sectors are located, so the average is taken over these six regencies only. Figures 1.30 and 1.31 show the new graphs, which point to the same qualitative findings as reported in the main figures.

foundation for understanding how flood risk and flood shocks interact with firm behavior by linking them to entry decision and input choice. The equilibrium analysis reveals that perceived flood risk, rather than actual flood shocks, has more significant impact on firm behavior. I conduct a counterfactual analysis in the spirit of building flood defenses to secure flood-prone regions and find that there are large gains in aggregate output from such an intervention, but equilibrium adjustments, in particular, upward pressure on wages and the entry of less productive firms, reduce these gains by half.

The theoretical framework developed in this chapter could be adapted to examine other aggregate and firm-level distortions generated by anticipation, such as AI adoption, technological disruptions, and policy-induced market frictions. Incorporating firm entry and exit dynamics into the model, while maintaining its tractability for policy analysis, could be useful in various contexts, such as assessing the impact of global trade disruptions, regional economic integration, and industry regulations.

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Asian Development Bank. 2019. *Policies to Support the Development of Indonesia's Manufacturing Sector During 2020–2024.* Technical report. A Joint ADB–BAPPENAS Report. https://dx.doi.org/10.22617/TCS199910-2.

Balboni, Clare. 2024. "In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities." *Working Paper.*

Balboni, Clare, Johannes Boehm, and Mazhar Waseem. 2023. "Firm adaptation and Production Networks: Structural Evidence from Extreme Weather Events in Pakistan." *Working Paper.*

Banerjee, Abhijit, and Esther Duflo. 2005. "Growth Theory through the Lens of Development Economics." Chap. 07 in *Handbook of Economic Growth,* 1st ed., edited by Philippe Aghion and Steven Durlauf, vol. 1, Part A, 473–552.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. "Cross-Country Differences in Productivity: The Role of Allocation and Selection." *American Economic Review* 103 (1): 305–34.

Bento, Pedro, and Diego Restuccia. 2017. "Misallocation, Establishment Size, and Productivity." *American Economic Journal: Macroeconomics* 9 (3): 267–303.

Besley, Timothy J., Isabelle A. Roland, and John Van Reenen. 2020. "The Aggregate Consequences of Default Risk: Evidence from Firm-level Data." *NBER Working Paper* 26686.

Bilal, Adrien, and Esteban Rossi-Hansberg. 2023. "Anticipating Climate Change Across the United States." *NBER Working Paper* 31323.

Blalock, Garrick, and Paul J. Gertler. 2008. "Welfare Gains from Foreign Direct Investment through Technology Transfer to Local Suppliers." *Journal of International Economics* 78 (2): 402–421.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies* rdae007.

Castro-Vincenzi, Juanma. 2024. "Climate Hazards and Resilience in the Global Car Industry." *Working Paper.*

Collard-Wexler, Allan, and Jan De Loecker. 2016. "Production Function Estimation and Capital Measurement Error." *NBER Working Paper* 22437.

Cruz, José-Luis, and Esteban Rossi-Hansberg. 2021. "The Economic Geography of Global Warming." *University of Chicago, Becker Friedman Institute for Economics Working Paper,* nos. 2021-130.

Desmet, Klaus, Robert E. Kopp, Scott A. Kulp, Dávid Krisztián Nagy, Michael Oppenheimer, Esteban Rossi-Hansberg, and Benjamin H. Strauss. 2021. "Evaluating the Economic Cost of Coastal Flooding." *American Economic Journal: Macroeconomics* 13 (2): 444–86.

Financial Services Authority of Indonesia. 2024. *Indonesia Banking Statistic - January 2024.* Technical report. https://www.ojk.go.id/en/kanal/perbankan/data-dan-statistik/statistik-perbankan-indonesia/Pages/Indonesia-Banking-Statistic---January-2024.aspx..

Geerolf, François. 2017. “A Theory of Pareto Distributions.” *Working Paper.*

Gopinath, Gita, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. 2017. “Capital Allocation and Productivity in South Europe*.” *The Quarterly Journal of Economics* 132 (4): 1915–1967.

Government of Republic of Indonesia. 2007. *Indonesia Country Report: Climate Variability and Climate Changes, and Their Implication.* Technical report. Ministry of Environment, Jakarta.

Hopenhayn, Hugo A. 1992. “Entry, Exit, and Firm Dynamics in Long Run Equilibrium.” *Econometrica* 60 (5): 1127–1150.

———. 2014. “Firms, Misallocation, and Aggregate Productivity: A Review.” *Annual Review of Economics* 6:735–770.

Hsiao, Allan. 2024. “Sea Level Rise and Urban Adaptation in Jakarta.” *Working Paper.*

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. “Misallocation and Manufacturing TFP in China and India.” *The Quarterly Journal of Economics* 124 (4): 1403–1448.

IPCC. 2023. *Climate Change 2023: Synthesis Report Summary for Policymakers.* Technical report. Intergovernmental Panel on Climate Change.

Islam, S., C. Chu, J. C. R. Smart, and L. Liew. 2019. “Integrating Disaster Risk Reduction and Climate Change Adaptation: A Systematic Literature Review.” *Climate and Development* 12 (3): 255–267.

Jia, Ruixue, Xiao Ma, and Victoria Wenxin Xie. 2022. “Expecting Floods: Firm Entry, Employment, and Aggregate Implications.” *NBER Working Paper* 30250.

Kocornik-Mina, Adriana, Thomas K. J. McDermott, Guy Michaels, and Ferdinand Rauch. 2020. “Flooded Cities.” *American Economic Journal: Applied Economics* 12 (2): 35–66.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies* 70 (2): 317–341.

Lucas, Robert E. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics* 9 (2): 508–523.

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *American Economic Review* 104 (2).

Najibi, Nasser, and Naresh Devineni. 2018. "Recent Trends in the Frequency and Duration of Global Floods." *Earth System Dynamics* 9 (2): 757–783.

Nath, Ishan. 2024. "Climate Change, the Food Problem, and the Challenge of Adaptation through Sectoral Reallocation." *Journal of Political Economy* forthcoming.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (6): 1263–1297.

Patel, Dev. 2024. "Floods." *Working Paper.*

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments." *Review of Economic Dynamics* 11 (4): 707–720.

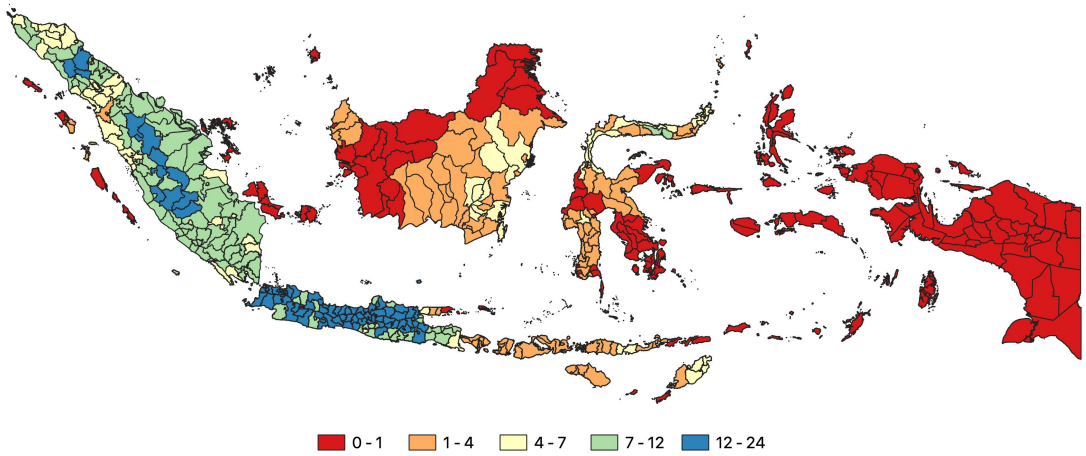
Ridhwan, Masagus M., Henri L.F. de Groot, Piet Rietveld, and Peter Nijkamp. 2012. "Regional Interest Rate Variations: Evidence from the Indonesian Credit Markets." *Tinbergen Institute Discussion Papers* 12-073/3.

Tellman, B., J. A. Sullivan, C. Kuhn, A. J. Kettner, C. S. Doyle, G. R. Brakenridge, T. A. Erickson, and D. A. Slayback. 2021. "Satellite Imaging Reveals Increased Proportion of Population Exposed to Floods." *Nature* 596:80–86.

World Risk Report. 2023. *World Risk Report 2023.* Technical report. Bündnis Entwicklung Hilft, Ruhr University Bochum – Institute for International Law of Peace and Conflict.

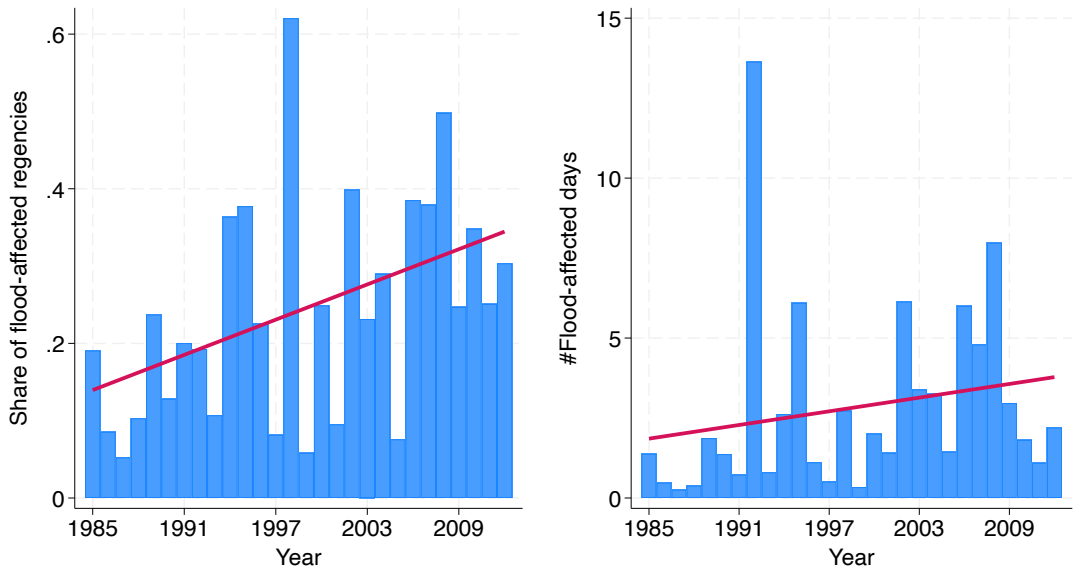
Main Figures and Tables

Figure 1.1: Total count of large floods affecting regencies in 1985-2012 period



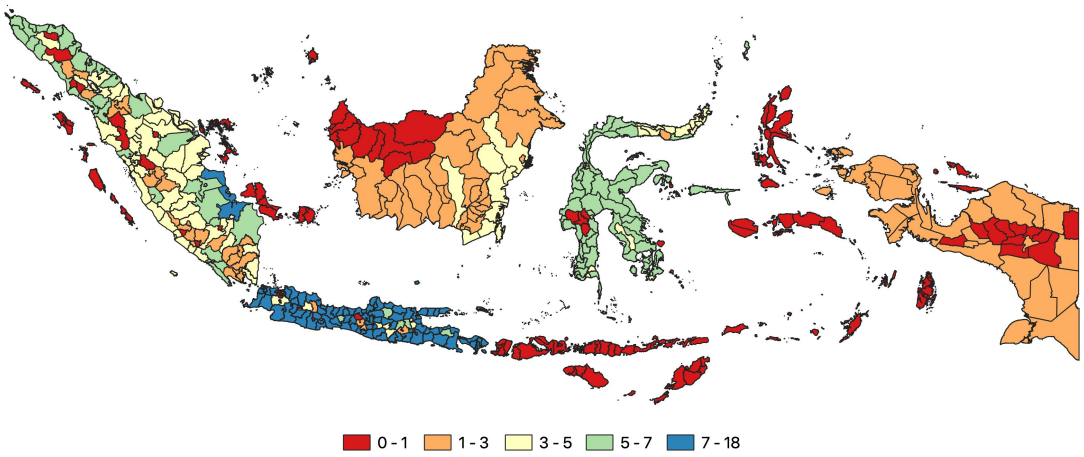
Notes: The map shows the total number of large flood events (as per the DFO archive of large flood events) that each Indonesian regency got affected by during the period 1985-2012. The internal boundaries are regency boundaries, and the legend entries represents number of large flood events experienced during the period 1985-2012. Regency boundaries correspond to the administrative divisions for the year 2020.

Figure 1.2: Number of flood-affected regencies and average count of flood days



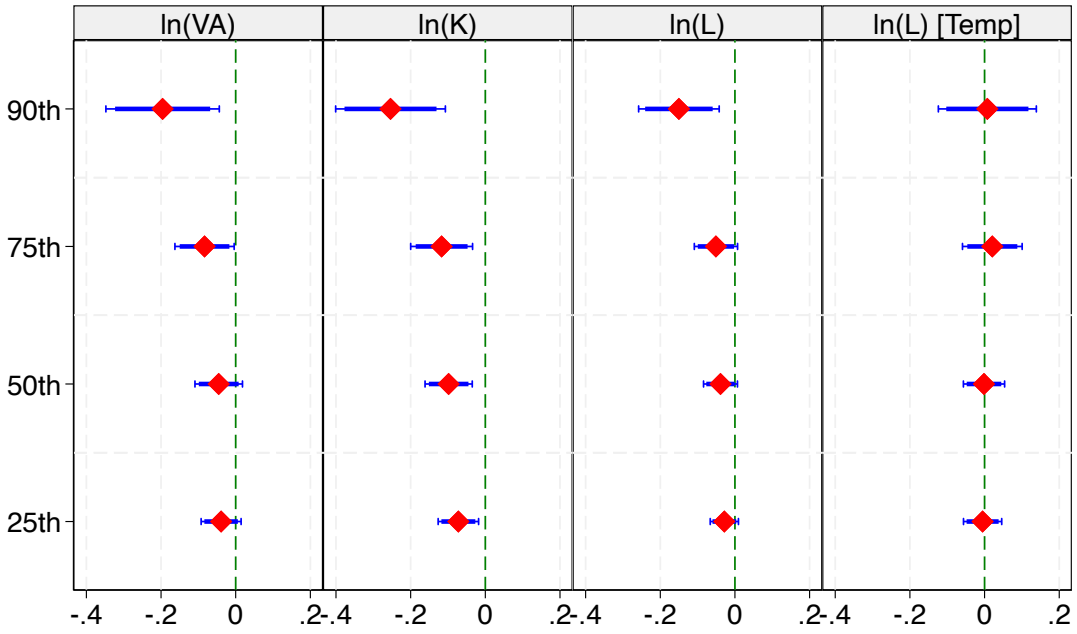
Notes: The graphs show the flooding trends within Indonesia using the information from the DFO archive of large flood events. The left (right) figure plots the count of flood-affected regencies (days) in each year for the period 1985-2012. Both the variables are trending positively over the years.

Figure 1.3: Total count of large floods based on satellite observation for 2002-18



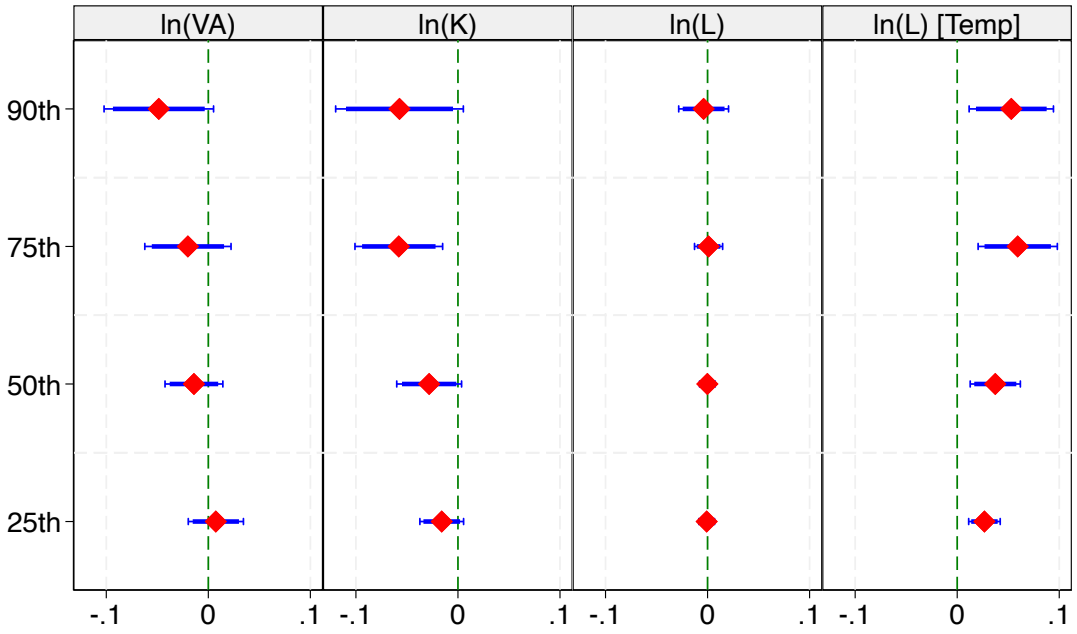
Notes: The map shows the total number of large flood events from the DFO archive of large flood events, which are confirmed using satellite observations in Tellman et al. (2021) for the period 2002-18. The internal boundaries are regency boundaries, and the legend entries represents number of large flood events experienced during the period 2002-18. Regency boundaries correspond to the administrative divisions for the year 2020.

Figure 1.4: Effect of flooding on sector-regency-level variables



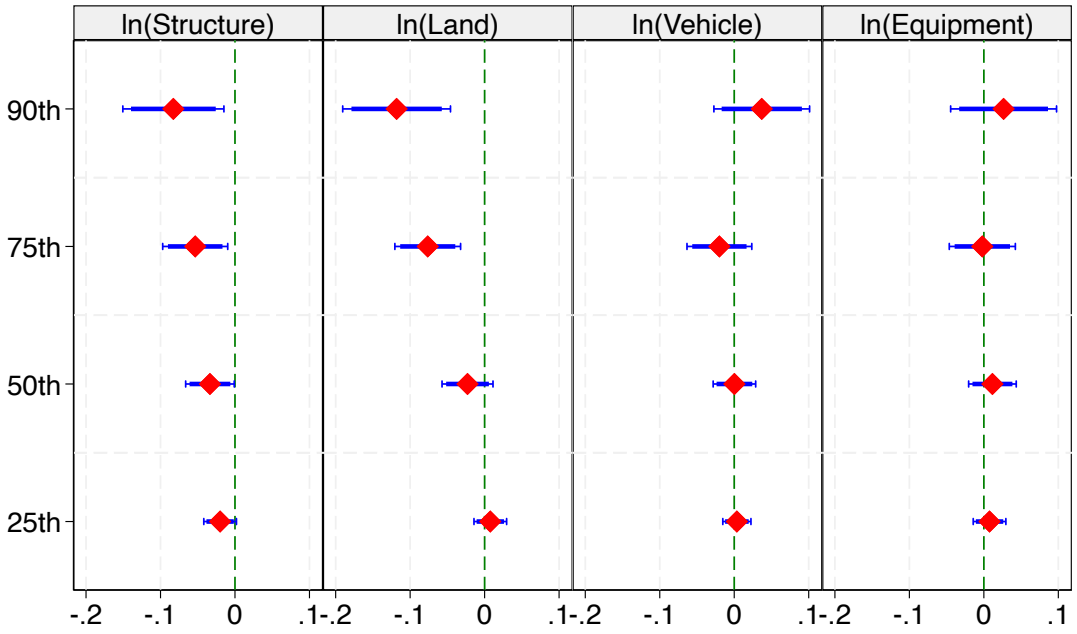
Notes: The graph presents the results of estimating Equation (1.1) for aggregate variables i.e., logarithms of value-added (left), capital stock (second-left), permanent labor employment (second-right), temporary labor employment (right). To get to the aggregate variables from firm-level information, following steps are undertaken. First, The un-logged version of all the monetary variables are deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values. Second, the tails on both ends of the resulting variables are trimmed by 1% for each year to address measurement error issues. Third, the variables are then summed across sector-regency for each year using labor share weights. Finally, the variables are log-transformed and used in the regressions. The labels on y-axis represent the percentiles of flood index for which dummy is used in the regression. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.5: Effect of flooding on firm-level variables



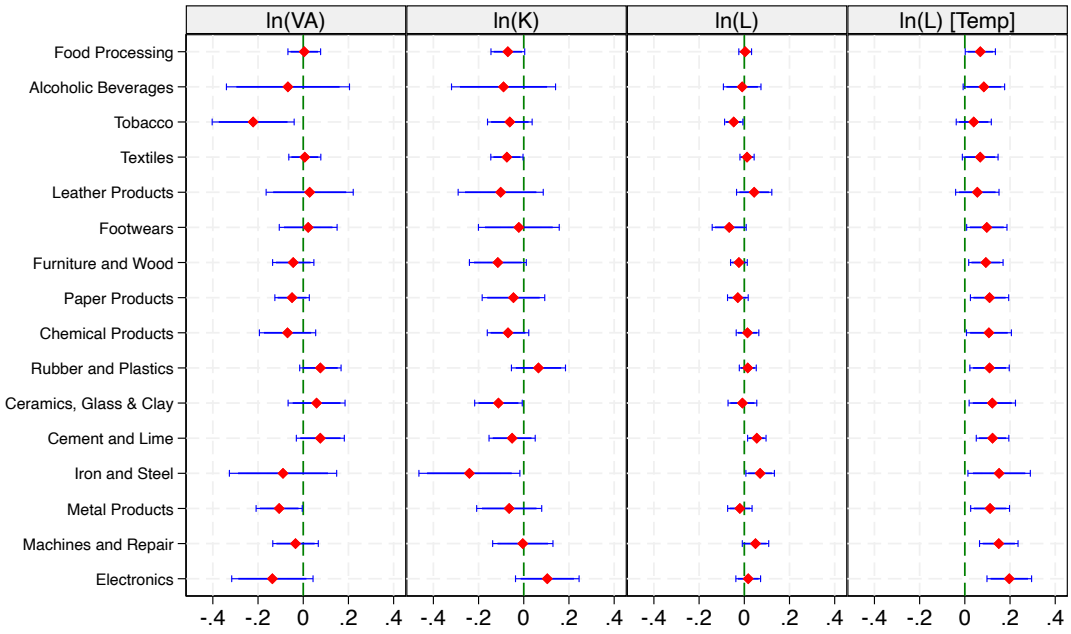
Notes: The graph presents the results of estimating Equation (1.2) for firm-level variables i.e., logarithms of value-added (left), capital stock (second-left), permanent labor employment (second-right), temporary labor employment (right). The un-logged version of all the monetary variables have been deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values and the log-transformed variables are trimmed by 1% for each year to address measurement error issues. The labels on y-axis represent the percentiles of flood index for which dummy is used in the regression. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.6: Effect of flooding on firm-level capital categories



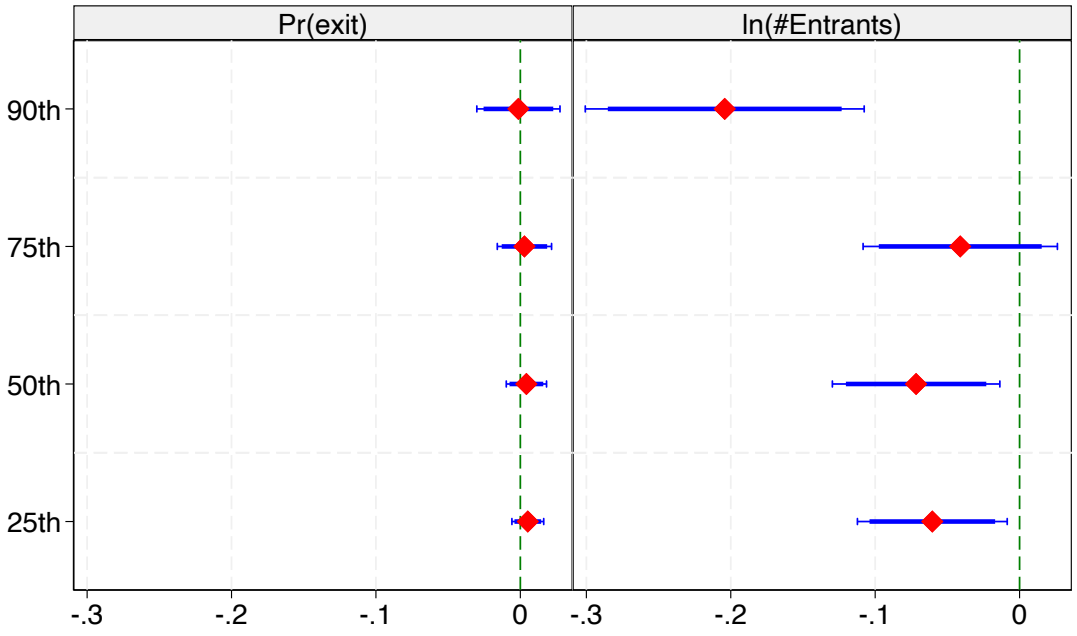
Notes: The graph presents the results of estimating Equation (1.2) for four different capital categories at the firm level. The un-logged version of all the monetary variables have been deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values and the log-transformed variables are trimmed by 1% for each year to address measurement error issues. Going left to right, first plot reports the results for value of structures, which include buildings and all man-made constructions to support the manufacturing activities within the firm. Second plot shows results on land, which is the total value of land occupied by the manufacturing firm. Third plot reports results on the value of vehicles and other transportation equipment owned by the firm. Last plot shows results for value of machinery and other production equipment employed in the firm. As mentioned in the data section, the reporting on different capital categories is not consistent over time, and that is why the number of observations are different across all four columns. The labels on y-axis represent the percentiles of flood index for which dummy is used in the regression. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.7: Effect of 90th percentile floods on firm-level variables by sectors



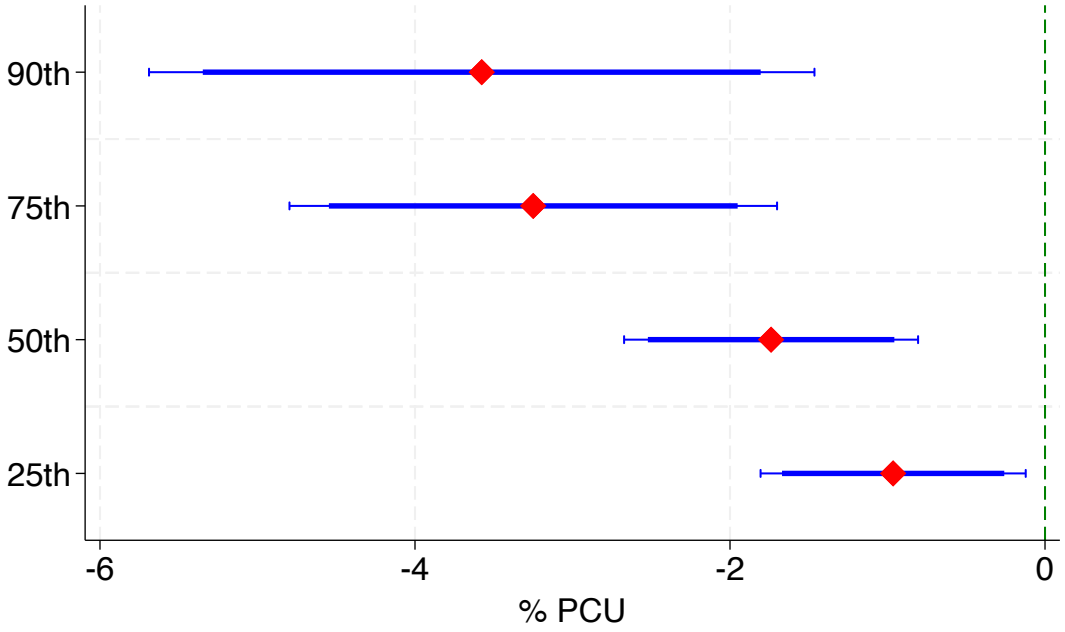
Notes: The graph presents the results of estimating Equation (1.3) for firm-level variables i.e., logarithms of value-added (left), capital stock (second-left), permanent labor employment (second-right), temporary labor employment (right) using the 90th percentile flood dummy. The un-logged version of all the monetary variables have been deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values and the log-transformed variables are trimmed by 1% for each year to address measurement error issues. The labels on y-axis represent the 2-digit ISIC manufacturing sectors. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.8: Effect of flooding on firm exit and entry



Notes: The graphs present results on firm exit and entry. Left graph presents the results of estimating Equation (1.4) for firm exit dummy, where the dummy variable takes a value of 1 in the last year of firm observation in the data. Right graph presents the results of estimating Equation (1.1) with the logarithm of number of new firms entering in a sector-regency in a given year as the dependent variable. The labels on y-axis represent the percentiles of flood index for which dummy is used in the regression. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.9: Effect of flooding on firm capacity utilization



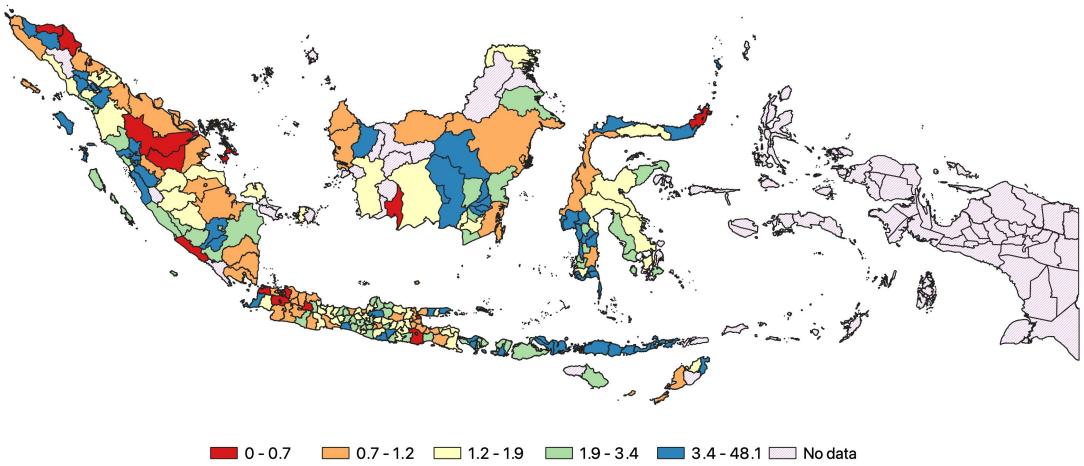
Notes: The graph presents the results of estimating Equation (1.2) for firm-level production capacity utilization (PCU). PCU measures the percentage of the potential firm capacity, in terms of production, that is realized in a given year. The labels on y-axis represent the percentiles of flood index for which dummy is used in the regression. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Table 1.1: Summary of model parameters

	(1) Parameter	(2) Level	(3) Value	(4) Method/Source
ϕ	Flooding shape	Regency-Year	-	MLE on firm-level PCU data
α	Output elasticity	Sector	-	PF estimation (Levinsohn and Petrin (2003))
η	Returns to scale	Sector	-	PF estimation (Levinsohn and Petrin (2003))
$\overline{\theta}$	Productivity scale	Regency	-	Aggregate regency manufacturing value-added
ξ	Productivity shape	Aggregate	4.514	MLE on firm-level value-added data

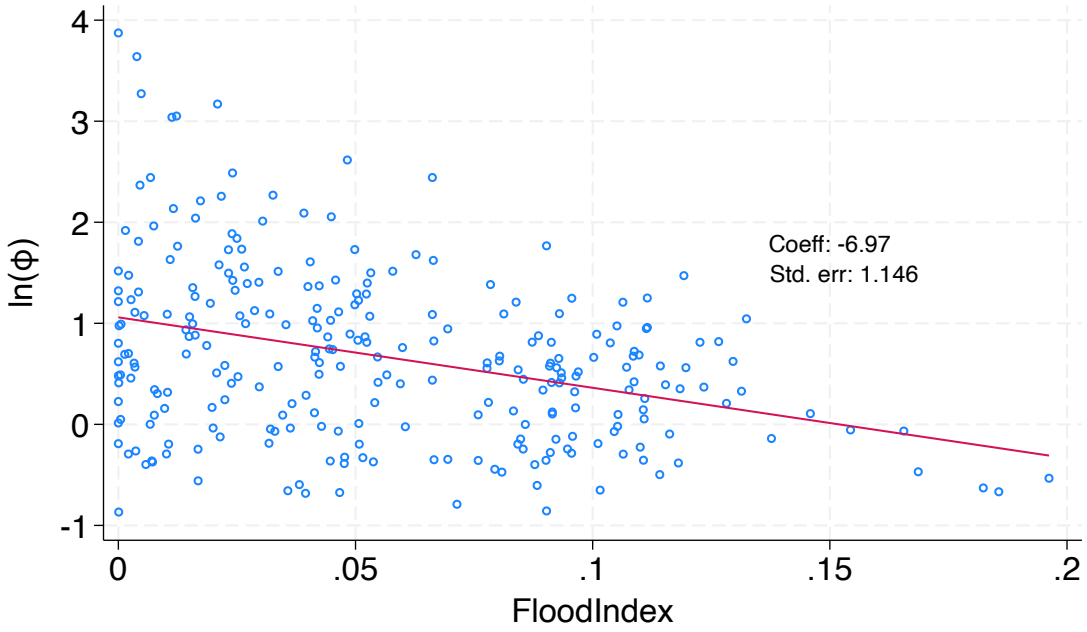
Notes: The table presents the summary of model parameters' estimation or calibration exercise. Values of some of the parameters are not included in the above table as they are too many in number to report. However, their estimation method/ calibration source is discussed in detailed in the Estimation section.

Figure 1.10: Distribution of flood exposure across regencies



Notes: The map shows the distribution of average (over years) regional flood exposure as captured by the regency-level shape parameters, ϕ_{rt}, estimated using the firm-level production capacity utilization data. In line with the properties of Pareto distribution, smaller value for a regency suggests that the regency, on average, faces more extreme flooding over time. The regencies for which flood exposure could not be estimated are shaded in pink and coded as "No data". Regency boundaries correspond to the administrative divisions for the year 1990.

Figure 1.11: Pareto tail exponent versus flood index across regencies



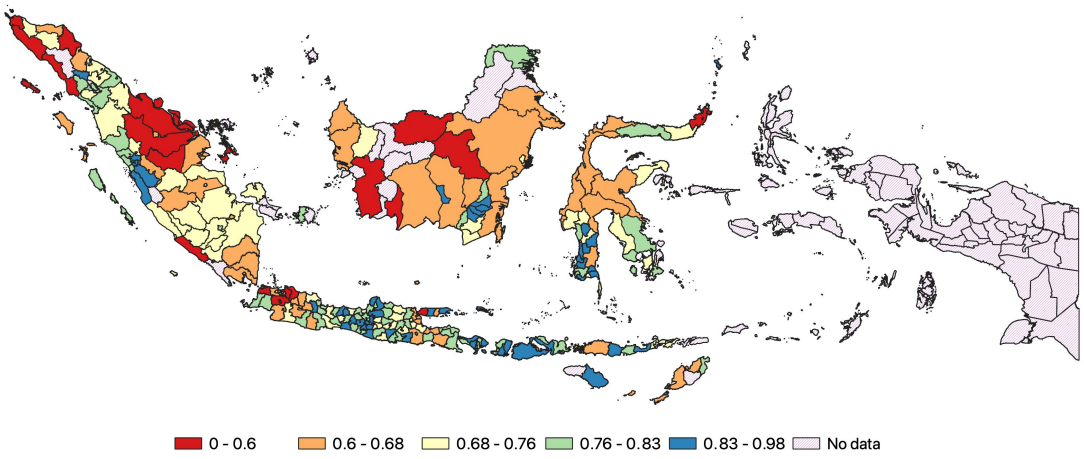
Notes: The graph plots the regency-level average regional flood exposure as captured by the regency-level shape parameters, ϕ_{rt} against the average regency-level flood index. The averages are taken across all the years in sample i.e., 1990-2012. Each dot represents one regency in Indonesia.

Table 1.2: Sectoral production function parameters

Industry name	3-digit ISIC	α_s	η_s
Food Processing	311	0.213	0.700
Food Processing 2	312	0.261	0.662
Cigarettes and Tobacco	314	0.253	0.487
Textiles	321	0.204	0.613
Leather Products	323	0.188	0.770
Manufacture of Footwear	324	0.150	0.668
Wood Products	331	0.230	0.710
Furniture	332	0.134	0.693
Paper Products	341	0.267	0.562
Paper Products, Finished	342	0.114	0.704
Chemical Products, Industrial	351	0.304	0.546
Chemical Products, Household	352	0.180	0.596
Rubber Products	355	0.104	0.625
Plastic Wares	356	0.229	0.652
Ceramics	361	0.341	0.593
Glass Products	362	0.295	0.705
Cement and Lime	363	0.223	0.687
Structural Clay Products	364	0.193	0.773
Other Non Metal Mineral Products	369	0.201	0.646
Basic Iron and Steel	371	0.265	0.742
Metal Products, Finished	381	0.194	0.744
Machines and Repair	382	0.289	0.707
Electronics	383	0.138	0.718
Motor Vehicles	384	0.268	0.601
Other Manufacturing	390	0.172	0.778

Notes: The table reports the computed values of production function parameters for each 3-digit ISIC sector. Using data from Table 1.7, the computation uses the following formulae for scale parameter, $\eta_s = ln(L)coeff + ln(K)coeff$ and index on capital, $\alpha_s = \frac{ln(K)coeff}{\eta_s}$.

Figure 1.12: Distribution of flood risk across regencies



Notes: The graph plots the distribution of flood risk as captured by τ_{srt} variable, where $\tau_{srt}(\phi) \equiv \frac{\phi_{rt}}{\phi_{rt}+\alpha_s\eta_s/(1-(1-\alpha_s)\eta_s)}$ captures distortions introduced in the optimal capital installation decisions due to flooding. Both the regency-level flood exposure, ϕ_{rt} and production function parameters for each 3-digit ISIC sector (α_s, η_s) are estimated. Lower values of τ_{srt} suggest larger capital distortions due to flooding. The regencies for which flood exposure could not be estimated are shaded in pink and coded as "No data". Regency boundaries correspond to the administrative divisions for the year 1990.

Table 1.3: Effect of flood shock and flood risk on capital and labor

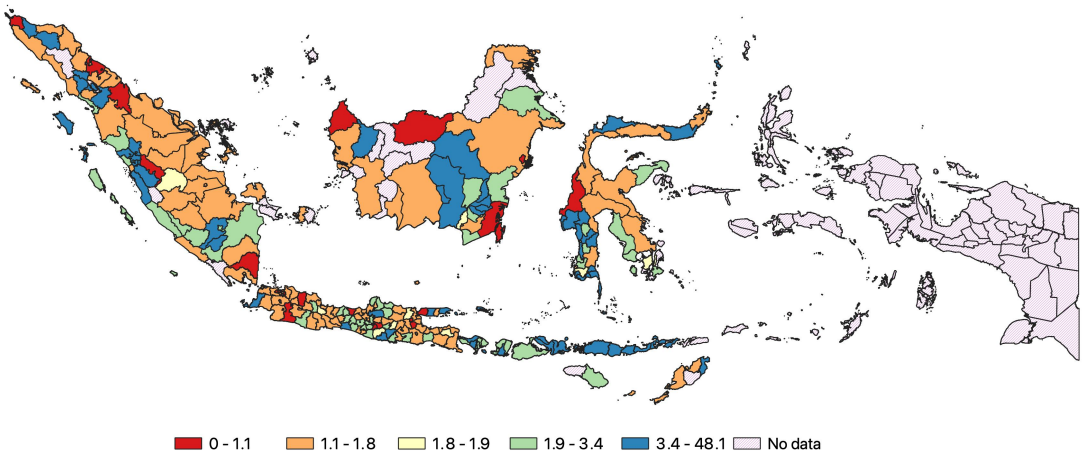
	(1) ln(K)	(2) ln(L)	(3) ln(K)	(4) ln(L)	(5) ln(K)	(6) ln(L)
ln(τ)	0.259***	-0.127***			0.249***	-0.129***
	(0.082)	(0.036)			(0.082)	(0.036)
ln(x)			-0.008***	-0.002***	-0.007***	-0.002***
			(0.001)	(0.000)	(0.001)	(0.000)
Observations	316,788	316,788	330,577	330,577	316,610	316,610
Dep. var mean	8.656	4.131	8.659	4.131	8.656	4.132
Firm FE	Y	Y	Y	Y	Y	Y
3-digit ISIC $\times$ year FE	Y	Y	Y	Y	Y	Y

Standard errors clustered at the sector-regency level are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

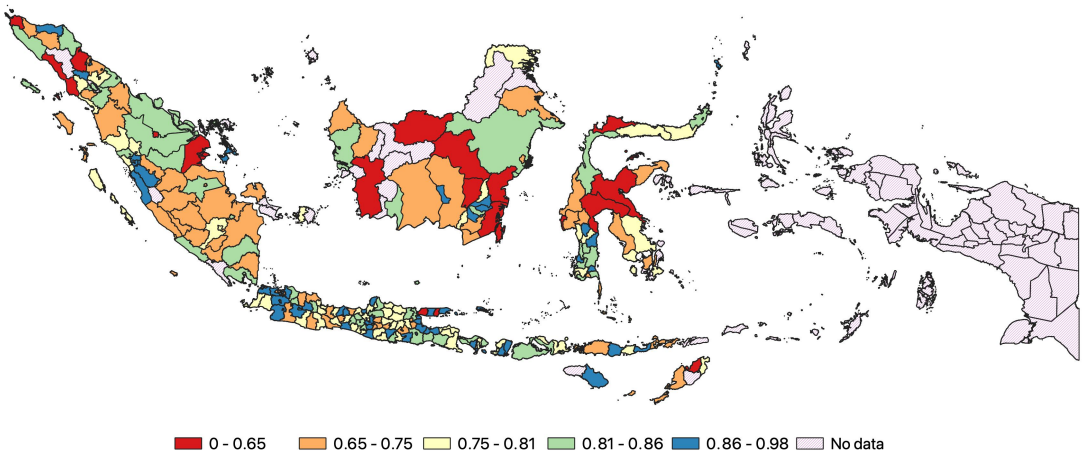
Notes: The table presents the results of estimating Equations (1.20) and (1.21) in Column 5 and 6 respectively. The regency-level flood exposure component of flood risk are estimated using the empirical distribution of PCU across firms within a regency for the past years. The un-logged version of value of capital stock has been deflated by the wholesale price index at the 5-digit ISIC level to reflect its real values and the log-transformed variables are trimmed by 1% for each year to address measurement error issues. Results reported both columns control for firm and sector $\times$ year fixed effects. Standard errors are clustered at the sector-regency level.

Figure 1.13: Distribution of flood exposure across regencies after flood defenses



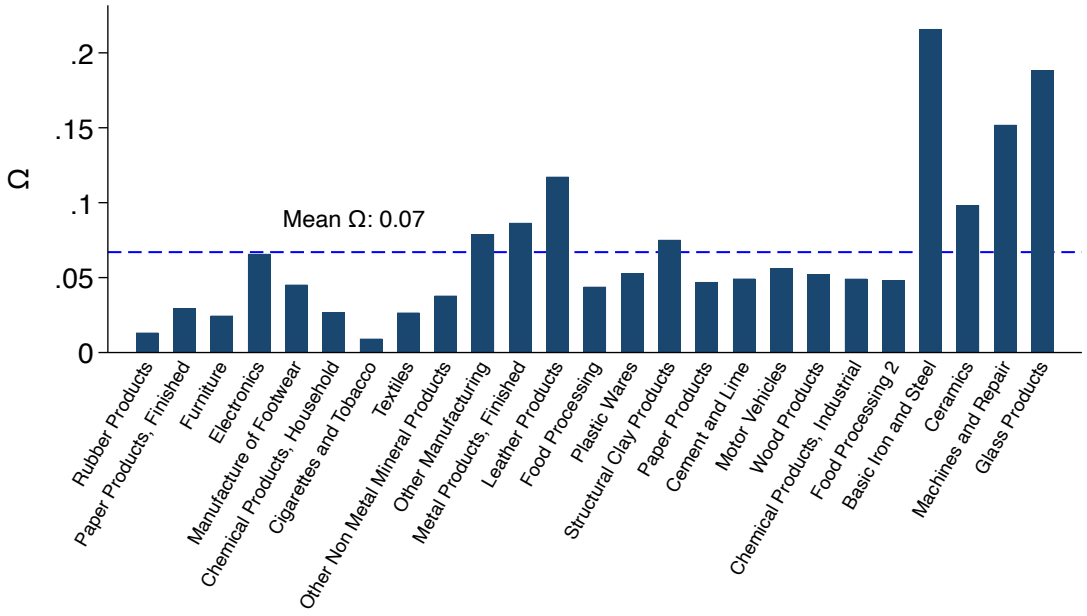
Notes: The map shows the distribution of average (over years) regional flood exposure as captured by the regency-level shape parameters, $\widetilde{\phi}_{rt}$, estimated using the firm-level production capacity utilization data after the installation of flood defenses. In line with the properties of Pareto distribution, smaller value for a regency suggests that the regency, on average, faces more extreme flooding over time. The regencies for which flood exposure could not be estimated are shaded in pink and coded as "No data". Regency boundaries correspond to the administrative divisions for the year 1990.

Figure 1.14: Distribution of flood risk across regencies after flood defenses



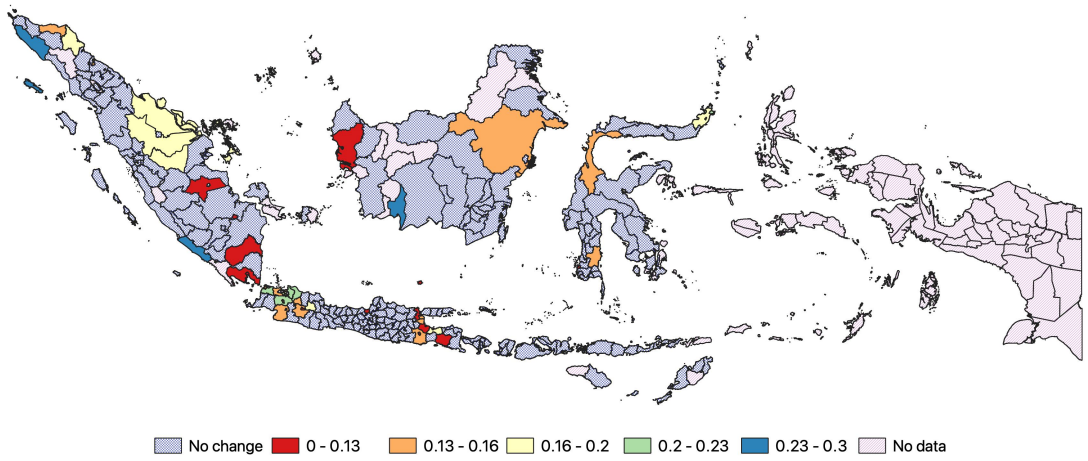
Notes: The graph plots the distribution of flood risk as captured by $\widetilde{\tau}_{srt}$ variable, where $\widetilde{\tau}_{srt}(\phi) \equiv \frac{\widetilde{\phi}_{rt}}{\widetilde{\phi}_{rt}+\alpha_s\eta_s/(1-(1-\alpha_s)\eta_s)}$ captures distortions introduced in the optimal capital installation decisions due to flooding after the installation of flood defenses. Both the regency-level flood exposure, $\widetilde{\phi}_{rt}$ and production function parameters for each 3-digit ISIC sector (α_s, η_s) are estimated. Lower values of $\widetilde{\tau}_{srt}$ suggest larger capital distortions due to flooding. The regencies for which flood exposure could not be estimated are shaded in pink and coded as "No data". Regency boundaries correspond to the administrative divisions for the year 1990.

Figure 1.15: Change in output across sectors due to flood risk



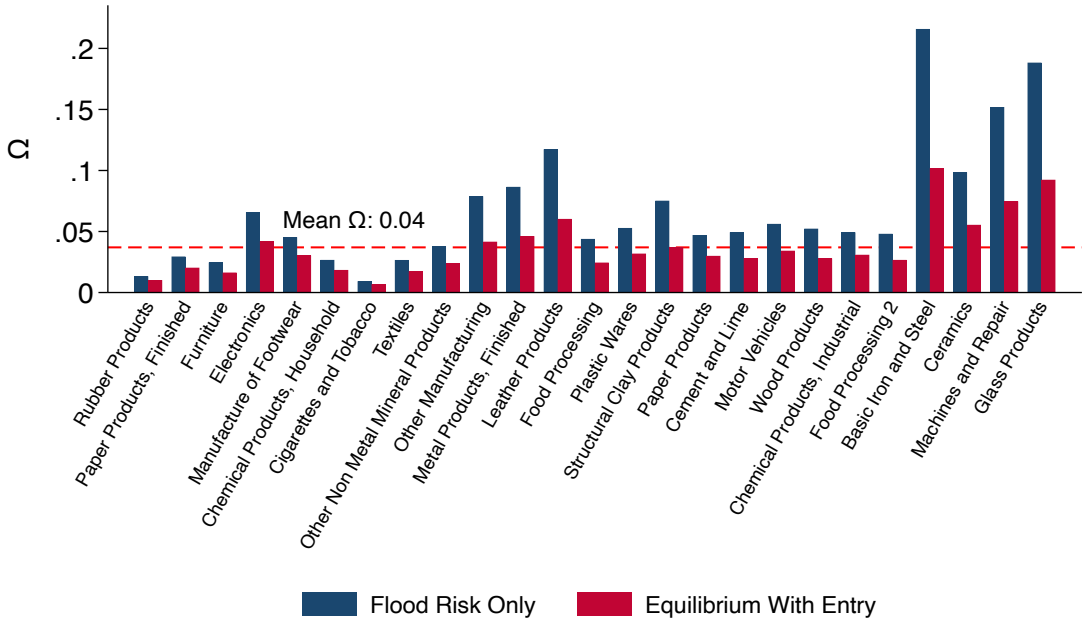
Notes: The graph plots the (log) change in aggregate output due to flood risk as outlined in Equation (1.22) across 3-digit ISIC sectors. This represents the Flood Risk Only scenario where in the counterfactual world with flood defenses, all the regencies above 80th percentile on the flood exposure distribution are assigned the median value of the distribution. The sectors are ranked from left to right in the increasing order of their respective capital intensities.

Figure 1.16: Change in output across regencies due to flood risk



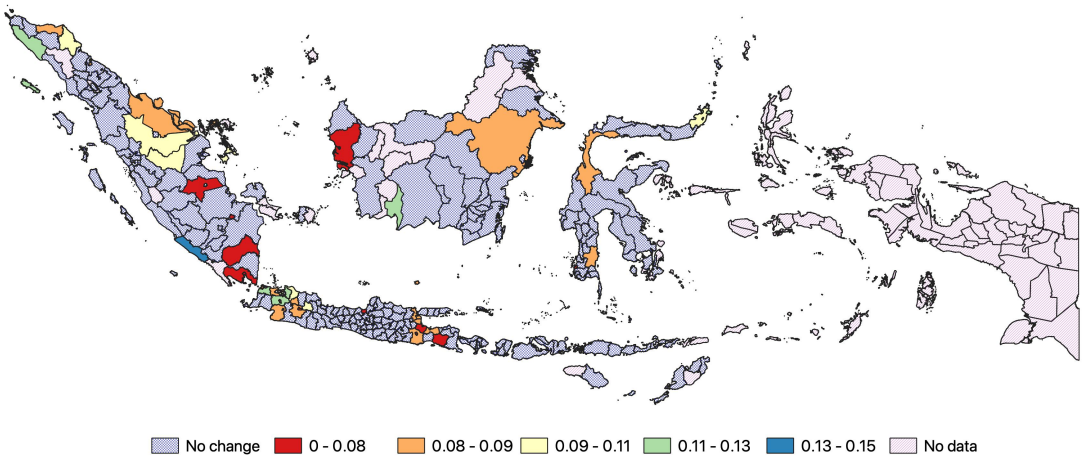
Notes: The map shows the (log) change in aggregate output due to flood risk as outlined in Equation (1.22) across regencies. This represents the Flood Risk Only scenario where in the counterfactual world with flood defenses, all the regencies above 80th percentile on the flood exposure distribution are assigned the median value of the distribution. The regencies for which flood exposure could not be estimated are shaded in pink and coded as "No data", while those regencies that remain unaffected by the flood defenses installation are shaded in blue and coded as "No change". The remaining legend values represent percentage change in aggregate output due to reduction in flood risk after the installation of flood defenses. Regency boundaries correspond to the administrative divisions for the year 1990.

Figure 1.17: Change in output across sectors due to flood risk and equilibrium



Notes: The graph plots the (log) change in aggregate output due to flood risk in blue and sum of (log) change in aggregate output due to flood risk and (log) change in aggregate output due to equilibrium wage adjustments and firm entry in red as outlined in Equation (1.22) across 3-digit ISIC sectors. In the counterfactual world with flood defenses, all the regencies above 80th percentile on the flood exposure distribution are assigned the median value of the distribution. The sectors are ranked from left to right in the increasing order of their respective capital intensities.

Figure 1.18: Change in output across regencies due to flood risk and equilibrium



Notes: The map shows the sum of (log) change in aggregate output due to flood risk and (log) change in aggregate output due to equilibrium wage adjustments and firm entry as outlined in Equation (1.22) across regencies. This represents the Equilibrium With Entry scenario where in the counterfactual world with flood defenses, all the regencies above 80th percentile on the flood exposure distribution are assigned the median value of the distribution. The regencies for which flood exposure could not be estimated are shaded in pink and coded as "No data", while those regencies that remain unaffected by the flood defenses installation are shaded in blue and coded as "No change". The remaining legend values represent percentage change in aggregate output due to reduction in flood risk and accounting for equilibrium adjustments after the installation of flood defenses. Regency boundaries correspond to the administrative divisions for the year 1990.

Appendix

1.8 Appendix: Long-run Effects of Flooding

1.8.1 Effect of Flooding on Economic Variables

Econometric Model

As discussed in the main section of the chapter, employing an event-study framework to estimate the long-term effects of flooding is not suitable in this context. Moreover, our interest lies not in assessing the impact of individual flood shocks, but rather in understanding the collective effects of flooding experienced by a region over an extended period of time. To achieve this objective, I look at the relationship between cumulative flooding and the long-term differences in aggregate variables and firm entry and exit. The long-run period is defined as starting from 1994 and ending in 2008, so it is representative of 15 years duration.[34]

First, cumulative flood shocks are obtained by aggregating innovations in flood index at the regency level over 1994-2008 period. Flood innovations are residuals obtained

34. There is no particular reason for choosing this epoch, except that it coincides with some of the major flood events in Indonesia, including the 2000 Sumatra floods, which killed 120 and affected around 600,000 people and 2007 Jakarta Floods, which resulted in $850 million worth of monetary loss, 80 deaths, and over 500,000 human displacements. Results are robust to choosing alternate epochs.

from the following linear regression on flood data for 1985-2012 period:[35]

$$FloodIndex_{rt} = \zeta_r + \chi_t + \varepsilon_{rt} \tag{1.23}$$

The residuals, $\widehat{\varepsilon}_{rt}$, are then summed across regencies over the period 1994-2008 to compute regency-level cumulative flood shocks as below:

$$CumulativeFlood_r = \sum_{t=1994}^{2008} \widehat{\varepsilon}_{rt}$$

I employ the first-difference (FD) estimator on the aggregate variables in 1994 and 2008, where the treatment variable is cumulative flood that takes zero as its initial value in 1994. I estimate the following econometric specification at the sector-regency or regency level:

$$\Delta y_{sr} = \upsilon + \beta \Delta CumulativeFlood_r + \nu_s + \varepsilon_{sr} \tag{1.24}$$

where Δy_{sr} denotes the difference in the logarithm of regency or sector-regency level value-added, capital stock, or labor employed between the years 1994 and 2008. β is the coefficient capturing relationship between the change in the cumulative flood shock from 0 to its end-of-period value and change in the aggregate variable. υ captures the linear time trend.[36]

Results and Discussion

I use the value at the 90th percentile from Table 1.9 for interpreting the strength of the relationship in all cases. Table 1.10 reports the results from estimating Equation (1.24).

35. This formulation is simply used to match the specifications employed for studying contemporaneous effect of flooding on economic variables. Results are robust to adding lags of flood index or regency-level linear time trends.

36. The estimation leverages cross-sectional variation in cumulative flood exposure across regencies. Since, the flood innovations always sum to zero for each regency over the complete 1985-2012 period, the realizations over the block of period 1994-2008 randomly sort regencies based on the realized exposures within this period. Figure 1.32 illustrates this point by showing the realized values of flood innovations in circles and their moving sums for different years.

Interpreting the magnitude using estimates from Columns 4-6, a 90th percentile cumulative flood shock at the regency level is associated with 18.9%, 20.9%, and 16.2% decrease in the aggregate value-added, capital stock, and labor employment respectively.

1.8.2 Effect of Flooding on Firm Exit and Entry

Econometric Model

To estimate the relationship between cumulative flood shocks and the exit decision of firms, I first define a cumulative flood shock at firm level by aggregating the flood index over the long-run analysis period for each firm. Obviously, a typical firm does not continue operating in all these years, so the last year of its observation in data is assumed to be the exit year. Under these assumptions, the firm-specific cumulative flood shock is defined as below:

$$CumulativeFlood_{irt} = \frac{1}{t - T^i_{start} + 1} \sum_{s=T^i_{start}}^{t \leq T^i_{end}} FloodIndex_{rs}$$

where T^i_{start} and T^i_{end} are the entry and exit years for firm i.[37] I then estimate the following relationship:

$$y_{isrt} = \upsilon + \beta CumulativeFlood_{irt} + \iota X_{isrt} + \zeta_r + \nu_{st} + \psi_{pt} + \varepsilon_{isrt} \tag{1.25}$$

where y_{isrt} is an exit dummy for firm i, belonging to 2-digit ISIC sector s, located in regency r, in year t and other terms have the same interpretation as Equation (1.4). Below is the summary statistics table on the cumulative shock variable.

37. Only those firms, which have their start and end years fall in the period 1994-2008 are included in the analysis.

Results and Discussion

Column 1 of Table 1.11 reports the results from estimating Equation (1.25) for all the firms that start and end their life in the period 1994-2008. Unlike temporary shocks, cumulative flood shocks do lead to firm exits in the long run. In particular, a 90th percentile cumulative flood shock increases the firm exit probability by 0.26%.

Columns 2 and 3 of Table 1.11 report the results from estimating Equation (1.24) with difference in the logarithm of number of firms operating in a regency or sector-regency between 2008 and 1994 as the dependent variable. Similar to the contemporaneous analysis, the evidence points towards the firms avoiding flood-prone locations when setting up their operations, a 90th percentile cumulative flood shock decreases the number of firms at the sector-regency (regency) level by 6.6% (8.5%).

1.9 Appendix: Figures and Tables

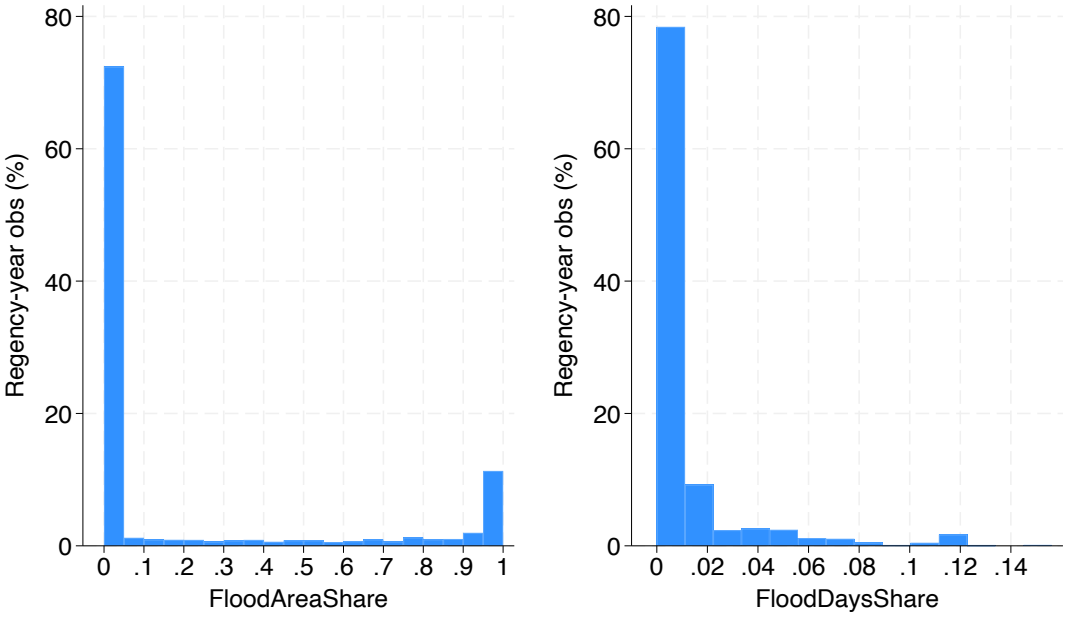
Table 1.4: Relationship between flooded-affected and flooded area share

	(1) Affected area	(2) Affected area	(3) Affected area
Flooded area	0.300	0.513***	1.422***
		(0.026)	(0.240)
Observations	21,074	21,074	21,074
Regency FE	-	-	Y
Flood FE	-	-	Y

Standard errors clustered at the regency level are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

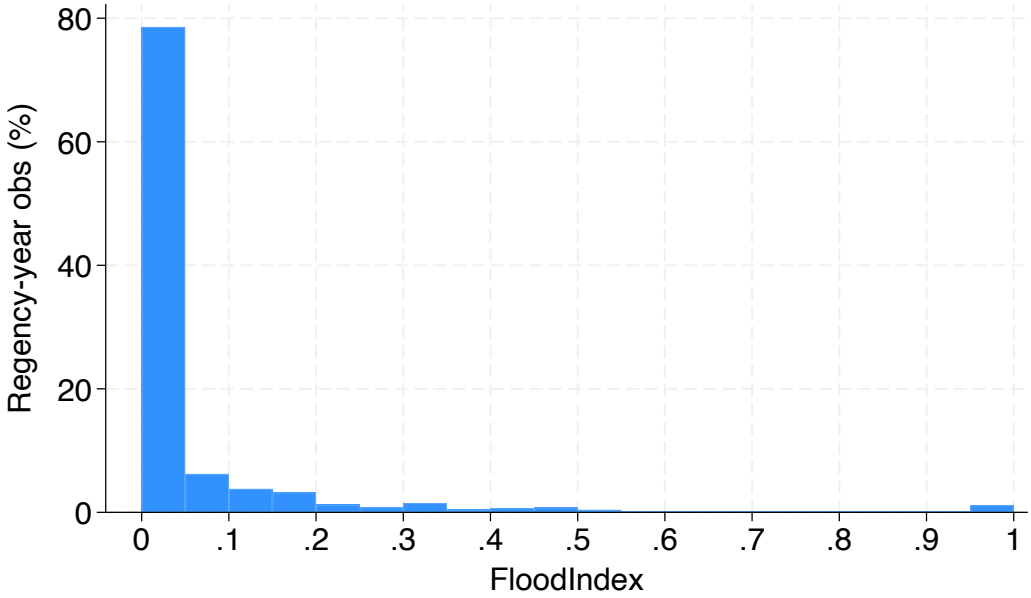
Notes: The table presents results on establishing relationship between flooded and flood-affected area in Indonesia. Flooded area metric at the regency level is constructed using detailed inundation maps available for 41 flood events within Indonesia. Flood-affected area within regencies for these 41 events are obtained from the polygons available in the DFO archive. Both these variables are then normalized by the total area of regency to represent area shares. Column 1 reports the Kendall's Tau-b coefficient of association between flooded and flood-affected area share. The coefficient ranges from -1 (perfect inversion) to +1 (perfect agreement) with 0 indicating no association. Column 2 reports the Somers' D coefficient, which also has the same range as Tau-b coefficient, but additionally it comes with standard errors on the coefficient of association that are generated using the jackknife variance calculation method. Column 3 reports the result of following a parametric approach by using fixed-effects regression analysis in which the flood-affected area is regressed on the flooded area share. Standard errors are clustered at the regency level.

Figure 1.19: Distribution of regencies on flood-affected area and days share



Notes: The graphs show the distribution of components of flood index across regency-year pairs. The left graph presents the distribution of regency-level flood-affected area share that is computed using the polygons provided in the DFO archive of large flood events. The right graph presents the share of days in a year that a regency remains flooded, where the number of days are computed using the start and end dates for each flood event. In case of multiple flood events affecting a regency in a year, the flood-affected area is the average across all flooding episodes and the flood days are total count of days that any area in the regency remains flooded.

Figure 1.20: Distribution of flood index across regency-year pairs



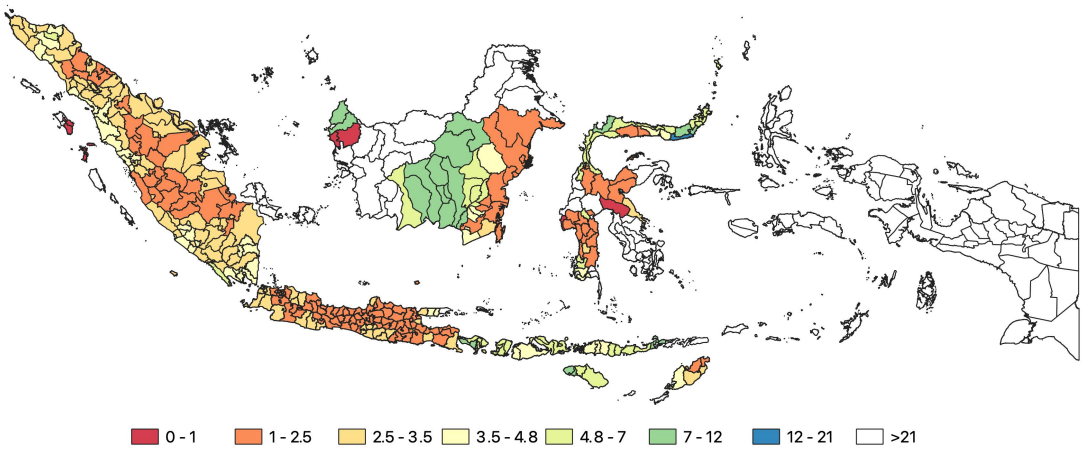
Notes: The graph shows the distribution of flood index across all regency-year pairs. Flood index is a rescaled product of flood-affected area share in a regency and flood days share in a year. The rescaling is done so that the index lies in the range 0 to 1. Most of the regency-year pairs remain unaffected by flooding with some extreme flood events affecting few of them located on the right tail of the distribution.

Table 1.5: Summary statistics on flood index

(1) Mean	(2) Std. Dev.	(3) 25th Pctile	(4) 50th Pctile	(5) 75th Pctile	(6) 90th Pctile
0.186	0.235	0.04	0.098	0.219	0.455

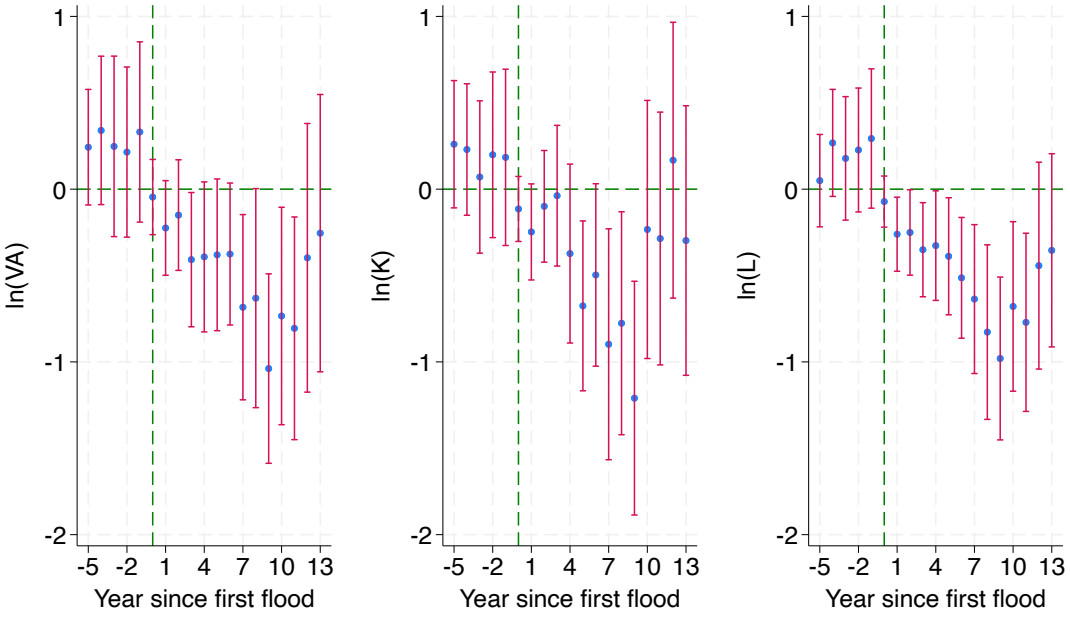
Notes: The table presents the summary statistics on flood index at the regency level that is defined by the rescaled product of flood-affected area share in a regency and flood days share in a year. Only those regency-year pairs that have non-zero flood index values are considered to derive these statistics.

Figure 1.21: Average time interval between two successive floods



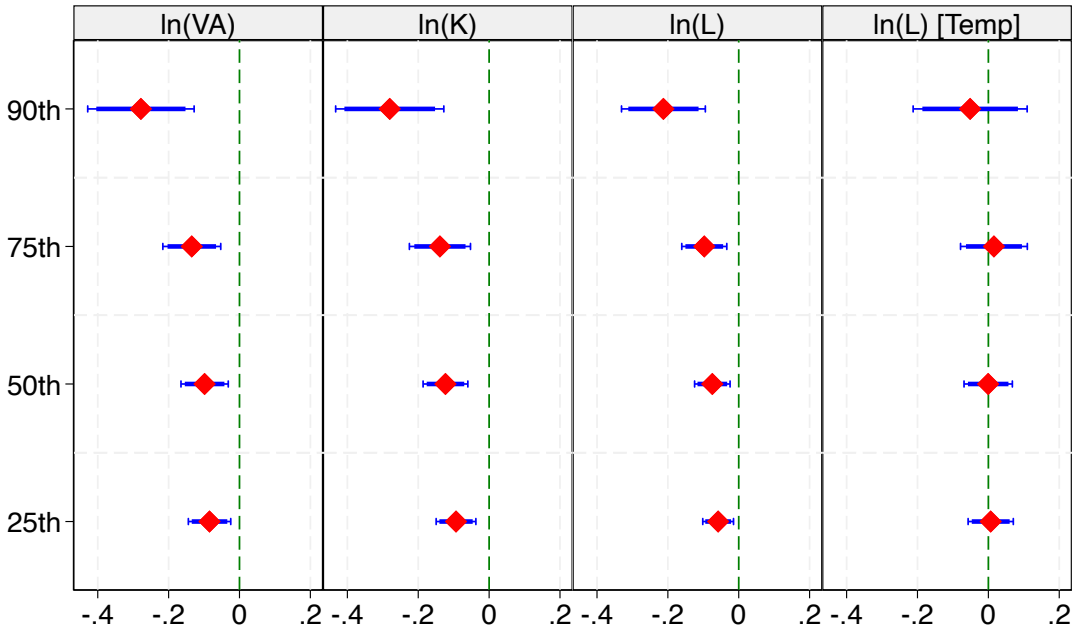
Notes: The map shows the average time interval in terms of years between two successive flood events in a regency during the period 1985-2012. The internal boundaries are regency boundaries, with the legends denoting the number of years between two successive flooding episodes. Most of the regencies located on the islands of Java and Sumatra witness a large flood almost every alternate year.

Figure 1.22: Effect of first flood on aggregate variables



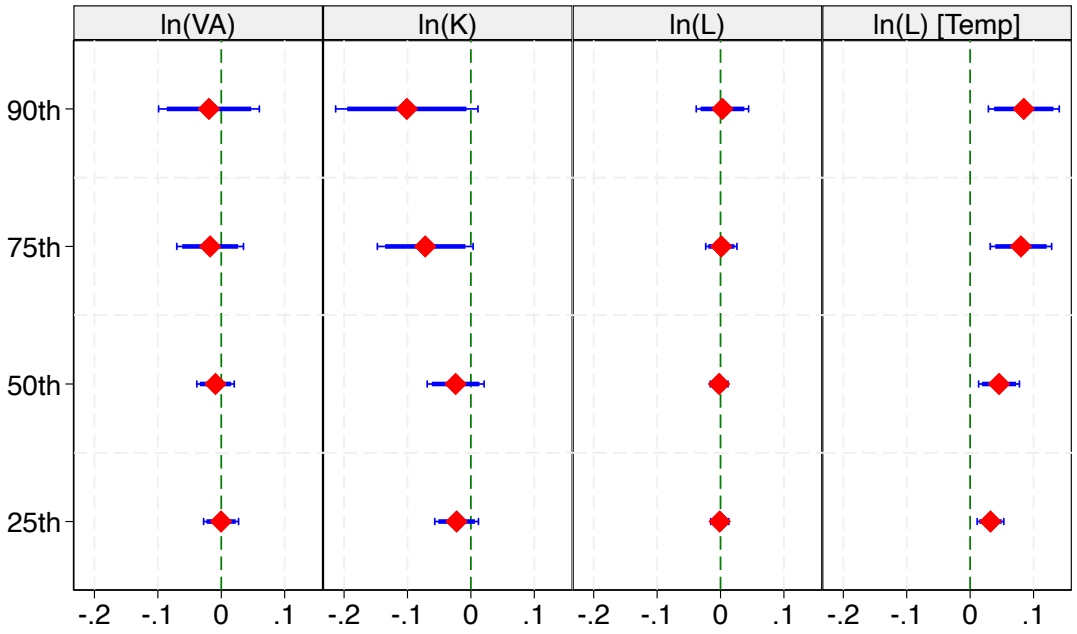
Notes: The graphs show the effect of "first" flood on the regency-level aggregate variables i.e., logarithms of total value-added (left), capital stock (middle), and labor employment (right) using the imputation-based difference-in-differences estimator proposed by Borusyak, Jaravel, and Spiess (2024). The first flood for a regency is defined as the first year in which the regency witnessed a flooding episode in the sample. Since the outcomes data starts in 1990 but the floods can be tracked since 1985, only those regencies that did not witness any flood event in the period 1985-89 are included in the analysis. To get to the aggregate variables from firm-level information, following steps are undertaken. First, The un-logged version of all the monetary variables are deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values. Second, the tails on both ends of the resulting variables are trimmed by 1% for each year to address measurement error issues. Third, the variables are then summed across regency for each year using labor share weights. Finally, the variables are log-transformed and used in the estimation. Standard errors are clustered at the regency level and whiskers on the point estimates show 90% confidence intervals.

Figure 1.23: Effect of flooding on sector-regency-level variables



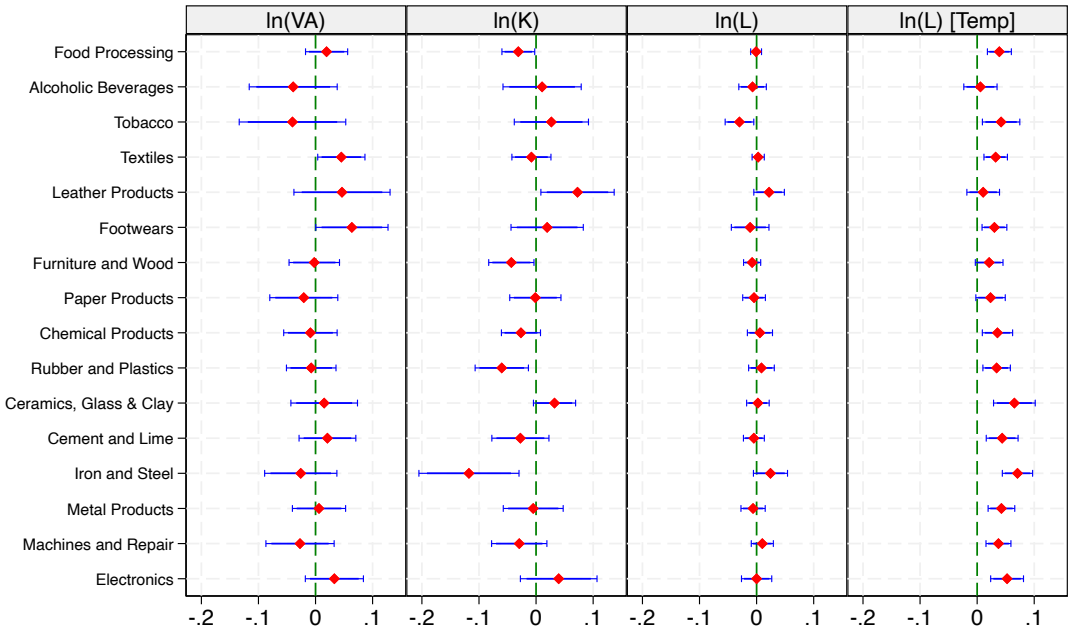
Notes: The graph presents the results of estimating Equation (1.1) for aggregate variables i.e., logarithms of value-added (left), capital stock (second-left), permanent labor employment (second-right), temporary labor employment (right) at the sector-regency level using only those regencies for which data is available for at least 20 years. To get to the aggregate variables from firm-level information, following steps are undertaken. First, The un-logged version of all the monetary variables are deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values. Second, the tails on both ends of the resulting variables are trimmed by 1% for each year to address measurement error issues. Third, the variables are then summed across sector-regency for each year using labor share weights. Finally, the variables are log-transformed and used in the regressions. The labels on y-axis represent the percentiles of flood index for which dummy is used in the regression. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.24: Effect of flooding on firm-level variables



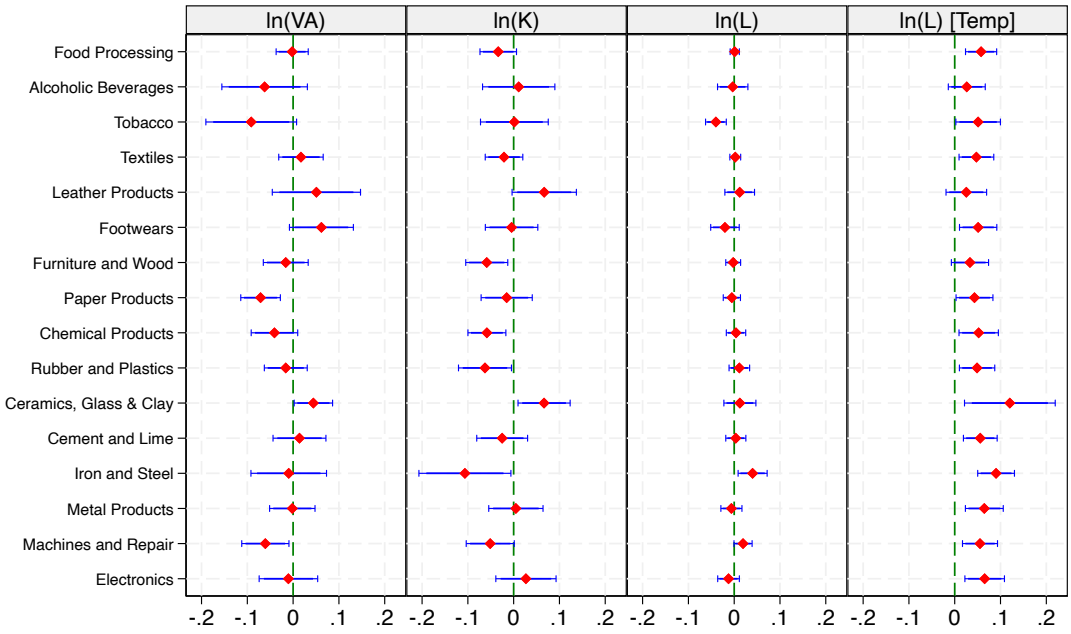
Notes: The graph presents the results of estimating Equation (1.2) for firm-level variables i.e., logarithms of value-added (left), capital stock (second-left), permanent labor employment (second-right), temporary labor employment (right) using only those firm observations for which data is available for at least 20 years. The un-logged version of all the monetary variables have been deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values and the log-transformed variables are trimmed by 1% for each year to address measurement error issues. The labels on y-axis represent the percentiles of flood index for which dummy is used in the regression. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.25: Effect of 25th percentile floods on firm-level variables by sectors



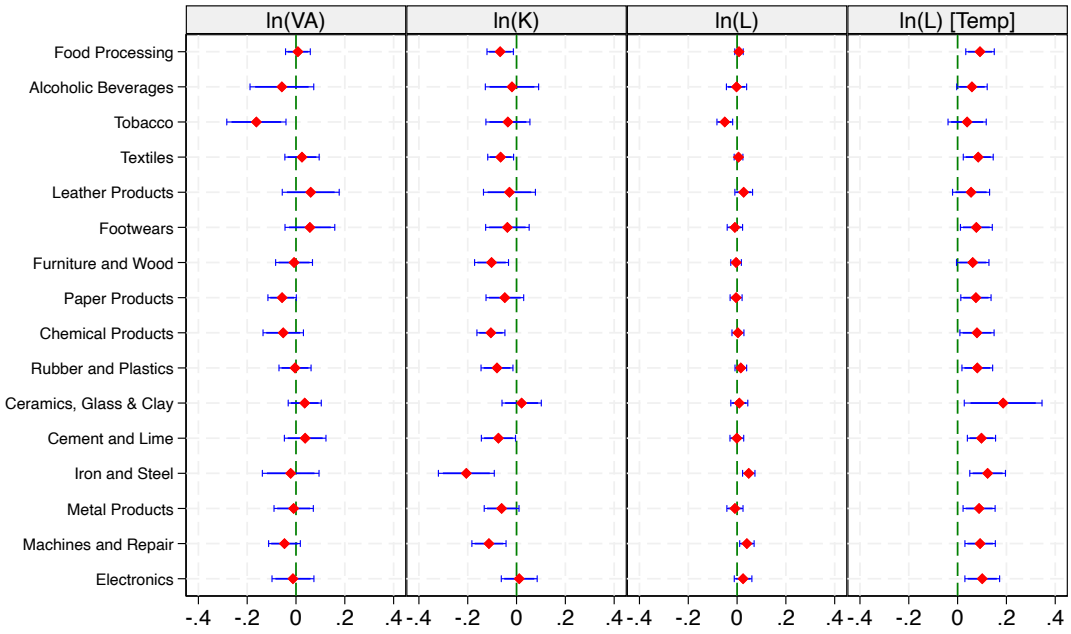
Notes: The graph presents the results of estimating Equation (1.3) for firm-level variables i.e., logarithms of value-added (left), capital stock (second-left), permanent labor employment (second-right), temporary labor employment (right) using the 25th percentile flood dummy. The un-logged version of all the monetary variables have been deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values and the log-transformed variables are trimmed by 1% for each year to address measurement error issues. The labels on y-axis represent the 2-digit ISIC manufacturing sectors. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.26: Effect of 50th percentile floods on firm-level variables by sectors



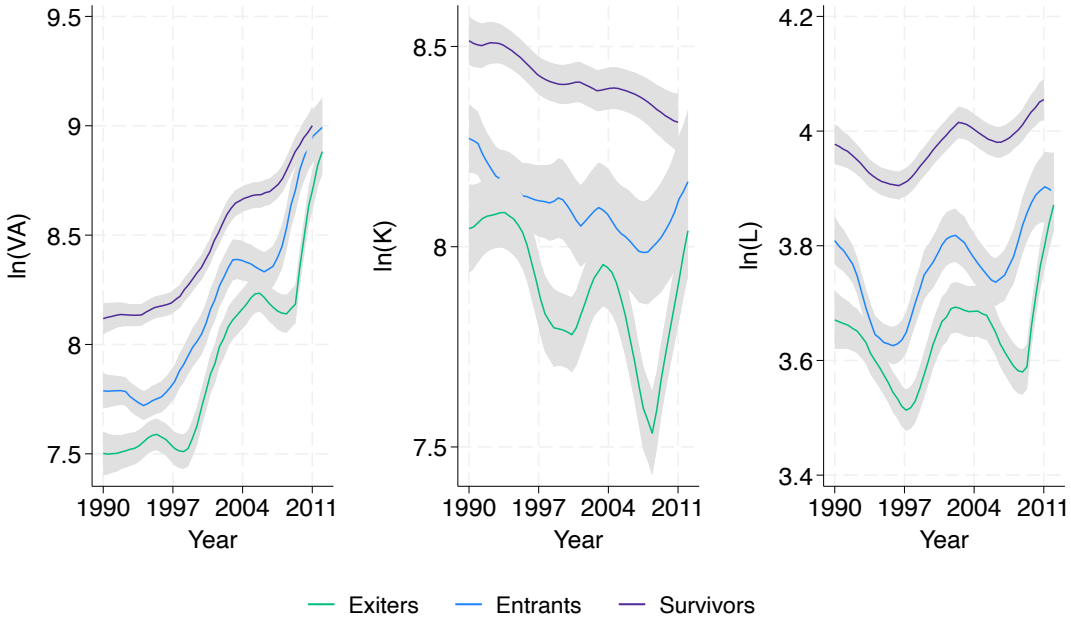
Notes: The graph presents the results of estimating Equation (1.3) for firm-level variables i.e., logarithms of value-added (left), capital stock (second-left), permanent labor employment (second-right), temporary labor employment (right) using the 50th percentile flood dummy. The un-logged version of all the monetary variables have been deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values and the log-transformed variables are trimmed by 1% for each year to address measurement error issues. The labels on y-axis represent the 2-digit ISIC manufacturing sectors. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.27: Effect of 75th percentile floods on firm-level variables by sectors



Notes: The graph presents the results of estimating Equation (1.3) for firm-level variables i.e., logarithms of value-added (left), capital stock (second-left), permanent labor employment (second-right), temporary labor employment (right) using the 75th percentile flood dummy. The un-logged version of all the monetary variables have been deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values and the log-transformed variables are trimmed by 1% for each year to address measurement error issues. The labels on y-axis represent the 2-digit ISIC manufacturing sectors. The control observations in all cases are regency-year pairs that are not flooded. 90 and 95% confidence intervals are shown in thick and thin blue lines respectively over the point estimates. Standard errors are clustered at the regency level.

Figure 1.28: Aggregate variables for entering, exiting, and surviving firms



Notes: The graphs plot the average of the three variables viz. logarithm of value-added (left), capital stock (middle), and labor employment (right) across regencies over time for three groups of firms: exiters, entrants, and survivors. A firm's year of exit is its last year of observation in the data, entry year is its first year of observation, and all the years in between are its years of survival.

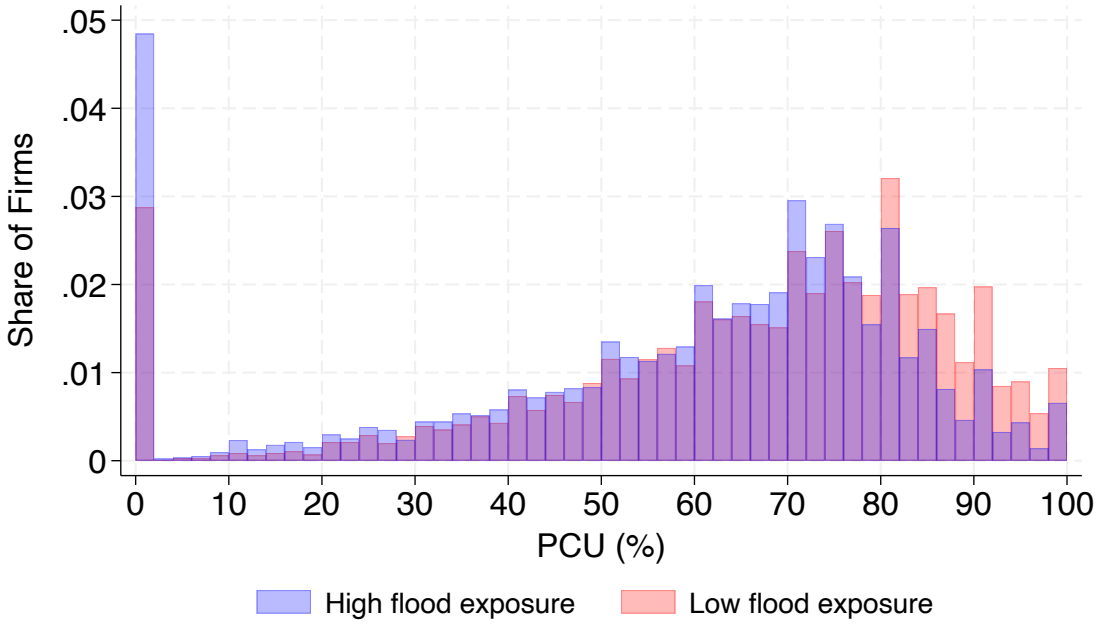
Table 1.6: Relationship between flooding and firm entry rate

	(1) Plant entry	(2) Plant entry
FloodIndex	-0.299**	
	(0.144)	
FloodRisk		-0.247*
		(0.131)
Observations	273	211
R-squared	0.025	0.031

Standard errors clustered at the regency level are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: The table presents the relationship between flood variables and regency-level firm entry rates. The entry rate is defined as the ratio of count of entrants in the current year over the count of survivors for the previous year. $FloodIndex$ is the average flood index at the regency level for the period 1990-2012 and has been rescaled to lie in the interval [0,1]. $FloodRisk$ variable reflects regency-level flood risk for the year 2013 as published by the IRBI in their annual report and also lie in the interval [0,1]. Each regency is given a risk score between 0 and 1 depending on its hazard profile, vulnerability index, and resilience to deal with destructive effects of flooding. The total number of regencies used in the flood risk analysis are smaller because the flood risk scores are unavailable for some regencies. However, the omission seems to be orthogonal to the flood risk, since some of the omitted regencies are also at high flood risk as outlined in the published report (IRBI 2013). Standard errors are clustered at the regency level.

Figure 1.29: Firm capacity utilization across low and high flood-prone regencies



Notes: The graph plots the distribution of firm-level production capacity utilization (PCU) across low and high flood-prone regencies in Indonesia. PCU is defined as the percentage of available production capacity utilized by a firm in each year. Flood exposure of a regency is the average flood index over the period 1990-2012. The distribution of regencies on average flood index is split into 20 quintiles. The last quintile is defined as high flood exposure and the first 10 quintiles are defined as low flood exposure regencies. This sampling choice ensures that around 20% of regencies fall in each of the two bins. The core qualitative finding in the above graph is robust to changing this sampling criteria.

Table 1.7: Sectoral output elasticities of capital and labor

Industry name	(1) 3-digit ISIC	(2) log(L) coeff	(3) log(L) se	(4) log(K) coeff	(5) log(K) se	(6) #Observations	(7) #Plants	(8) #Years (avg)
Food Processing	311	.551	.0002	.149	.0001	40914	5955	6.9
Food Processing 2	312	.489	.0003	.173	.0002	31821	4391	7.2
Cigarettes and Tobacco	314	.364	.0008	.123	.0011	16494	2638	6.3
Textiles	321	.488	.0003	.125	.0002	34911	4855	7.2
Leather Products	323	.625	.0011	.145	.0015	3094	469	6.6
Manufacture of Footwear	324	.568	.0007	.1	.0007	5491	886	6.2
Wood Products	331	.547	.0002	.163	.0001	22945	3877	5.9
Furniture	332	.6	.0003	.093	.0001	20661	3421	6
Paper Products	341	.412	.0018	.15	.001	5162	660	7.8
Paper Products, Finished	342	.624	.0011	.08	.0003	8522	1139	7.5
Chemical Products, Industrial	351	.38	.0016	.166	.0006	5706	781	7.3
Chemical Products, Household	352	.489	.0008	.107	.0004	8897	1006	8.8
Rubber Products	355	.56	.0008	.065	.0004	6494	734	8.8
Plastic Wares	356	.503	.0004	.149	.0002	14934	1979	7.5
Ceramics	361	.391	.0036	.202	.0026	1282	142	9
Glass Products	362	.497	.0026	.208	.0064	945	128	7.4
Cement and Lime	363	.534	.0014	.153	.0006	7965	1221	6.5
Structural Clay Products	364	.624	.0007	.149	.0002	15704	1983	7.9
Other Non Metal Mineral Products	369	.516	.0018	.13	.0007	4499	700	6.4
Basic Iron and Steel	371	.545	.003	.197	.002	2685	346	7.8
Metal Products, Finished	381	.6	.0006	.144	.0003	13941	1910	7.3
Machines and Repair	382	.503	.0025	.204	.0034	4475	593	7.5
Electronics	383	.619	.0011	.099	.0015	4389	736	6
Motor Vehicles	384	.44	.0012	.161	.0006	7959	1070	7.4
Other Manufacturing	390	.644	.0005	.134	.0004	8111	1272	6.4

Notes: The table presents the production function estimation results for each 3-digit ISIC sector by employing the Levinsohn and Petrin (2003) methodology in Stata through the `prodest` package. Columns 2 & 3 (4 & 5) report output elasticity of labor (capital) coefficient and standard errors respectively. Column 6 reports the total number of observations used in the estimation with Column 7 and 8 reporting statistics on number of firms used and average number of years observed for each firm in the estimation.

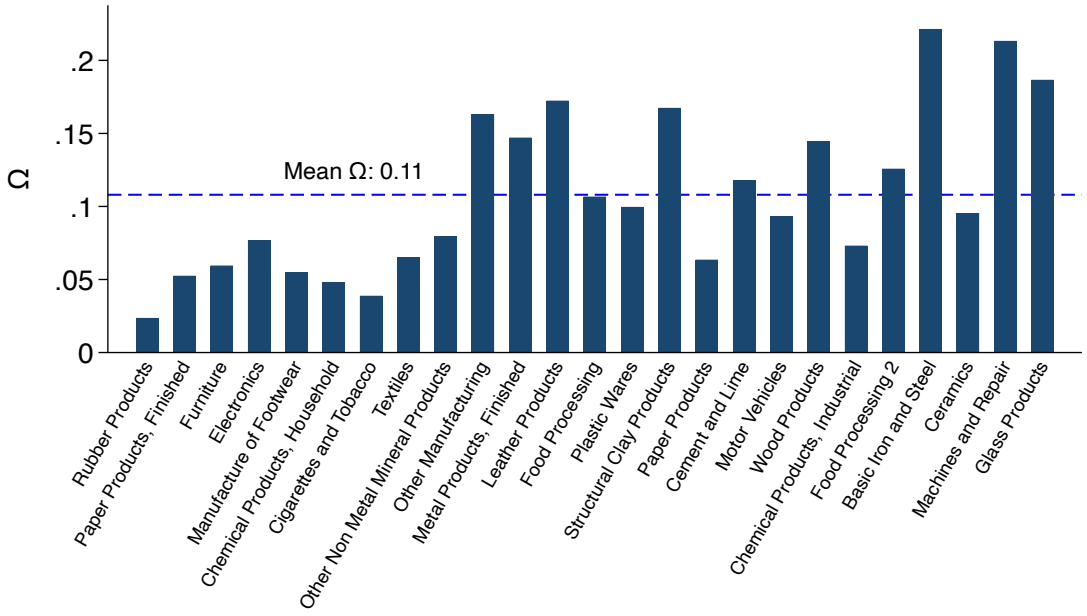
Table 1.8: Effect of flooding on firm capacity utilization by firm size

	(1)	(2)
	% PCU	% PCU
FloodIndex	-3.064**	-3.078**
	(1.268)	(1.301)
Large (L) Firm × FloodIndex	0.960	
	(0.648)	
Large (K) Firm × FloodIndex		1.080
		(0.777)
Observations	330,580	330,580
Adj R-squared	0.296	0.296
Dep. var mean	68.349	68.349
Firm FE	Y	Y
Province × year FE	Y	Y
2-digit ISIC × year FE	Y	Y
Plant-level controls	Y	Y

Standard errors clustered at the regency level are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

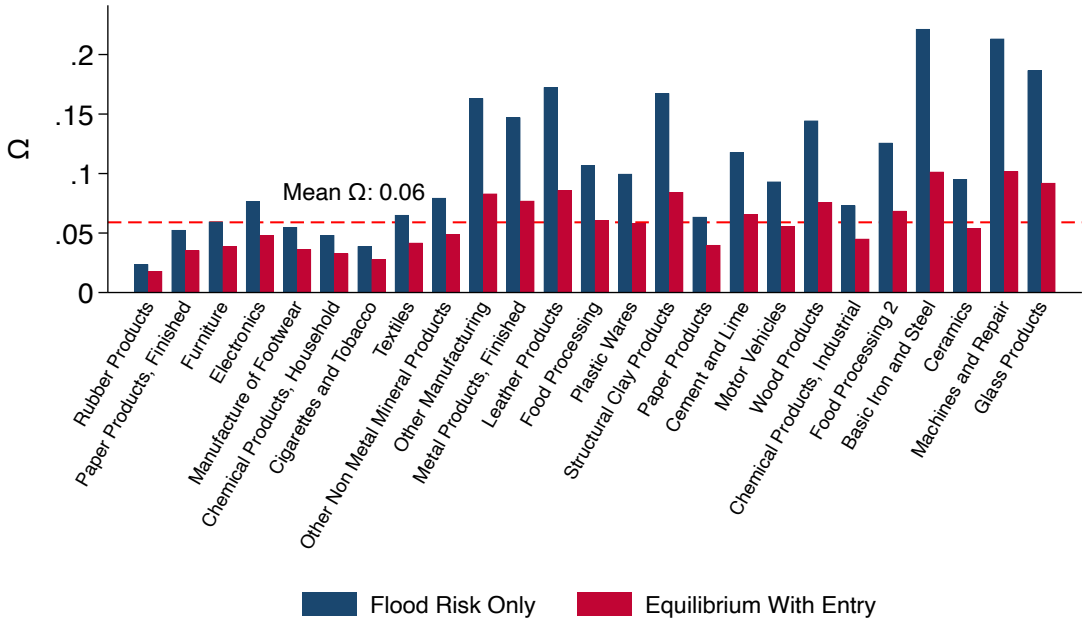
Notes: The table presents the results of estimating an interaction version of Equation (1.2) for firm-level production capacity utilization (PCU), where the flood index is interacted with the firm size dummy. The firm size dummy in Column 1 (2) uses average (over years) labor employment (capital stock) for each firm. PCU measures the percentage of the potential firm capacity, in terms of production, that is realized in a given year. Results reported in both the columns control for firm age controls, firm, province × year, and sector × year fixed effects. Standard errors are clustered at the regency level.

Figure 1.30: Change in aggregate output across sectors due to flood risk



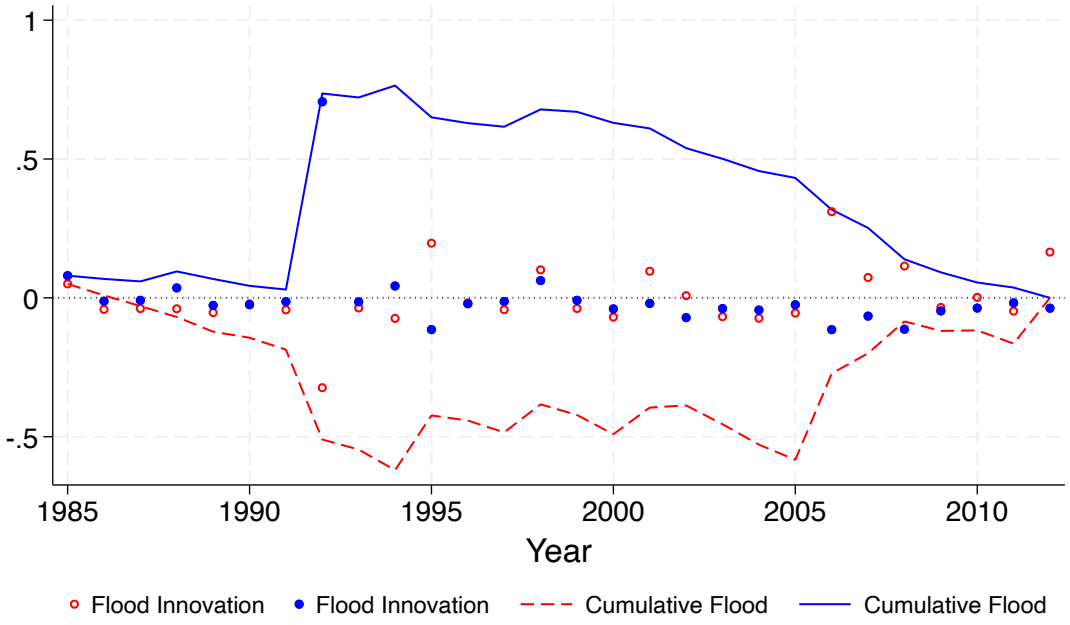
Notes: The graph plots the (log) change in aggregate output due to flood risk as outlined in Equation (1.22) across 3-digit ISIC sectors, keeping only six regencies in which all 25 3-digit ISIC sectors are situated. This represents the Flood Risk Only scenario where in the counterfactual world with flood defenses, all the regencies above 80th percentile on the flood exposure distribution are assigned the median value of the distribution. The sectors are ranked from left to right in the increasing order of their respective capital intensities.

Figure 1.31: Change in output across sectors due to flood risk and equilibrium



Notes: The graph plots the (log) change in aggregate output due to flood risk in blue and sum of (log) change in aggregate output due to flood risk and (log) change in aggregate output due to equilibrium wage adjustments and firm entry in red as outlined in Equation (1.22) across 3-digit ISIC sectors, keeping only six regencies in which all 25 3-digit ISIC sectors are situated. In the counterfactual world with flood defenses, all the regencies above 80th percentile on the flood exposure distribution are assigned the median value of the distribution. The sectors are ranked from left to right in the increasing order of their respective capital intensities.

Figure 1.32: Flood innovations and cumulative flood shocks



Notes: The graph presents the evolution of cumulative flood shock variable over the years for two sample regencies in Indonesia. Circles represent flood innovations, which are generated as residuals from estimating Equation (1.23) for all the regencies in the period 1985-2012. The lines show the running sum of these flood innovations for each regency over time. Red (hollow circle and dashed line) represents the Bogor regency, which experienced runs of low flooding during this period. One the other hand, Yogyakarta city, represented in blue (solid circle and solid line) experienced runs of high flooding.

Table 1.9: Summary statistics on cumulative flood variables

(1)	(2)	(3)	(4)	(5)
Mean	Std. Dev.	50th Pctile	90th Pctile	95th Pctile
Panel 1: Regency-level Cumulative Flood Shocks				
0.009	0.36	-0.027	0.348	0.69
Panel 2: Firm-level Cumulative Flood Shocks				
0.07	0.092	0.041	0.197	0.252

Notes: The table presents the summary statistics on cumulative flood variables used in the reduced-form analysis of long-run effects of flooding. Panel 1 and Panel 2 report statistics on cumulative flood variables at the regency and plant level respectively.

Table 1.10: Long-run effect of flooding on aggregate variables

	(1) D.ln(VA)	(2) D.ln(K)	(3) D.ln(L)	(4) D.ln(VA)	(5) D.ln(K)	(6) D.ln(L)
D.CumulativeFlood	-0.563**	-0.703***	-0.803***	-0.544***	-0.600***	-0.464***
	(0.229)	(0.253)	(0.172)	(0.146)	(0.166)	(0.123)
Observations	246	246	246	1,320	1,320	1,320
R-squared	0.012	0.017	0.042	0.062	0.058	0.073
2-digit ISIC FE	-	-	-	Y	Y	Y

Standard errors clustered at the regency level are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: The table presents the results of estimating Equation (1.24) using the first-difference estimator for aggregate variables i.e., logarithms of total value-added, capital stock, and labor employment at the regency or sector-regency level for the years 1994 and 2008. To get to the aggregate variables from firm-level information, following steps are undertaken. First, The un-logged version of all the monetary variables are deflated by the wholesale price index at the 5-digit ISIC level to reflect their real values. Second, the tails on both ends of the resulting variables are trimmed by 1% for each year to address measurement error issues. Third, the variables are then summed across regency or sector-regency for each year. Finally, the variables are log-transformed and used in the regressions. Columns 1, 2 and 3 (4, 5, and 6) show the results for value-added, capital stock, and labor employment respectively when the firm data is collapsed at the regency (sector-regency) level. Standard errors are clustered at the regency level.

Table 1.11: Long-run effect of flooding on firm exit and entry

	(1) Pr(exit)	(2) D.ln(#Plants)	(3) D.ln(#Plants)
CumulativeFlood	0.013*		
	(0.007)		
D.CumulativeFlood		-0.189***	-0.245***
		(0.070)	(0.085)
Observations	166,176	1,474	254
Adj R-squared	0.076	0.049	0.011
Regency FE	Y	-	-
2-digit ISIC FE	-	Y	-
Province × year FE	Y	-	-
2-digit ISIC × year FE	Y	-	-

Standard errors clustered at the regency level are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Column 1 presents the results of estimating Equation (1.25) for firm exit dummy where the dummy variable takes a value of 1 in the last year of the firm observation in the data. All the firms that start and end their operations in the period 1994-2008 are included in the estimation. Columns 2 and 3 present the results of estimating Equation (1.24) with the difference in the logarithm of number of firms operating in a regency or sector-regency for years 1994 and 2008 as the dependent variable. Column 1 controls for regency, province × year, and sector × year fixed effects. Column 2 controls for industry fixed effects. Standard errors are clustered at the regency level.

1.10 Appendix: Detailed Proofs of the Theory

1.10.1 Flood Risk (τ_{srt})

The distribution of the share variable x is as follows:

$$G_{rt}(x) = \begin{cases} 1 - \left(\frac{1}{x}\right)^{\phi_{rt}} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

Firms maximize expected profits by choosing the optimal capital to install in a given period taking expectations on the random variable x_{it}. Firm's optimization problem is as below:

$$K_{it} = \text{argmax} \left\{ \Gamma_{it} \mathbb{E} \left[\left(\frac{K_{it}}{x_{it}} \right)^{\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} \right] - \rho K_{it} \right\}$$

where $\Gamma_{it}(\theta, w) \equiv [1-(1-\alpha_s)\eta_s] \, \theta_i^{\frac{1}{1-(1-\alpha_s)\eta_s}} \left\{ \frac{w_t}{(1-\alpha_s)\eta_s} \right\}^{-\frac{(1-\alpha_s)\eta_s}{1-(1-\alpha_s)\eta_s}}$.

The above problem can be written as:

$$K_{it} = \text{argmax} \left\{ \Gamma_{it} K_{it}^{\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} \int_1^{\infty} x_{it}^{-\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} g(x_{it}) dx_{it} - \rho K_{it} \right\}$$

Putting the p.d.f of share distribution follows:

$$K_{it} = \text{argmax} \left\{ \Gamma_{it} K_{it}^{\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} \int_1^{\infty} x_{it}^{-\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} \phi_{rt} x_{it}^{-\phi_{rt}-1} dx_{it} - \rho K_{it} \right\}$$

$$= \text{argmax} \left\{ \Gamma_{it} K_{it}^{\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} \phi_{rt} \int_1^{\infty} x_{it}^{-\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}-\phi_{rt}-1} dx_{it} - \rho K_{it} \right\}$$

$$= \text{argmax} \left\{ \frac{\phi_{rt}}{\phi_{rt} + \frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} \Gamma_{it} K_{it}^{\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} - \rho K_{it} \right\}$$

$$= \text{argmax} \left\{ \tau_{srt} \Gamma_{it} K_{it}^{\frac{\alpha_s \eta_s}{1-(1-\alpha_s)\eta_s}} - \rho K_{it} \right\}$$

1.10.2 Expected Aggregate Equilibrium Output $(\overline{Y}_{srt})$

The general expression for aggregate output is as follows:

$$\overline{Y}_{srt} = \int_{\theta^*_{srt}}^{\infty} Y_{it}(\theta)\mu_{srt}(\theta)d\theta$$

Putting the expression for equilibrium firm-level output from Equation (1.12):

$$\overline{Y}_{srt} = \Lambda_{st}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}} \int_{\theta^*_{srt}}^{\infty} \theta^{\frac{1}{1-\eta_s}}\mu_{srt}(\theta)d\theta$$

Adding the equilibrium productivity distribution from Equation (1.15):

$$\overline{Y}_{srt} = \Lambda_{st}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}} \frac{\xi\overline{\theta}_r^{\xi}}{1-H_r(\theta^*_{srt})} \int_{\theta^*_{srt}}^{\infty} \theta^{\frac{1}{1-\eta_s}-\xi-1}d\theta$$

Integrating the above, follows:

$$\overline{Y}_{srt} = \Lambda_{st}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}} \frac{\xi\overline{\theta}_r^{\xi}(1-\eta_s)}{\xi(1-\eta_s)-1} \frac{(\theta^*_{srt})^{\frac{1}{1-\eta_s}-\xi}}{1-H_r(\theta^*_{srt})}$$

Using the initial productivity distribution to compute $(1-H_r(\theta^*_{srt}))$ and assuming $(\xi(1-\eta_s) > 1)$, the formula for aggregate output in Equation (1.16) is obtained as below:

$$\overline{Y}_{srt} = \Lambda_{st}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}} \frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1} (\theta^*_{srt})^{\frac{1}{1-\eta_s}}$$

1.10.3 Labor Market Clearing (w_t)

The labor market clearing takes place at the regency level with the total labor employed by all the firms equal to the aggregate (exogenous, time-invariant) labor supply in the regency as follows:

$$\overline{L}_t = \int_{i\in t} Ldi$$

Using Equation (1.11), LHS can be expanded as follows:

$$\overline{L}_t = \sum_{r=1}^{R}\sum_{s=1}^{S}\int \frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\theta_i^{\frac{1}{1-\eta_s}}x_{it}^{-\frac{\alpha_s\eta_s}{1-(1-\alpha_s)\eta_s}}f(x,\theta)dxd\theta$$

Using the independence of stochastic processes for θ and x, $f(x,\theta)$ can written as the product of respective densities as follows:

$$\overline{L} = \sum_{r=1}^{R}\sum_{s=1}^{S}\int_{\theta^*_{srt}}^{\infty}\int_{1}^{\infty}\frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\theta_i^{\frac{1}{1-\eta_s}}x_{it}^{-\frac{\alpha_s\eta_s}{1-(1-\alpha_s)\eta_s}}g(x)\mu_{srt}(\theta)dxd\theta$$

$$= \sum_{r=1}^{R}\sum_{s=1}^{S}\frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\left(\int_{\theta^*_{srt}}^{\infty}\theta_i^{\frac{1}{1-\eta_s}}\mu_{srt}(\theta)d\theta\right)\left(\int_{1}^{\infty}x_{it}^{-\frac{\alpha_s\eta_s}{1-(1-\alpha_s)\eta_s}}g(x)dx\right)$$

Using the definitions of the respective distributions, the above can be written as follows:

$$\overline{L} = \sum_{r=1}^{R}\sum_{s=1}^{S}\frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\left(\frac{\xi\overline{\theta}_r^{\xi}}{1-H_r(\theta^*_{srt})}\int_{\theta^*_{srt}}^{\infty}\theta_i^{\frac{1}{1-\eta_s}-\xi-1}d\theta\right)\left(\int_{1}^{\infty}x_{it}^{-\frac{\alpha_s\eta_s}{1-(1-\alpha_s)\eta_s}}\phi_{rt}x_{it}^{-\phi_{rt}-1}dx\right)$$

$$= \sum_{r=1}^{R}\sum_{s=1}^{S}\frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\left(\frac{\xi\overline{\theta}_r^{\xi}}{1-H_r(\theta^*_{srt})}\int_{\theta^*_{srt}}^{\infty}\theta_i^{\frac{1}{1-\eta_s}-\xi-1}d\theta\right)\left(\phi_{rt}\int_{1}^{\infty}x_{it}^{-\frac{\alpha_s\eta_s}{1-(1-\alpha_s)\eta_s}-\phi_{rt}-1}dx\right)$$

$$= \sum_{r=1}^{R}\sum_{s=1}^{S}\frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\left(\frac{\xi\overline{\theta}_r^{\xi}(1-\eta_s)}{\xi(1-\eta_s)-1}\frac{(\theta^*_{srt})^{\frac{1}{1-\eta_s}-\xi}}{1-H_r(\theta^*_{srt})}\right)\left(\frac{\phi_{rt}}{\phi_{rt}+\frac{\alpha_s\eta_s}{1-(1-\alpha_s)\eta_s}}\right)$$

$$= \sum_{r=1}^{R}\sum_{s=1}^{S}\frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}^{\frac{1-\eta+\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1}(\theta^*_{srt})^{\frac{1}{1-\eta_s}}$$

Expanding the θ^*_{srt} using Equation (1.17) follows:

$$\overline{L}_t = \sum_{r=1}^{R}\sum_{s=1}^{S}\frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}^{\frac{1-\eta+\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}\frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1}\left\{\frac{f}{[1-(1-\alpha_s)\eta_s-\alpha_s\eta_s\tau_{srt}]\,\tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}}\Lambda_{st}}\right\}$$

$$= \sum_{r=1}^{R}\sum_{s=1}^{S}\frac{(1-\alpha_s)\eta_s}{w_t}\tau_{srt}\frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1}\left\{\frac{f}{1-(1-\alpha_s)\eta_s-\alpha_s\eta_s\tau_{srt}}\right\}$$

The above delivers the equilibrium wage equation as follows:

$$w_t = \frac{f}{\overline{\overline{L}}} \sum_{r=1}^{R} \sum_{s=1}^{S} \frac{(1-\alpha_s)\eta_s \tau_{srt}}{1-(1-\alpha_s)\eta_s - \alpha_s \eta_s \tau_{srt}} \frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1}$$

The expression for cutoff productivity derived in Equation (1.17) combined with the definition of Λ_{st} gives:

$$\theta^*_{srt} = \left\{ \frac{f}{\left[1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt}\right] \tau_{srt}^{\frac{\alpha_s\eta_s}{1-\eta_s}} \left\{\frac{w_t}{(1-\alpha_s)\eta_s}\right\}^{-\frac{(1-\alpha_s)\eta_s}{1-\eta_s}} \left\{\frac{\rho}{\alpha_s\eta_s}\right\}^{-\frac{\alpha_s\eta_s}{1-\eta_s}}} \right\}^{1-\eta_s}$$

$$= \frac{f^{1-\eta_s} \rho^{\alpha_s\eta_s} w_t^{(1-\alpha_s)\eta_s}}{\left(1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt}\right)^{1-\eta_s} \tau_{srt}^{\alpha_s\eta_s} \left((1-\alpha_s)\eta_s\right)^{(1-\alpha_s)\eta_s} \left(\alpha_s\eta_s\right)^{\alpha_s\eta_s}}$$

Endogenizing wages using the equilibrium expression derived above:

$$\theta^*_{srt} = \frac{f^{1-\eta_s} \rho^{\alpha_s\eta_s}}{\left(1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt}\right)^{1-\eta_s} \tau_{srt}^{\alpha_s\eta_s} \left((1-\alpha_s)\eta_s\right)^{(1-\alpha_s)\eta_s} \left(\alpha_s\eta_s\right)^{\alpha_s\eta_s}} \times \left\{ \frac{f}{\overline{\overline{L}}} \sum_{r=1}^{R} \sum_{s=1}^{S} \frac{(1-\alpha_s)\eta_s\tau_{srt}}{1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt}} \frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1} \right\}^{(1-\alpha_s)\eta_s}$$

Simplifying further delivers the equilibrium cutoff productivity expression:

$$\theta^*_{srt} = \frac{f^{1-\alpha_s\eta_s}}{\left((1-\alpha_s)\eta_s\overline{L}\right)^{(1-\alpha_s)\eta_s}} \left(\frac{\rho}{\alpha_s\eta_s\tau_{srt}}\right)^{\alpha_s\eta_s} \times \left\{\frac{1}{1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt}}\right\}^{1-\eta_s} \times \left\{ \sum_{r=1}^{R} \sum_{s=1}^{S} \frac{(1-\alpha_s)\eta_s\tau_{srt}}{1-(1-\alpha_s)\eta_s - \alpha_s\eta_s\tau_{srt}} \frac{\xi(1-\eta_s)}{\xi(1-\eta_s)-1} \right\}^{(1-\alpha_s)\eta_s}$$

1.10.4 MLE Estimator for Regency Flood Exposure (ϕ_{rt})

The distribution of the share variable x is as follows:

$$G_{rt}(x) = \begin{cases} 1 - \left(\frac{1}{x}\right)^{\phi_{rt}} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

This gives the p.d.f.

$$g_{rt}(x) = \frac{\phi_{rt}}{x^{\phi_{rt}+1}}$$

The likelihood function for a sample of firms in period t $(x_{1t}, x_{2t}, x_{3t}, \cdots, x_{Nt})$ located in regency r can be written as:

$$L(\phi) = \prod_{i=1}^{N_{rt}} \frac{\phi_{rt}}{x_{it}^{\phi_{rt}+1}}$$

The log-likelihood function becomes:

$$\ln(L(\phi)) = N_{rt} \ln(\phi_{rt}) - (\phi_{rt} + 1) \sum_{i=1}^{N_{rt}} \ln(x_{it})$$

To find the MLE for ϕ_{rt}, take the derivative of the log-likelihood with respect to ϕ_{rt}:

$$\frac{d \ln(L(\phi))}{d\phi} = \frac{N_{rt}}{\phi_{rt}} - \sum_{i=1}^{N_{rt}} \ln(x_{it})$$

Setting the derivative to zero delivers the estimator:

$$\widehat{\phi}_{rt} = \frac{N_{rt}}{\sum_{i=1}^{N_{rt}} \ln(x_{it})}$$

Chapter 2

Air Pollution and the Case for a Green Transition

2.1 Introduction

There is now widespread recognition that phasing out coal-fired power is a central plank of the green transition towards renewable energy. But there is also much concern that the pace of change is too slow, most often blamed on the failure of political will. Moreover, some countries continue to invest in coal-fired power plants and are even building new ones. Coal-fired power is not just bad for carbon emissions, it is detrimental to air quality with negative consequences for public health (see, for example, Lelieveld et al. 2015). This is worse when plants are located near dense population centers.[1] But it also implies that some benefits from closing coal-fired power should be both rapid and local so that local political processes can play a role in spearheading the green transition.

However, even though individual citizens may suffer the consequences of air pollution, it has long been argued that without increasing the political salience of the issue, public action may not take place (e.g. Crenson 1971; Singh and Thachil 2023, for the

1. There are at least ten thermal power plants in states of Punjab, Haryana, and Uttar Pradesh that are located in the vicinity of New Delhi, which is the most densely populated city of India. *Source:* Economic Times - Energy News, 4 June, 2021.

US and India respectively). Moreover, one way to galvanize such action is to provide evidence of collective benefits from closing down coal-fired power plants. This links to the increased interest in measuring environmental damages alongside studying ways to adapt to and mitigate their consequences (see, for example, Stern 2007; Aghion et al. 2019). Research in environmental psychology tries to uncover relationships between individual characteristics and incentives, location attributes, and perceptions towards damages, and how these interact with governance and politics (Whitmarsh 2008; Egan and Mullin 2017; Poortinga et al. 2019). Some of these studies have established correlations using variations in existing datasets at the state or city level (Howe et al. 2015; Konisky, Hughes, and Kaylor 2016) and others have leveraged more granular analysis using bespoke local surveys (Kaiser 2006; Bogner and Wiseman 1999).

However, such issues are rarely studied in low- and middle-income countries where data availability is more limited. Yet, the damages due to air pollution and climate change are argued to be disproportionately higher in the Global South (Cruz and Rossi-Hansberg 2021). Furthermore, the growth in coal-fired power in recent years has also predominantly been in the middle-income countries. This makes studying such contexts even more relevant.

This chapter has two main aims. First, we study the link between air quality perceptions and coal-fired power to show that citizens do appear to notice the detrimental effects of this polluting technology. Second, we use data on life satisfaction to construct a measure of the benefits of closing down coal-fired power stations and replacing them with renewables. We show that air quality benefits alone can be used to make the case for a green transition.

The chapter takes advantage of a unique dataset, which provides geocodes of the locations of survey respondents in 51 countries covered in the Gallup World Poll, most of which are low- and middle-income countries. Using the precise locations of interviews, we could construct a measure of proximity to coal-fired power stations. We find robust evidence that those respondents who live closer to an operational coal-fired power plant

express greater air quality dissatisfaction compared to others in the same country/region who are farther away from an operational coal-fired power plant. The link between dissatisfaction and proximity to coal-fired power cannot be explained by a priming effect since respondents were not asked about coal-fired power prior to answering the air quality question.

We then construct an equivalent variation (EV) measure by combining questions on life and air quality satisfaction with income to construct a monetary value of the benefit of switching to renewable technologies with the same electricity-generating capacity. This is motivated by on-going programs of investment in renewables as an alternative to coal-fired power plants either towards retirement or conversion into natural gas plants (Davis, Holladay, and Sims 2022).[2] Moreover, since R&D investments in energy storage technologies promise finding a way of balancing out supply and demand,[3] the transition looks technologically feasible in the near future. We find that just looking at air quality benefits yields a strong case for replacing coal-fired power with clean energy.

We use these estimates to show that the air quality satisfaction benefits from closing the "top" 25 coal-fired power stations in our sample of countries are large enough to justify their closure, even without factoring in the carbon-reduction benefits. We also use our estimated benefits "out of sample", i.e., for countries that are not in our survey data, projecting the valuations of air quality and finding a similarly strong case for closing coal-fired power stations elsewhere based solely on the air quality benefits.

These findings provide a new window on the case for phasing out coal-fired power since it stresses that this is coming from citizens' own perceptions rather than expert opinion. It therefore compliments approaches that estimate public health benefits from reducing reliance on coal-fired power, such as Lelieveld et al. (2019) which attributes 65% of excess global mortality to fossil fuel-related emissions, with significant hetero-

2. Coal will account for 85% of U.S. electricity generating capacity retirements in 2022. *Source:* US Energy Information Administration

3. In 2019, around 80% of all public energy R&D spending was on low-carbon technologies – energy efficiency, CCUS, renewables, nuclear, hydrogen, energy storage, and cross-cutting issues such as smart grids. *Source:* IEA World Energy Investment Report, 2020

geneity across regions.

Those who focus on climate change imperatives often refer to air quality improvement as a co-benefit from low-carbon investments (see, for example, Stern 2016) and that coal generation has a negative value added when accounting for the external social costs of the air pollution it produces (Muller, Mendelsohn, and Nordhaus 2011). But, when it comes to politics, it can be first-order due to its visibility. However, individuals might be aware of poor ambient air quality without being able to attribute it to their proximity to a coal-fired power station and, even if they are aware of it, they need not know about collective benefits, which are obtained by aggregating across individuals, as we do here. Ultimately, domestic and international policies to reduce carbon emissions are likely to be encouraged if citizens, firms, and civil society demand change. As stressed in Besley and Persson (2023), facilitating a green transition requires citizens as voters and consumers to embrace green values. Citizens' perceptions of the need for change are likely to be the key drivers in increasing the salience of policy issues in this area where global debates about abstract notions, like climate change, may not readily cut through.

The remainder of the chapter is organized as follows. In the next section, we discuss the data that we use. In Section 2.3, we establish a robust empirical link between a survey respondent's proximity to a coal-fired station and their satisfaction with air quality. The policy implications of these findings are developed in Section 2.4. Section 2.5 contains some concluding comments.

2.2 Data

2.2.1 Geocoded Gallup World Poll

The outcomes data is taken from the Gallup World Poll, a nationally-representative annual survey of citizens which began data collection in 2006 and represents around 99% of the world's adult population living in more than 160 countries. We only use the 2019 data in which we are given access to geocoded data for a sample of countries where

face-to-face interviews were undertaken. This excludes the US and a majority of Western European countries with phone surveys as shown in the top panel of Figure 2.5 in the Appendix. For the sample countries, we have exact latitudes and longitudes of the interview clusters and we use them to measure the distance of survey locations from the nearest coal-fired power plant. This gives a sample of 17,964 surveys from 51 countries listed in Table 2.11 and mapped in the bottom panel of Figure 2.5 in the Appendix. The main outcome variable is a binary indicator of the survey respondent's dissatisfaction with ambient air quality. The exact question (translated into English) is: *"In the city or area where you live, are you satisfied or dissatisfied with the quality of air?"*

We also use survey responses to a question on current life satisfaction as a proxy for overall wellbeing. It asks respondents to rate their present life on an eleven-point scale from 0 ("the worst possible life") to 10 ("the best possible life"). This measure of life satisfaction is popular among researchers and has been used extensively to make cross-country comparisons of wellbeing, particularly for less-developed countries (Deaton 2008; Kahneman and Deaton 2010). Apart from these two "outcome" variables, we also use controls for education, age, income, gender, and whether or not they have children under 15 years of age (also from the Gallup World Poll). We also make use of a different, but related, attitudinal survey based on a subset of countries included in the Gallup World Poll: the Lloyd's Register Foundation World Risk Poll.[4] Here also, we restrict the sample to 51 countries of the main analysis.

2.2.2 Global Energy Monitor Coal Plants Tracker

Data on coal-fired power plants come from the Global Coal Plant Tracker (GCPT) database released by the Global Energy Monitor (GEM).[5] This is freely-available data that tracks all coal-fired generating units, which are 30 MW or larger, in different stages of operation

4. In this survey, 150,000 interviews were done by Gallup in 142 countries in 2019 to measure the risk perceptions around climate change, pollution, food, cyber security, etc. (LRF 2020).

5. GCPT provides information on coal-fired power units from around the world generating 30 megawatts and above. It catalogues every operating coal-fired generating unit, every new unit proposed since 2010, and every unit retired since 2000. *Source:* Global Coal Plant Tracker - Global Energy Monitor

across the world and provides units' precise locations in terms of latitudes and longitudes and other characteristics, such as capacity, annual CO_2 emissions, etc. At present, it has detailed information on 13,412 coal units located in 108 countries. Of the total reported units, 6,613 units are operational, and these generate more than 2 million megawatts of power and produce 12 trillion kilograms of CO_2 each year. The database makes available rich data on other energy sources also, such as natural gas, wind, and solar and heavy industries, such as iron and steel. Figures 2.6 and 2.7 in the Appendix show the distribution of operational and planned units respectively for coal, solar, and wind energy generation across 51 countries that constitute our main analysis sample.

We also use remote-sensing data on vegetation cover and pollutant concentration from the NASA Earth Observations project for each survey location and a 1 km $\times$ 1 km grid population count from the Gridded Population of the World v4 (GPWv4) database for the year 2020 to compute the population estimates.

2.3 Air Quality Dissatisfaction and Coal-Fired Power

Our first step is to show that there is a robust link between air quality dissatisfaction and proximity to coal-fired power plants. This observation underpins the policy exercise that we turn to in the next section. We first lay out the empirical approach, and then develop some core results and explore their robustness.

2.3.1 Approach

Suppose that air quality dissatisfaction, *AirDiss*, for an individual, i, located near a coal plant, c, surveyed in location, ℓ, can be explained as follows:

$$AirDiss_{i\ell} = \alpha\delta_{ic} + \tau_i + \varepsilon_{i\ell} \tag{2.1}$$

where δ_{ic} is i's distance to the nearest operating coal-fired power plant, c, and τ_i represents

unobserved idiosyncratic distaste for air pollution. Obviously, we cannot estimate this exact relationship in practice because we observe each individual only once in the data, but it has value for the discussion to go forward.

If coal plants were randomly assigned to different locations, or equivalently, if individuals chose to locate randomly across different locations, then OLS would give us an unbiased estimate of α, i.e., how, on average, distance from the nearest coal-fired power plant is related to perceived ambient air quality.

If policymakers may choose to locate coal-fired power stations where opposition is lowest, i.e., where people are less concerned about pollution or people who care strongly about pollution move away from locations where there is heavy air pollution then OLS could underestimate the negative impact of coal-fired power on the general population. So we think of our results as a lower bound on the effect.[6]

Our core results come from supposing that $\tau_i = \beta \mathbf{X}_{i\ell} + \eta_\ell$, where $\mathbf{X}$ contains the geocode (latitude$\times$longitude)-level and individual-level controls, and where η_ℓ are region fixed effects, either at the country (admin 0) or state/province (admin 1) level. We then estimate the following core equation using OLS:

$$AirDiss_{i\ell} = \alpha \delta_{ic} + \beta \mathbf{X}_{i\ell} + \eta_\ell + \varepsilon_{i\ell} \tag{2.2}$$

Prior research on perceptions and actual impacts lead us to expect a larger effect on households, which are closer to coal-fired power stations (Zhang et al. 2022; Datt et al. 2023). We therefore present our main findings for three distance bands: 0-40 km,

6. More formally, note that

$$\widehat{\alpha}_{OLS} = \frac{cov\left(AirDiss_{i\ell}, \delta_{ic}\right)}{var\left(\delta_{ic}\right)} = \frac{cov\left(\alpha\delta_{ic} + \tau_i + \varepsilon_{i\ell}, \delta_{ic}\right)}{var\left(\delta_{ic}\right)} = \alpha + \frac{cov\left(\tau_i, \delta_{ic}\right)}{var\left(\delta_{ic}\right)}$$

If $cov(\varepsilon_{i\ell}, \delta_{ic}) = 0$, then any bias in OLS comes from the final term representing the correlation between unobserved tolerance for air pollution and the location of coal-fired power stations. The two sources of biases that we have mentioned would lead use to expect that $cov(\tau_i, \delta_{ic}) > 0$, implying that the estimate of α is, if anything, biased downwards as an estimate of the average relationship between being located close to a coal-fired power station and air quality dissatisfaction. In the Appendix Section 2.6, we report results from an instrumental variables strategy which, consistent with this, finds much larger estimates of the relationship between coal-fired power and air quality perceptions.

40-80 km, and 80-120 km, which are distances between a survey location and the nearest coal-fired power plant.[7]

2.3.2 Core Findings

Table 2.1 reports the results.[8] In Columns 1, 2, and 3 we use country fixed effects while those in Columns 4, 5, and 6 use state/province fixed effects. Columns 1 and 4 are for distance band 0-40 km, 2 and 5 for 40-80 km, and 3 and 6 for 80-120 km. The results in Columns 1 and 4 confirm our hypothesis that α is negative, i.e., air quality dissatisfaction is negatively correlated with distance from the nearest coal plant for respondents located within 40 km of a coal-fired power plant.[9]

The core results are robust to changing the range of distance i.e., starting from 0 km and ending at 60 km as the upper limit of domain. However, there is no effect of distance on perception when using 40-80 km or 80-120 km distance bins, thereby suggesting that the "immediate" effect is local (Ha et al. 2015).[10]

Table 2.1 also gives suggestive evidence that "elite" opinion is geared towards some form of climate action as evidenced in the gradient on education level; individuals with higher education levels tend to be significantly more dissatisfied compared to the less educated ones, *ceteris paribus*. This significant result, along with mixed patterns on age

7. We look at the concentration of pollutants around the operational coal-fired power plants to check if people's perceptions are not totally off the actual level of air pollution. We rely on remote-sensing data on pollutant concentration from NASA Earth Observations and Donkelaar et al. (2021). Figure 2.8 in the Appendix reports the mean $PM_{2.5}$ and NO_2 concentration in different distance bins relative to a coal power plant. The pollutant level goes down as one moves away from coal plant locations.

8. The OLS estimation uses a linear probability model, which might be a strong assumption given the binary nature of the dependent variable. We test the robustness of OLS results by estimating a logistic regression model with region fixed effects alongside the same controls as in the OLS specification. Results reported in Table 2.12 in the Appendix suggest that the OLS estimates are robust to relaxing the linearity assumption.

9. We also run a specification using Equation (2.2) with a general measure of health problems as the dependent variable. The exact survey question is: *Do you have any health problems that prevent you from doing any of the things that people of your age normally can do?* This is a portmanteau health question, and as expected, we do not detect any significant effect of our main regressor, δ.

10. Throughout the chapter, we report region-clustered heteroskedasticity-robust standard errors. However, following Conley (1999); Conley (2008), which allow for spatial correlation in the errors across neighboring areas with distances less than a specified threshold, we report results in Table 2.13 in the Appendix with spatial clusters defined at 5 km distance threshold. The results are essentially identical with slightly smaller standard errors.

group and income, has been documented in other studies, which use different attitudes datasets (Dechezleprêtre et al. 2022).

Taken together these results suggest that the mere existence of coal-fired power stations nearby do indeed affect perceptions of air quality negatively.[11] Also, to reiterate, we expect these to be lower bound estimates, so the actual effects could be much larger.

2.3.3 Robustness and Additional Findings

We now present a range of additional results that explore the validity of our core findings. We first show that the granularity from using geocoded data is essential to our findings. We then ask whether the core results are reflected in individual climate risk perceptions rather than air quality satisfaction. As a "placebo" test, we check whether non-operational power stations have a similar effect on air quality perceptions to those that are operational. To ensure that this effect is coming from coal-fired power plants, we also check whether the location of other polluting industries, such as iron and steel production, have similar links to air quality dissatisfaction. We also look for heterogeneous effects based on whether survey respondents are located upwind or downwind from operational coal-fired power plants. Finally, we use a semi-parametric approach to examine the validity of the distance cutoff used in our core results.

Data aggregated at regional level

A unique feature of the analysis is being able to use spatially granular data. Most previous work has used much less granular data. To show that this is important, we contrast our core findings with results using data aggregated to the region level. While we have a less clear-cut way of measuring survey respondents' proximity to coal-fired power stations, it does permit a longer time period to be studied since we can now use the World Poll for

11. To see if there is a relationship between the level of emissions and air quality dissatisfaction conditional on distance, we estimate Equation (2.2) and include an interaction of the distance regressor and the nearest plant-level annual CO_2 emissions. We find that the interaction term is not statistically significant, as reported in Table 2.14 in the Appendix. This highlights that, in our case, distance is a "sufficient statistic" to explain the effect of coal plants on people's perceptions.

all years rather than just 2019, the year for which we have geocoded data. However, to maintain comparability, we will use the same 51 countries as in our main analysis.

We experiment with different ways of defining exposure at a regional level. Our first measure is the number of operational coal-fired plants in a region in a given year divided by the total area of the region. This can be constructed without knowing specifically where a respondent lives. The second measure is closer to what we use in Equation (2.2), and is the logarithm of the average distance between all survey geocodes and the nearest operational coal-fired power plant at the region level for survey locations that are within 40 km of the plant in 2019.[12]

Results using aggregated data, which is reported in Table 2.2, show no significant relationship between any of the two measures of exposure to coal-fired power defined at the regional level and the average air quality dissatisfaction in a region. Even though the coefficients are not statistically significant, it is interesting to note that the coefficient on the second exposure variable, which is our closest counterpart to the main results reported in Table 2.1, is of the same order of magnitude as in the core results.[13]

These findings underline the value of using spatially granular data to assess the impact of coal-fired power on air quality dissatisfaction. Even our best estimate of exposure to coal-fired power based on aggregation to the region level is much less precisely estimated than what we find with precise locations.

Risk perceptions

Data from the World Risk Poll allow us to estimate a similar specification to Equation (2.2) but with the left-hand side variable now being individual risk assessments on pollution and climate. Table 2.3 reports the results.

12. For this to be an accurate exposure measure for all years, the locations of the sample collected in 2019 needs to be similar to that in other years.

13. The results in Table 2.2 also show that the magnitude of the coefficient on the exposure to coal-fired power is not sensitive to the inclusion of year fixed effects. This is also shown in Figure 2.9 in the Appendix. It suggests stable air quality perceptions over time across sample countries, thereby allaying some concerns around using only a single cross-section for 2019 for our core results.

Whether we use admin-0 or admin-1 fixed effects, we find that, as before, a significant negative relationship exists between an individual's location relative to the nearest coal power plant and their pollution risk perception when they are located within the 0-40 km distance band. However, there is no such relationship when we look at perception of risk of climate change damages.[14] This suggests that air quality perception is more linked to a visible source of risk and not linked to climate change *per se*, something that we return to when think of possible political economy implications.

Placebo tests

We report two kinds of placebo test. First, we should not expect a relationship between perceptions of air quality and *retired* (closed) or *planned* (for the future) coal-fired power plants in new locations i.e., plants that are no longer operational[15] or have been announced, at a pre-permit or permit stage of commissioning. Second, we do not expect the proximity to coal-fired power plants to be associated with reduced perceptions of other environmental amenities, such as water quality.

The results are in Table 2.4 and, as expected, the coefficients on distance are not significantly different from zero. Similarly, the effect of distance from the nearest operational coal-fired power plant on water quality dissatisfaction is also insignificant, thereby confirming our placebo hypothesis.[16]

Other polluting industries

Iron and steel production plants tend to be located near coal-fired power plants and are also a major source of local air pollution. We now see whether they have similar effects on air quality dissatisfaction.

We use the GEM database as a source of geolocation for iron and steel plants around

14. Results for 40-80 km and 80-120 km distance band are reported in Table 2.15 in the Appendix.
15. Units that have been permanently decommissioned or converted to another fuel are classified as retired while units that have been deactivated or put into an inactive state but are not retired are called mothballed units.
16. Results for 40-80 km and 80-120 km distance band are reported in Table 2.16 in the Appendix.

the world. In our core specification, we add the logarithm of the straight-line distance between an operational coal-fired power plant in the analysis and the nearest iron and steel plant, when estimating Equation (2.2). Table 2.5 reports the results. The point estimates for distance between the survey locations and coal plants remain same as from Table 2.1. In addition, the coefficient estimates for the new control variable, though smaller in magnitude, are also negative, suggesting that iron and steel plants affect air quality perceptions, but with smaller magnitude compared to coal-fired power plants.

Wind direction

Wind transports air pollutants across space and previous work has found it to be a source of heterogeneity when looking at the effects of pollution (see, for example, Deryugina et al. 2019). In the case of coal-fired power plants, we expect that areas lying downwind from the plants will receive more pollution.

We exploit cross-sectional variation in the wind direction to see whether this is a source of heterogeneity. To do so, we use the so-called u- and v-component of wind, which are wind velocities in two orthogonal directions, to derive the resultant wind direction vector at each coal-fired power plant location for all the survey geocodes located in its domain of influence.[17]

We re-estimate Equation (2.2) with the distance variable interacted with a downwind dummy that takes a value of 1 if a survey geocode is located in the domain of influence of a coal-fired power plant. Table 2.6 reports the regression results. The estimates on the downwind dummy suggest that being in the downwind direction of an operational coal power plant does not have a significant effect on local air pollution perceptions. However,

17. We use the monthly averaged u- and v-component of wind at 10 meter elevation from ground surface for single pressure level using the global version from ERA5 Climate Data Store. We do the further averaging over the monthly data for years 2015-19 to arrive at one u- and v-component for each coal plant location. To define the domain of influence i.e., wind buffer zones for each coal plant, we use the 0-40 km distance band, same as earlier, but also employ angular restriction viz. 60°, 90° and 120° angular width with the wind direction vector defining the central azimuth. All the survey geocodes that fall in the buffer zone are classified as downwind points. Figure 2.10 in the Appendix shows the buffer zones for 60° angular restriction and 40 km distance band for operational coal power plants located in some parts of the Indian subcontinent.

under strong restrictions on the domain of influence i.e., within a 0-40 km distance band and 60° angle, individuals located in downwind areas do show some tendency to express more dissatisfaction with ambient air quality, as Column 4 shows.[18] We have also looked at whether wind direction affects actual pollution measured using the $PM_{2.5}$ concentration at the geocode level. Here we also find no significant effect.[19]

Distance cutoff

Our core measure of distance focused on survey respondents residing in areas, which are less than 40 km from the nearest operational coal-fired power plant. Those who live further away do not appear to show higher levels of air quality dissatisfaction.

To explore the validity of the 40 km cutoff, Figure 2.1 shows the result of estimating a semi-parametric locally-smoothed polynomial to show how air quality dissatisfaction varies with distance. It demonstrates that air quality dissatisfaction decays, essentially to zero at around 20 km from coal power plants. Using this as our core distance measure would, however, give us a much smaller number of survey respondents, only 6% of the survey respondents live within 20 km of a coal-fired power plant compared to 13% living within 40 km. So, we are likely to get more statistical power at 40 km.[20]

2.4 Policy Implications

We have now established that perceptions of ambient air quality are indeed related to the proximity to coal-fired power plants. We use this observation to calibrate a measure of the

18. Note that we are using annual averages on wind direction, thereby removing seasonal and almost entire idiosyncratic variations that could be more important for shaping perceptions. Wind direction predictions at coal plant locations may also be measured with error due to intervening convection and radiation currents due to coal plants' operations itself (see, for example, Balboni, Burgess, and Olken 2021, which reports null effects on the propagation of forest fires.).

19. See Table 2.17 in the Appendix where we use actual pollution levels i.e. $PM_{2.5}$ concentration, and still find null effects.

20. As a further robustness check, we run our main regressions for the 0-20 km bandwidth to see whether our results continue to hold. Table 2.18 in the Appendix shows that both the main and placebo results do continue to hold even though we lose some statistical significance on the main results due to the smaller number of observations from which we are trying to identify the effect.

air quality satisfaction benefits of closing down coal-fired power stations for the approximately 1.12 billion people living within 40 km of an operational coal-fired power plant in our sample of countries and the 2.18 billion (about one-third of the global population) in the world as a whole.

The policy analysis proceeds in three steps. First, we construct the equivalent variation (EV) of increasing air quality satisfaction from the survey data using life satisfaction responses. Second, we aggregate this across the affected population. Third, we obtain a ballpark measure of the cost of replacing coal-fired power generation with a non-polluting source, such as solar or wind energy, and compare this to the benefits.

This approach builds on the large existing literature that links life satisfaction to the value of "amenities", for example, Layard, Mayraz, and Nickell (2008); Kahneman and Deaton (2010), a sub-strand of which has focused on valuing natural disasters and environmental amenities (Luechinger and Raschky 2009; Frey and Stutzer 2002; Frey, Luechinger, and Stutzer 2010; Welsch 2006). Data limitations mean that the scope of these studies has been limited to the US and parts of Europe.[21]

To construct an EV measure, we first show a negative correlation between a standard life-satisfaction measure from the Gallup survey data and air quality dissatisfaction. Since income and well-being are correlated, this can be used to calibrate the marginal rate of substitution between money and air quality dissatisfaction that can be used to create a benefit measure, which can be compared with the cost of clean energy transition. This method can be used to measure aggregate benefits but could also be deployed to gauge much more disaggregated, plant-level benefits, based on the affected local population.

21. The correlation between objective and perceived air quality is not always strong (Liu, Cranshaw, and Roseway 2020), and, arguably, it is the latter that matters most for economic decision-making and political activism (Chasco and Gallo 2013).

2.4.1 Approach

We first estimate the determinants of life satisfaction by OLS using the following econometric specification:[22]

$$LifeSat_{i\ell} = \psi \log(AirDiss_{i\ell}) + \phi \log(Income_{i\ell}) + \beta \mathbf{X}_{i\ell} + \eta_{\ell} + \varepsilon_{i\ell} \tag{2.3}$$

where the dependent variable, $LifeSat_{i\ell}$, is the life satisfaction score on a 0-10 Cantril ladder for individual i in location ℓ, η_{ℓ} controls for region fixed effects, *Income* stands for household income in 1000 USD, *AirDiss* is air quality dissatisfaction that takes value 2 (1) if individual is dissatisfied (satisfied) with ambient air quality, and $\mathbf{X}$ is a vector of controls, which are the same as in our previous specifications.

We use the estimates of ϕ and ψ to quantify the relationship between income and air quality dissatisfaction with life satisfaction. Equation (2.3) is estimated for all 51 countries in our sample.[23] The results are reported in Table 2.7.[24] To be cautious, we consider upper and lower bound estimates, from a 95% confidence interval, rather than just point estimates.[25]

Our EV measure, denoted by e, uses a reference level of air quality based on a Cobb-Douglas utility function and is defined in a standard way, as the amount of money that an individual would need to obtain the reference air quality dissatisfaction level, $\widetilde{AirDiss} < AirDiss$. This is given by:

22. There is no consensus in the literature on the exact econometric equation that should be used here, but the majority of previous work in this vein has used a specification similar to ours. The coefficient on logarithm of income is precisely estimated and is around 0.5, which lies well-within the bounds estimated in the existing literature (Layard, Mayraz, and Nickell 2008).

23. As in Section 2.3, there is a potential concern about downward bias due to selection issues here also. Some studies using a life satisfaction approach for air pollution have used IV approaches and tend to find IV estimates that are significantly larger than those found using OLS (Luechinger 2010).

24. We also estimate Equation (2.3) using actual pollution level i.e., $PM_{2.5}$ concentration at the geocode level to see whether respondents' perceptions appear "misguided". Results reported in Table 2.19 in the Appendix suggest that they are not, as the coefficient on actual pollutant level is also negative.

25. Figure 2.11 in the Appendix shows 95% confidence interval bounds on ϕ and ψ estimates for each of the 51 countries in our main sample. There appears to be a fair amount of heterogeneity in preferences across countries (Falk et al. 2018). However, this is less true for air quality preferences than income preferences.

$$\psi \log(\widetilde{AirDiss}) + \phi \log(Income - e) = \psi \log(AirDiss) + \phi \log(Income)$$

which implies that the equivalent variation is:

$$e = Income \left[1 - \exp \left\{ \frac{\psi}{\phi} \log \left(\frac{AirDiss}{\widetilde{AirDiss}} \right) \right\} \right] \tag{2.4}$$

To estimate e in Equation (2.4), we use the parameter estimate for $\frac{\psi}{\phi}$ from Column 2 of Table 2.7.[26] and the average level of dissatisfaction outside the 0-40 km distance band for the 51 countries in the core sample. The results are in Column 6 of Table 2.8 where we report results for both point estimates and at the upper and lower bounds of the 95% confidence interval from Column 2 of Table 2.7.

An attractive feature of our approach compared to standard stated preference evaluations is that it is not based on any kind of proposed hypothetical change in air quality. Hence, we believe that these estimates are less susceptible to concerns about hypothetical bias using such approaches.[27]

To obtain an Aggregate Equivalent Variation (AEV hereafter), we scale up the individual values using the measure of affected population i.e., those located within 40 km of operational coal-fired power plants. Our core results are for the world and use the population figures reported in Column 7 of Table 2.8, adjusted for the household-size to get to the total *residences* within 40 km of coal-fired plants. Multiplying this by e, we obtain our estimate of the global AEV, which we report in Column 9 of Table 2.8. We will also produce plant-level AEVs using the population that resides within 40 km of any given plant.

26. Since life satisfaction has no obvious cardinality, we follow Ferreri-Carbonell and Frijters (2004) and test the robustness of our results by estimating ordered logit models with region fixed effects alongside the same controls as in the OLS specification. The results from this exercise are in Table 2.20 in the Appendix. Our estimate of $\frac{\psi}{\phi}$ in this case is -1.047 which is close to the value of -0.989 that we get from the OLS estimation. We use the OLS estimates in the analysis that follows.

27. Such biases have been widely studied (see, for example, Murphy et al. 2005; Penn and Hu 2018).

To represent a green transition, we consider a thought experiment in which coal-fired power plants are replaced with either solar or wind farms of equivalent generation capacity over a certain period of time. To give a ballpark estimate of the costs involved, we use the total power generation capacity of coal plants and the source-specific average global Levelized Cost of Energy (LCOE).[28] We extract country-level LCOE estimates of coal, solar, and onshore wind energy from a variety of sources, which include the International Renewable Energy Agency, International Energy Agency, country reports, etc. We assume a gradual "linear" transition over twenty-five years where 4% of coal-fired power production is replaced by solar or wind in each year.[29] [30]

2.4.2 Findings

Global benefits

In Figure 2.2, we show the aggregate benefits over time for the twenty-five year time horizon for the entire world, discounted at a constant rate of 2% per year.[31] As well as point estimates, we give a shaded area for the upper and lower bounds of the global AEV. It is striking that, even at the lower bound, and only considering air quality benefits, a

28. LCOE is a popular measure to estimate the costs associated with renewables technology projects. It measures lifetime costs divided by energy production and accounts for the present value of the total cost of building and operating a power plant over an assumed lifetime. This measure allows a comparison of different technologies of unequal life spans, project size, different capital cost, risk, return, capacity factor, and capacity for each of the respective sources. Figure 2.12 in the Appendix shows the LCOE for all 51 countries in our sample; the per unit cost of energy generation is highest in the coal sector for most of the countries.

29. Fulfilling highly variable grid demand requires reliable sources of energy, such as coal and natural gas, which can supply just enough power to match both peak and off-peak demands without wasting energy whereas renewable sources suffer from uncertain fluctuations due to weather conditions. Advancement in energy storage technology is important and, apart from advancement in electrochemical storage technology, R&D investments are being made in less conventional ways to store energy, such as mechanical storage using liquid CO_2, thermal storage, and chemical storage using hydrogen. *Source:* The Economist, Technology Quarterly, June 25, 2022.

30. In light of "excess" coal power capacity in many countries, including China (Lin, Kahrl, and Liu 2018), making a transition could also pay dividends in other forms also i.e., by overcoming the sunk cost fallacy around investments in coal-fired power. Indonesia's path to green transition is getting blocked due to large sunk investments from Japan and China on coal-fired power plants in the country. *Source:* IEEFA.org

31. Following Stern (2007), there is also a debate about the correct discount rate; using 2% annual discount rate is in line with many existing studies such as Hassler, Krusell, and Nycander (2016); Nordhaus (2014).

green energy transition on the global scale is worthwhile. Moreover, these results are not particularly sensitive to the exact choice of discount factor.[32]

We have made no adjustment for the possibility that any additional fiscal burden could be costly if the transition were publicly financed. However, we do not view this as a major issue since the cost is of the order of only 1% of annual household income.[33] Hence, even as a tax-financed proposition, our proposed green transition looks feasible.

Plant-specific benefits

In practice, the decisions that policymakers will have to make to bring about a green transition will involve deciding whether to decommission specific coal-fired power plants (see, for example, Tong et al. 2021, for a discussion of the strategic importance of population density in scheduling plant retirements). Our approach allows us to construct plant-specific benefits using the AEV for those living within 40 km of any given plant. Hence, Table 2.9 presents a "league table" of the "top" 25 coal-fired power plants based on the affected population for our sample of 51 countries ranked by the total population affected by poor air quality. It is noteworthy that most of the plants on this list are in India and China, the two most populous countries in the world.[34]

Table 2.9 also presents the benefits and the costs of closing down each power station while replacing them with either wind or solar farms of equivalent generation capacities. In line with the country-level results, we find that for these highly polluting power stations, air quality benefits alone are in excess of the costs even at the lower bound estimates for gross benefits of closing them.

We can also look at the benefits from closing coal-fired power stations in countries, which are not in our sample of 51 countries by using our estimates of $\frac{\psi}{\phi}$ to estimate

32. We have established the robustness using alternative discount rates. Please see Figure 2.13 in the Appendix.

33. Figure 2.14 in the Appendix shows the values over the transition period of 25 years.

34. Table 2.21 in the Appendix looks at the plants by affected population for the world as a whole. Most of the plants are again located in China and India, and 16 out of 25 plants repeat from the previous list. Moreover, all the new plants, which are now on the list, are located in China.

benefits for these countries. Specifically, we take operational coal power plants across the globe in 2019 outside of the 51 countries in our survey sample with Table 2.10 giving a list of the top 25 most polluting coal plants for this sample. It is notable that most of the plants in this sample are located in Germany and Japan. Although the plant-level gross benefits are somewhat smaller for these plants compared to those in Table 2.9, the air quality benefits at the lower bound estimates are still able to generate positive net benefits for all plants. Thus, our finding about ambient air quality provides a potentially compelling case to close these power stations too.

As a final step, Figure 2.3 gives the plant-level net benefits for *all* operational coal-fired power plants across the world in 2019. It gives a good sense of the distribution of benefits and makes it clear that replacing coal plants with solar and wind generation units would be beneficial in almost all cases, even if we use the lower bound estimates of the net benefits of air quality improvement.[35]

2.4.3 Lessons for Political Economy

Creating a green transition that moves away from coal-fired power requires a political process, and whether having a high net benefit, as represented by our AEV measures, is sufficient to generate public action depends on the politics of decision-making. Our findings on aggregate benefits can be thought of as an input into policymaking via whatever process is in place.

To sharpen things further, we consider two countries, China and India, which. as we saw earlier, are home to most of the plants with large affected populations. For these two large countries, it makes sense to look at benefits using country-specific parameters.[36]

35. There is a growing evidence base in the engineering literature on estimating the costs of replacing fossil fuels with renewable energy generation. It suggests that the transition is unlikely to be one-to-one. Bolsona, Prieto, and Patzeka (2022) estimates that replacing 1W of fossil fuel is equivalent to installing 4W of solar capacity or 2W of wind power. We use these estimates to inflate our global LCOE values and re-plot Figure 2.3. The new plot in Figure 2.15 in the Appendix suggests that, though smaller, net benefits continue to be positive for the majority of coal plants around the world. Also, the net benefits for more plants are now negative.

36. Please see Table 2.22 in the Appendix for country-specific parameters values.

We now find that the gains from a green transition based on air quality dissatisfaction are lower in India than China mainly due to differences in estimated preference parameters.[37] This finding could explain why even if they have a political voice, Indian citizens may put less pressure on their government to reduce dependence on coal-fired power, while policy action by Chinese political elites could be justified to their citizens more easily given our finding. Either way, drawing conclusions on the potential for public action based on the findings depends critically on how such findings land in policy discussions, and the political salience of air pollution is an issue (see, for example, Crenson 1971; Singh and Thachil 2023).

Heterogeneity by education level is also interesting since those who are politically active in all kinds of governance systems tend to be more educated. Our main findings assume that $\frac{AirDiss}{\widetilde{AirDiss}}$ is common across all education categories and is set at the global level. The differences in EV are mostly guided by differences in income levels across education categories, with only small proportions of these differences explained by variation in preferences, i.e., $\frac{\psi}{\phi}$ ratio across the categories as reported in Table 2.25 in the Appendix. Again, using Equation (2.4), we find that the EV for better air quality satisfaction among highly educated individuals is more than double that of those with only primary or intermediate-level education, as reported in Table 2.26 in the Appendix. This too may be relevant in political economy terms across a range of political systems given how important elite opinion is in policymaking.

2.4.4 Further Issues

Comparison to alternative approaches We now compare our estimates to those that are obtained using Contingent Valuation Methods (CVM) and Revealed Preference (RP) approaches. We find that our estimates lie in between these two.

37. Please see Table 2.23 in the Appendix for country-specific AEV values. Figure 2.16 in the Appendix gives the benefits and costs over time for each country. Note though that the air quality benefits tend to go up substantially in India when we re-compute benefits with global preference parameters as reported in Panel 2 of Table 2.24 in the Appendix.

CVM methods, relying on survey responses, are widely used in environmental impact assessment more generally (Arrow et al. 1993; Hanemann 1994).[38] One well-known critique of such methods is that by asking directly about negative impacts of something like coal-fired power survey respondents are "primed" to think about something negative. However, the Gallup World Poll surveys do not even mention coal-fired power in the survey instruments, let alone prime respondents about it. Due to this fact, our study does not suffer from various issues raised in Diamond and Hausman (1994). To benchmark our findings against CVM studies, we use a value of $247.95 per tonne of CO_2 emissions, taken from a survey of CVM studies (Mitchell and Carson 1989).[39] Using this, we find that the aggregate benefit from eliminating coal-fired power is about 1.828 trillion USD,[40] which is 42% (215%) higher than the upper (lower) bound of our global AEV estimates reported in Table 2.8.

There is also a body of work that estimates the value of clean air using RP approach (Chay and Greenstone 2005; Ito and Zhang 2020). To compare our AEV estimates to the those obtained using RP methods, we use the lower bound estimates valued at $19.84 per tonne of CO_2 emissions from Rodemeier (2023).[41] Using this, we obtain the aggregate global benefits to be about 0.146 trillion USD,[42] which is about a quarter of the lower bound aggregate AEV estimates reported in Table 2.8. Nonetheless, with the estimates of the social cost of carbon being revised upwards, more than quadrupling in the last 10 years (Tol 2022), we expect estimates based on RP approaches to be larger in the future.

38. Such studies have also been used to study coal-fired power, e.g. Chikkatur, Chaudhary, and Sagar (2011); Wang and Mullahy (2006).

39. The average social cost of carbon is 200 EUR per 'ton' of CO_2 emitted, which when converted to USD per 'tonne' using the average 2020 EUR-USD exchange rate of 1.125 and the tonne-to-ton conversion factor of 1.102, comes out to be $247.95 per tonne of CO_2 emissions.

40. The total annual emissions in 2019 from the operational coal-fired power plants located in the sample of 51 countries was 7.371 billion tonnes. We take the product of these total emissions and $247.95 to get to the aggregate monetary benefits.

41. The lower bound of the social cost of carbon emissions is 16 EUR per 'ton' of CO_2 emitted, which when converted to USD per 'tonne' using the average 2020 EUR-USD exchange rate of 1.125 and the tonne-to-ton conversion factor of 1.102, comes out to be $19.84 per tonne of CO_2 emissions.

42. The total annual emissions in 2019 from the operational coal-fired power plants located in the sample of 51 countries was 7.371 billion tonnes. We take the product of this these emissions and $19.84 to get to the monetary equivalent of RP-based aggregate benefits.

Adding carbon benefits Coal-fired power generation is one of the biggest sources of CO_2 emissions across the world, accounting for nearly 30% of total annual global emissions with the lion's share coming from Asia.[43] Therefore, shutting down coal-fired power plants has an additional dividend in terms of carbon-reduction benefits that could help mitigate the climate change problem (Greenstone and Looney 2012).

There is much debate about the appropriate Social Cost of Carbon (SCC) to use, with widely different numbers available (Tol 2022).[44] We therefore assume lower and upper bound values of \$20 and \$100 per ton of CO_2 for our estimated benefits. Recent work estimates that the carbon benefits from a global closure of coal-fired plants is of the order of 80 trillion USD (Adrian, Bolton, and Kleinnijenhuis 2022) using a SCC value of \$75 per ton of CO_2 (Parry, Black, and Vernon 2021).

Figure 2.4 adds in the carbon-reduction benefits for a twenty-five year horizon using a 2% annual discount rate. The area covered by the upper and lower bounds on air quality benefits are shaded, but we have not shown the upper bound of carbon-reduction benefits since this, combined with air quality benefits, dwarfs other estimates. Not surprisingly, this further strengthens the case for a green energy transition.[45]

The cost of air quality deterioration, using our measure of benefits, may be lower in the future if governments move coal-fired plants away from densely populated areas to please voters. There is some evidence that this is happening: planned (future) coal plants are, on average, located farther away from large population centers when compared to the existing ones.[46]

43. Global energy-related emissions was around 33.1 Gt CO_2 in 2018; the power sector accounted for nearly two-thirds of emissions growth. Coal use in power alone surpassed 10 Gt CO_2. China, India, and the US accounted for 85% of the net increase in emissions, while emissions declined for Germany, Japan, Mexico, France and the UK. *Source:* Global Energy & CO_2 Status Report 2019

44. Although there has been more recent work on estimating these costs for specific cases, such as on human mortality and labor productivity, we do not use them as they are only partial SCC estimates (Carleton et al. 2022).

45. We can also look at plant-level net benefits after adding the carbon-reduction benefits; please see Figure 2.17 in the Appendix. The net benefits from closing almost every coal-fired power plant on the earth is positive.

46. On average, an existing operational coal plant affects 3,457,731 individuals, while a typical planned plant, which was non-existent in 2019, is expected to affect 1,328,480 individuals.

Other effects Those who depend on the coal economy, directly or indirectly, tend to express lower dissatisfaction with its existence (Eyer and Kahn 2020). Employment concerns could be important for shaping citizens' debates and policy design around a green energy transition. However, as Table 2.27 shows, it is unclear that clean energy would lead to aggregate job losses, which would depend, in part, on whether the cost of energy is higher or lower in an age of renewables as new firms tend to locate in areas with lower energy prices and where labor is available (Kahn and Mansur 2013). Nonetheless, the employment effects could still be distortionary at the local level, especially when low-skilled individuals are dependent on coal sector and allied activities. There is also a potential threat of intensive mining of aluminium, silicon, lithium, and cobalt, which are used in many forms of renewable energy generation. One also cannot discount the adverse health effects of some renewables, such as noise pollution generated by wind turbines (Zou 2017).

2.5 Conclusion

Some, but not all, countries are phasing out coal-fired electricity generation. This is, in part, motivated by concerns about climate change, but air pollution concerns are also important. As well as showing a direct link to air quality dissatisfaction, we find that citizens are more attentive to risk being framed as "pollution risk" rather than "climate risk" when they live in proximity to coal-fired power plants. Together these findings suggest that downgrading air quality to the status of a "secondary" benefit may be an error when analyzing drivers of the political economy of climate change, since air quality and local pollution may be more tangible issues.

To reinforce this message, we have used survey data to construct measures of benefits from improving air quality. By using geocoded perceptions data that we match to the location of coal-fired power stations, we have computed estimates of the benefit from phasing out coal-fired power plants based on air quality dissatisfaction. Being able to do

this for countries in the Global South, where expanding generation capacity is likely to be greatest in the years to come, is particularly important. These findings are particularly relevant to countries like China and India, which are home to many of the largest coal-fired power systems. The analysis suggests that air quality benefits alone (without factoring in carbon-reduction benefits) can make a credible case for phasing out coal-fired power in such places.

References

Adrian, Tobias, Patrick Bolton, and Alissa M. Kleinnijenhuis. 2022. "The Great Carbon Arbitrage." *IMF Working Paper,* nos. 2022/107.

Aghion, Philippe, Cameron Hepburn, Alexander Teytelboym, and Dimitri Zenghelis. 2019. "Path Dependence, Innovation and The Economics of Climate Change." *Handbook on Green Growth.*

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. 1993. "Report of The NOAA Panel on Contingent Valuation." *Fed. Regist.* 58:4601–14.

Balboni, Clare, Robin Burgess, and Benjamin A. Olken. 2021. "The Origins and Control of Forest Fires in the Tropics." *LSE Working Paper.*

Besley, Timothy J., and Torsten Persson. 2023. "The Political Economics of Green Transitions." *Quarterly Journal of Economics,* forthcoming.

Bogner, Franz X., and Michael Wiseman. 1999. "Toward Measuring Adolescent Environmental Perception." *European Psychologist* 4 (3): 139–151.

Bolsona, Natanael, Pedro Prieto, and Tadeusz Patzeka. 2022. "Capacity Factors for Electrical Power Generation from Renewable and Non-renewable Sources." *Proceedings of National Academy of Sciences* 119 (52).

Carleton, Tamma, Amir Jina, Michael Delgado, Michael Greenstone, Trevor Houser, Solomon Hsiang, et al. 2022. "Valuing The Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits." *Quarterly Journal of Economics* 137 (4): 2037–2105.

Chasco, Coro, and Julie Le Gallo. 2013. "The Impact of Objective and Subjective Measures of Air Quality and Noise on House Prices: A Multilevel Approach for Downtown Madrid." *Economic Geography* 89 (2): 127–148.

Chay, Kenneth Y., and Michael Greenstone. 2005. "Does Air Quality Matter? Evidence From The Housing Market." *Journal of Political Economy* 113 (2): 376–424.

Chikkatur, Ananth P., Ankur Chaudhary, and Ambuj D. Sagar. 2011. "Coal Power Impacts, Technology, and Policy: Connecting The Dots." *Annual Review of Environment and Resources* 36:101–138.

Conley, T. G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92 (1): 1–45.

———. 2008. *Spatial Econometrics.* In Microeconometrics, edited by Steven N. Durlauf / Lawrence E. Blume, 303–13. London: Palgrave Macmillan.

Crenson, Matthew. 1971. *The Un-Politics of Air Pollution: A Study of Non-Decisionmaking in The Cities.* John Hopkins Press.

Cruz, José-Luis, and Esteban Rossi-Hansberg. 2021. "The Economic Geography of Global Warming." *University of Chicago, Becker Friedman Institute for Economics Working Paper,* nos. 2021-130.

Datt, Gaurav, Pushkar Maitra, Nidhiya Menon, and Ranjan Ray. 2023. "Impact of Pollution from Coal on the Anemic Status of Children and Women: Evidence from India." *Journal of Development Studies,* forthcoming.

Davis, R. J., J. S. Holladay, and C Sims. 2022. "Coal-fired power Plant Retirements in the United States." *Environmental and Energy Policy and the Economy* 3 (1): 4–36.

Deaton, Angus. 2008. "Income, Health, and Well-Being Around The World: Evidence From The Gallup World Poll." *Journal of Economic Perspectives* 22 (2): 53–72.

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. 2022. "Fighting Climate Change: International Attitudes Toward Climate Policies." *NBER Working Paper Series,* no. 30265.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif. 2019. "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction." *American Economic Review* 109 (12): 4178–4219.

Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives* 8 (4): 45–64.

Donkelaar, Aaron van, et al. 2021. "Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty." *Environ. Sci. Technol.* 55 (22): 15287–15300.

Egan, Patrick J., and Megan Mullin. 2017. "Climate Change: US Public Opinion." *Annual Review of Political Science* 20:209–27.

Eyer, Jonathan, and Matthew E. Kahn. 2020. "Prolonging Coal's Sunset: Local Demand for Local Supply." *Regional Science and Urban Economics* 81 (103487).

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics* 133 (4): 1645–1692.

Ferreri-Carbonell, Ada, and Paul Frijters. 2004. "How Important is Methodology for The Estimates of The Determinants of Happiness." *The Economic Journal* 114 (497): 641–659.

Frey, Bruno S., Simon Luechinger, and Alois Stutzer. 2010. "The Life Satisfaction Approach to Environmental Valuation." *Annual Review of Resource Economics* 2:139–160.

Frey, Bruno S., and Alois Stutzer. 2002. "What Can Economists Learn From Happiness Research?" *Journal of Economic Literature* 40 (2): 402–435.

Greenstone, M., and A. Looney. 2012. "Paying Too Much for Energy? The True Costs of Our Energy Choices." *Daedalus* 141 (2): 10–30.

Ha, Sandie, Hui Hu, Jeffrey Roth, Haidong Kan, and Xiaohui Xu. 2015. "Associations Between Residential Proximity to Power Plants and Adverse Birth Outcomes." *American Journal of Epidemiology* 182 (3): 215–224.

Hanemann, W. Michael. 1994. "Valuing The Environment Through Contingent Valuation." *Journal of Economic Perspectives* 8 (4): 19–43.

Hassler, John, Per Krusell, and Jonas Nycander. 2016. "Climate Policy." *Economic Policy* 31 (87): 503–558.

Howe, Peter D., Matto Mildenberger, Jennifer R. Marlon, and Anthony Leiserowitz. 2015. "Geographic Variation in Opinions on Climate Change at State and Local Scales in The USA." *Nature Climate Change* 5 (6): 596–603.

Ito, Koichiro, and Shuang Zhang. 2020. "Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China." *Journal of Political Economy* 128 (5): 1627–1672.

Kahn, Matthew E., and Erin T. Mansur. 2013. "Do Local Energy Prices and Regulation Affect The Geographic Concentration of Employment?" *Journal of Public Economics* 101:105–114.

Kahneman, Daniel, and Angus Deaton. 2010. "High Income Improves Evaluation of Life but Not Emotional Well-Being." *Proceedings of the National Academy of Sciences* 107 (38): 16489–16493.

Kaiser, Florian G. 2006. "A General Measure of Ecological Behavior." *Journal of Applied Social Psychology* 28 (5): 395–422.

Konisky, David M., Llewelyn Hughes, and Charles H. Kaylor. 2016. "Extreme Weather Events and Climate Change Concern." *Climatic Change* 134 (4): 533–547.

Layard, Richard, Guy Mayraz, and Stephen Nickell. 2008. "The Marginal Utility of Income." *Journal of Public Economics* 92 (8-9): 1846–1857.

Lelieveld, J., J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer. 2015. "The Contribution of Outdoor Air Pollution Sources to Premature Mortality on a Global Scale." *Nature* 525:367–371.

Lelieveld, J., K. Klingmüller, A. Pozzer, R. T. Burnett, A. Haines, and V. Ramanathan. 2019. "Effects of Fossil Fuel and Total Anthropogenic Emission Removal on Public Health and Climate." *Proceedings of National Academy of Sciences* 116 (15): 7192–7197.

Lin, Jiang, Fredrich Kahrl, and Xu Liu. 2018. "A Regional Analysis of Excess Capacity in China's Power Systems." *Resources, Conservation and Recycling* 129:93–101.

Liu, Szu-Yu, Justin Cranshaw, and Asta Roseway. 2020. "Making Air Quality Data Meaningful: Coupling Objective Measurement With Subjective Experience Through Narration." In *Proceedings of the 2020 ACM Designing Interactive Systems Conference,* 1313–1326.

LRF. 2020. *World Risk Poll 2019*. Lloyd's Register Foundation. https://wrp.lrfoundation.org.uk/2019-world-risk-poll/data-resources/.

Luechinger, Simon. 2010. “Life Satisfaction and Transboundary Air Pollution.” *Economics Letters* 107:4–6.

Luechinger, Simon, and Paul A. Raschky. 2009. “Valuing Flood Disasters Using The Life Satisfaction Approach.” *Journal of Public Economics* 93 (3-4): 620–633.

Mitchell, Robert C., and Richard T. Carson. 1989. “Using Surveys to Value Public Goods: The Contingent Valuation Method.”

Muller, N. Z., R. Mendelsohn, and W. Nordhaus. 2011. “Environmental Accounting for Pollution in the United States Economy.” *American Economic Review* 101 (5): 1649–1675.

Nordhaus, William. 2014. “Estimates of The Social Cost of Carbon: Concepts and Results from The DICE-2013R Model and Alternative Approaches.” *Journal of the Association of Environmental and Resource Economists* 1 (1/2): 273–312.

Parry, Ian W.H., Simon Black, and Nate Vernon. 2021. “Still Not Getting Energy Prices Right: A Global and Country Update of Fossil Fuel Subsidies.” *IMF Working Paper,* nos. 2021/236.

Poortinga, Wouter, Lorraine Whitmarsh, Linda Steg, Gisela Böhm, and Stephen Fisher. 2019. “Climate Change Perceptions and Their Individual-Level Determinants: A Cross-European Analysis.” *Global Environmental Change* 55:25–35.

Rodemeier, Matthias. 2023. “Willingness to Pay for Carbon Mitigation: Field Evidence from the Market for Carbon Offsets.” *IZA Discussion Paper,* no. 15939.

Singh, Shikhar, and Tariq Thachil. 2023. “Why Citizens Don’t Hold Politicians Accountable for Air Pollution.” *OSF Preprints.*

Stern, Nicholas. 2007. *The Economics of Climate Change: The Stern Review*. Cambridge University Press.

Stern, Nicholas. 2016. *Why Are We Waiting? The Logic, Urgency, and Promise of Tackling Climate Change*. MIT Press.

Tol, Richard S. J. 2022. *Estimates of The Social Cost of Carbon Have Increased Over Time*. https://arxiv.org/abs/2105.03656.

Tong, Dan, Guannan Geng, Qiang Zhang, Jing Cheng, Xinying Qin, Chaopeng Hong, Kebin He, and Steven J. Davis. 2021. "Health Co-benefits of Climate Change Mitigation Depend on Strategic Power Plant Retirements and Pollution Controls." *Nature Climate Change* 11 (2): 1077–1083.

Wang, Hong, and John Mullahy. 2006. "Willingness to Pay for Reducing Fatal Risk by Improving Air Quality: A Contingent Valuation Study in Chongqing, China." *Science of The Total Environment* 367 (1): 50–57.

Welsch, Heinz. 2006. "Environment and Happiness: Valuation of Air Pollution Using Life Satisfaction Data." *Ecological Economics* 58 (4): 801–813.

Whitmarsh, Lorraine. 2008. "Are Flood Victims More Concerned About Climate Change Than Other People? The Role of Direct Experience in Risk Perception and Behavioural Response." *Journal of Risk Research* 11 (3): 351–374.

Zhang, Charlie H., Lonnie Sears, John V. Myers, Guy N. Brock, Clara G. Sears, and Kristina M. Zierold. 2022. "Proximity to Coal-Fired Power Plants and Neurobehavioral Symptoms in Children." *Journal of Exposure Science & Environmental Epidemiology* 32 (1): 124–134.

Zou, Eric. 2017. "Wind Turbine Syndrome: The Impact of Wind Farms on Suicide." *Working Paper*. HPSA SATC Income Climate 36.

Main Figures and Tables

Table 2.1: Results for air quality dissatisfaction and operational plants location

	(1) Air Diss	(2) Air Diss	(3) Air Diss	(4) Air Diss	(5) Air Diss	(6) Air Diss
Geocode's log dist from nearest plant	-0.044***	-0.056	-0.094	-0.039***	-0.020	-0.111
	(0.0106)	(0.0407)	(0.0617)	(0.0106)	(0.0372)	(0.0837)
Geocode's vegetation index	-0.097**	-0.097*	-0.084	-0.063*	-0.104**	-0.139*
	(0.0327)	(0.0455)	(0.0473)	(0.0297)	(0.0395)	(0.0580)
Geocode area is urban	0.106***	0.144***	0.142***	0.089***	0.120***	0.125***
	(0.0215)	(0.0248)	(0.0359)	(0.0203)	(0.0172)	(0.0261)
Respondent's age is 26-60 years	0.020	0.016	0.027**	0.015	0.022*	0.030**
	(0.0104)	(0.0101)	(0.0082)	(0.0099)	(0.0090)	(0.0099)
Respondent's age is more than 60 years	-0.022	0.011	0.018	-0.020	0.017	0.027*
	(0.0150)	(0.0123)	(0.0125)	(0.0128)	(0.0119)	(0.0132)
Respondent's gender is male	-0.018*	-0.020*	-0.016*	-0.015*	-0.015*	-0.012
	(0.0089)	(0.0081)	(0.0064)	(0.0072)	(0.0068)	(0.0071)
Respondent's education is intermediate	0.057***	0.039*	0.037**	0.059***	0.036***	0.035***
	(0.0102)	(0.0150)	(0.0131)	(0.0100)	(0.0103)	(0.0100)
Respondent's education is high	0.089***	0.066***	0.059**	0.089***	0.059***	0.062***
	(0.0151)	(0.0173)	(0.0217)	(0.0142)	(0.0169)	(0.0159)
Log annual hh income in '000 USD	-0.006	-0.003	-0.009	-0.004	-0.006	-0.010*
	(0.0054)	(0.0052)	(0.0049)	(0.0050)	(0.0042)	(0.0047)
Respondent has children under 15 yrs	0.004	0.000	0.010	0.001	0.001	0.008
	(0.0077)	(0.0093)	(0.0111)	(0.0077)	(0.0078)	(0.0091)
Number of observations	17,964	16,461	13,137	17,964	16,461	13,137
Adj R-squared	0.128	0.092	0.110	0.179	0.167	0.162
Mean of dependent variable	0.327	0.249	0.240	0.327	0.249	0.240
Region fixed effects	Admin-0	Admin-0	Admin-0	Admin-1	Admin-1	Admin-1
Distance band	0-40 km	40-80 km	80-120 km	0-40 km	40-80 km	80-120 km

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2) for operational coal-fired power plants. The sample used in each column is defined by the distance band i.e., how far the survey location is relative to the nearest coal power plant. Table 2.11 provides the list of countries that are used in the main specification i.e., 0-40 km distance band and results are reported in Columns 1 and 4. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for first three columns and state/province/admin-1 level for last three columns. Columns 1-3 and Columns 4-6 control for admin-0 and admin-1 fixed effects respectively. The dependent variable, *Air Diss*, is a shorthand for Air Quality Dissatisfaction, which takes value 1 (0) if the surveyed individual is dissatisfied (satisfied) with the ambient air quality. The main variable of interest is geocode's logarithm of distance from the nearest plant, which is the straight-line distance between the survey and nearest coal plant location. Vegetation index measures green cover for survey location and urban is a dummy variable for urban area classification. The regression also controls for the respondent's age group (young/middle-aged/old), gender (male/female), education level (primary/intermediate/high), logarithm of household income in 1000 USD, and whether the respondent has children under 15 years of age.

Table 2.2: Results for regional exposure to operational plants

	(1) Air Diss	(2) Air Diss	(3) Air Diss	(4) Air Diss
#Coal plants over total area of region	-2.337	-1.870		
	(1.7254)	(1.4962)		
Log avg. region-level distance from coal plant			-0.015	-0.015
			(0.0112)	(0.0111)
Regional vegetation index	-0.299*	-0.046	-0.124	-0.101
	(0.1247)	(0.1219)	(0.0752)	(0.0770)
Area is urban	0.150***	0.149***	0.180***	0.180***
	(0.0103)	(0.0103)	(0.0206)	(0.0204)
Respondent's age is 26-60 years	0.003	0.003	0.003	0.003
	(0.0025)	(0.0025)	(0.0033)	(0.0034)
Respondent's age is more than 60 years	-0.033***	-0.032***	-0.032***	-0.031***
	(0.0041)	(0.0040)	(0.0059)	(0.0057)
Respondent's gender is male	-0.016***	-0.016***	-0.018***	-0.018***
	(0.0027)	(0.0027)	(0.0047)	(0.0046)
Respondent's education is intermediate	0.032***	0.033***	0.034**	0.036**
	(0.0040)	(0.0041)	(0.0101)	(0.0105)
Respondent's education is high	0.072***	0.074***	0.076***	0.079***
	(0.0055)	(0.0055)	(0.0123)	(0.0129)
Log annual hh income in '000 USD	-0.001	-0.000	0.002	0.003
	(0.0018)	(0.0018)	(0.0043)	(0.0047)
Respondent has children under 15 yrs	-0.001	-0.001	-0.001	-0.002
	(0.0024)	(0.0024)	(0.0028)	(0.0027)
Number of observations	340,657	340,657	340,657	340,657
Adj R-squared	0.141	0.142	0.118	0.119
Mean of dependent variable	0.288	0.288	0.288	0.288
Region fixed effects	Admin-1	Admin-1	Admin-0	Admin-0
Time fixed effects	-	Year	-	Year
Years included	2009-20	2009-20	2009-20	2009-20

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2) for operational coal-fired power plants where δ is replaced by an "exposure" variable, which is either (i) the number of coal plants per square kilometers of area of region or (ii) logarithm of average distance of survey geocodes from the nearest operational coal-fired power plant at the region level in 2019. Columns 1-2 and 3-4 use exposure variable (i) and (ii) respectively. All the regressions use the sample of 51 countries in the main analysis, as given in Table 2.11. Standard errors, which are reported in parentheses, are clustered at admin-1 level for Columns 1-2 and at admin-0 level for the remaining ones. Columns 2 and 4 control for year fixed effects. The dependent variable, *Air Diss*, is a shorthand for Air Quality Dissatisfaction, which takes value 1 (0) if the surveyed individual is dissatisfied (satisfied) with the ambient air quality. Please refer to Table 2.1 notes for details on other variables.

Table 2.3: Risk assessment results for operational plants

	(1) Poll Risk	(2) Poll Risk	(3) Clim Risk	(4) Clim Risk
Geocode's log dist from nearest plant	-0.005**	-0.006*	0.005	0.006
	(0.0018)	(0.0027)	(0.0044)	(0.0054)
Geocode's vegetation index	0.004	0.010*	0.023	0.021
	(0.0036)	(0.0050)	(0.0183)	(0.0181)
Geocode area is urban	-0.002	-0.004	-0.021*	-0.016*
	(0.0032)	(0.0043)	(0.0098)	(0.0080)
Respondent's age is 26-60 years	0.000	-0.001	0.008	0.006
	(0.0029)	(0.0029)	(0.0068)	(0.0049)
Respondent's age is more than 60 years	-0.004	-0.004	0.012	0.014*
	(0.0044)	(0.0037)	(0.0083)	(0.0067)
Respondent's gender is male	-0.003	-0.003	-0.003	-0.004
	(0.0020)	(0.0022)	(0.0057)	(0.0046)
Respondent's education is intermediate	0.003	0.003	-0.003	-0.004
	(0.0023)	(0.0025)	(0.0082)	(0.0062)
Respondent's education is high	0.008*	0.008*	0.009	0.006
	(0.0042)	(0.0040)	(0.0070)	(0.0081)
Log annual hh income in '000 USD	-0.000	-0.000	0.002	0.004
	(0.0016)	(0.0016)	(0.0031)	(0.0023)
Respondent has children under 15 yrs	0.001	0.002	-0.001	-0.001
	(0.0022)	(0.0025)	(0.0043)	(0.0047)
Number of observations	15,117	15,117	15,117	15,117
Adj R-squared	0.031	0.030	0.036	0.061
Mean of dependent variable	0.016	0.016	0.062	0.062
Region fixed effects	Admin-0	Admin-1	Admin-0	Admin-1
Distance band	0-40 km	0-40 km	0-40 km	0-40 km

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2). The sample used in each column is defined by the distance band i.e., how far the survey location is relative to the nearest coal power plant. Table 2.11 provides the list of countries that are used in the main specification i.e., 0-40 km distance band. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for Columns 1 and 3 and state/province/admin-1 level for remaining columns. Columns 1 and 3 and Columns 2 and 4 control for admin-0 and admin-1 fixed effects respectively. The dependent variables, *Poll Risk* and *Clim Risk*, are shorthands for Pollution Risk and Climate Risk respectively. Poll Risk/Clim Risk take value 1 (0) if the surveyed individual does (does not) considers pollution/climate as one of the two major sources of risks to their safety in daily life. The main variable of interest is the geocode's logarithm of distance from the nearest plant, which is the straight-line distance between the survey and nearest coal plant location. Please refer to Table 2.1 notes for details on other variables.

Table 2.4: Placebo results for non-operational plants and water quality perception

	(1) Air Diss	(2) Air Diss	(3) Air Diss	(4) Air Diss	(5) Water Diss
Geocode's log dist from nearest plant	0.004	-0.001	-0.045	-0.015	-0.012
	(0.0162)	(0.0199)	(0.0344)	(0.0290)	(0.0099)
Geocode's vegetation index	-0.141*	-0.039	-0.479**	-0.420	-0.023
	(0.0612)	(0.0774)	(0.1178)	(0.2328)	(0.0450)
Geocode area is urban	0.108*	0.117**	0.046	0.070	0.011
	(0.0401)	(0.0390)	(0.0320)	(0.0645)	(0.0160)
Respondent's age is 26-60 years	0.026	0.011	-0.006	0.009	0.036***
	(0.0244)	(0.0261)	(0.0194)	(0.0324)	(0.0094)
Respondent's age is more than 60 years	0.021	0.010	-0.047	-0.026	0.001
	(0.0240)	(0.0347)	(0.0275)	(0.0322)	(0.0117)
Respondent's gender is male	-0.022	-0.019	-0.027*	-0.029	-0.019**
	(0.0183)	(0.0241)	(0.0090)	(0.0200)	(0.0071)
Respondent's education is intermediate	0.023	0.015	0.068*	0.073**	0.036***
	(0.0274)	(0.0231)	(0.0295)	(0.0224)	(0.0100)
Respondent's education is high	-0.002	-0.015	0.077*	0.066	0.057***
	(0.0378)	(0.0323)	(0.0253)	(0.0351)	(0.0134)
Log annual hh income in '000 USD	-0.022	-0.015	-0.015	-0.015	-0.006
	(0.0132)	(0.0124)	(0.0081)	(0.0097)	(0.0050)
Respondent has children under 15 yrs	-0.000	0.009	-0.016	-0.041	-0.005
	(0.0236)	(0.0190)	(0.0231)	(0.0303)	(0.0079)
Number of observations	2,948	2,948	2,317	2,317	18,027
Adj R-squared	0.059	0.114	0.125	0.192	0.106
Mean of dependent variable	0.284	0.284	0.291	0.291	0.280
Region fixed effects	Admin-0	Admin-1	Admin-0	Admin-1	Admin-1
Distance band	0-40 km	0-40 km	0-40 km	0-40 km	0-40 km
Status of plant operation	Planned	Planned	Retired	Retired	Operational

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2) separately for planned and retired and mothballed coal-fired power plants and for water quality dissatisfaction. The sample used in each column is defined by the distance band i.e., how far the survey location is relative to the nearest coal power plant. Table 2.11 provides the list of countries that are used in the main specification i.e., 0-40 km distance band. Columns 1-2 and Columns 3-4 report results for planned and retired plants respectively and Column 5 reports results for water quality instead of air quality dissatisfaction. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for Columns 1 and 3 and at state/province/admin-1 level for remaining columns. Columns 1 and 3 control for admin-0 fixed effects and remaining columns control for admin-1 fixed effects. The dependent variable, *Air (Water) Diss*, is a shorthand for Air (Water) Quality Dissatisfaction, which takes value 1 (0) if the surveyed individual is dissatisfied (satisfied) with the ambient air(water) quality. The main variable of interest is geocode's logarithm of distance from the nearest plant, which is the straight-line distance between the survey and nearest coal plant location. Please refer to Table 2.1 notes for details on other variables.

Table 2.5: Results for operational plants with iron and steel plants' distance control

	(1) Air Diss	(2) Air Diss
Geocode's log dist from nearest plant	-0.045***	-0.038***
	(0.0122)	(0.0107)
Coal plant's log dist from nearest steel plant	-0.023***	-0.018*
	(0.0081)	(0.0108)
Geocode's vegetation index	-0.094***	-0.062**
	(0.0313)	(0.0295)
Geocode area is urban	0.100***	0.084***
	(0.0211)	(0.0207)
Respondent's age is 26-60 years	0.019*	0.016
	(0.0102)	(0.0099)
Respondent's age is more than 60 years	-0.021	-0.020
	(0.0149)	(0.0128)
Respondent's gender is male	-0.017*	-0.016**
	(0.0090)	(0.0072)
Respondent's education is intermediate	0.055***	0.059***
	(0.0099)	(0.0100)
Respondent's education is high	0.089***	0.090***
	(0.0147)	(0.0141)
Log annual hh income in '000 USD	-0.008	-0.004
	(0.0055)	(0.0050)
Respondent has children under 15 yrs	0.005	0.001
	(0.0076)	(0.0077)
Number of observations	17,964	17,964
Adj R-squared	0.131	0.179
Mean of dependent variable	0.327	0.327
Region fixed effects	Admin-0	Admin-1
Distance band	0-40 km	0-40 km

Region-clustered robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents OLS estimates using the specification in Equation (2.2) but including an additional control variable for logarithm of distance between coal plants and iron and steel plant i.e., how far an operational coal power plant is from the nearest iron and steel production unit, for the distance band 0-40 km. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for Columns 1 and at state/province/admin-1 level for Column 2. Columns 1 control for admin-0 fixed effects and Column 2 for admin-1 fixed effects. The dependent variable, *Air(Water) Diss*, is a shorthand for Air(Water) Quality Dissatisfaction, which takes value 1 (0) if the surveyed individual is dissatisfied (satisfied) with the ambient air(water) quality. The main variable of interest is geocode's logarithm of distance from the nearest plant, which is the straight-line distance between the survey and nearest coal plant location. Please refer to Table 2.1 notes for details on other variables.

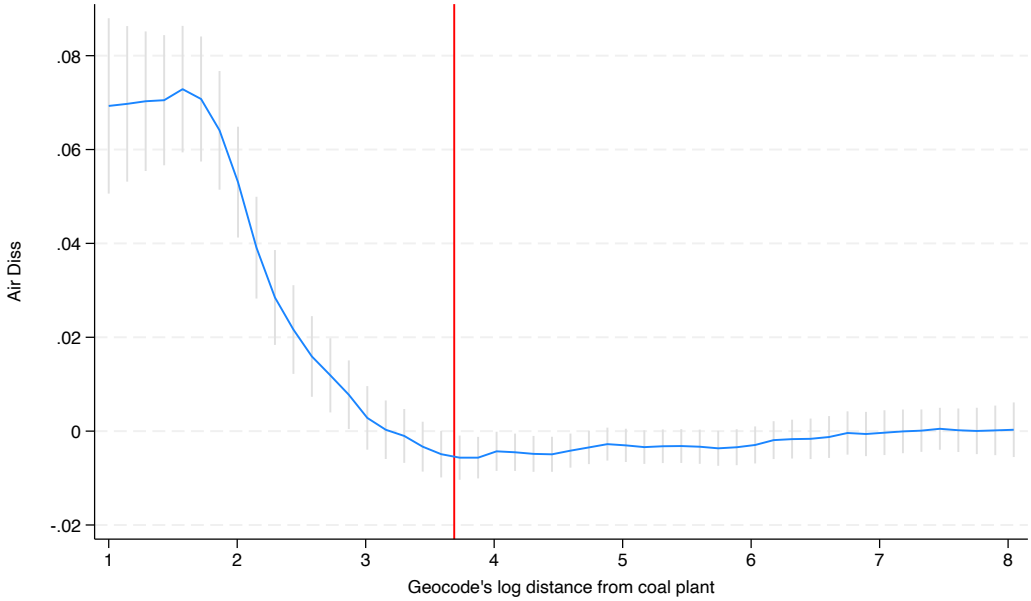
Table 2.6: Results for operational plants with wind direction interaction

	(1) Air Diss	(2) Air Diss	(3) Air Diss	(4) Air Diss	(5) Air Diss	(6) Air Diss
Geocode's log dist from nearest plant	-0.050***	-0.047***	-0.051***	-0.049***	-0.047***	-0.051***
	(0.0124)	(0.0124)	(0.0124)	(0.0123)	(0.0127)	(0.0130)
Downwind of plant	-0.046	-0.009	-0.029	-0.097	-0.064	-0.073
	(0.0645)	(0.0630)	(0.0585)	(0.0523)	(0.0521)	(0.0469)
Downwind of plant × Geocode's log dist from nearest plant	0.026	0.014	0.017	0.040*	0.027	0.025
	(0.0225)	(0.0221)	(0.0207)	(0.0188)	(0.0185)	(0.0163)
Geocode's vegetation index	-0.096**	-0.098**	-0.097**	-0.060*	-0.061*	-0.063*
	(0.0319)	(0.0322)	(0.0330)	(0.0294)	(0.0295)	(0.0299)
Geocode area is urban	0.107***	0.106***	0.107***	0.089***	0.089***	0.089***
	(0.0216)	(0.0215)	(0.0215)	(0.0204)	(0.0203)	(0.0204)
Respondent's age is 26-60 years	0.020	0.020	0.019	0.016	0.016	0.015
	(0.0104)	(0.0105)	(0.0104)	(0.0099)	(0.0099)	(0.0099)
Respondent's age is more than 60 years	-0.021	-0.021	-0.022	-0.020	-0.020	-0.020
	(0.0152)	(0.0152)	(0.0151)	(0.0128)	(0.0128)	(0.0128)
Respondent's gender is male	-0.018*	-0.018*	-0.018*	-0.016*	-0.016*	-0.016*
	(0.0088)	(0.0088)	(0.0088)	(0.0072)	(0.0072)	(0.0072)
Respondent's education is intermediate	0.057***	0.057***	0.057***	0.059***	0.059***	0.059***
	(0.0101)	(0.0100)	(0.0102)	(0.0100)	(0.0100)	(0.0100)
Respondent's education is high	0.088***	0.089***	0.089***	0.089***	0.089***	0.089***
	(0.0150)	(0.0149)	(0.0149)	(0.0142)	(0.0142)	(0.0142)
Log annual hh income in '000 USD	-0.006	-0.006	-0.006	-0.004	-0.004	-0.004
	(0.0054)	(0.0054)	(0.0054)	(0.0050)	(0.0050)	(0.0050)
Respondent has children under 15 yrs	0.003	0.003	0.003	0.001	0.001	0.001
	(0.0076)	(0.0077)	(0.0077)	(0.0077)	(0.0077)	(0.0077)
Number of observations	17,964	17,964	17,964	17,964	17,964	17,964
Adj R-squared	0.129	0.129	0.128	0.179	0.179	0.179
Mean of dependent variable	0.327	0.327	0.327	0.327	0.327	0.327
Region fixed effects	Admin-0	Admin-0	Admin-0	Admin-1	Admin-1	Admin-1
Distance band	0-40 km	0-40 km	0-40 km	0-40 km	0-40 km	0-40 km
Wind direction angular buffer	60°	90°	120°	60°	90°	120°

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2) for operational coal-fired power plants but interacting δ with a dummy for downwind direction of coal-fired power plant. The sample used in each column is defined by the distance band 0-40 km and the angular buffer around the coal-fired power plant i.e., all survey locations that are located within 40 km and falling in the angular buffer of either 60°, 90° or 120° of an operational coal power plant. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for Columns 1-3 and state/province/admin-1 level for remaining columns. Columns 1-3 control for admin-0 fixed effects and remaining columns control for admin-1 fixed effects. The dependent variable, *Air Diss*, is a short-hand for Air Quality Dissatisfaction, which takes value 1 (0) if the surveyed individual is dissatisfied (satisfied) with the ambient air quality. Geocode's logarithm of distance from the nearest plant is a measure of straight-line distance between the survey location and nearest coal plant location. Wind direction is a dummy, which takes value of 1 if the survey geocode falls in the downwind buffer region of a coal-fired power plant, and that varies based on the angular threshold used. Please refer to Table 2.1 notes for details on other variables.

Figure 2.1: Effect of distance from operational plants on air quality dissatisfaction



Notes: The graph above shows local polynomial regression results with 90% confidence intervals spikes for the effect of logarithm of distance of geocode from an operational coal plant on the residualized value of air quality dissatisfaction that is obtained after running an OLS similar to Equation (2.2) but without the distance regressor. The red line shows our chosen distance threshold of 40 km. We censor the distance values, which are less than "e" i.e., 2.718 km to avoid issues due to small sample in the left tail of the distance distribution. The dependent variable, *Air Diss*, is a shorthand for Air Quality Dissatisfaction, which takes value 1 (0) if the surveyed individual is dissatisfied (satisfied) with the ambient air quality. The main regressor, geocode's logarithm of distance from the nearest plant, is the straight-line distance between the survey and nearest coal plant location.

Table 2.7: Life satisfaction regression results for operational plants

	(1) Life Sat	(2) Life Sat
Log air quality dissatisfaction	-0.482***	-0.469***
	[-0.643,-0.321]	[-0.611,-0.326]
Geocode's vegetation index	-0.041	0.010
	[-0.310,0.227]	[-0.226,0.247]
Geocode area is urban	0.097	0.107
	[-0.037,0.232]	[-0.041,0.255]
Respondent's age is 26-60 years	-0.331***	-0.377***
	[-0.454,-0.209]	[-0.481,-0.272]
Respondent's age is more than 60 years	-0.431**	-0.467***
	[-0.746,-0.115]	[-0.623,-0.311]
Respondent's gender is male	-0.166*	-0.159***
	[-0.317,-0.016]	[-0.252,-0.067]
Respondent's education is intermediate	0.313***	0.328***
	[0.158,0.468]	[0.203,0.452]
Respondent's education is high	0.669***	0.703***
	[0.523,0.815]	[0.543,0.863]
Log annual hh income in '000 USD	0.489***	0.474***
	[0.357,0.620]	[0.404,0.543]
Respondent has children under 15 yrs	-0.023	0.031
	[-0.161,0.115]	[-0.062,0.124]
Number of observations	17,701	17,701
Adj R-squared	0.203	0.238
Mean of dependent variable	5.411	5.411
Mean household income in USD	14855	14855
Region fixed effects	Admin-0	Admin-1
Countries included	Global	Global

95% confidence interval in brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

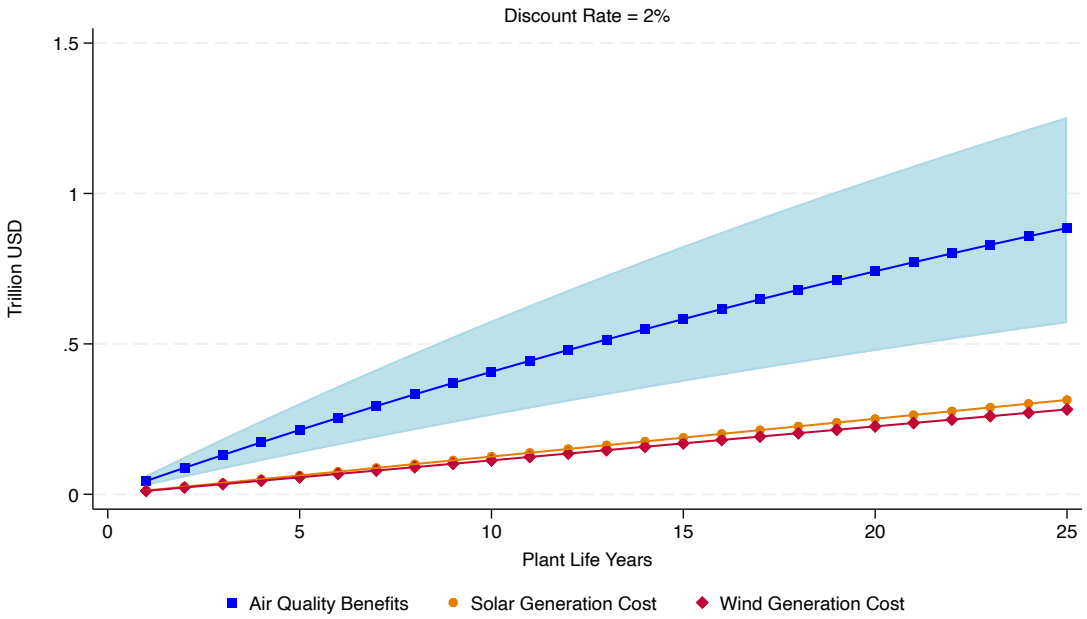
Notes: This table presents estimates using the specification in Equation (2.3) for operational coal-fired power plants. The sample used in each column is defined by distance band 0-40 km i.e., survey locations that are located within 40 km distance from the nearest coal power plant. Table 2.11 provide the list of countries from which sample surveys are used in this specification. 95% confidence interval bounds are reported in square brackets. Column 1 controls for admin-0 fixed effects while Column 2 controls for admin-1 fixed effects. The dependent variable, *Life Sat*, is a shorthand for life satisfaction, which takes values between 0 ("the worst possible life") and 10 ("the best possible life") based on what surveyed individuals report as their current life satisfaction. The main variables of interest are logarithm of air quality dissatisfaction and logarithm of annual household income. The first variable takes value 2 (1) if an individual is dissatisfied (satisfied) with ambient air quality and the second variable is logarithm of household reported total annual income in 1000 USD. Please refer to Table 2.1 notes for details on other variables.

Table 2.8: Aggregate equivalent variation results

(1) Estimate Type	(2) ψ	(3) ϕ	(4) y (in \$)	(5) $AirDiss/\widetilde{AirDiss}$	(6) e (in \$)	(7) Affected Population	(8) HH Size (# persons)	(9) AEV (in tril. \$)
Point estimate	-0.469	0.474	14855	1.37	3948	1,120,626,356	4.9	0.903
Lower bound	-0.326	0.543	14855	1.37	2539	1,120,626,356	4.9	0.581
Upper bound	-0.611	0.404	14855	1.37	5591	1,120,626,356	4.9	1.279

Notes: The three rows correspond to point estimates and lower and upper bounds of 95% confidence intervals of ψ and ϕ parameters respectively. Estimates on logarithm of annual household income, ϕ, logarithm of air quality dissatisfaction, ψ, and average income, y, are taken from Table 2.7. $\frac{AirDiss}{\widetilde{AirDiss}}$ is the ratio of air quality dissatisfaction level in the 0-40 km distance band and that outside of the band. e is the equivalent variation computed using Equation (2.4). The population data comes from the Gridded Population of the World, v4 (GPWv4) database for year 2020. AEV is generated by multiplying e with the population estimate downscaled by the number of persons living in a typical household, which is taken from the Area Database v4.1 of the Global Data Lab.

Figure 2.2: Aggregate air quality benefits and costs of closing operational plants



Notes: Chart shows the cost-benefit analysis results for all 51 countries combined as listed in Table 2.11. The policy experiment entails phasing out coal-fired power at a constant rate of 4% per year and replacing that freed capacity with solar or wind generation over a period of 25 years. The blue line represents point estimates of air quality benefits with the shaded area showing upper and lower bounds on the estimates. The costs of solar and wind energy generation are calculated by multiplying their respective source-specific average global LCOE values in USD/kWh with the total excess energy demand because of closing of coal plants. All the costs and benefits are expressed in present-discounted value terms with the annual discount rate set at 2% per year.

Table 2.9: In-sample top-25 coal power stations based on affected population

(1) Country	(2) State/Province	(3) Name of Plant	(4) Population	(5) Ann. Emission (in mil. tons)	(6) Capacity (in MW)	(7) Plant Life (in years)	(8) Solar Cost (in mil. $)	(9) Wind Cost (in mil. $)	(10) Gross Benefits (in mil. $)	(11) Gross Benefits LB (in mil. $)
India	Delhi	Rajghat Delhi	30871582	0.8	135	2	78.64	70.72	24874.62	15998.17
China	Shanghai	Wujing	29394098	2.8	600	14	349.52	314.31	23684.15	15232.51
China	Shanghai	Shanghai Gaoqiao	26608464	0.9	150	9	87.38	78.58	21439.64	13788.95
China	Shanghai	Baoshan Works	24979817	5.9	1050	11	611.67	550.04	20127.36	12944.96
India	West Bengal	Budge Budge	23684622	4.1	750	23	436.91	392.89	19083.77	12273.77
India	West Bengal	Southern CESC	23486539	0.8	135	11	78.64	70.72	18924.16	12171.12
India	West Bengal	Titagarh	23426320	1.2	240	5	139.81	125.72	18875.64	12139.91
India	Haryana	Faridabad	22755274	0.9	165	2	96.12	86.43	18334.95	11792.16
China	Guangdong	Guangzhou Refinery	22396021	1.0	200	28	116.51	104.77	18045.48	11605.99
Indonesia	West Java	Cikarang Babelan	21297338	1.4	280	38	163.11	146.68	17160.23	11036.64
India	Maharashtra	Trombay	21296044	4.0	810	16	471.86	424.32	17159.18	11035.97
China	Guangdong	Guangzhou Lixin	20995940	2.8	660	33	384.48	345.74	16917.37	10880.45
China	Guangdong	Mawan	20927798	9.4	1940	19	1130.13	1016.27	16862.47	10845.13
China	Guangdong	Lee & Man Paper	20536522	1.3	216	28	125.83	113.15	16547.20	10642.37
India	Uttar Pradesh	National Capital Dadri	19695645	9.0	1820	14	1060.22	953.40	15869.67	10206.61
Thailand	Bangkok	Bangkok HSFC Plant	18440092	0.2	36	29	20.97	18.86	14858.01	9555.96
Russia	Moscow Oblast	Moscow CHP-22	15602338	6.0	1160	7	675.75	607.66	12571.51	8085.39
Vietnam	Dong Nai	Nhon Trach Formosa	13878848	2.2	450	32	262.14	235.73	11182.81	7192.25
Indonesia	Banten	Banten Lontar	13412602	4.3	945	35	550.50	495.04	10807.14	6950.63
Pakistan	Sindh	Port Qasim EPC	12929133	5.2	1320	39	768.95	691.48	10417.58	6700.09
China	Guangdong	Sanshui Hengyi	12899233	5.0	1200	32	699.05	628.62	10393.49	6684.60
China	Guangdong	Dongguan Jianhui	12595530	0.3	50	29	29.13	26.19	10148.79	6527.21
China	Hebei	Sanhe Yanjiao	12573655	6.4	1300	30	757.30	681.00	10131.16	6515.88
China	Tianjin	Junliangcheng	12239822	4.6	1050	13	611.67	550.04	9862.18	6342.88
China	Tianjin	Tianjin Northeast	12096624	3.0	660	36	384.48	345.74	9746.79	6268.67

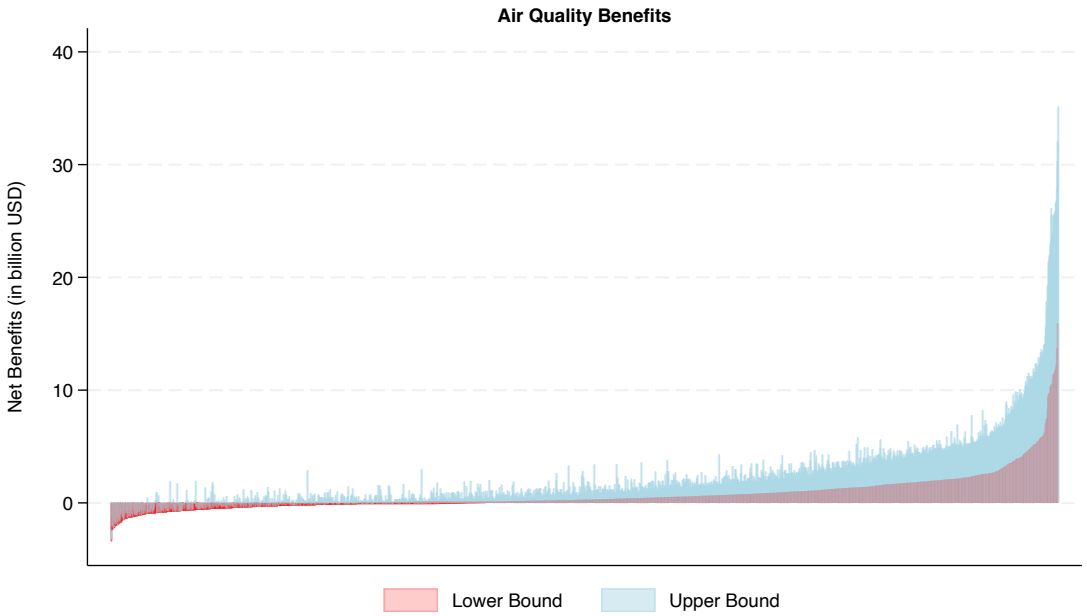
Notes: The table lists 25 coal power stations in our sample in decreasing order of total population affected, which is reported in Column 4. The population figures are the total number of individuals located within 40 km of respective plants. Solar and wind costs report the cost of green transition through solar and wind technology respectively. These costs are calculated by using source-specific average global LCOE values and respective capacities of coal plants as reported in Column 6. Air quality benefits in Column 10 are computed by multiplying EV values, which are computed by using Equation (2.4), with the total number of residences in the 0-40 km distance band. For EV calculations, global parameter values for ψ, ϕ, $\frac{AirDiss}{\widetilde{AirDiss}}$, and y are used. Column 11 reports the lower bound on the gross air quality benefits from shutting down each of the listed plants.

Table 2.10: Out-of-sample top-25 coal power stations based on affected population

(1) Country	(2) State/Province	(3) Name of Plant	(4) Population	(5) Ann. Emission (in mil. tons)	(6) Capacity (in MW)	(7) Plant Life (in years)	(8) Solar Cost (in mil. $)	(9) Wind Cost (in mil. $)	(10) Gross Benefits (in mil. $)	(11) Gross Benefits LB (in mil. $)
Japan	Kanto	Isogo	19357188	5.0	1200	27	699.05	628.62	15596.96	10031.22
Hong Kong	Hong Kong	Castle Peak	18885140	20.4	4110	7	2394.24	2153.02	15216.61	9786.59
Japan	Kansai	Kobe	11970257	5.8	1400	24	815.56	733.39	9644.97	6203.19
Japan	Kansai	Nadahama Works	11295782	0.4	67	23	39.03	35.10	9101.52	5853.66
China	Tianjin	Dagang Oilfield	10829386	10.4	2000	10	1165.08	1047.70	8725.72	5611.97
Taiwan	Taipei	Shu-Lin	10181423	0.3	52	15	30.29	27.24	8203.63	5276.18
Taiwan	Taipei	Linkou Plant TP	10168361	0.2	36	11	20.97	18.86	8193.11	5269.41
Taiwan	Taoyuan	Jinshin	10098233	0.6	114	21	66.41	59.72	8136.60	5233.07
Taiwan	Taoyuan	Hwa Ya Cogen	10096528	1.6	300	25	174.76	157.15	8135.23	5232.19
Taiwan	Taipei	Linkou Power	9582077	9.0	2400	38	1398.10	1257.24	7720.71	4965.59
Japan	Chubu	Tokai Kyodo	7561710	0.9	149	11	86.80	78.05	6092.81	3918.60
Japan	Chubu	Meinan Kyodo Energy	7162755	0.1	31	39	18.06	16.24	5771.35	3711.86
Japan	Chubu	Nagoya	6499765	1.3	259	30	150.88	135.68	5237.15	3368.29
Germany	North Rhine-Westphalia	Krefeld-Uerdingen	5408234	0.7	120	7	69.90	62.86	4357.66	2802.64
Germany	North Rhine-Westphalia	Herne	5312472	2.3	500	10	291.27	261.92	4280.50	2753.01
United Kingdom	England	Fiddler's Ferry	5294138	10.4	2132	2	1241.98	1116.84	4265.73	2743.51
Germany	North Rhine-Westphalia	Cologne-Merkenich	5090692	0.5	85	31	49.52	44.53	4101.80	2638.08
Germany	North Rhine-Westphalia	Chempark Leverkusen	5088491	0.6	112	7	65.24	58.67	4100.03	2636.94
Germany	North Rhine-Westphalia	Scholven	4996383	3.8	740	4	431.08	387.65	4025.81	2589.21
Germany	North Rhine-Westphalia	Buer	4975825	0.4	76	6	44.27	39.81	4009.25	2578.56
Japan	Chubu	Hekinan	4964135	18.0	4100	17	2388.41	2147.78	3999.83	2572.50
Japan	Chubu	MC Shiohama Energy	4962023	0.2	34	29	19.81	17.81	3998.12	2571.40
Germany	North Rhine-Westphalia	Neurath	4879265	18.6	4112	14	2395.40	2154.06	3931.44	2528.52
Germany	North Rhine-Westphalia	Duisburg-Walsum	4848081	2.9	790	34	460.21	413.84	3906.32	2512.36
Germany	North Rhine-Westphalia	Niederaussem	4775255	14.7	2933	10	1708.59	1536.45	3847.64	2474.62

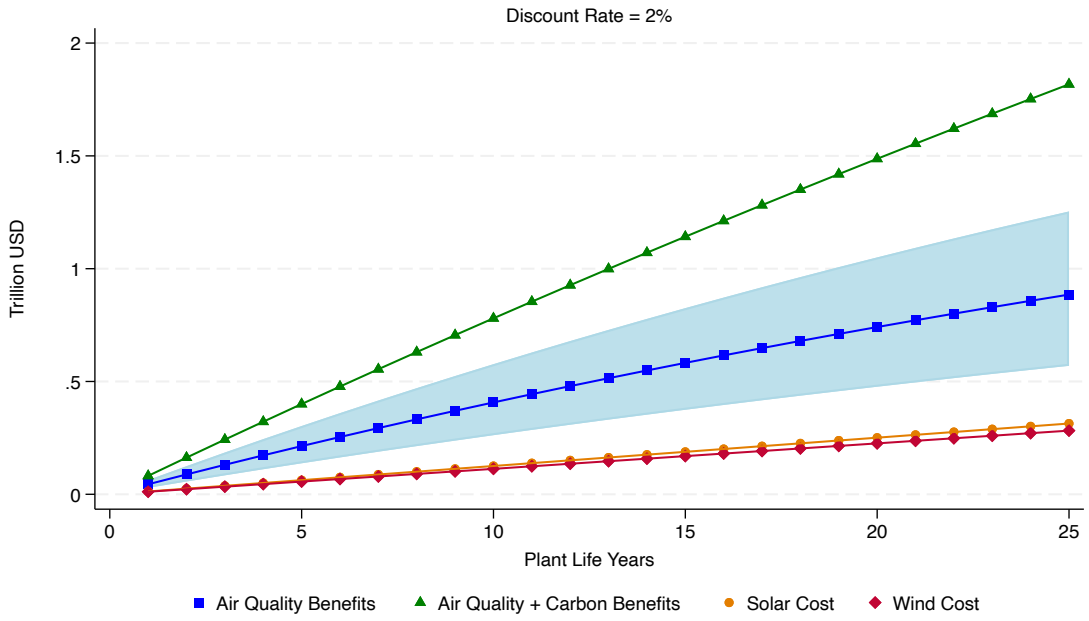
Notes: The table lists 25 coal power stations in the rest of the world i.e., countries outside our 51 country sample in the decreasing order of total population affected, which is reported in Column 4. The population figures are total number of individuals located within 40 km of respective plants. Solar and wind costs report the cost of green transition through solar and wind technology respectively. These costs are calculated by using source-specific average global LCOE values and respective capacity of coal plants as reported in Column 6. Air quality benefits in Column 10 are computed by multiplying EV values, which are computed by using Equation (2.4), with the total number of residences in the 0-40 km distance band. For EV calculation, global parameter values for ψ, ϕ, $\frac{\widehat{AirDiss}}{AirDiss}$, and y are used. Column 11 reports the lower bound on the gross air quality benefits from shutting down each of the listed plants.

Figure 2.3: Plant-level net air quality benefits from closing operational plants



Notes: Chart shows the net benefits from closing all the operational coal-fired power in 2019 located across the whole world. The parameter values for ψ, ϕ, $\frac{AirDiss}{AirDiss}$, and y are taken from the global estimates using all 51 countries combined. The policy experiment entails phasing out coal-fired power and replacing that freed capacity with 50% solar and 50% wind generation. The costs of solar and wind energy generation are calculated by multiplying respective source-specific global average LCOE values in USD/kWh with the total energy demand.

Figure 2.4: Aggregate benefits and costs of closing operational plants



Notes: Chart shows the cost-benefit analysis results after accounting for carbon-reduction benefits. The green line shows the lower bound of carbon benefits added to the air quality benefits. Please refer to Figure 2.2 notes for more details.

Appendix

2.6 Appendix: Instrumental Variables Strategy

We discuss how an IV approach may address the concerns about the selection of power-plant locations and/or migration patterns of citizens based on air quality preferences. We propose two instruments for coal-fired power station locations based on the need to supply such power stations with coal. They are (i) the logarithm of distance of survey locations from the nearest railroad and (ii) the logarithm of distance of survey locations from the nearest body of water, such as a lake, river, or sea. The first instrument picks up an important transportation linkage since the majority of coal worldwide is transported using railways. A small but significant fraction of coal transportation uses coal barges and other sea vessels (National Research Council 2007). This is picked up in our second instrument. Proximity to water may also increase the reliability of water supply and eases waste treatment. We show below that these variables are strongly predictive of coal-fired power station locations.

To construct these instruments, we use global geo-referenced data on railways and locations of water-bodies. The source of the railways network shapefile is the World Food Program-Logistics Cluster[47], which brings together various sources such as Open-StreetMap, American Digital Cartography, Global Discovery, etc. To get the location of water-bodies, we combine data from multiple sources[48] to create an "amalgam" water-

47. This program works to ensure effective and efficient humanitarian response by optimizing logistics during times of disasters and other emergencies. It also acts as a provider of last resort for shared logistics services across the world.

48. Three data layers: (i) linear water showing lines of rivers, streams, and canals from ESRI, (ii) a shape-

bodies shapefile.

We also need a plausible exclusion restriction, i.e., that these two instrumental variables predict perceptions of pollution, conditional on covariates, only through the first-stage channel. Given that we have two instruments, we can use a formal test of over-identification. However, beyond this formal approach, we believe that it is plausible *a priori* to think that the exclusion restriction holds as there is no obvious reason to expect proximity to railroads or water-bodies to affect air quality perceptions. Railways that run on diesel are much less polluting than coal-fired power, and nearly 30% of the global railways network has now been electrified. So, it is highly unlikely that there is a direct effect of railway locations on air quality.[49]

More formally, we write the selection equation for δ as follows:

$$\delta_{ic} = \lambda \tau_i + \gamma_i z_\ell + \nu_{ic} \tag{2.5}$$

where z are factors, which affect distance other than taste for pollution, i.e., "instruments" for δ_{ic}. We allow γ, the relationship between z_ℓ and δ_{ic}, to be heterogeneous. We cannot estimate this relationship in practice because we only observe an individual once.

Now consider an IV estimator of α where we put in $\widehat{\delta}_{ic}$, as in the first-stage prediction of δ, under the 2SLS routine. Then, using Equations (2.1) and (2.5)

$$\widehat{\alpha}_{IV} = \frac{cov\left(z_\ell, AirDiss_{i\ell}\right)}{cov\left(z_\ell, \delta_{ic}\right)} = \frac{cov\left(z_\ell, \alpha\left[\lambda \tau_i + \gamma_i z_\ell + \nu_{ic}\right] + \tau_i + \varepsilon_{i\ell}\right)}{cov\left(z_\ell, \lambda \tau_i + \gamma_i z_\ell + \nu_{ic}\right)} = \alpha \tag{2.6}$$

as long as $cov(\tau_i, z_\ell) = 0$. Then the difference between OLS and IV is

$$\widehat{\alpha}_{OLS} - \widehat{\alpha}_{IV} = \frac{cov\left(\tau_i, \delta_{ic}\right)}{var\left(\delta_{ic}\right)} \tag{2.7}$$

file for major rivers from UNESCO World-wide Hydrogeological Mapping and Assessment Program, and (iii) an ocean coastline shapefile from the North American Cartographic Information Society are merged using the spatial join tool in ArcGIS software.

49. Railways emit less than 1% of all transport NO_2 emissions and less than 0.5% of transport PM_{10} emissions. *Source:* European Environment Agency

Given $\alpha < 0$, a larger magnitude IV coefficient (relative to OLS) is plausible if $cov(\tau_i, \delta_{ic}) > 0$, i.e., those with more distaste for air pollution are less likely to locate to areas with high pollution, the selection issue at hand.

Having explained how an IV strategy could remove the OLS bias towards finding null effects, we estimate the following specification for households located in distance band 0-40 km from an operational coal-fired power plant:

$$AirDiss_{i\ell} = \alpha_{IV}\widehat{\delta}_{ic} + \beta \mathbf{X}_{i\ell} + \eta_\ell + \varepsilon_{i\ell} \tag{2.8}$$

where $\mathbf{X}$ contains geocode (latitude$\times$longitude)-level and individual-level controls and $\widehat{\delta}_{i\ell}$ is predicted from the first-stage using the vector of instruments, Ω:

$$\delta_{ic} = \theta \Omega_{i\ell} + \xi \mathbf{X}_{i\ell} + \zeta_\ell + \nu_{i\ell}. \tag{2.9}$$

In this case, we expect α_{IV} to be negative and larger in magnitude compared to α.

The results are reported in Table 2.28. Columns 1 and 2 use country fixed effects and Columns 3 and 4 use state fixed effects. Columns 1 and 3 employ only the survey location's logarithm of distance from nearest railroad as an instrument, while Columns 2 and 4 use both nearest railroad and body of water distances as instruments. As hypothesized, α_{IV} is negative in all four specifications and has a magnitude nearly eight times that of α, which is reported in Table 2.1 and obtained by estimating Equation (2.2) using OLS.

Large values of first-stage Kleibergen-Paap F-statistics and Kleibergen-Paap LM statistics suggest that these are strong instruments. Moreover, for over-identified cases with two instruments, the over-identifying restrictions are valid as evidenced from low Hansen J-test statistics.[50] As a robustness test, we do the same IV estimation for retired plants. First-stage and reduced-form results are reported in Table 2.30 in the Appendix. As expected, the first-stage results are significant i.e., railroads and water-bodies predict retired coal plants locations, but reduced-form results are insignificant, meaning that distance

50. The first-stage and reduced-form results are presented in Table 2.29 in the Appendix.

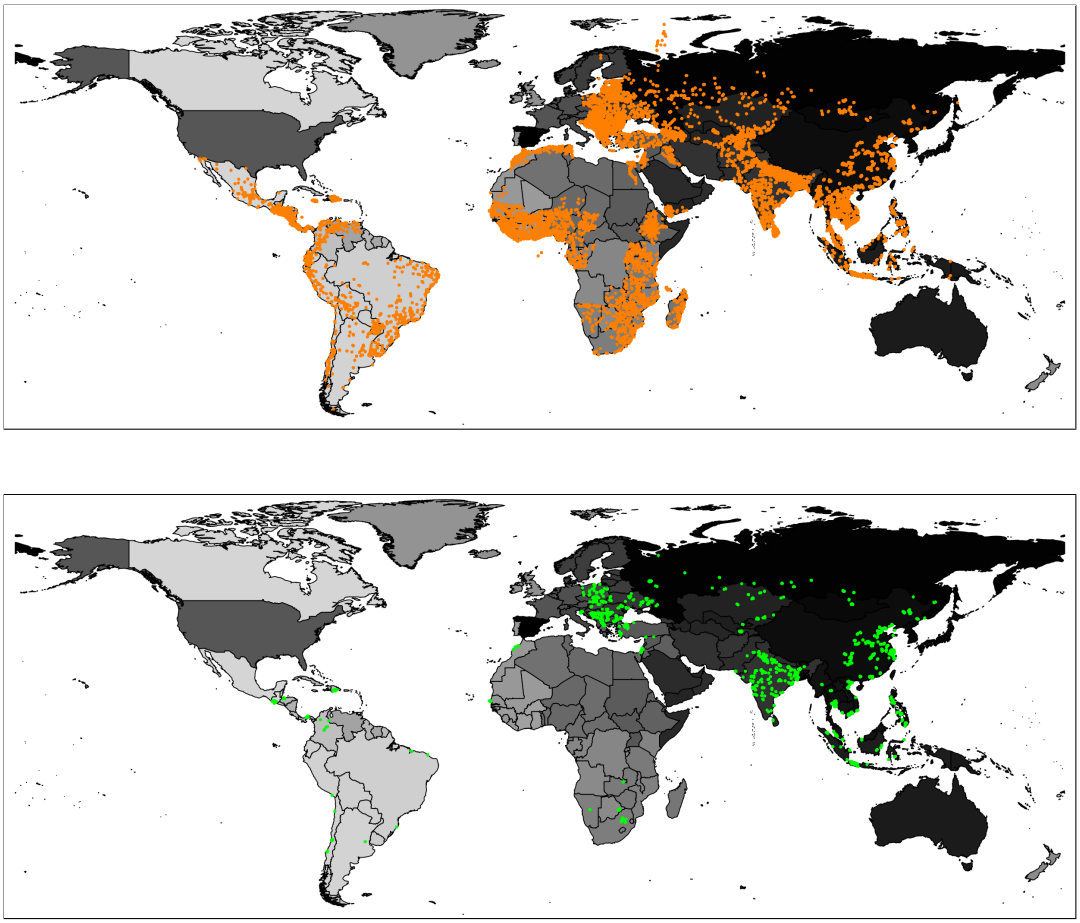
from railroads and water-bodies does not impact air quality perceptions.

These findings give credence to a causal interpretation of a link between air quality perception and proximity to coal-fired power plants. The difference in magnitude between OLS and IV estimates also highlights the potential importance of selection-bias if citizens who value air quality choose to locate further away from coal plants even though these areas are likely to be richer neighborhoods with higher overall life satisfaction.[51] This is plausible since, once a government sets up a coal plant in an area, it could bring other socio-economic and cultural activities into the area.

51. Please see Figures 2.18 and 2.19 in the Appendix.

2.7 Appendix: Figures and Tables

Figure 2.5: 2019 Gallup World Poll survey geocodes



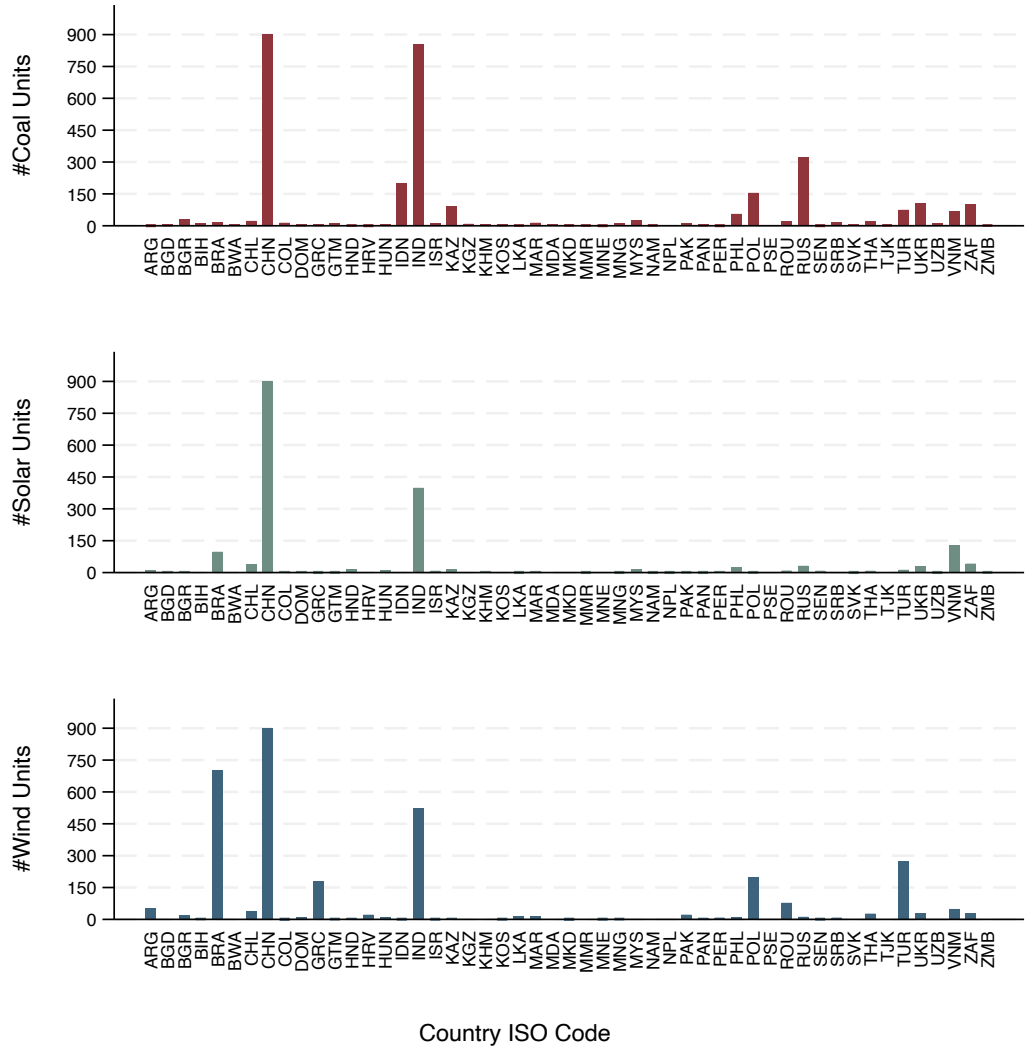
Notes: Top map shows all the surveys (in orange dots) where precise GPS coordinates were recorded in the 2019 round of the Gallup World Poll, a total of 138,242 surveys spread across 140+ countries worldwide. Bottom map shows the subset of surveys (in green dots) that are located in the 0-40 km distance band from an operational coal-fired power plant and this subset has been used in the main analysis, a total of 17,964 surveys, covering 51 countries listed in Table 2.11.

Table 2.11: List of countries in the main analysis

No.	ISO	Country
1	ARG	Argentina
2	BGD	Bangladesh
3	BIH	Bosnia and Herzegovina
4	BWA	Botswana
5	BRA	Brazil
6	BGR	Bulgaria
7	KHM	Cambodia
8	CHL	Chile
9	CHN	China
10	COL	Colombia
11	HRV	Croatia
12	DOM	Dominican Republic
13	GRC	Greece
14	GTM	Guatemala
15	HND	Honduras
16	HUN	Hungary
17	IND	India
18	IDN	Indonesia
19	ISR	Israel
20	KAZ	Kazakhstan
21	KOS	Kosovo
22	KGZ	Kyrgyzstan
23	MYS	Malaysia
24	MDA	Moldova
25	MNG	Mongolia
26	MNE	Montenegro
27	MAR	Morocco
28	MMR	Myanmar
29	NAM	Namibia
30	NPL	Nepal
31	MKD	North Macedonia
32	PAK	Pakistan
33	PSE	Palestine
34	PAN	Panama
35	PER	Peru
36	PHL	Philippines
37	POL	Poland
38	ROU	Romania
39	RUS	Russia
40	SEN	Senegal
41	SRB	Serbia
42	SVK	Slovakia
43	ZAF	South Africa
44	LKA	Sri Lanka
45	TJK	Tajikistan
46	THA	Thailand
47	TUR	Turkey
48	UKR	Ukraine
49	UZB	Uzbekistan
50	VNM	Vietnam
51	ZMB	Zambia

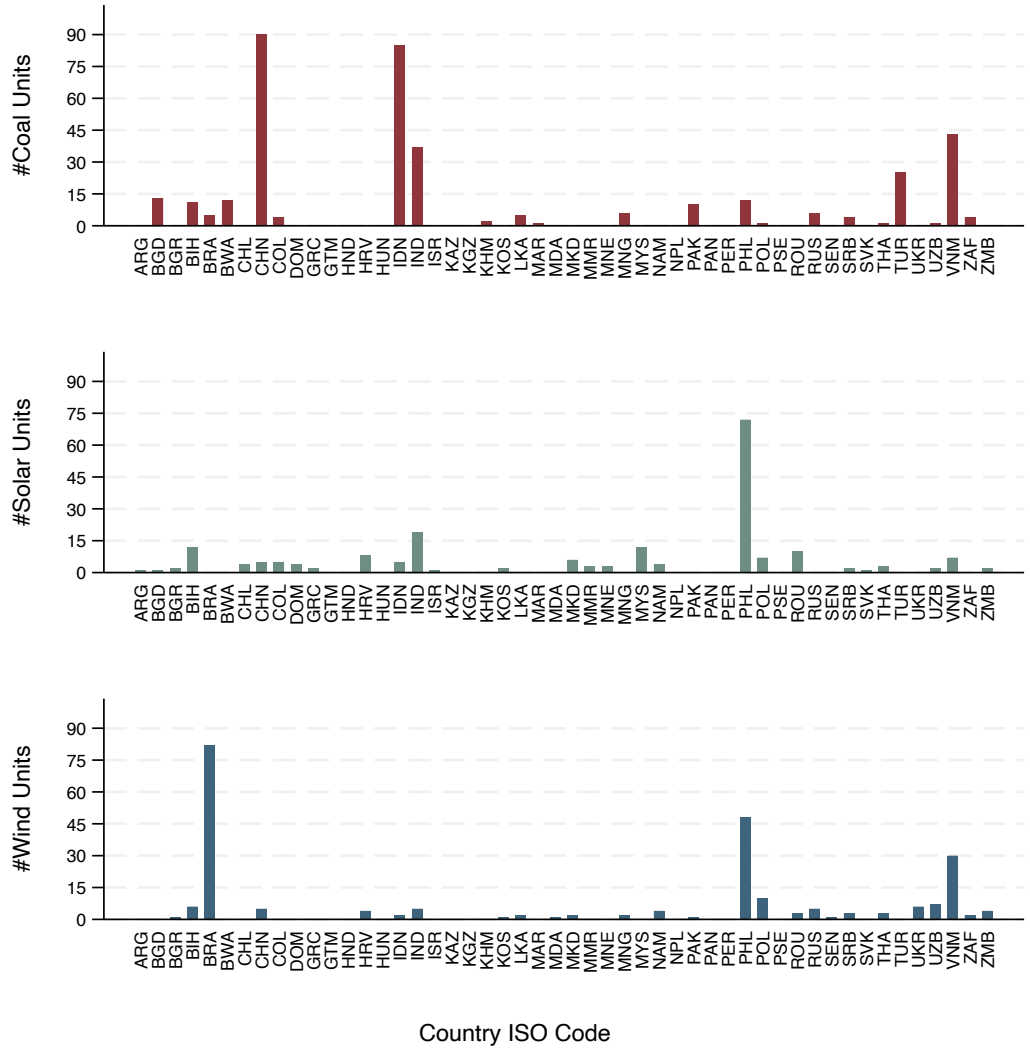
Notes: These countries contain the sample of surveys that are used in the main analysis. Some of the survey locations within these countries qualify under the distance band 0-40 km i.e., survey locations that are located within 40 km of the nearest operational coal-fired power plants. Bottom panel of Figure 2.5 maps the geocodes of these survey locations.

Figure 2.6: Distribution of operational energy sources in sample countries



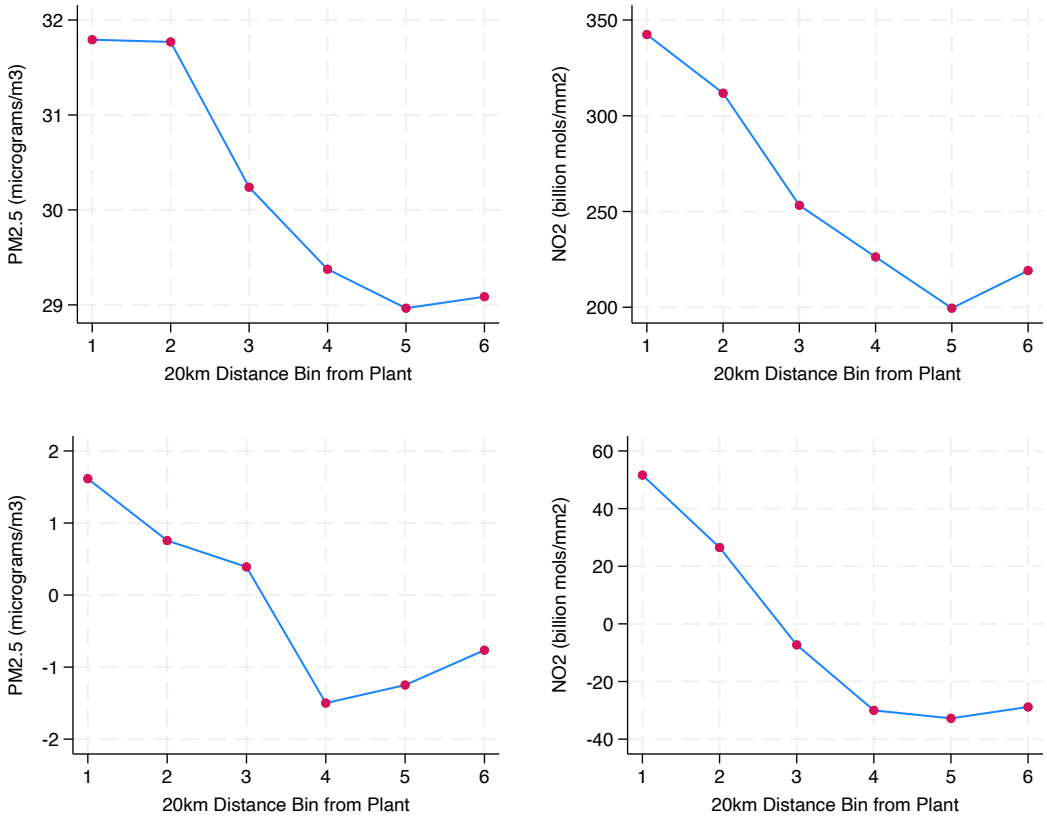
Notes: The graph shows the count of operational coal plants (top), solar farms (middle), and wind farms (bottom) for 51 countries in the main sample as listed in Table 2.11. The number of units have been capped at 900 for display purpose, thereby censoring all units counts for China (CHN). The actual count of operational coal, solar, and wind units for CHN are 2990, 3782, and 2663 respectively.

Figure 2.7: Distribution of planned energy sources in sample countries



Notes: The graph shows the count of planned coal plants (top), solar farms (middle), and wind farms (bottom) for 51 countries in the main sample as listed in Table 2.11. The planned category includes plants/farms which are in the "announced", "pre-permit", or "permitted" stage of commissioning. The number of units have been capped at 90 for display purpose, thereby censoring coal units count for China (CHN). The actual count of planned coal units for CHN is 292.

Figure 2.8: Air pollution level indicators around operational plants



Notes: The label on x-axis should be multiplied by 20 to get the distance bin of the survey location from the nearest coal plant. Top panel charts present raw means from the data using the pollutant concentration at each geocode in the respective distance bin and the bottom panel demeans all those observations of the country fixed effects.

Table 2.12: Conditional logit estimation results for operational plants

	(1) Air Diss	(2) Air Diss	(3) Air Diss	(4) Air Diss	(5) Air Diss	(6) Air Diss
Geocode's log dist from nearest plant	-0.230***	-0.349	-0.621	-0.225***	-0.137	-0.827
	(0.0493)	(0.2423)	(0.3918)	(0.0594)	(0.2426)	(0.5689)
Geocode's vegetation index	-0.493**	-0.514*	-0.531*	-0.349*	-0.635**	-0.955**
	(0.1677)	(0.2375)	(0.2636)	(0.1623)	(0.2320)	(0.3363)
Geocode area is urban	0.536***	0.765***	0.768***	0.473***	0.691***	0.724***
	(0.1008)	(0.1163)	(0.1885)	(0.1037)	(0.0935)	(0.1480)
Respondent's age is 26-60 years	0.097	0.093	0.170***	0.079	0.140*	0.199**
	(0.0548)	(0.0621)	(0.0493)	(0.0547)	(0.0582)	(0.0663)
Respondent's age is more than 60 years	-0.112	0.066	0.114	-0.119	0.105	0.174*
	(0.0806)	(0.0730)	(0.0769)	(0.0715)	(0.0769)	(0.0885)
Respondent's gender is male	-0.095*	-0.118*	-0.096*	-0.088*	-0.093*	-0.076
	(0.0439)	(0.0458)	(0.0405)	(0.0398)	(0.0439)	(0.0466)
Respondent's education is intermediate	0.312***	0.246**	0.231**	0.335***	0.236***	0.237***
	(0.0560)	(0.0874)	(0.0758)	(0.0575)	(0.0676)	(0.0644)
Respondent's education is high	0.453***	0.374***	0.342**	0.484***	0.361***	0.391***
	(0.0724)	(0.0934)	(0.1210)	(0.0749)	(0.1009)	(0.0943)
Log annual hh income in '000 USD	-0.025	-0.013	-0.053	-0.018	-0.038	-0.063*
	(0.0283)	(0.0316)	(0.0305)	(0.0273)	(0.0267)	(0.0295)
Respondent has children under 15 yrs	0.017	-0.001	0.059	0.007	0.009	0.050
	(0.0396)	(0.0546)	(0.0685)	(0.0430)	(0.0498)	(0.0593)
Number of observations	17,964	16,452	13,108	17,729	16,033	12,567
Pseudo R-squared	0.028	0.027	0.024	0.018	0.017	0.020
Log likelihood	-9,994	-8,310	-6,353	-8,969	-7,206	-5,527
Region fixed effects	Admin-0	Admin-0	Admin-0	Admin-1	Admin-1	Admin-1
Distance band	0-40 km	40-80 km	80-120 km	0-40 km	40-80 km	80-120 km

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table above reports results for conditional logistical model estimation with fixed effects corresponding to OLS estimation results reported in Table 2.1. We implement a robust estimation for fixed effects conditional logit models using the estimator proposed by Baetschmann et al. (2020).

Table 2.13: Results with spatial clustering for operational plants

	(1) Air Diss	(2) Air Diss	(3) Air Diss	(4) Air Diss	(5) Air Diss	(6) Air Diss
Geocode's log dist from nearest plant	-0.044***	-0.056	-0.094	-0.039***	-0.020	-0.111
	(0.0095)	(0.0325)	(0.0621)	(0.0090)	(0.0317)	(0.0648)
Geocode's vegetation index	-0.097**	-0.097**	-0.084*	-0.063*	-0.104**	-0.139**
	(0.0327)	(0.0373)	(0.0402)	(0.0287)	(0.0342)	(0.0491)
Geocode area is urban	0.106***	0.144***	0.142***	0.089***	0.120***	0.125***
	(0.0152)	(0.0157)	(0.0200)	(0.0166)	(0.0147)	(0.0194)
Respondent's age is 26-60 years	0.020*	0.016	0.027**	0.015	0.022**	0.030**
	(0.0091)	(0.0088)	(0.0095)	(0.0090)	(0.0082)	(0.0094)
Respondent's age is more than 60 years	-0.022	0.011	0.018	-0.020	0.017	0.027*
	(0.0122)	(0.0122)	(0.0130)	(0.0117)	(0.0119)	(0.0126)
Respondent's gender is male	-0.018*	-0.020**	-0.016*	-0.015*	-0.015*	-0.012
	(0.0071)	(0.0073)	(0.0070)	(0.0069)	(0.0068)	(0.0070)
Respondent's education is intermediate	0.057***	0.039***	0.037***	0.059***	0.036***	0.035***
	(0.0092)	(0.0092)	(0.0098)	(0.0089)	(0.0087)	(0.0091)
Respondent's education is high	0.089***	0.066***	0.059***	0.089***	0.059***	0.062***
	(0.0132)	(0.0155)	(0.0150)	(0.0129)	(0.0139)	(0.0135)
Log annual hh income in '000 USD	-0.006	-0.003	-0.009	-0.004	-0.006	-0.010*
	(0.0047)	(0.0046)	(0.0049)	(0.0045)	(0.0043)	(0.0048)
Respondent has children under 15 yrs	0.004	0.000	0.010	0.001	0.001	0.008
	(0.0077)	(0.0083)	(0.0087)	(0.0074)	(0.0079)	(0.0088)
Number of observations	17,964	16,461	13,137	17,964	16,461	13,137
Adj R-squared	0.032	0.030	0.025	0.018	0.016	0.018
Region fixed effects	Admin-0	Admin-0	Admin-0	Admin-1	Admin-1	Admin-1
Distance band	0-40 km	40-80 km	80-120 km	0-40 km	40-80 km	80-120 km

Heteroskedasticity- and Autocorrelation-Consistent standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2) for operational coal-fired power plants. The sample used in each column is defined by the distance band i.e., how far the survey location is relative to the nearest coal power plant. Table 2.11 provides the list of countries that are used in the main specification i.e., 0-40 km distance band and results are reported in Columns 1 and 4. Standard errors, which are reported in parentheses, are clustered spatially using the distance threshold of 5 km, following Conley (1999) and Conley (2008). Columns 1-3 and Columns 4-6 control for admin-0 and admin-1 fixed effects respectively. The dependent variable, *Air Diss*, is a shorthand for Air Quality Dissatisfaction, which takes value 1 (0) if the surveyed individual is dissatisfied (satisfied) with the ambient air quality. The main variable of interest is geocode's logarithm of distance from the nearest plant, which is the straight-line distance between the survey and nearest coal plant location. Please refer to Table 2.1 notes for details on other variables.

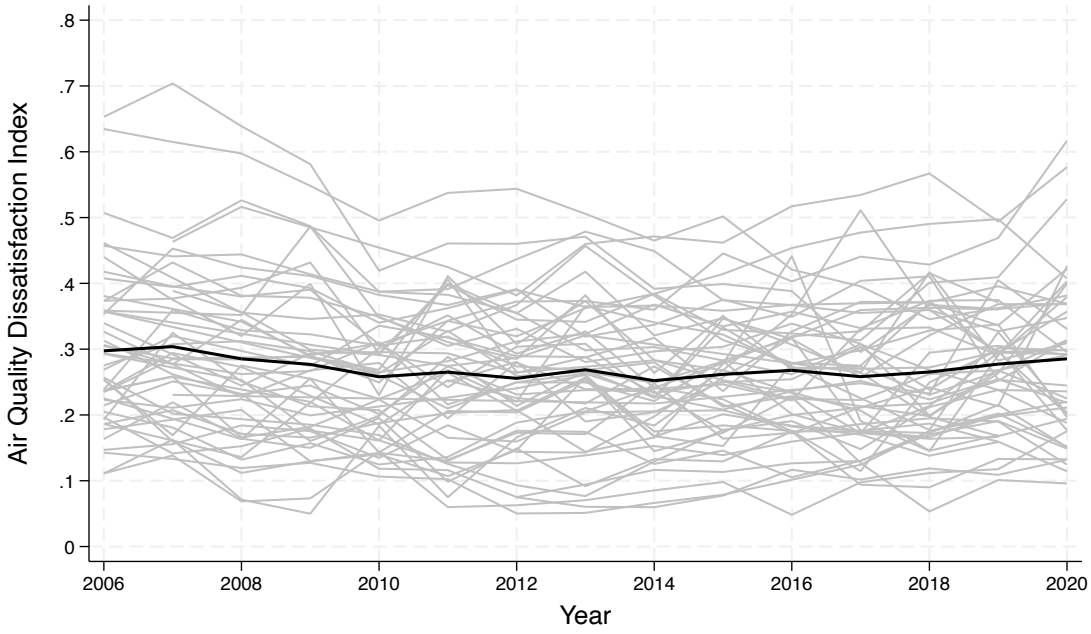
Table 2.14: Results with CO_2 interaction for operational plants

	(1) Air Diss	(2) Air Diss	(3) Air Diss	(4) Air Diss
Geocode's log dist from nearest plant	-0.042**	-0.046***	-0.036*	-0.039**
	(0.0128)	(0.0136)	(0.0143)	(0.0148)
Annual CO2 emission	0.005	-0.008		
	(0.0102)	(0.0087)		
Geocode's log dist from nearest plant × Annual CO2 emission	-0.001	0.003		
	(0.0030)	(0.0027)		
High CO2 emission			0.070	0.021
			(0.0745)	(0.0676)
High CO2 emission × Geocode's log dist from nearest plant			-0.017	0.001
			(0.0234)	(0.0221)
Geocode's vegetation index	-0.097**	-0.064*	-0.097**	-0.063*
	(0.0330)	(0.0300)	(0.0324)	(0.0299)
Geocode area is urban	0.107***	0.088***	0.107***	0.089***
	(0.0219)	(0.0205)	(0.0216)	(0.0204)
Respondent's age is 26-60 years	0.020	0.015	0.019	0.015
	(0.0103)	(0.0099)	(0.0103)	(0.0098)
Respondent's age is more than 60 years	-0.021	-0.021	-0.021	-0.020
	(0.0149)	(0.0128)	(0.0149)	(0.0127)
Respondent's gender is male	-0.018	-0.015*	-0.018*	-0.016*
	(0.0090)	(0.0073)	(0.0091)	(0.0073)
Respondent's education is intermediate	0.057***	0.058***	0.057***	0.058***
	(0.0102)	(0.0100)	(0.0102)	(0.0100)
Respondent's education is high	0.089***	0.089***	0.090***	0.089***
	(0.0152)	(0.0142)	(0.0149)	(0.0142)
Log annual hh income in '000 USD	-0.006	-0.004	-0.006	-0.004
	(0.0054)	(0.0050)	(0.0054)	(0.0050)
Respondent has children under 15 yrs	0.004	0.001	0.004	0.001
	(0.0076)	(0.0077)	(0.0077)	(0.0077)
Number of observations	17,964	17,964	17,964	17,964
Adj R-squared	0.128	0.179	0.128	0.179
Mean of dependent variable	0.327	0.327	0.327	0.327
Region fixed effects	Admin-0	Admin-1	Admin-0	Admin-1

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2) for operational coal-fired power plants but interacting δ with either a discrete or continuous measure of annual CO_2 emission from all the units of the nearest coal power plant. The sample used in each column is defined by the distance band 0-40 km i.e., all survey locations that are located within 40 km of an operational coal power plant. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for Columns 1 and 3 and state/province/admin-1 level for remaining columns. Columns 1 and 3 control for admin-0 fixed effects and remaining columns control for admin-1 fixed effects. The dependent variable, *Air Diss*, is a shorthand for Air Quality Dissatisfaction, which takes value 1 (0) if the surveyed individual is dissatisfied (satisfied) with the ambient air quality. Geocode's logarithm of distance from the nearest plant is a measure of straight-line distance between the survey location and nearest coal plant location. Annual CO_2 emission is measured in million tonnes per year and high (low) CO_2 emission correspond to above (below) median plant-level emissions. Please refer to Table 2.1 notes for details on other variables.

Figure 2.9: Air quality dissatisfaction trends across sample countries



Notes: Each gray line represents one country from the list of countries in Table 2.11. Each point on the line is generated by taking the average of all individuals in a country-year. The black line represents the average across all the 51 countries for each year.

Table 2.15: Risk assessments for 40-80 km and 80-120 km distance bands

	(1) Poll Risk	(2) Poll Risk	(3) Poll Risk	(4) Poll Risk	(5) Clim Risk	(6) Clim Risk	(7) Clim Risk	(8) Clim Risk
Geocode's log dist from nearest plant	0.002	-0.006	0.006	-0.021	0.022	-0.042	0.007	-0.022
	(0.0039)	(0.0102)	(0.0054)	(0.0112)	(0.0171)	(0.0288)	(0.0184)	(0.0358)
Geocode's vegetation index	0.001	-0.006	0.007	0.007	-0.045*	-0.000	-0.026	0.003
	(0.0048)	(0.0054)	(0.0067)	(0.0047)	(0.0224)	(0.0181)	(0.0270)	(0.0232)
Geocode area is urban	0.000	-0.005	0.002	-0.004	-0.020*	-0.011	-0.018*	-0.010
	(0.0031)	(0.0039)	(0.0035)	(0.0030)	(0.0074)	(0.0099)	(0.0075)	(0.0103)
Respondent's age is 26-60 years	-0.001	0.001	-0.001	0.001	0.008	0.008	0.008	0.007
	(0.0022)	(0.0028)	(0.0024)	(0.0026)	(0.0059)	(0.0059)	(0.0057)	(0.0059)
Respondent's age is more than 60 years	-0.003	-0.004	-0.002	-0.003	0.013	0.010	0.015	0.012
	(0.0029)	(0.0031)	(0.0028)	(0.0026)	(0.0104)	(0.0087)	(0.0082)	(0.0074)
Respondent's gender is male	-0.000	0.001	-0.000	0.001	-0.011*	0.001	-0.012**	0.002
	(0.0019)	(0.0017)	(0.0018)	(0.0018)	(0.0046)	(0.0041)	(0.0044)	(0.0044)
Respondent's education is intermediate	0.002	0.003	0.002	0.003	0.007	-0.003	0.009	-0.002
	(0.0023)	(0.0015)	(0.0023)	(0.0019)	(0.0076)	(0.0052)	(0.0067)	(0.0056)
Respondent's education is high	0.003	0.003	0.004	0.003	0.010	-0.004	0.011	-0.003
	(0.0031)	(0.0038)	(0.0035)	(0.0035)	(0.0116)	(0.0083)	(0.0094)	(0.0079)
Log annual hh income in '000 USD	0.001	0.002*	0.001	0.001	0.001	0.006	0.003	0.005
	(0.0012)	(0.0007)	(0.0012)	(0.0008)	(0.0031)	(0.0035)	(0.0028)	(0.0029)
Respondent has children under 15 yrs	-0.001	-0.000	-0.001	0.000	-0.004	-0.005	-0.005	-0.005
	(0.0021)	(0.0023)	(0.0019)	(0.0021)	(0.0049)	(0.0041)	(0.0052)	(0.0053)
Number of observations	14,128	11,307	14,128	11,307	14,128	11,307	14,128	11,307
Adj R-squared	0.008	0.009	0.014	0.026	0.033	0.034	0.062	0.062
Mean of dependent variable	0.011	0.009	0.011	0.009	0.061	0.050	0.061	0.050
Region fixed effects	Admin-0	Admin-0	Admin-1	Admin-1	Admin-0	Admin-0	Admin-1	Admin-1
Distance band	40-80 km	80-120 km	40-80 km	80-120 km	40-80 km	80-120 km	40-80 km	80-120 km

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2). The sample used in each column is defined by the distance band i.e., how far the survey location is relative to the nearest coal power plant. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for Columns 1-2 and 5-6 and state/province/admin-1 level for remaining columns. Columns 1-2 and 5-6 control for admin-0 fixed effects and remaining control for admin-1 fixed effects. Please refer to Table 2.3 notes for more details.

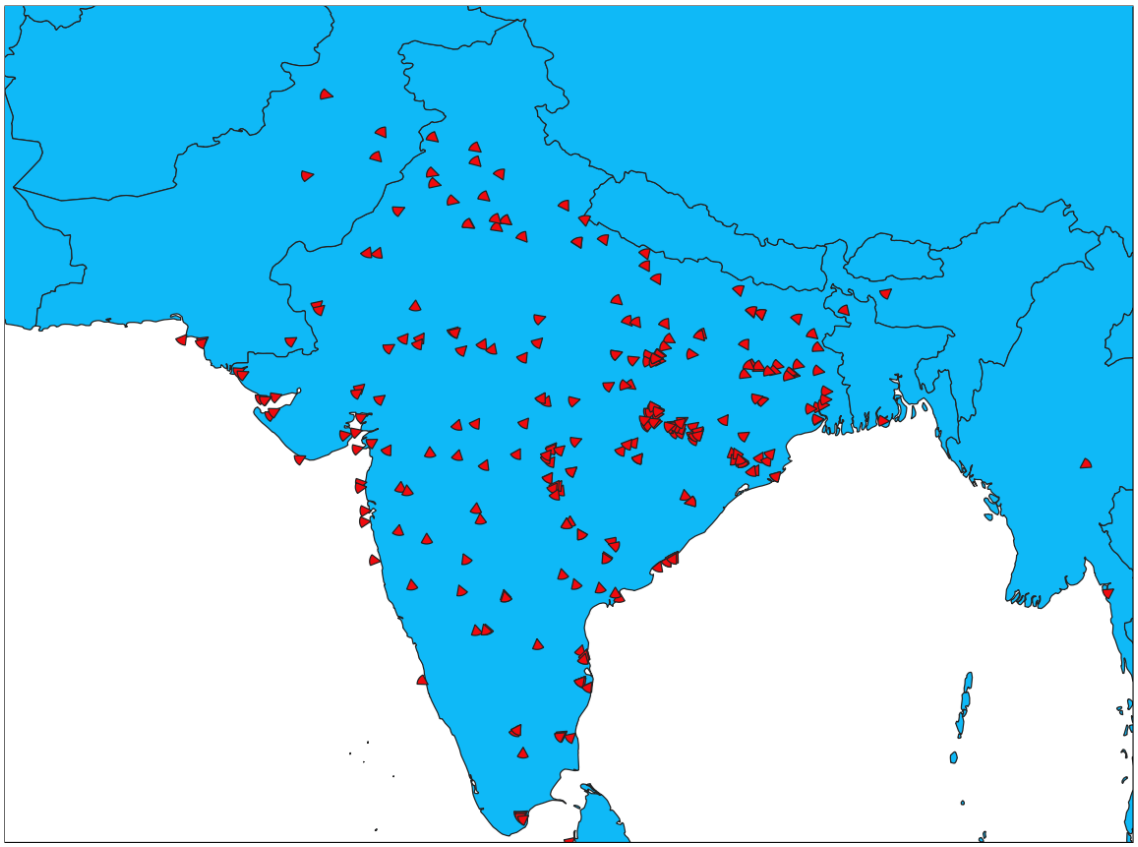
Table 2.16: Placebo results for 40-80 km and 80-120 km distance bands

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Air Diss	Air Diss	Air Diss	Air Diss	Air Diss	Air Diss	Air Diss	Air Diss	Water Diss	Water Diss
Geocode's log dist from nearest plant	-0.065	0.020	-0.077	0.142	-0.115	0.193*	-0.144	0.301*	-0.050	-0.117
	(0.0576)	(0.1206)	(0.0596)	(0.1150)	(0.0998)	(0.0925)	(0.0895)	(0.1225)	(0.0393)	(0.0849)
Geocode's vegetation index	-0.078	0.030	-0.100*	-0.136	-0.171	-0.111	0.009	-0.276***	-0.080	-0.106
	(0.0544)	(0.1053)	(0.0457)	(0.0722)	(0.0878)	(0.1594)	(0.0887)	(0.0674)	(0.0603)	(0.0546)
Geocode area is urban	0.114**	0.128*	0.085**	0.174***	0.161	0.074	0.132*	0.110**	0.029	0.022
	(0.0400)	(0.0494)	(0.0265)	(0.0482)	(0.0823)	(0.0523)	(0.0556)	(0.0374)	(0.0171)	(0.0208)
Respondent's age is 26-60 years	0.047*	0.036*	0.037*	0.036**	0.015	0.017	0.019	0.020	0.019	0.026*
	(0.0196)	(0.0145)	(0.0150)	(0.0135)	(0.0207)	(0.0167)	(0.0284)	(0.0163)	(0.0105)	(0.0120)
Respondent's age is more than 60 years	0.031	0.027	0.017	0.034	-0.054	-0.003	-0.019	0.007	0.001	0.020
	(0.0339)	(0.0314)	(0.0285)	(0.0250)	(0.0344)	(0.0157)	(0.0384)	(0.0191)	(0.0135)	(0.0157)
Respondent's gender is male	-0.015	0.002	-0.016	0.002	-0.008	-0.004	0.002	-0.005	0.002	-0.009
	(0.0138)	(0.0119)	(0.0146)	(0.0127)	(0.0131)	(0.0093)	(0.0155)	(0.0115)	(0.0068)	(0.0080)
Respondent's education is intermediate	0.046*	0.020	0.033*	0.006	0.020	0.057*	0.022	0.051**	0.032**	0.024
	(0.0197)	(0.0134)	(0.0166)	(0.0141)	(0.0217)	(0.0272)	(0.0200)	(0.0181)	(0.0098)	(0.0121)
Respondent's education is high	0.030	0.022	0.013	0.034	0.052	0.026	0.038	0.029	0.056***	0.041*
	(0.0397)	(0.0423)	(0.0339)	(0.0291)	(0.0431)	(0.0266)	(0.0262)	(0.0217)	(0.0142)	(0.0179)
Log annual hh income in '000 USD	-0.007	0.000	-0.007	-0.008	-0.019*	-0.018*	-0.014	-0.023**	-0.017***	-0.020***
	(0.0061)	(0.0091)	(0.0052)	(0.0088)	(0.0073)	(0.0080)	(0.0098)	(0.0085)	(0.0047)	(0.0055)
Respondent has children under 15 yrs	0.002	0.023	-0.002	0.009	0.012	0.023	0.021	0.028	0.009	0.020*
	(0.0126)	(0.0147)	(0.0119)	(0.0127)	(0.0172)	(0.0227)	(0.0164)	(0.0144)	(0.0091)	(0.0086)
Number of observations	5,903	6,361	5,903	6,361	3,608	4,162	3,608	4,162	16,549	13,241
Adj R-squared	0.067	0.041	0.116	0.120	0.113	0.061	0.190	0.180	0.123	0.133
Mean of dependent variable	0.280	0.234	0.280	0.234	0.260	0.230	0.260	0.230	0.271	0.303
Region fixed effects	Admin-0	Admin-0	Admin-1	Admin-1	Admin-0	Admin-0	Admin-1	Admin-1	Admin-1	Admin-1
Distance band	40-80 km	80-120 km	40-80 km	80-120 km	40-80 km	80-120 km	40-80 km	80-120 km	40-80 km	80-120 km
Status of plant operation	Planned	Planned	Planned	Planned	Retired	Retired	Retired	Retired	Operational	Operational

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2) separately for planned and retired and mothballed coal-fired power plants. The sample used in each column is defined by the distance band i.e., how far the survey location is from the nearest coal power plant. Columns 1-4 and Columns 5-8 report the results for planned and retired plants respectively. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for Columns 1-2 and 5-6 and at state/province/admin-1 level for remaining columns. Columns 1-2 and 5-6 control for admin-0 fixed effects and remaining control for admin-1 fixed effects. Refer to Table 2.4 notes for more details.

Figure 2.10: Wind buffer zones for operational plants



Notes: The figure shows buffer zones for influence of wind using an angular restriction of 60° and a distance restriction of 40 km. The direction of the central azimuth through red sectors indicates the annual wind direction for the plants located at specific geolocations on the map.

Table 2.17: Results for wind direction using $PM_{2.5}$ concentration

	(1) PM2.5 Conc.	(2) PM2.5 Conc.	(3) PM2.5 Conc.	(4) PM2.5 Conc.	(5) PM2.5 Conc.	(6) PM2.5 Conc.
Downwind of plant	2.793	2.712	2.698	0.158	0.514	1.186
	(1.6611)	(1.5273)	(1.4113)	(0.6451)	(0.5481)	(0.6112)
Geocode's vegetation index	-1.235	-1.265	-1.096	-0.897	-0.893	-0.800
	(1.6470)	(1.6092)	(1.6285)	(0.9519)	(0.9533)	(0.9271)
Geocode area is urban	-0.251	-0.230	-0.149	0.631*	0.622*	0.628*
	(0.8499)	(0.8325)	(0.8057)	(0.3115)	(0.3107)	(0.3102)
Respondent's age is 26-60 years	-0.360	-0.377	-0.413	0.131	0.132	0.127
	(0.6591)	(0.6723)	(0.6888)	(0.1228)	(0.1224)	(0.1192)
Respondent's age is more than 60 years	-0.006	0.017	-0.042	0.171	0.171	0.151
	(0.2807)	(0.2670)	(0.2889)	(0.1692)	(0.1690)	(0.1625)
Respondent's gender is male	0.180	0.163	0.175	-0.078	-0.080	-0.079
	(0.2972)	(0.2961)	(0.2915)	(0.0772)	(0.0766)	(0.0770)
Respondent's education is intermediate	0.330	0.332	0.353	0.185*	0.186*	0.195*
	(0.2521)	(0.2400)	(0.2456)	(0.0917)	(0.0921)	(0.0928)
Respondent's education is high	0.142	0.162	0.161	0.007	0.008	0.005
	(0.1899)	(0.1916)	(0.1900)	(0.1192)	(0.1180)	(0.1182)
Log annual hh income in '000 USD	-0.335	-0.322	-0.332	0.113	0.113	0.109
	(0.2466)	(0.2452)	(0.2453)	(0.0933)	(0.0932)	(0.0910)
Respondent has children under 15 yrs	1.161	1.166	1.169	-0.065	-0.067	-0.072
	(0.9607)	(0.9749)	(0.9771)	(0.0956)	(0.0948)	(0.0940)
Number of observations	18,147	18,147	18,147	18,147	18,147	18,147
Adj R-squared	0.703	0.704	0.704	0.949	0.949	0.949
Mean of dependent variable	32.026	32.026	32.026	32.026	32.026	32.026
Region fixed effects	Admin-0	Admin-0	Admin-0	Admin-1	Admin-1	Admin-1
Distance band	0-40 km	0-40 km	0-40 km	0-40 km	0-40 km	0-40 km
Wind direction angular buffer	60°	90°	120°	60°	90°	120°

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates by regressing $PM_{2.5}$ concentration at geocode level on the downwind dummy for operational coal-fired power plants. The sample used in each column is defined by the distance band 0-40 km and the angular buffer around the coal-fired power plant i.e., all survey locations that are located within 40 km and falling in the angular buffer of either 60°, 90° or 120° of an operational coal power plant. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for Columns 1-3 and state/province/admin-1 level for remaining columns. Columns 1-3 control for admin-0 fixed effects and remaining columns control for admin-1 fixed effects. "Downwind of plant" is a dummy, which takes value of 1 if the survey geocode falls in the downwind buffer region of a coal-fired power plant, and that varies based on the angular threshold used. Please refer to Table 2.1 notes for details on other variables.

Table 2.18: Results for 0-20 km distance band

	(1) Air Diss	(2) Air Diss	(3) Air Diss	(4) Air Diss	(5) Air Diss	(6) Air Diss
Geocode's log dist from nearest plant	-0.037*	-0.038*	-0.001	-0.035	-0.066	-0.034
	(0.0147)	(0.0150)	(0.0233)	(0.0283)	(0.0534)	(0.0401)
Geocode's vegetation index	-0.019	-0.009	-0.115	0.081	-0.492**	-0.503
	(0.0289)	(0.0376)	(0.0833)	(0.0807)	(0.1190)	(0.2773)
Geocode area is urban	0.092**	0.074*	0.071	0.035	0.077	0.115
	(0.0318)	(0.0322)	(0.0402)	(0.0599)	(0.0459)	(0.1048)
Respondent's age is 26-60 years	0.031*	0.023	0.032	0.030	-0.015	0.023
	(0.0122)	(0.0153)	(0.0326)	(0.0377)	(0.0237)	(0.0382)
Respondent's age is more than 60 years	-0.003	-0.003	0.082	0.084	-0.053	0.006
	(0.0147)	(0.0188)	(0.0474)	(0.0517)	(0.0289)	(0.0400)
Respondent's gender is male	-0.025	-0.021*	-0.028	-0.024	-0.015	-0.019
	(0.0128)	(0.0099)	(0.0297)	(0.0314)	(0.0206)	(0.0308)
Respondent's education is intermediate	0.064***	0.069***	0.052	0.045	0.068	0.081*
	(0.0131)	(0.0144)	(0.0447)	(0.0367)	(0.0459)	(0.0322)
Respondent's education is high	0.090***	0.094***	0.037	0.032	0.079	0.075
	(0.0166)	(0.0155)	(0.0736)	(0.0563)	(0.0452)	(0.0417)
Log annual hh income in '000 USD	-0.012	-0.011	-0.020	-0.011	-0.001	0.003
	(0.0062)	(0.0070)	(0.0245)	(0.0255)	(0.0027)	(0.0126)
Respondent has children under 15 yrs	0.008	0.008	-0.001	0.011	-0.019	-0.061
	(0.0094)	(0.0110)	(0.0220)	(0.0345)	(0.0348)	(0.0420)
Number of observations	8,356	8,356	1,032	1,032	1,352	1,352
Adj R-squared	0.169	0.230	0.066	0.115	0.172	0.253
Mean of dependent variable	0.383	0.383	0.249	0.249	0.352	0.352
Region fixed effects	Admin-0	Admin-1	Admin-0	Admin-1	Admin-0	Admin-1
Distance band	0-20 km	0-20 km	0-20 km	0-20 km	0-20 km	0-20 km
Status of plant operation	Operational	Operational	Planned	Planned	Retired	Retired

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents OLS estimates using the specification in Equation (2.2) for operational, planned, and retired and mothballed coal-fired power plants. The sample used in each column is defined by the distance band 0-20 km. Columns 1-2, Columns 3-4, and Columns 5-6 report the results for operational, planned, and retired plants respectively. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for Columns 1, 3 and 5 and at state/province/admin-1 level for remaining columns. Columns 1, 3 and 5 control for admin-0 fixed effects and remaining control for admin-1 fixed effects. Refer to Table 2.4 notes for more details.

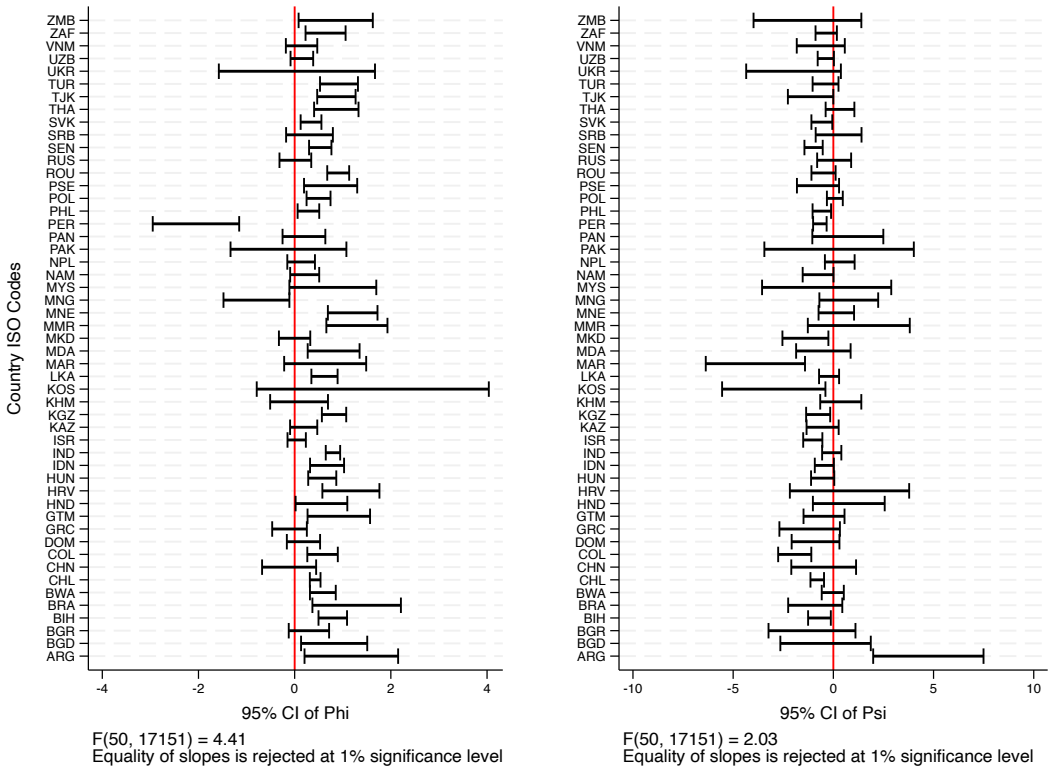
Table 2.19: Life satisfaction results using $PM_{2.5}$ concentration

	(1) Life Sat	(2) Life Sat
Geocode's PM2.5 concentration in $\mu g/m^3$	-0.007 [-0.018,0.003]	-0.015** [-0.027,-0.004]
Geocode's vegetation index	-0.014 [-0.293,0.265]	0.044 [-0.196,0.284]
Geocode area is urban	0.061 [-0.076,0.198]	0.087 [-0.063,0.237]
Respondent's age is 26-60 years	-0.334*** [-0.453,-0.214]	-0.372*** [-0.477,-0.268]
Respondent's age is more than 60 years	-0.429** [-0.744,-0.114]	-0.464*** [-0.621,-0.307]
Respondent's gender is male	-0.154* [-0.303,-0.006]	-0.152** [-0.243,-0.060]
Respondent's education is intermediate	0.300*** [0.135,0.465]	0.316*** [0.191,0.441]
Respondent's education is high	0.644*** [0.492,0.795]	0.676*** [0.518,0.835]
Log annual hh income in '000 USD	0.487*** [0.358,0.615]	0.477*** [0.408,0.546]
Respondent has children under 15 yrs	-0.020 [-0.145,0.105]	0.022 [-0.071,0.115]
Number of observations	17,869	17,869
Adj R-squared	0.199	0.234
Mean of dependent variable	5.405	5.405
Mean household income in USD	14810	14810
Region fixed effects	Admin-0	Admin-1
Countries included	Global	Global

95% confidence interval in brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents estimates using the specification in Equation (2.3) for operational coal-fired power plants. The sample used in each column is defined by distance band 0-40 km i.e., survey locations that are located within 40 km distance from the nearest coal power plant. Table 2.11 provide the list of countries from which sample surveys are used in this specification. 95% confidence interval bounds are reported in square brackets. Column 1 controls for admin-0 fixed effects while Column 2 controls for admin-1 fixed effects. The dependent variable, *Life Sat*, is a shorthand for life satisfaction, which takes values between 0 ("the worst possible life") and 10 ("the best possible life") based on what surveyed individuals report as their current life satisfaction. The main variables of interest are $PM_{2.5}$ concentration at the geocode level and logarithm of annual household income. The first variable takes value 2 (1) if an individual is dissatisfied (satisfied) with ambient air quality and the second variable is logarithm of household reported total annual income in 1000 USD. Please refer to Table 2.1 notes for details on other variables.

Figure 2.11: Estimates of ϕ and ψ parameters for sample countries



Notes: The chart shows 95% confidence interval for ϕ and ψ estimates for each of the 51 countries in the main sample by running a pooled regression with country interactions corresponding to Equation (2.3). Equality of slopes across countries for both ϕ and ψ is rejected at 1% significance level, thereby highlighting the heterogeneous effect of both income and air quality satisfaction on overall life satisfaction across countries.

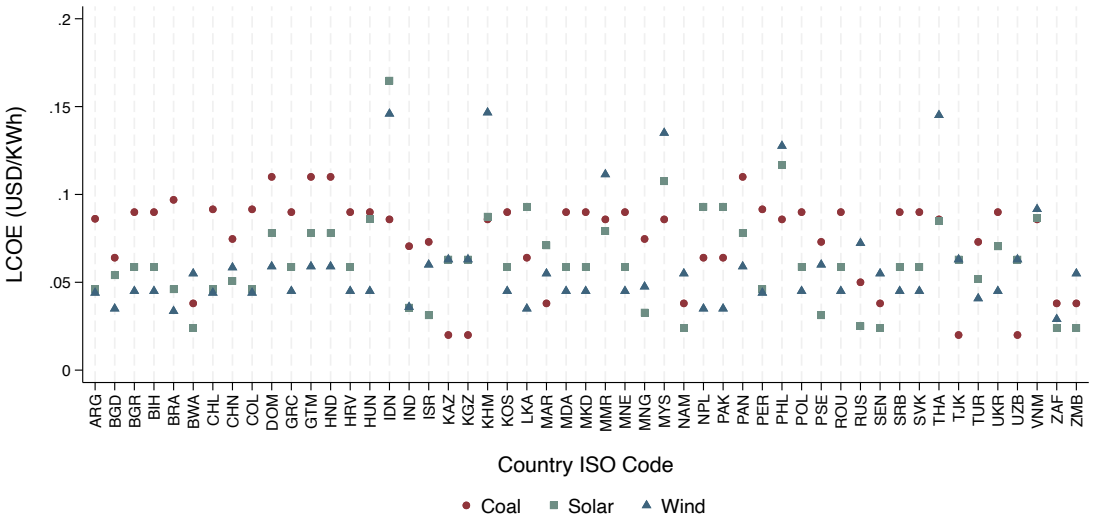
Table 2.20: Ordered logit estimation results for life satisfaction

	(1) Life Sat
Log air quality dissatisfaction	-0.395*** [-0.511,-0.279]
Geocode's vegetation index	0.020 [-0.155,0.195]
Geocode area is urban	0.093 [-0.029,0.215]
Respondent's age is 26-60 years	-0.301*** [-0.384,-0.219]
Respondent's age is more than 60 years	-0.397*** [-0.529,-0.264]
Respondent's gender is male	-0.133*** [-0.210,-0.057]
Respondent's education is intermediate	0.247*** [0.146,0.348]
Respondent's education is high	0.608*** [0.472,0.744]
Log annual hh income in '000 USD	0.377*** [0.318,0.436]
Respondent has children under 15 yrs	0.031 [-0.047,0.108]
Number of observations	163,029
Pseudo R-squared	0.034
Log likelihood	-61,047
Mean of dependent variable	5.411
Mean household income in USD	14855
Region fixed effects	Admin-1
Countries included	Global

95% confidence interval in brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

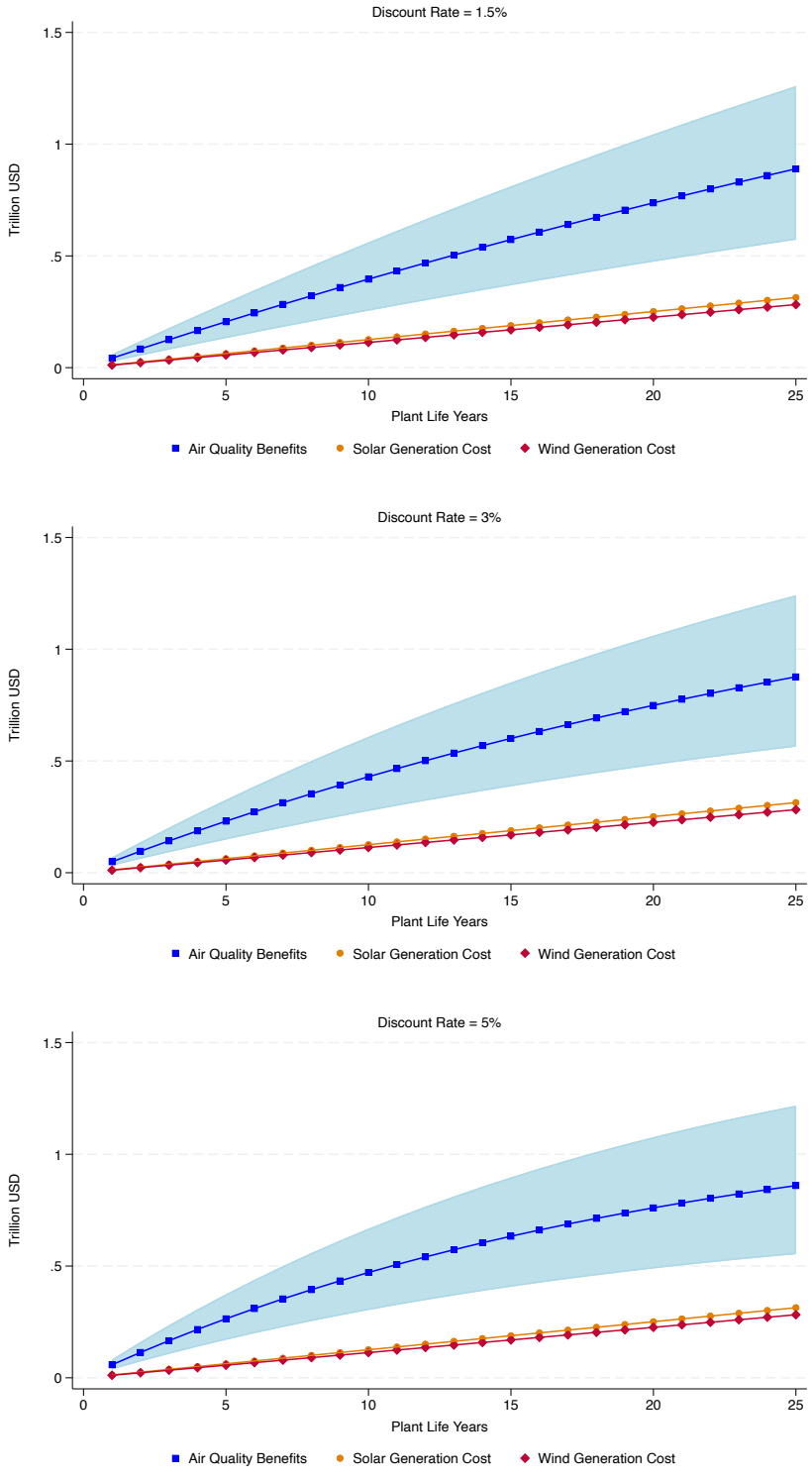
Notes: The table above reports results for ordered logistical model estimation with fixed effects corresponding to OLS estimation results reported in Table 2.7. We implement a robust estimation for fixed effects ordered logit models using the estimator proposed by Baetschmann et al. (2020).

Figure 2.12: Unit cost of energy for different generation technologies



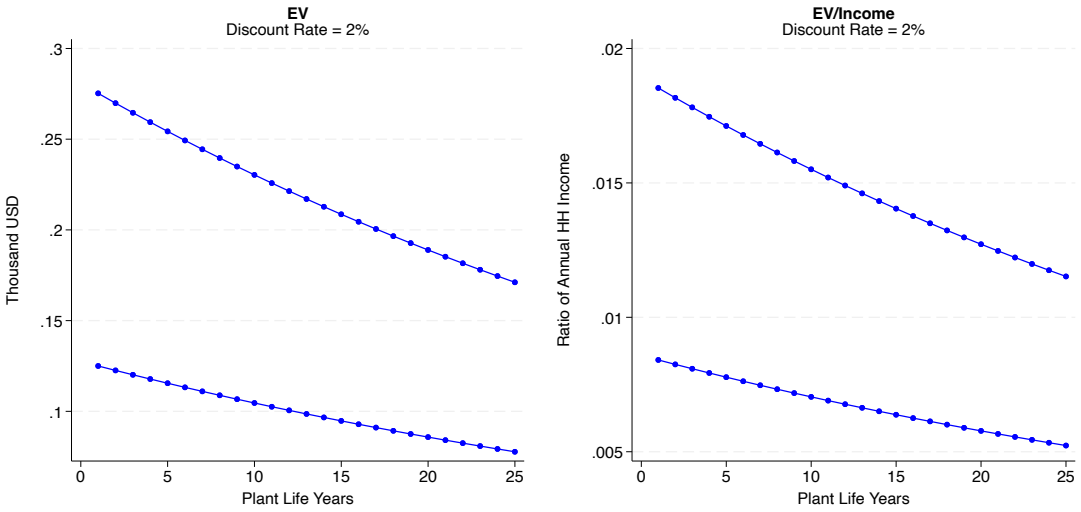
Notes: The graph shows LCOE values for all 51 countries in the main sample as listed in Table 2.11. LCOE measures lifetime costs divided by energy production. It accounts for present value of the total cost of building and operating a power plant over an assumed lifetime. This measure allows comparison of different technologies (e.g., wind, solar, coal) of unequal life spans, project size, different capital cost, risk, return, and capacities for each of the respective sources. LCOE also accounts for different capacity factors across energy sources and plants.

Figure 2.13: Cost-benefit analysis for alternate discount rates



Notes: Top/mid/bottom row show results for 1.5/3/5% discount rate. Refer to Figure 2.2 for more details.

Figure 2.14: EV and EV/Income during transition project life cycle



Notes: The chart shows the present-discounted value of estimated EV and EV to annual household income ratio in left and right plots respectively assuming an annual discount rate of 2% for an energy transition project life cycle of 25 years.

Table 2.21: Combined top-25 coal power stations based on affected population

(1) Country	(2) State/Province	(3) Name of Plant	(4) Population	(5) Ann. Emission (in mil. tons)	(6) Capacity (in MW)	(7) Plant Life (in years)	(8) Solar Cost (in mil. $)	(9) Wind Cost (in mil. $)	(10) Gross Benefits (in mil. $)	(11) Gross Benefits LB (in mil. $)
India	Delhi	Rajghat Delhi	30871582	0.8	135	2	78.64	70.72	24874.62	15998.17
China	Shanghai	Wujing	29310080	11.9	2537.5	17	1478.20	1329.26	23616.45	15188.97
China	Shanghai	Shanghai Gaoqiao	26608464	0.9	150	9	87.38	78.58	21439.64	13788.95
China	Shanghai	Waigaoqiao	25449414	22.6	5240	22	3052.51	2744.96	20505.74	13188.31
China	Shanghai	Baoshan Works	24979818	5.9	1050	11	611.67	550.04	20127.36	12944.96
China	Shanghai	Shidongkou	24205972	17.6	3820	13	2225.30	2001.10	19503.84	12543.94
India	West Bengal	Budge Budge	23684622	4.1	750	23	436.91	392.89	19083.77	12273.77
India	West Bengal	Southern CESC	23486538	0.8	135	11	78.64	70.72	18924.16	12171.12
India	West Bengal	Titagarh	23426320	1.2	240	5	139.81	125.72	18875.64	12139.91
China	Guangdong	Hengyun-D	22829176	3.2	660	28	384.48	345.74	18394.49	11830.46
China	Guangdong	Hengyun-C	22803540	2.3	420	5	244.67	220.02	18373.84	11817.18
India	Haryana	Faridabad	22755274	0.9	165	2	96.12	86.43	18334.95	11792.16
China	Guangdong	Jiulong Paper Mill	22686148	3.2	620	25	361.17	324.79	18279.25	11756.34
China	Guangdong	Yuehua Huangpu	22633900	3.4	660	2	384.48	345.74	18237.15	11729.27
China	Guangdong	Guangzhou Refinery	22396020	1.0	200	28	116.51	104.77	18045.48	11605.99
Indonesia	West Java	Cikarang Babelan	21297338	1.4	280	38	163.11	146.68	17160.23	11036.64
India	Maharashtra	Trombay	21296044	4.0	810	12	471.86	424.32	17159.18	11035.97
China	Guangdong	Guangzhou Lixin	20995940	2.8	660	33	384.48	345.74	16917.37	10880.45
China	Guangdong	Mawan	20927798	9.4	1940	19	1130.13	1016.27	16862.47	10845.13
China	Guangdong	Guangzhou Nansha	20716364	2.8	600	30	349.52	314.31	16692.11	10735.57
China	Guangdong	Lee & Man Paper	20536522	1.3	216	28	125.83	113.15	16547.20	10642.37
China	Guangdong	Shunde Desheng	20437078	2.8	600	29	349.52	314.31	16467.07	10590.83
India	Uttar Pradesh	National Capital Dadri	19695644	4.8	840	14	489.33	440.03	15869.67	10206.61
Japan	Kanto	Isogo	19357188	5.0	1200	27	699.05	628.62	15596.96	10031.22
China	Guangdong	Dongtang Plant	19029100	1.5	285	2	166.02	149.3	15332.60	9861.20

Notes: The table lists the top 25 coal power stations in the world in decreasing order of total population affected, which is reported in Column 4. The population figures are the total number of individuals located within 40 km of respective plants. Solar and wind costs report the cost of green transition through solar and wind technology respectively. These costs are calculated by using source-specific average global LCOE values and respective capacities of coal plants as reported in Column 6. Air quality benefits in Column 10 are computed by multiplying EV values, which are computed by using Equation (2.4), with the total number of residences in 0-40 km distance band. For EV calculations, global parameter values for ψ, ϕ, $\frac{AirDiss}{\widetilde{AirDiss}}$, and y are used. Column 11 reports the lower bound on the gross air quality benefits from shutting down each of the listed plants.

Figure 2.15: Plant-level net air quality benefits from closing operational plants

Notes: Chart shows the net benefits from closing all the operational coal-fired power in 2019 located across the whole world. The parameter values for ψ, ϕ, $\frac{AirDiss}{AirDiss}$, and y are taken from the global estimates using all 51 countries combined. The policy experiment entails phasing out coal-fired power and replacing that freed capacity with 50% solar and 50% wind generation. The costs of solar and wind energy generation are calculated by multiplying respective source-specific global average LCOE values in USD/kWh with the total energy demand. The LCOE values for solar and wind are inflated by a factor of 4 and 2 respectively.

Table 2.22: Life satisfaction regression results for India and China

	(1) Life Sat	(2) Life Sat	(3) Life Sat	(4) Life Sat
Log air quality dissatisfaction	-0.080	-0.803***	-0.124	-0.646**
	[-0.553,0.393]	[-1.137,-0.469]	[-0.709,0.461]	[-1.051,-0.241]
Geocode's vegetation index	-0.363	-0.973**	-0.038	-0.331
	[-1.224,0.497]	[-1.635,-0.311]	[-1.142,1.066]	[-1.430,0.768]
Geocode area is urban	0.352*	0.018	0.118	0.130
	[0.066,0.637]	[-0.219,0.254]	[-0.413,0.650]	[-0.447,0.708]
Respondent's age is 26-60 years	-0.181	-0.017	-0.414**	-0.121
	[-0.475,0.113]	[-0.279,0.246]	[-0.679,-0.150]	[-0.392,0.149]
Respondent's age is more than 60 years	-0.474*	0.550**	-0.730**	0.409*
	[-0.902,-0.047]	[0.200,0.899]	[-1.174,-0.285]	[0.017,0.800]
Respondent's gender is male	-0.345**	0.142	-0.183	0.187
	[-0.604,-0.086]	[-0.054,0.337]	[-0.484,0.118]	[-0.065,0.438]
Respondent's education is intermediate	0.586***	0.253*	0.332*	0.267*
	[0.291,0.880]	[0.029,0.477]	[0.008,0.655]	[0.041,0.492]
Respondent's education is high	0.708**	0.424*	0.545	0.544***
	[0.200,1.216]	[0.075,0.774]	[-0.065,1.155]	[0.266,0.822]
Log annual hh income in '000 USD	0.797***	0.427***	0.681***	0.454***
	[0.649,0.944]	[0.317,0.536]	[0.512,0.850]	[0.309,0.599]
Respondent has children under 15 yrs	-0.297*	-0.122	-0.025	-0.068
	[-0.549,-0.045]	[-0.324,0.079]	[-0.202,0.152]	[-0.285,0.149]
Number of observations	2,131	2,099	2,131	2,099
Adj R-squared	0.093	0.072	0.171	0.127
Mean of dependent variable	3.262	5.213	3.262	5.213
Mean household income in USD	4626	19365	4626	19365
Region fixed effects	-	-	Admin-1	Admin-1
Countries included	India	China	India	China

95% confidence interval in brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

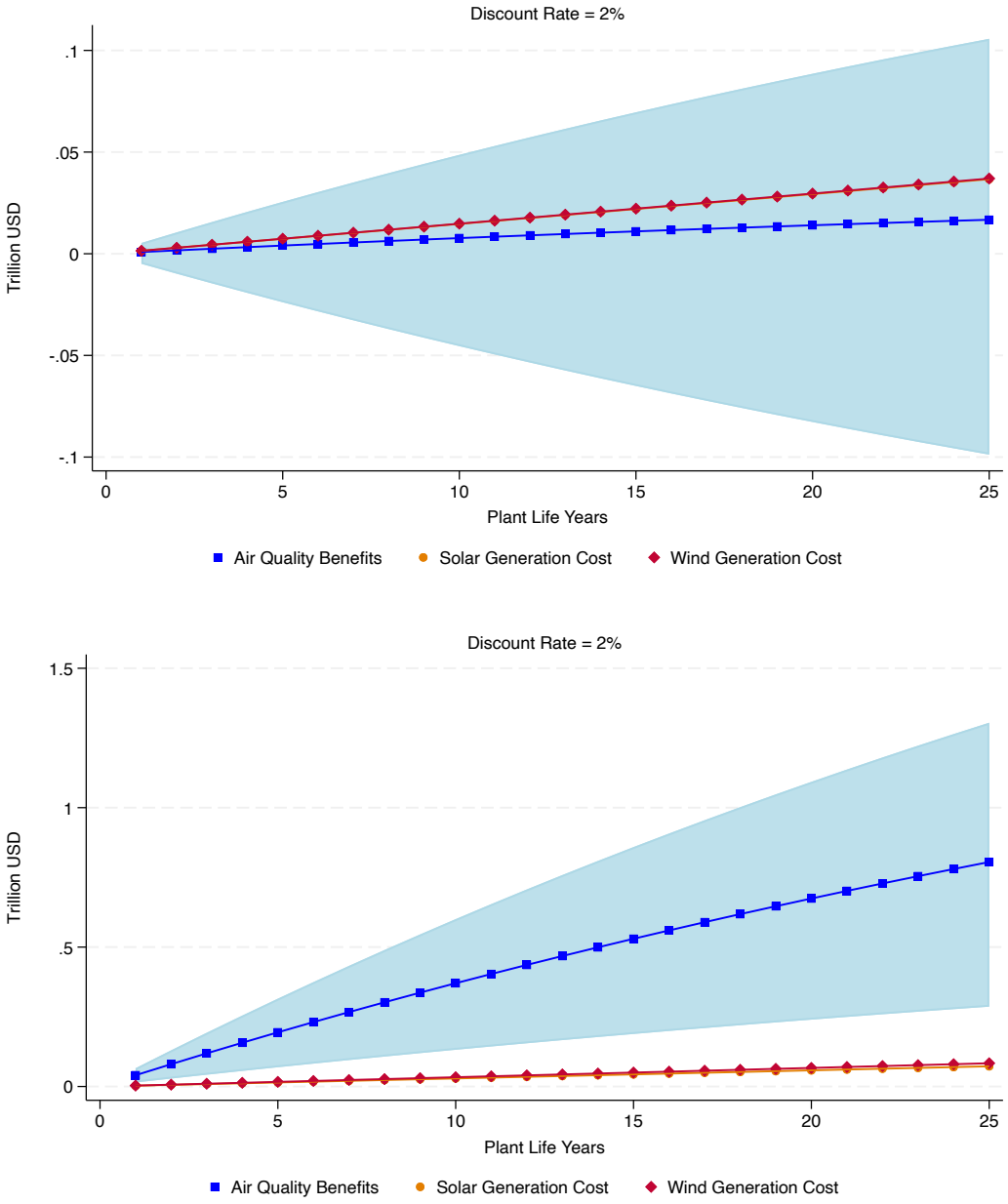
Notes: This table presents estimates using the specification in Equation (2.3) for operational coal-fired power plants in India and China. The sample used in each column is defined by distance band 0-40 km i.e., survey locations that are located within a 40 km distance from the nearest coal power plant. 95% confidence interval bounds are reported in square brackets. Columns 3 and 4 control for admin-1 fixed effects. The dependent variable, *Life Sat*, is a shorthand for life satisfaction, which takes values between 0 ("the worst possible life") and 10 ("the best possible life") based on what surveyed individuals reports as their current life satisfaction. The main variables of interest are logarithm of air quality dissatisfaction and logarithm of annual household income. The first variable takes value 2 (1) if an individual is dissatisfied (satisfied) with ambient air quality and the second variable is logarithm of household reported total annual income in 1000 USD. Please refer to Table 2.1 notes for details on other variables.

Table 2.23: Aggregate equivalent variation results for India and China

(1) Geographical Category	(2) ψ	(3) ϕ	(4) y (in \$)	(5) $AirDiss/\widetilde{AirDiss}$	(6) e (in \$)	(7) Affected Population	(8) HH Size (# persons)	(9) AEV (in tril. \$)
Panel 1: Point estimates								
India	-0.124	0.681	4626	1.38	264	375,939,467	5.8	0.017
China	-0.646	0.454	19365	1.62	9617	374,225,419	4.4	0.818
Panel 2: $\underline{\gamma}$ and $\underline{\beta}$								
India	-0.709	0.512	4626	1.38	1665	375,939,467	5.8	0.108
China	-1.051	0.309	19365	1.62	15612	374,225,419	4.4	1.328
Panel 3: $\overline{\gamma}$ and $\overline{\beta}$								
India	0.461	0.850	4626	1.38	-883	375,939,467	5.8	-0.057
China	-0.241	0.599	19365	1.62	3416	374,225,419	4.4	0.291

Notes: The three rows correspond to point estimates and lower and upper bounds of 95% confidence intervals of ψ and ϕ parameters respectively. Estimates on logarithm of annual household income, ϕ, logarithm of air quality dissatisfaction, ψ, and average income, y, are taken from Columns 3 and 4 of Table 2.22 for respective countries. $\frac{AirDiss}{\widetilde{AirDiss}}$ is the ratio of air quality dissatisfaction level in the 0-40 km distance band and that outside of the band for each country. e is the equivalent variation computed using Equation (2.4). The population is computed by adding the number of individuals living in a circle of radius 40 km around each coal plant. The population data comes from the Gridded Population of the World, v4 (GPWv4) database for year 2020. AEV is generated by multiplying e with population estimates downscaled by the number of persons living in a typical household taken from the Area Database v4.1 of the Global Data Lab.

Figure 2.16: Cost-benefit analysis results for India and China



Notes: Charts show the cost-benefit analysis results for India (top) and China (bottom). The blue line represents point estimates of air quality benefits with the shaded area showing upper and lower bounds on the estimates calculated using country-specific parameter values. The costs of solar and wind energy generation are calculated by multiplying their respective source-geography-specific LCOE values in USD/kWh with the total excess energy demand because of closing of coal plants. Please refer to Figure 2.2 for more details.

Table 2.24: Total benefits of energy transition for different regions

(1) Geographical Category	(2) Gross Benefits (in tril. $)	(3) Net Benefits (in tril. $)	(4) Gross Benefits LB (in tril. $)	(5) Net Benefits LB (in tril. $)
Panel 1: Actual parameters				
Global	.903	.605	.581	.283
India	.017	-.02	-.057	-.094
China	.821	.743	.292	.214
Panel 2: Global preference parameters				
Global	.903	.605	.581	.283
India	.081	.044	.053	.016
China	.628	.555	.416	.338

Notes: The table reports gross and net benefits of closing coal plants in different geographical categories using point estimates for the respective categories in Columns 2 and 3 respectively. Columns 4 and 5 report the lower bound on the benefits. The policy experiment entails phasing out coal-fired power at a constant rate of 4% per year and replacing that freed capacity with 50% solar and 50% wind generation over a period of 25 years. The benefits shown here are for the last year i.e., 25th year of plant operation. The costs of solar and wind energy generation are calculated by multiplying their respective source-geography-specific LCOE values in USD/kWh with the total excess energy demand because of closing of coal plants. Panel 1 reports results when respective parameter values for each category is used to calculate benefits, while in Panel 2, we use Global category parameter values of ψ and ϕ for all categories.

Table 2.25: Life satisfaction regression results for education categories

	(1) Life Sat	(2) Life Sat	(3) Life Sat	(4) Life Sat	(5) Life Sat	(6) Life Sat
Log air quality dissatisfaction	-0.621***	-0.447***	-0.468***	-0.650***	-0.407***	-0.511***
	[-0.922,-0.320]	[-0.647,-0.247]	[-0.734,-0.202]	[-0.914,-0.386]	[-0.586,-0.229]	[-0.771,-0.251]
Geocode's vegetation index	-0.413	0.106	0.006	-0.184	0.036	0.236
	[-1.090,0.263]	[-0.090,0.303]	[-0.440,0.452]	[-0.800,0.431]	[-0.208,0.280]	[-0.206,0.678]
Geocode area is urban	-0.043	0.134	0.178	-0.084	0.170*	0.233
	[-0.244,0.157]	[-0.012,0.280]	[-0.070,0.426]	[-0.340,0.173]	[0.014,0.327]	[-0.038,0.504]
Respondent's age is 26-60 years	-0.561***	-0.305***	-0.087	-0.608***	-0.335***	-0.204*
	[-0.844,-0.277]	[-0.426,-0.185]	[-0.312,0.138]	[-0.816,-0.400]	[-0.452,-0.219]	[-0.395,-0.013]
Respondent's age is more than 60 years	-0.315	-0.575***	-0.426**	-0.353**	-0.615***	-0.494**
	[-0.669,0.039]	[-0.894,-0.255]	[-0.732,-0.121]	[-0.611,-0.095]	[-0.812,-0.418]	[-0.809,-0.178]
Respondent's gender is male	-0.227	-0.153	-0.145	-0.219*	-0.148*	-0.131
	[-0.482,0.027]	[-0.317,0.012]	[-0.298,0.008]	[-0.394,-0.044]	[-0.269,-0.028]	[-0.275,0.012]
Log annual hh income in '000 USD	0.565***	0.481***	0.393***	0.549***	0.456***	0.391***
	[0.418,0.711]	[0.344,0.619]	[0.204,0.582]	[0.452,0.645]	[0.361,0.550]	[0.248,0.534]
Respondent has children under 15 yrs	-0.176*	0.043	-0.011	-0.065	0.058	0.022
	[-0.312,-0.040]	[-0.104,0.190]	[-0.204,0.181]	[-0.221,0.090]	[-0.075,0.192]	[-0.133,0.177]
Number of observations	5,572	9,166	2,957	5,547	9,161	2,911
Adj R-squared	0.190	0.155	0.166	0.229	0.182	0.213
Mean of dependent variable	4.665	5.611	6.196	4.666	5.610	6.190
Mean household income in USD	8872	15291	24735	8865	15289	24810
Region fixed effects	Admin-0	Admin-0	Admin-0	Admin-1	Admin-1	Admin-1
Countries included	Global	Global	Global	Global	Global	Global
Education level	Primary	Intermediate	High	Primary	Intermediate	High

95% confidence interval in brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

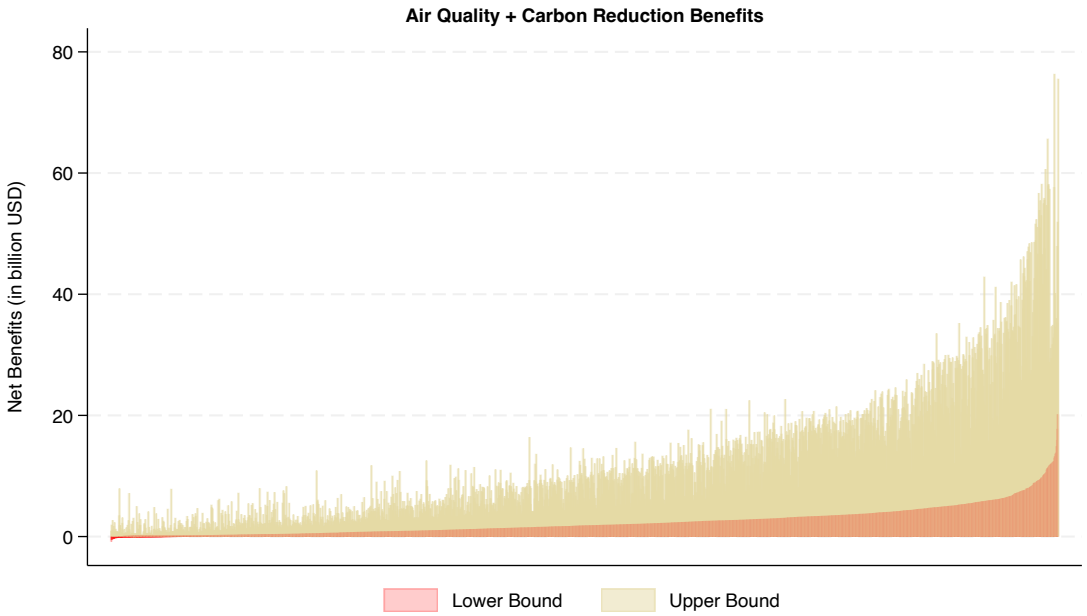
Notes: This table presents estimates using the specification in Equation (2.3) for operational coal-fired power plants for each education group separately. The sample used in each column is defined by distance band 0-40 km i.e., survey locations that are located within a 40 km distance from the nearest coal power plant. Table 2.11 provides the list of countries from which sample surveys are used in this specification. 95% confidence interval bounds are reported in square brackets. Columns 1-3 control for admin-0 fixed effects while Columns 4-6 control for admin-1 fixed effects. The dependent variable, *Life Sat*, is a shorthand for life satisfaction, which takes values between 0 ("the worst possible life") and 10 ("the best possible life") based on what surveyed individuals report as their current life satisfaction. The main variables of interest are logarithm of air quality dissatisfaction and logarithm of annual household income. The first variable takes value 2 (1) if an individual is dissatisfied (satisfied) with ambient air quality and the second variable is logarithm of household reported total annual income in 1000 USD. Please refer to Table 2.1 notes for details on other variables.

Table 2.26: Equivalent variation results for education categories

(1) Education Category	(2) ψ	(3) ϕ	(4) y (in \$)	(5) $AirDiss/\widetilde{AirDiss}$	(6) e (in \$)
Panel 1: Point estimates					
Primary	-0.650	0.549	8865	1.37	2758
Intermediate	-0.407	0.456	15289	1.37	3745
High	-0.511	0.391	24810	1.37	8368
Panel 2: $\underline{\psi}$ and $\underline{\phi}$					
Primary	-0.914	0.452	8865	1.37	4175
Intermediate	-0.586	0.361	15289	1.37	6117
High	-0.771	0.248	24810	1.37	15487
Panel 3: $\overline{\psi}$ and $\overline{\phi}$					
Primary	-0.386	0.645	8865	1.37	1522
Intermediate	-0.229	0.550	15289	1.37	1878
High	-0.251	0.534	24810	1.37	3413

Notes: The three panels correspond to point estimates and lower and upper bounds of 95% confidence intervals of ψ and ϕ parameters respectively. Estimates on logarithm of annual household income, ϕ, logarithm of air quality dissatisfaction, ψ, and average income, y, are taken from Columns 4, 5, and 6 of Table 2.25 for respective education categories. $\frac{AirDiss}{\widetilde{AirDiss}}$ is the ratio of average air quality dissatisfaction level in the 0-40 km distance band to that outside of the 40 km band for global category. e is the equivalent variation computed using Equation (2.4).

Figure 2.17: Plant-level net benefits from closing operational plants



Notes: Chart shows the sum of net air quality and carbon benefits from closing all the operational coal-fired power in 2019 across the whole world. The parameter values for ψ, ϕ, $\frac{AirDiss}{\widetilde{AirDiss}}$, and y are taken from the global estimates using all 51 countries combined. The policy experiment entails phasing out coal-fired power and replacing that freed capacity with 50% solar and 50% wind generation. The costs of solar and wind energy generation are calculated by multiplying respective source-specific average global LCOE values in USD/kWh with the total energy demand.

Table 2.27: Employment in energy generation sectors for sample countries

ISO	Country	Solar			Wind			Coal		
		Jobs (000)	Capacity (MW)	Jobs/MW	Jobs (000)	Capacity (MW)	Jobs/MW	Jobs (000)	Capacity (MW)	Jobs/MW
ARG	Argentina	2.2	764.1	2.9	1.7	2623.9	0.6			
BGD	Bangladesh	110	284	387.3	0.1	2.9	34.5			
BIH	Bosnia and Herzegovina	0.1	34.9	1.7	0.2	135.0	1.5			2.8
BWA	Botswana	0.04	5.9	6.5	0.04	170.2	0.3			
BRA	Brazil	68	7879.2	8.6	40.2	17198.3	2.3			
BGR	Bulgaria	1	1097.4	0.9	0.5	702.8	0.8	55.3	3733	14.8
KHM	Cambodia	7.1	315.0	22.4	0.005	0.3	20.6			
CHL	Chile	7.1	3205.4	2.2	7.5	2149	3.5			
CHN	China	2300	253417.8	9.1	550	282112.7	2	3209	1064400	3
COL	Colombia	0.4	85.5	4.2	2.1	18.4	114	44.3	1633.5	27.1
HRV	Croatia	0.1	108.5	0.5	2.3	801.3	2.9			2.8
DOM	Dominican Republic	0.3	385.6	0.8	0.3	370.3	0.8			
GRC	Greece	6.1	3287.7	1.9	6.8	4119.3	1.7	6.1	4337	1.4
GTM	Guatemala	0.1	100.8	0.8	0.1	107.4	0.8			
HND	Honduras	0.4	514	0.8	0.2	241.3	0.8			
HUN	Hungary	8.9	2131	4.2	0.8	321	2.5	2.2	783	2.8
IND	India	163.5	39042.7	4.2	44	38558.6	1.1	416.2	231900	1.8
IDN	Indonesia	4.2	185.3	22.4	3.2	154.3	20.6	240	40200	6
ISR	Israel	2.3	2230	1	0.1	27.3	3.7			
KAZ	Kazakhstan	5	1718.6	2.9	2.6	486.3	5.3	29.7	12986	2.3
KOS	Kosovo	0.1	10	6.3	0.02	32	0.5			2.8
KGZ	Kyrgyzstan	0.03	584.3	0.1	0.9	162.5	5.3			
MYS	Malaysia	54.9	1482.6	37	7.7	374.6	20.6			
MDA	Moldova	0.01	4.3	2.4	0.1	37	1.6			2.8
MNG	Mongolia	0.04	89.6	0.4	0.1	156	0.6			
MNE	Montenegro	0.01	6	1.7	0.9	118	7.6			2.8
MAR	Morocco	1	194	5.2	3.5	1405	2.5			
MMR	Myanmar	1.9	84.5	22.4	0.0001	0.006	20.6			
NAM	Namibia	0.5	145	3.2	0.001	5.2	0.3			
NPL	Nepal	0.1	66.9	2.2	0.0002	0.2	1.0			
MKD	North Macedonia	0.9	94.2	9.6	0.03	37.0	0.8			2.8
PAK	Pakistan	1.9	860.3	2.2	1	1235.9	0.8			
PSE	Palestine	0.1	116.8	1	0.1	27.3	3.7			
PAN	Panama	0.2	242.1	0.8	0.2	270	0.7			
PER	Peru	0.4	334.8	1.1	0.3	409	0.7			
PHL	Philippines	41	1057.9	38.8	23.8	442.9	53.7			
POL	Poland	29.4	3955	7.4	9.7	6298.3	1.5	91.4	27244	3.4
ROU	Romania	1	1382.5	0.7	2.3	3012.5	0.8	16	4465	3.6
RUS	Russia	3.5	1427.8	2.5	12	945.3	12.7	150.1	41800	3.6
SEN	Senegal	1.1	171	6.5	0.04	158.7	0.3			
SRB	Serbia	0.1	30.5	3	0.1	398	0.2	18.4	5314	3.5
SVK	Slovakia	0.2	535	0.4	0.007	3	2.2	2.4	926	2.6
ZAF	South Africa	21.5	5489.6	3.9	18.8	2516	7.5	74.8	43400	1.7
LKA	Sri Lanka	0.8	370.9	2.2	2.7	179	15.1			
TJK	Tajikistan	0.9	584.3	1.5	0.9	162.5	5.3			
THA	Thailand	18.7	2982.6	6.3	2	1506.8	1.3	0.9	5933	0.1
TUR	Turkey	7.7	6667.4	1.2	23	8832.4	2.6	51.8	19700	2.6
UKR	Ukraine	29.8	7331	4.1	3.8	1402	2.7	44.3	21842	2
UZB	Uzbekistan	0.005	3.5	1.5	0.004	0.8	5.3			
VNM	Vietnam	126.3	16660.5	7.6	3.5	518	6.8	86.4	20917	4.1
ZMB	Zambia	1.2	96.4	12.4	0.043	170.2	0.3			

Notes: The table reports country-level estimates of jobs present in different energy generation sectors. We could not come up with estimates for the coal sector of all the countries and that is why there are blanks in the table. Also, estimates for some of the countries are imputed from nearby countries. For example, for Jobs/MW of wind for Kyrgyzstan, Tajikistan, and Uzbekistan, we use the estimates for Kazakhstan as it is a neighboring country to all three of them. References used for deriving the numbers include country-level estimates from different energy accounting estimates from IRENA, IEA, etc.

Table 2.28: IV results for air quality dissatisfaction and operational plants location

	(1) Air Diss	(2) Air Diss	(3) Air Diss	(4) Air Diss
Geocode's log dist from nearest plant	-0.441**	-0.324***	-0.305**	-0.301**
	(0.1413)	(0.0889)	(0.1057)	(0.0978)
Geocode's vegetation index	0.078	0.026	0.053	0.051
	(0.0714)	(0.0531)	(0.0547)	(0.0520)
Geocode area is urban	0.013	0.040	0.023	0.024
	(0.0456)	(0.0357)	(0.0347)	(0.0325)
Respondent's age is 26-60 years	0.023	0.022*	0.019	0.019
	(0.0116)	(0.0109)	(0.0107)	(0.0108)
Respondent's age is more than 60 years	-0.021	-0.021	-0.018	-0.018
	(0.0193)	(0.0176)	(0.0135)	(0.0135)
Respondent's gender is male	-0.010	-0.013	-0.014	-0.014
	(0.0123)	(0.0110)	(0.0077)	(0.0077)
Respondent's education is intermediate	0.054***	0.055***	0.054***	0.055***
	(0.0123)	(0.0111)	(0.0106)	(0.0106)
Respondent's education is high	0.064**	0.071***	0.075***	0.075***
	(0.0213)	(0.0190)	(0.0155)	(0.0154)
Log annual hh income in '000 USD	-0.009	-0.008	-0.009	-0.009
	(0.0087)	(0.0074)	(0.0058)	(0.0057)
Respondent has children under 15 yrs	0.010	0.008	0.007	0.007
	(0.0104)	(0.0093)	(0.0083)	(0.0082)
Number of observations	17,964	17,964	17,964	17,964
Under-id LM test statistic	8.743	8.787	13.172	15.084
Under-id LM test p-value	0.003	0.012	0.000	0.001
Weak-id F statistic (first stage)	16.302	11.888	15.872	9.404
Hansen J test statistic	0.000	1.553	0.000	0.006
Hansen J test p-value		0.213		0.939
Mean of dependent variable	0.327	0.327	0.327	0.327
Number of instruments	1	2	1	2
Region fixed effects	Admin-0	Admin-0	Admin-1	Admin-1

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents IV estimates using the specification in Equation (2.8) for operational coal-fired power plants. The two instruments used are: (i) logarithm of distance of survey locations from nearest railroad and (ii) logarithm of distance of survey locations from nearest water-body. Columns 1 and 3 use instrument (i) only, while Columns 2 and 4 use both instruments. The sample used in each column is defined by distance band 0-40 km i.e., survey locations that are located within 40 km distance from the nearest coal power plant. Table 2.11 provides the list of countries for which sample surveys are used in this specification. Standard errors, which are reported in parentheses, are clustered at country/admin-0 level for the first two columns and state/province/admin-1 level for the last two columns. Columns 1-2 and Columns 3-4 control for admin-0 and admin-1 fixed effects respectively. The dependent variable, *Air Diss*, is a shorthand for Air Quality Dissatisfaction, which takes value 1 (0) if the surveyed individual is dissatisfied (satisfied) with the ambient air quality. The main variable of interest is geocode's logarithm of distance from the nearest plant, which is the straight-line distance between the survey and nearest coal plant location. Please refer to Table 2.1 notes for details on other variables. First-stage and reduced-form results are reported in Table 2.29 in the Appendix.

Table 2.29: First-stage and reduced-form results for operational plants

	(1)	(2)	(3)	(4)
Geocode's log dist from nearest railroad	-0.020***	-0.020***	-0.017***	-0.017***
	(0.0038)	(0.0037)	(0.0045)	(0.0045)
Geocode's vegetation index	-0.118***	-0.115***	-0.079**	-0.068*
	(0.0313)	(0.0325)	(0.0283)	(0.0282)
Geocode area is urban	0.102***	0.101***	0.086***	0.084***
	(0.0225)	(0.0234)	(0.0219)	(0.0220)
Respondent's age is 26-60 years	0.018	0.018	0.015	0.015
	(0.0108)	(0.0108)	(0.0099)	(0.0099)
Respondent's age is more than 60 years	-0.023	-0.023	-0.020	-0.020
	(0.0154)	(0.0154)	(0.0128)	(0.0128)
Respondent's gender is male	-0.018*	-0.018*	-0.016*	-0.016*
	(0.0090)	(0.0091)	(0.0072)	(0.0072)
Respondent's education is intermediate	0.055***	0.055***	0.058***	0.058***
	(0.0103)	(0.0103)	(0.0100)	(0.0100)
Respondent's education is high	0.090***	0.090***	0.091***	0.091***
	(0.0157)	(0.0158)	(0.0145)	(0.0145)
Log annual hh income in '000 USD	-0.007	-0.007	-0.003	-0.003
	(0.0053)	(0.0053)	(0.0050)	(0.0050)
Respondent has children under 15 yrs	0.004	0.004	0.001	0.001
	(0.0075)	(0.0075)	(0.0078)	(0.0078)
Geocode's log dist from nearest waterbody		-0.002		-0.010
		(0.0071)		(0.0062)
Geocode's log dist from nearest railroad	0.045***	0.046***	0.056***	0.055***
	(0.0112)	(0.0110)	(0.0142)	(0.0141)
Geocode's vegetation index	0.443**	0.394*	0.432***	0.394***
	(0.1655)	(0.1591)	(0.0921)	(0.0908)
Geocode area is urban	-0.202**	-0.189**	-0.208***	-0.201***
	(0.0680)	(0.0694)	(0.0561)	(0.0569)
Respondent's age is 26-60 years	0.010	0.009	0.015	0.015
	(0.0175)	(0.0177)	(0.0134)	(0.0135)
Respondent's age is more than 60 years	0.004	0.000	0.007	0.006
	(0.0255)	(0.0260)	(0.0190)	(0.0191)
Respondent's gender is male	0.017	0.018	0.005	0.007
	(0.0132)	(0.0129)	(0.0084)	(0.0083)
Respondent's education is intermediate	-0.002	-0.001	-0.012	-0.013
	(0.0213)	(0.0202)	(0.0164)	(0.0163)
Respondent's education is high	-0.059**	-0.057*	-0.051*	-0.051*
	(0.0225)	(0.0225)	(0.0227)	(0.0228)
Log annual hh income in '000 USD	-0.003	-0.004	-0.018*	-0.018*
	(0.0148)	(0.0141)	(0.0087)	(0.0087)
Respondent has children under 15 yrs	0.013	0.013	0.019	0.018
	(0.0146)	(0.0141)	(0.0118)	(0.0117)
Geocode's log dist from nearest waterbody		0.040**		0.036
		(0.0157)		(0.0223)
Observations	17964	17964	17964	17964

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Top table reports reduced-form results and bottom reports first-stage results of IV regression using Equation (2.8). The columns correspond to Table 2.28, which reports IV results.

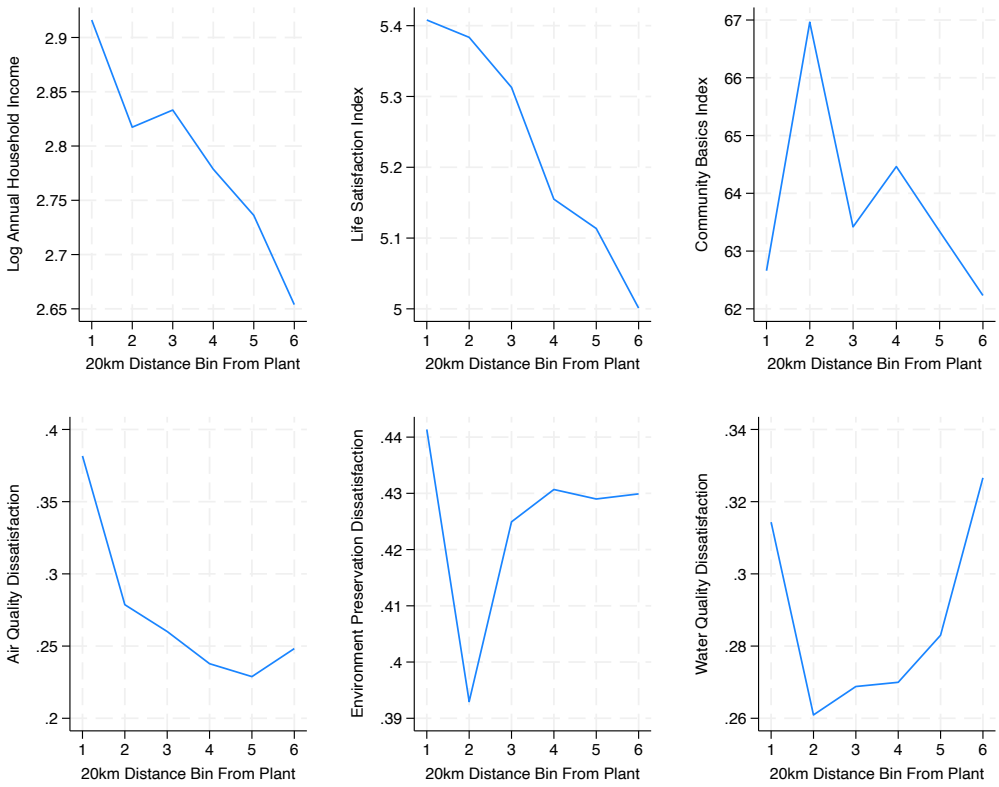
Table 2.30: First-stage and reduced-form results for retired plants

	(1)	(2)	(3)	(4)
Geocode's log dist from nearest railroad	-0.009	-0.009	-0.005	-0.005
	(0.0057)	(0.0055)	(0.0087)	(0.0088)
Geocode's vegetation index	-0.551***	-0.551***	-0.444	-0.450
	(0.1248)	(0.1403)	(0.2403)	(0.2449)
Geocode area is urban	0.064	0.064	0.074	0.074
	(0.0344)	(0.0344)	(0.0568)	(0.0564)
Respondent's age is 26-60 years	-0.005	-0.005	0.010	0.010
	(0.0192)	(0.0193)	(0.0325)	(0.0324)
Respondent's age is more than 60 years	-0.046	-0.046	-0.024	-0.025
	(0.0268)	(0.0265)	(0.0327)	(0.0328)
Respondent's gender is male	-0.028**	-0.028**	-0.030	-0.030
	(0.0105)	(0.0106)	(0.0205)	(0.0205)
Respondent's education is intermediate	0.070**	0.070**	0.074***	0.074***
	(0.0269)	(0.0265)	(0.0219)	(0.0217)
Respondent's education is high	0.078**	0.078**	0.067	0.067
	(0.0270)	(0.0266)	(0.0356)	(0.0349)
Log annual hh income in '000 USD	-0.016*	-0.016*	-0.015	-0.015
	(0.0071)	(0.0074)	(0.0095)	(0.0095)
Respondent has children under 15 yrs	-0.016	-0.016	-0.042	-0.042
	(0.0253)	(0.0253)	(0.0301)	(0.0300)
Geocode's log dist from nearest waterbody		0.000		0.003
		(0.0183)		(0.0158)
Geocode's log dist from nearest railroad	0.153***	0.153***	0.152**	0.149**
	(0.0440)	(0.0438)	(0.0471)	(0.0464)
Geocode's vegetation index	1.623	1.654	2.150**	2.264**
	(0.9679)	(1.0040)	(0.7958)	(0.8063)
Geocode area is urban	-0.432**	-0.432**	-0.365**	-0.370**
	(0.1430)	(0.1422)	(0.1111)	(0.1126)
Respondent's age is 26-60 years	-0.027	-0.025	-0.048	-0.048
	(0.0488)	(0.0505)	(0.0430)	(0.0433)
Respondent's age is more than 60 years	-0.018	-0.014	-0.088	-0.080
	(0.0794)	(0.0853)	(0.0578)	(0.0599)
Respondent's gender is male	0.031	0.032	0.045	0.048
	(0.0470)	(0.0462)	(0.0297)	(0.0296)
Respondent's education is intermediate	-0.044	-0.045	-0.068	-0.071
	(0.0359)	(0.0352)	(0.0517)	(0.0501)
Respondent's education is high	-0.033	-0.035	-0.044	-0.054
	(0.0570)	(0.0545)	(0.0517)	(0.0486)
Log annual hh income in '000 USD	-0.000	-0.001	0.001	0.001
	(0.0313)	(0.0294)	(0.0248)	(0.0246)
Respondent has children under 15 yrs	0.019	0.018	0.055	0.056
	(0.0427)	(0.0437)	(0.0348)	(0.0337)
Geocode's log dist from nearest waterbody		-0.015		-0.061
		(0.0391)		(0.0709)
Observations	2317	2317	2317	2317

Region-clustered robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

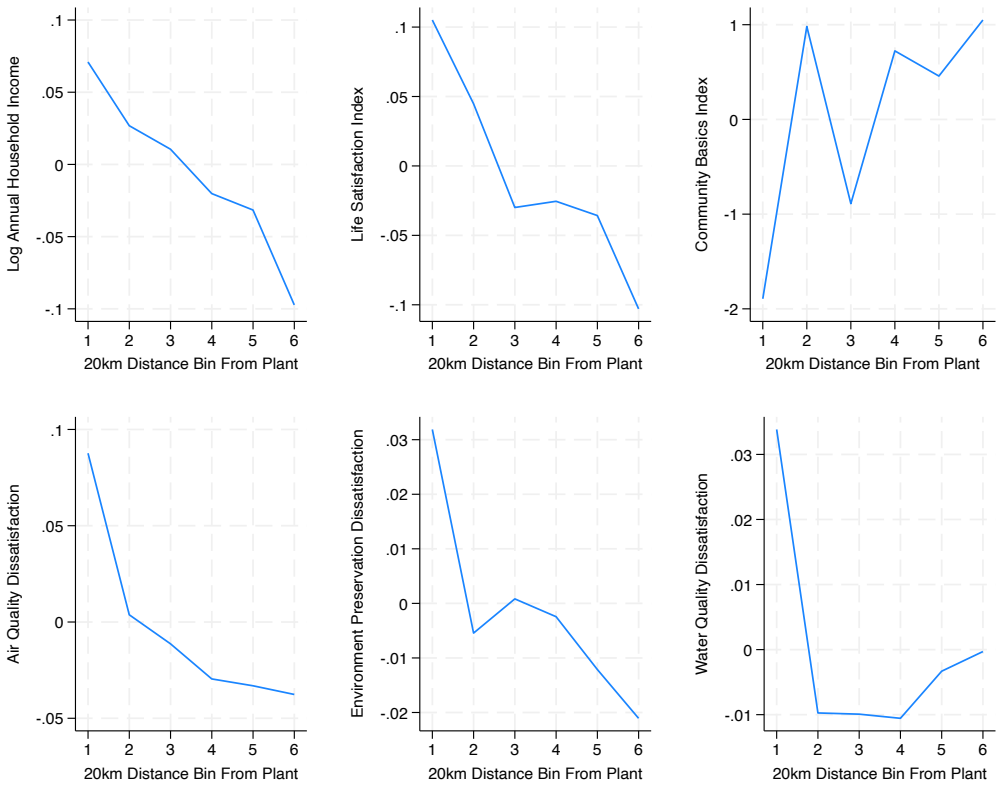
Notes: Top table reports reduced-form results and bottom reports first-stage results of IV regression using Equation (2.8) for retired plants.

Figure 2.18: Descriptive plots - I



Notes: All the variables are taken from the 2019 Gallup World Poll. The label on x-axis should be multiplied by 20 to get the distance bin of the survey location from the nearest coal plant.

Figure 2.19: Descriptive plots - II



Notes: All the variables are taken from the 2019 Gallup World Poll. The label on x-axis should be multiplied by 20 to get the distance bin of the survey location from the nearest coal plant. The estimates on y-axis have been demeaned of country fixed effects.

Chapter 3

Climate Protests, Public Awareness, and Electoral Outcomes

3.1 Introduction

Mass movements are often precursors to major political change. At times, elites concede reforms when faced with revolutionary threats to their authority to govern and extract rents from the population (Acemoglu and Robinson 2006). For example, the extension of the voting franchise in Europe in the 19th century was caused by several revolutions in different parts of Europe (Aidt and Jensen 2014). The classic work of Charles Tilly notes that sustained contentious activity has, on many occasions, led to new policies or even new regimes (Tilly and Tarrow 2015). However, not all such movements lead to favorable outcomes for citizens, especially when considering the medium to long run economic impacts (Collins and Margo 2007). On the other hand, voters could also swing towards pro-reform political parties if the threat of violence and disruption associated with such mass protests is significant (Aidt and Franck 2015). Even though participation in these costly movements could suffer from the free-rider problem and the groups might not be representative of the population, they continue to be a prominent channel to express dissent with leadership and policies (Lohmann 1993). The interest in the anal-

ysis of public demonstration by looking at it as a resource for marginalized groups to exert influence on policymakers is not new (see, for example, Lipsky 1968). However, modern day protests are more sophisticated in terms of their planning and organization due to a range of communication channels and other resources readily available to the protesters. Moreover, conditional on the the size of actual crowd joining a protest, the degree of interaction with the Internet and other mass media channels also decides how impactful a protest would be (McCarthy, McPhail, and Smith 1996). One form of such protests that has seen rapid growth in the 21st century is related to environmental issues, such as pollution and climate change. As depicted in Figure 3.1, climate protests have become more frequent and widespread globally. These protests appear to coincide with the growing public interest in climate change and corresponding media coverage. Specifically, Figure 3.2 demonstrates a clear co-movement between key environmental events and the frequency of Google searches, US news media coverage, and TV airtime focused on climate change. While a correlation is evident, the establishment of a causal link between protest, public attitudes, and policy preferences remains elusive for researchers. Moreover, evidence on the ability of climate protests to engender any form of climate action is particularly limited. Identifying a causal relationship between protests and these outcomes would pave the way for exploring the conditions, forms, and channels through which climate protests might effectively raise public awareness and influence policymakers to take a pro-environment stance. This chapter delves into these questions.

We first examine whether climate protests in a given location increase climate awareness among individuals who live in that location using instances of several thousand protests across the United States. We utilize data on search queries from Google Trends and media coverage from the Global Database of Events, Language, and Tone (GDELT) project to construct measures of climate awareness and activism. Our findings suggest that following a climate-related protest, there is a significant increase in search queries and media attention on climate change-related topics. This finding suggests protests can elevate climate change as a salient public issue and raise awareness. However, the short-

term nature of these outcome variables raises questions about the lasting impact of the heightened awareness and its potential to translate into concrete policy changes.

To address this limitation, we next present evidence suggesting that climate-related protests influence citizens' electoral voting decisions (Fabel et al. 2022). More precisely, in the wake of the widespread *Fridays for Future* protests in March of 2019, voters in Europe showed more support for the local Green parties in their respective regions during the European Parliamentary (EP) elections held in May of 2019. Interestingly, we also observe a rise in support for the radical-right parties, aligning with the notion that protests, while raising awareness and support for climate concerns, can simultaneously polarize opinions (Djourelova et al. 2024). The support could come from individuals who oppose the protesters' methods and demands or those whose daily lives are directly affected by the protests, thereby potentially deepening societal divisions over climate issues.

Finally, we investigate the impact of protests on policymaking. Through textual analysis of speeches by elected UK Members of Parliament (MPs), we find a positive correlation between protest frequency and the intensity of climate-related discussions among MPs. This finding highlights the power of protests to elevate climate issues on the legislative agenda, ensuring they receive due attention from policymakers. By bringing public pressure to the forefront, protests can encourage legislators to address previously neglected environmental concerns.

Protests are coordinated expressions of dissatisfaction with certain elements or phenomena in society, but they are also costly due to economic losses caused by disruptions. Previous research in this vein has found that political protests could be instrumental in reforming policies, but the effects could be conditional on the scope of these protests. For example, Madestam et al. (2013) finds that larger initial Tea Party protests led to greater local movement organization, increased public support for Tea Party positions, higher voter turnout and Republican vote shares, and even made incumbent Congress members vote more conservatively. However, Gethin and Pons (2024) conduct a comprehensive

study of 14 social movements in the U.S. (2017–2022) and find that most protests generated online engagement but produced only modest, often negligible, changes in aggregate public opinion or political attitudes; only exceptionally large movements (e.g. Black Lives Matter) had discernible impacts. We complement these studies by exploiting rich media coverage databases available now to draw more robust inferences. Moreover, our focus is on climate-related protests, which are different from general protests, in the sense that the rank-and-file opinion is still not fully geared towards pro-environment policies (Besley and Hussain 2023). The research is also related to the work on understanding the importance of electronic media in changing social outcomes (Kearney and Levine 2015). In addition, papers highlighting issue attention on climate action across geography and culture suggest that a one-size-fits-all policy is neither optimal nor feasible for tackling climate change (Hase et al. 2021). The findings in this chapter shed some light on this point also by exploiting media attention measures.

Numerous global movements that have sparked political change seem to occur in tandem with protests, yet it remains initially unclear whether these protests merely mirror broader societal unrest or actively contribute to instigating change. In this chapter, we demonstrate the efficacy of climate activism, particularly protests, in eliciting tangible climate action. We examine diverse indicators, including heightened civil society response (reflected in media and internet activity), increased support for pro-environment parties, and even potential backlash in the form of support for the radical-right groups. Additionally, we analyze parliamentary speeches, finding a greater likelihood of legislators addressing climate-related topics after protests.

The remainder of the chapter is organized as follows. In the next section, we discuss the data that we use. In Section 3.3, we establish a robust empirical link between awareness about climate change and protests. The implications on voting and policy discussions are developed in Sections 3.4 and 3.5 respectively. Section 3.6 offers some concluding remarks.

3.2 Data

This section outlines the datasets used in the analysis, which focuses on climate protests within the US and Europe, where climate issues are particularly prominent.

Climate Protests: We utilize several datasets on climate protests, each with its own strengths and limitations:

- *CountLove:* This dataset is a comprehensive record of protests across the United States, providing details such as date, location, cause, and estimated number of attendees for each event. It spans from January 20, 2017, to January 31, 2021, providing a broad temporal window to analyze protest activities and their implications on public engagement with various causes, including climate change.

- *ACLED:* The Armed Conflict Location and Event Data (ACLED) project database covers a wider geographical area, covering both the United States and Europe, from January 2020 to the present. It tracks a variety of political violence and protest events, offering detailed information, such as date, location, participating group, and type of each event. One limitation of this database is that the attendance figures are not reported, thereby inhibiting any analysis using this margin.

- *Fridays for Future (FFF) Protests:* In the latter part of our analysis, particularly when examining the influence of protests on voting behavior, we focus on worldwide climate protests organized under the Fridays for Future movement on March 15, 2019.[1] Covering 131 countries and 2350 cities, it is a rich source of data on climate protests held about two months before the 2019 EP elections voting. The reporting rate on protest intensity (the number of attendees gathered at each protest location), however, is low (less than 38%). We aggregate this protest occurrence

1. A youth-led, global climate-strike movement, Fridays for Future organizes climate strikes across the world to put moral pressure on policymakers to take action against global warming.

data at the EU NUTS 3 region level. Out of 12,955 NUTS 3 regions, 6,967 had at least one protest organized in one of their constituent towns.

Google Trends: We leverage Google Trends, a tool that tracks the popularity of Google search queries across different regions and languages over time. We use the daily and weekly search intensity for the 'climate change' topic[2] to gauge public interest and concerns regarding environmental issues. Using this footprint of the populace's environmental concerns, we can look at how societal interest in this area fluctuates in response to climate protests and broader environmental movements.

GDELT Media Coverage: The Global Database of Events, Language, and Tone (GDELT) is a comprehensive archive, cataloging a vast spectrum of media outputs worldwide to track events, linguistic patterns, and emotional tones across many languages. Leveraging this resource, we craft two specific indicators to assess climate salience, awareness, and activism within the US. First, we develop a measure of the presence of 'climate change' and 'global warming' terms in print media by using the proportion of news articles that address these topics; we call this "coverage". Second, we use television news broadcasts, measuring the percentage of airtime devoted to these issues, with the GDELT data allowing for a precise breakdown into 15-second intervals, we call this "airtime". These measures collectively offer a granular view of media engagement with climate change issues.

Parliamentary Speeches: To analyze the parliamentary discourse on climate change, we utilize the Hansard records, which are the official transcripts of UK parliamentary debates available on the Hansard website[3]. We use the complete set of these records, as collected, harmonized, and compiled by Shamsi (2024). Hansard provides a near-

2. Google Trends offers two distinct search options: topic and keyword. A keyword search targets exact matches within search queries, focusing on the specific word or phrase as it appears. In contrast, a topic search interprets the keyword as a broader concept, capturing related terms, synonyms, and relevant queries that fall under the same general subject.

3. https://hansard.parliament.uk

verbatim account of parliamentary discussions with minor edits for clarity and accuracy. Through textual analysis of these records, we aim to uncover the prominence and evolution of climate change discussions among policymakers.

European Parliamentary Elections Voting: Data on vote shares for different political parties and voter turnout in the 2019 European Parliamentary elections comes from the European NUTS level Election Dataset (EU-NED).[4] This dataset contains reports on the national parliamentary elections in all current EU member-states, the UK, Norway, Turkey, and Switzerland over the period 1990-2020. It also includes coverage of the European parliamentary elections for all the EU member states and the UK. Election results are reported at the lowest level of aggregation (i.e., NUTS 3) wherever possible. Using this data, we construct vote shares obtained by 300 political parties across 1085 NUTS 3 regions.

Precipitation: We obtain the data on precipitation from the ERA5-Land dataset, which is a *gridded* reanalysis product and records hourly precipitation at a spatial resolution of $0.1° \times 0.1°$.[5] To construct the rainfall variable used in the analysis, we compute the average precipitation between 12PM and 4PM local time on the day of the FFF protest i.e., March 15, 2019, two months before the EP election voting dates i.e., May 23-26, 2019. We also construct a long-run average precipitation variable by computing the average monthly precipitation in March through the years 2005 to 2018.

3.3 Awareness and Attitudes

Protest is a potential channel that drives public opinion and attitudes on range of issues, including politics, government performance, and policy (Garcia 2023). Moreover,

4. The European NUTS Level Election Database provides national and European parliamentary election results on the level of Eurostat's NUTS 2 and NUTS 3 administrative units. It is optimized for combination with Eurostat's Regional Database. *Source:* EU-NED.

5. Data can be accessed here.

these effects could be persistent in the long run, thereby impacting future policymaking and electoral outcomes (Mazumder 2018). In this section, we explore the first link in the chain of impact by looking at whether climate protest events are associated with increased online search activity and media coverage related to climate change. We begin by examining potential co-movement at the national level between protest dates and measures of public awareness of climate change. Specifically, we estimate the following econometric relationship for protests in the US:

$$y_t = \alpha + \beta Protest_t + \gamma X_t + \varepsilon_t \tag{3.1}$$

where, y_t is either Google search intensity or media coverage outcome variables in week t. $Protest_t$ is either the number of climate-related protests or the number of attendees in climate-related protests in week t across the US. X_t are controls, which include linear and quadratic time trends, and seasonal effects captured by month-fixed effects. Data on protests is from CountLove. The findings are presented in Table 3.1.

The outcome variable in the first two columns is the Google search intensity for the climate change topic. Columns 3-4 and 5-6 look at the print media coverage and television news airtime respectively of climate change-related news. In terms of magnitude, using estimates from Columns 1 and 2, an additional protest event leads to a 0.02 standard deviation increase in the Google search intensity, while an additional attendee increases it by 0.16 standard deviation. Similarly, using Columns 3, 4, 5 and 6, each additional protest event increases news coverage and TV airtime by 0.01 standard deviation, while an additional attendee increases them by 0.08 and 0.09 standard deviation respectively. The outcome variable for the final four columns assesses the coverage in print media and television news without reference to "protests" or similar events. This distinction aims to determine if increases in news or TV coverage are attributable solely to reporting on protests or if they transcend beyond those news stories. Previous research has found evidence of strong complementarities between protests and media attention when looking at the "success" of a protest (McCarthy, McPhail, and Smith 1996), even when

those protests take place under authoritarian regimes (Tertytchnaya and Lankina 2020). Moreover, it's not just whether protests get covered, but how that coverage unfolds and redirects public conversation leads to an "agenda-setting" (Wasow 2020). The analysis suggests that protests, measured by the total number of protests or attendee counts, are strongly associated with various indicators of climate change awareness. Moreover, even after excluding direct protest coverage news, protests influence mass media coverage, thereby suggesting that protests influence media independently of the mention of events itself.

Additionally, we examine the responsiveness of media to protests broken down by various media outlets. As depicted in Figure 3.3, outlets with a more liberal editorial stance, such as the BBC, MSNBC, and Al Jazeera, appear more inclined to cover these events. These analyses suggest that climate protests lead to increased public awareness of the climate change phenomenon.

However, this relationship is not necessarily causal, as both protests and coverage could be driven by a third variable such as broader political, social, or environmental events (e.g., natural disasters, policy changes) that independently increase both the likelihood of protests and public interest in climate change. Alternatively, it is possible that heightened public awareness and concern about climate change might lead to the organization of climate protests. In this reverse causality scenario, increased media coverage could signal a growing public concern, which motivates activists to stage protests.

To better identify a causal link and address some of these concerns, we refine our analysis by shifting from the national level to a more granular geographical level, specifically the Designated Market Area (DMA) in the US. A DMA is a region where the population has access to the same set of television and radio stations, thereby forming a distinct media market.[6] This approach enables controlling for time fixed effects and

6. Defined by Nielsen, DMAs categorize specific areas where individuals receive identical media content, which is crucial for television advertising and audience measurement. Advertisers and marketers leverage DMAs to tailor their advertising campaigns to specific geographic locales, ensuring that messages reach the designated audience within those areas. The United States comprises over 200 DMAs, ranging from small rural communities to extensive metropolitan areas.

location fixed effects. The time fixed effects account for national-level shocks, such as macroeconomic shocks and federal policy changes that could simultaneously influence both the occurrence of protests and the outcome variables, while the location fixed effects control for time-invariant characteristics specific to a location, such as long-standing political preferences, socio-economic status, and cultural forces in place. By doing so, the analysis facilitates a "within" region and "across" time comparison, thereby addressing some of the potential confoundedness in the previous analysis.

One potential concern with this new two-way fixed effects approach is the presence of transient, unobserved regional shocks, which could simultaneously drive the likelihood of protests in a region and affect the outcome variable. To tackle this issue, we draw from the methodology proposed by Madestam et al. (2013), which utilizes variations in rainfall intensity on protest days as a source of exogenous variation affecting protest attendance. The underlying premise is that, given an expected probability of rainfall, the actual occurrence of rain acts as an external factor that likely diminishes protest attendance.

The CountLove dataset, however, is unsuitable for this type of analysis for two main reasons. First, it lacks attendance data, i.e., the number of people attending the protest, for most protests. Second, it does not provide precise location data, preventing us from accurately determining whether it rained at the protest site on the day of the protest. Despite these limitations, we can still validate the negative correlation between rainfall and attendance, which is the key idea behind this identification strategy, using a subset of the *Fridays for Future* protests that report attendance (Figure 3.4). This correlation suggests that rainfall could serve as a viable instrument for measuring attendance.

Given the limitations of the CountLove dataset, we turn to the ACLED dataset, which, although, does not provide attendance figures, employs a more systematic data collection methodology and includes geolocation information. The geolocation allows us to match protest events with rainfall data. While the absence of attendance data precludes a direct instrumental variables (IV) approach, we can instead examine whether protests that occur on rainy days have a different impact compared to those on non-rainy days. This approach

treats rainfall as a quasi-experimental instrument to induce variation in protest intensity, allowing us to explore its causal effect even without directly observing attendance. This alternative approach maintains the core logic of the original method but limits our ability to precisely interpret the coefficient magnitudes. We estimate the following econometric specification for DMA i in day t:

$$y_{it} = \alpha_i + \eta_t + \beta Protest_{it} + \delta Protest_{it} \times Precipitation_{it} + \varepsilon_{it} \tag{3.2}$$

Results are reported in Table 3.2. The regression analysis is conducted on a daily basis. The dependent variable is akin to the Google search intensity defined in the previous analysis but this time it is constructed at the daily frequency. The treatment variable, $Protest_{it}$, is the number of climate protests during day t in location i. The results suggest a diminished effect of protests on search intensity during increased rainfall, as evidenced by a negative coefficient for the interaction between precipitation and the occurrence of protests. This finding supports the hypothesis that protests significantly affect public interest in climate change-related information.

Our analysis thus far emphasizes the immediate impact of protests on public awareness and attitudes. However, a more critical question remains: do protests ultimately influence policy? On one hand, protests that garner significant public interest and media coverage can exert pressure on policymakers and corporations, potentially leading to meaningful outcomes, such as new legislation, regulatory changes, or corporate commitments (see, for example, Wasow 2020). On the other hand, while protests may raise public engagement with climate-related issues, they do not always translate into concrete actions or policy shifts, particularly when political will is lacking, the protests are not sustained, or they face strong opposition from powerful interests. The effectiveness of protests in driving policy change often hinges on a complex interplay of factors, including the political context, the responsiveness of institutions, and the strategic actions of the protesters.

The next section examines whether protests influence citizens' voting decisions and

policy discussions. This analysis will help determine whether protests can move beyond generating short-term public interest to drive long-term policy changes.

3.4 Election Voting

The literature has argued for the agenda-setting power of protest i.e., parliamentary, governmental, and legislative actions are impacted by the number and size of demonstrations (see, for example, Walgrave and Vliegenthart 2012, a study on 1993-2000 Belgium). However, other studies, such as Gethin and Pons (2024), have found limited impact of protest activities on electoral behavior and outcomes. Along these lines, this section investigates the effect of protests on voting behavior. To conduct this analysis, our previous research design requires some modification. While the hypothesis that rainfall reduces protest attendance is logical and supported by our previous findings, it presents a challenge when examining long-term outcomes or those that are averaged over a longer time frame. While rainfall may affect the likelihood of a specific protest, but over a longer time frame it could also influence the likelihood and timing of subsequent protests. That is, if sudden rain leads to low attendance, organizers may reschedule the protest for a later date. In an extreme case, every rainfall-hampered protest is offset by another one later in time, thereby challenging the use of rainfall as a reliable instrumental variable for protests. To overcome this problem, we concentrate on protests that were pre-announced at the national level and were shortly followed by an election. This approach ensures that rainfall will not affect the probability of future protests, thereby strengthening our results.

In line with this approach, we use the widespread protests organized under FFF across all of Europe. These were a set of international demonstrations to demand action from political leaders to prevent climate change and to phase out subsidies for fossil fuel industries. The largest set of protests on March 15, 2019 gathered over one million protesters in 2,200 strike instances organized in major cities across 125 countries.[7] Conveniently

7. The set of protests is available https://fridaysforfuture.org/what-we-do/strike-statistics/.

for our analysis, the timing of these demonstrations aligned with the 2019 EP elections. This presents a unique opportunity to investigate whether these climate-related protests, which engulfed the whole of Europe, had any tangible effect on the electoral performance of pro-environment political parties. The act of voting is a citizen-driven expression of definitive action with significant implications for both present and future policymaking. Therefore, this offers us an opportunity to estimate the effects of protests on a more substantive indicator of public activism.

Using voting data at the level of NUTS 3 regions, we estimate the effect of climate protests on the vote shares of different political parties. In particular, we estimate the following econometric specification:

$$y_i = \alpha + \beta Protest_i + \delta Protest_i \times Precipitation_i + \gamma X_i + \varepsilon_i \tag{3.3}$$

Results are reported in Table 3.3. In terms of magnitude, an FFF protest on a non-rainy day in a given NUTS3 region increases vote share of Green party by 5.2 percentage points, but an average rainfall (of 2.9 mm) during the 12PM to 4PM time period reduces this gain by 1 percentage point. Now, we do a qualitative analysis of these numbers in more detail. First, the estimate of β aligns with our expectations. Climate protests are more likely in areas with strong Green Party support, as these regions typically share the environmental focus of the protests. They are less common in Conservative or Christian Democrat areas, where priorities may differ or focus on other social issues. Conversely, in liberal regions, the alignment with progressive environmental policies encourages such activism. Interestingly, areas with higher radical-right support might also see more climate protests, plausibly as expressions of retaliation to the radical right's strong stance against pro-environment policies.

However, the focal point of our analysis is the interaction coefficient δ, which shows a negative value for both Green and radical-right parties. This implies that in regions with less rain, where protests are likely to have higher attendance, support for these groups tends to increase. This trend might be explained by the direct emphasis on environmental

issues in such protests, which aligns with and bolsters support for Green parties. Conversely, increased backing for radical right parties could be due to a perception of these protests as disruptive and a challenge to the social order and traditional values, resonating with the radical right's focus on stability and nationalism (Ketchley and El-Rayyes 2021). This kind of polarizing effect of demonstrations has been noted in other contexts also. For example, *March for Science* rallies in the United States polarized the support for scientists and their research, where liberals' attitudes became more positive and conservatives' attitudes became more negative after the rallies (Motta 2018). Radical right parties leverage these protests to highlight issues of national sovereignty and traditional values, aligning with their agenda and possibly increasing their vote share. This interplay illustrates the dual impact of environmental activism: it can galvanize support for pro-environment policy but simultaneously polarize other voters, something that is also highlighted in Djourelova et al. (2024). In addition, the data indicates that protests also increase overall voter turnout.

These regressions also control for the probability of rainfall in NUTS 3 regions. This is to exploit weather variation across counties with similar baseline likelihoods of rainfall on the protest day. We can control for the rainfall probability flexibly by including dummy variables corresponding to the deciles in the historical rainfall probability distribution. Employing rainfall percentiles as a measure instead of absolute rainfall in millimeters yields results that are qualitatively consistent with our primary findings. Results are reported in Table 3.5 in the Appendix. One potential concern might be that the results are driven by compositional changes across parties i.e., due to different NUTS3 regions used in each analysis. To address this concern, we rerun the analysis with only those NUTS3 regions that consists of candidates from all the three major parties: Green, Radical Left, and Radical Right. Results reported in Table 3.6 in the Appendix show that the findings hold even in this balanced case.[8]

8. The number of observations across the columns in Table 3.6 varies due to multiple candidates from the same party type in each NUTS3 region. We address this concern by using the average vote share across all candidates for each party type. Results reported in Table 3.7 suggest that the results are robust to this

The findings thus far suggest that climate protests have the potential to generate public interest on climate issues. In the long run, protests affect the vote shares of political parties, particularly increasing those of the Green parties. Nevertheless, protest advocates often theorize that protests can also directly pressure policymakers – an aspect we are yet to examine. Demonstrating how protests can directly impact the supply of politics would provide a more comprehensive understanding of their impact. We explore this in the following section, while acknowledging that the relationship between voters and policymakers is inherently intertwined: voters influence policymakers through electoral pressure, while policymakers shape voter preferences through their policy choices and the options they present.

3.5 Parliamentary Discussions

In the previous section, we studied the impact of climate protests on electoral behavior. However, even if these protests could aid pro-environment parties to win elections, it is non-trivial that environment-friendly policies would follow through, a key objective of such protests. Research has documented instances where such activities have won support in the legislature (see, for example, Walgrave and Vliegenthart 2012; Gillion 2012) and facilitated institutionalized grievance redressal (Robertson and Teitelbaum 2011). Berry and Sobieraj (2014) note that U.S. Congress members often reference high-profile protests, such as climate marches, in speeches to signal responsiveness and Porta (2013) report evidence of European anti-austerity protests leading to policy debates. To this end, in this section, we explore the relationship between the occurrence of protests in a constituency and the degree to which its MP discusses climate-related issues in the legislature. To implement this, we leverage textual analysis on the speeches made by MPs within the UK Parliament by creating two indicators for each constituency that shed light on the nuances of political rhetoric circulating in policy circles regarding climate change

change.

issues.

Discussion density: This indicator measures the frequency of climate-related keywords in MPs' parliamentary speeches, capturing the focus on environmental issues. Keywords and bigrams are selected to cover a wide range of climate terms [9]. Their occurrences are counted and normalized against the total word count of the MP's annual speeches, creating a standardized frequency measure. This metric objectively assesses MPs' emphasis on climate topics in their legislative discussions, acting as a gauge for thematic focus.

Valence measure: This metric evaluates the sentiment in parliamentary discourse by analyzing the context around relevant keywords and bigrams in MPs' speeches. Using the `NLTK` library's `SentimentIntensityAnalyzer`, each identified segment (10 words before and after a keyword or a bigram) is scored for sentiment, ranging from -1 (negative) to +1 (positive). This approach provides a nuanced understanding of the emotional and evaluative tones in parliamentary discussions on climate, offering an aggregate sentiment score that reflects MPs' attitudes towards climate issues.

Upon constructing these indicators, we proceed to perform regression analysis, employing these metrics as dependent variables against the annual count of protests in the constituency, with a focus on data post 2019, the period for which protest data is available. We also incorporate a control for the baseline average discussion density measure. This adjustment is made to account for the preexisting levels of awareness and interest in climate-related issues within a constituency, acknowledging that such a baseline could influence both the occurrence of protests and the frequency of parliamentary discussions on environmental issues (for instance, by leading to the election of MPs with a stronger environmental agenda). We estimate the following econometric specification:

$$y_i = \alpha + \beta \cdot \text{Protest count}_i + \gamma X_i + \varepsilon_i \tag{3.4}$$

9. Keywords include environment, climate change, global warming, biodiversity, carbon footprint, sustainability, greenhouse effect, carbon emissions, climate policy, fossil fuels, energy efficiency, renewable energy, carbon neutral, and paris agreement.

Results are reported in Table 3.4. Column 1 suggests a strong positive impact of protests on the discussion density of climate-related issues. The subsequent Columns 2-4 report disaggregated results by the MPs' party affiliation, and suggest that Labor MPs exhibit the highest level of responsiveness towards protests. Columns 5-8 extend this examination to the valence measure, analyzing the emotional and evaluative tone of MPs' discussions on climate issues. A clear pattern emerges here, suggesting a positive shift in the sentiment surrounding climate discussions in correlation with increased protest activity, with Labor MPs again showing a more pronounced reaction compared to others.

These results suggest that protests can serve as a vital mechanism for elevating climate issues on the political agenda, particularly within parties and regions more predisposed to environmental activism. Such findings underscore the potential of grassroots activism to shape political discourse and action on climate change, influencing political agenda and priorities. The differential responsiveness highlights the importance of understanding party-specific dynamics when assessing responsiveness within this party to public demands for action on environmental issues.

However, these findings should be interpreted with caution, as the analysis lacks a robust method to isolate exogenous variation in the frequency of protests at the constituency level. In this context, the use of exogenous variation, such as rainfall shocks, is not feasible. While rainfall may influence immediate protest turnout, its broader impact on protest dynamics and the number of protests over an extended period is less certain and could introduce confounding variables into the analysis.

3.6 Conclusion

The frequency of protests against climate change and human-induced environmental damages has been rising in both the US and Europe. However, it is far from clear whether this form of climate activism could induce or aid pro-environmental policy reforms. In this chapter, using Google Trends search intensity and GDELT media coverage measures,

we first document that protests generate significant public engagement and media attention in the short-run. Then, to look at more long-run effects, we leverage the exogenous variation induced by rainfall shocks around Fridays for Future strikes and see their impact on the vote shares of different political parties in the 2019 European Parliamentary elections. We find that vote shares of Green parties in different NUTS 3 regions saw a significant increase following the strikes. Furthermore, we provide suggestive evidence that these protests influence policy discussions at the constituency level in the UK Parliament, as reflected in the content of speeches by Members of Parliament.

The findings highlight the potent role of climate protests in not only increasing public awareness and media discourse on climate issues, but also influencing tangible political outcomes. The observed increase in Green Party vote shares following the Fridays for Future strikes indicates a shift in voter preferences toward environmental priorities, driven by grassroots activism. However, the concurrent rise in support for radical right parties suggests that these movements may also provoke a backlash among certain segments of the electorate, highlighting the importance for organizers to consider strategies that mitigate potential counterproductive effects. Additionally, the observable shift in parliamentary discussions towards more climate-focused narratives suggests that the echoes of the streets are reaching the halls of power, potentially paving the way for more robust climate policies.

These results contribute to a better understanding of the relationship between public mobilization, grassroots activism, and policy formation, emphasizing the role of civic engagement in addressing global climate challenges.

References

Acemoglu, Daron, and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy.* Cambridge University Press.

Aidt, Toke S., and Raphaël Franck. 2015. "Democratization Under the Threat of Revolution: Evidence From the Great Reform Act of 1832." *Econometrica* 83 (2): 505–547.

Aidt, Toke S., and Peter S. Jensen. 2014. "Workers of the world, unite! Franchise extensions and the threat of revolution in Europe, 1820–1938." *European Economic Review* 72:52–75.

Berry, Jeffrey M., and Sarah Sobieraj. 2014. *The Outrage Industry: Political Opinion Media and the New Incivility.* Oxford University Press.

Besley, Tim, and Azhar Hussain. 2023. "Global Gains from a Green Energy Transition: Evidence on Coal-Fired Power and Air Quality Dissatisfaction." *CEPR Discussion Paper* 18046.

Collins, William, and Robert Margo. 2007. "The Economic Aftermath of the 1960s Riots in American Cities: Evidence from Property Values." *The Journal of Economic History* 67 (4): 849–883.

Djourelova, Milena, Ruben Durante, Elliot Motte, and Eleonora Patacchini. 2024. "Experience, narratives, and climate change beliefs." *SSRN Electron. J.*

Fabel, Marc, Matthias Flückiger, Markus Ludwig, Helmut Rainer, Maria Waldinger, and Sebastian Wichert. 2022. *The Power of Youth: Did the "Fridays for Future" Climate Movement Trickle-Up to Influence, Voters, Politicians, and the Media?* CESifo Working Paper Series 9742. CESifo.

Garcia, Carla. 2023. "The Effects of Protest on Public Opinion: Evidence from Colombia." *MIT Political Science Department Research Paper* 2023-3.

Gethin, Amory, and Vincent Pons. 2024. *Social Movements and Public Opinion in the United States.* Working Paper, Working Paper Series 32342. National Bureau of Economic Research.

Gillion, Daniel Q. 2012. "Protest and Congressional Behavior: Assessing Racial and Ethnic Minority Protests in the District." *The Journal of Politics* 74 (4): 950–962.

Hase, Valerie, Daniela Mahl, Mike S. Schäfer, and Tobias R. Keller. 2021. "Climate change in news media across the globe: An automated analysis of issue attention and themes in climate change coverage in 10 countries (2006–2018)." *Global Environmental Change* 70 (102353).

Kearney, Melissa S, and Phillip B Levine. 2015. "Media influences on social outcomes: The impact of MTV's 16 and pregnant on teen childbearing." *American Economic Review* 105 (12): 3597–3632.

Ketchley, Neil, and Thoraya El-Rayyes. 2021. "Unpopular Protest: Mass Mobilization and Attitudes to Democracy in Post-Mubarak Egypt." *The Journal of Politics* 83 (1).

Lipsky, Michael. 1968. "Protest as a Political Resource." *American Political Science Review* 62 (4): 1144–1158.

Lohmann, Susanne. 1993. "A Signaling Model of Informative and Manipulative Political Action." *The American Political Science Review* 87 (2): 319–333.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. 2013. "Do political protests matter? evidence from the tea party movement." *The Quarterly Journal of Economics* 128 (4): 1633–1685.

Mazumder, Soumyajit. 2018. "The Persistent Effect of U.S. Civil Rights Protests on Political Attitudes." *American Journal of Political Science* 62 (4): 922–935.

McCarthy, John David, Clark McPhail, and Jackie Smith. 1996. "Images of protest: Dimensions of selection bias in media coverage of Washington demonstrations, 1982 and 1991." *American Sociological Review* 61 (3): 478–499.

Motta, Matthew. 2018. "The Polarizing Effect of the March for Science on Attitudes toward Scientists." *Political Science & Politics* 51 (4).

Porta, Donatella Della. 2013. *Clandestine Political Violence.* Cambridge University Press.

Robertson, Graeme B., and Emmanuel Teitelbaum. 2011. "Foreign Direct Investment, Regime Type, and Labor Protest in Developing Countries." *American Journal of Political Science* 55 (3): 665–677.

Shamsi, Javad. 2024. "Immigration and political realignment." *Available at SSRN 4748438.*

Tertytchnaya, Katerina, and Tomila Lankina. 2020. "Electoral Protests and Political Attitudes under Electoral Authoritarianism Katerina Tertytchnaya and Tomila Lankina." *The Journal of Politics* 82 (1).

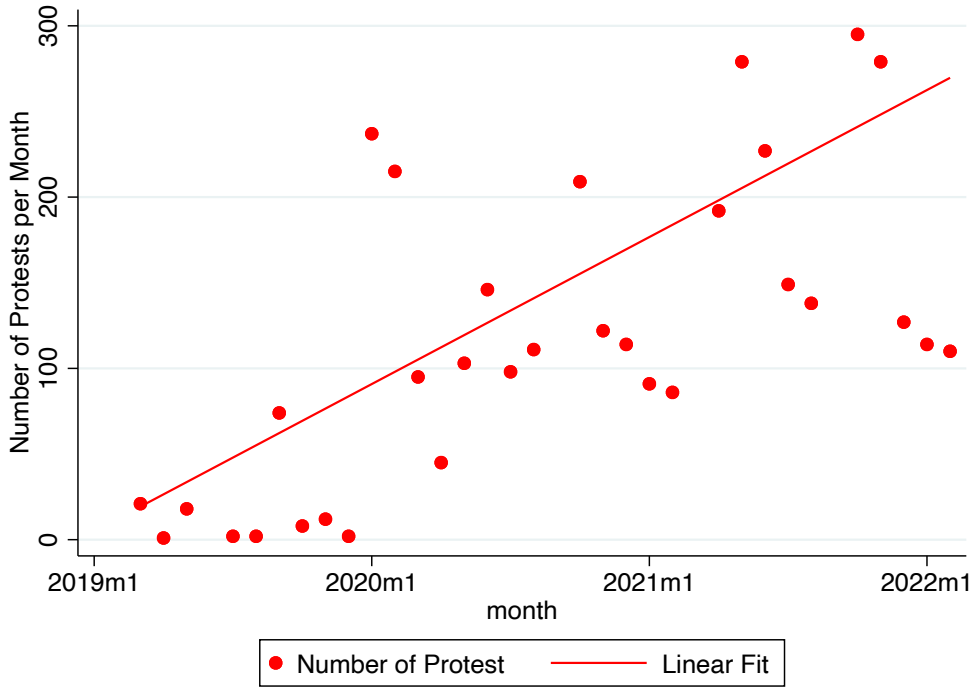
Tilly, Charles, and Sidney Tarrow. 2015. *Contentious Politics.* Oxford University Press.

Walgrave, Stefaan, and Rens Vliegenthart. 2012. "The Complex Agenda-Setting Power of Protest: Demonstrations, Media, Parliament, Government, and Legislation in Belgium, 1993-2000." *Mobilization: An International Quarterly* 17 (2): 129–156.

Wasow, Omar. 2020. "Agenda Seeding: How 1960s Black Protests Moved Elites, Public Opinion and Voting." *American Political Science Review* 114 (3): 638–659.

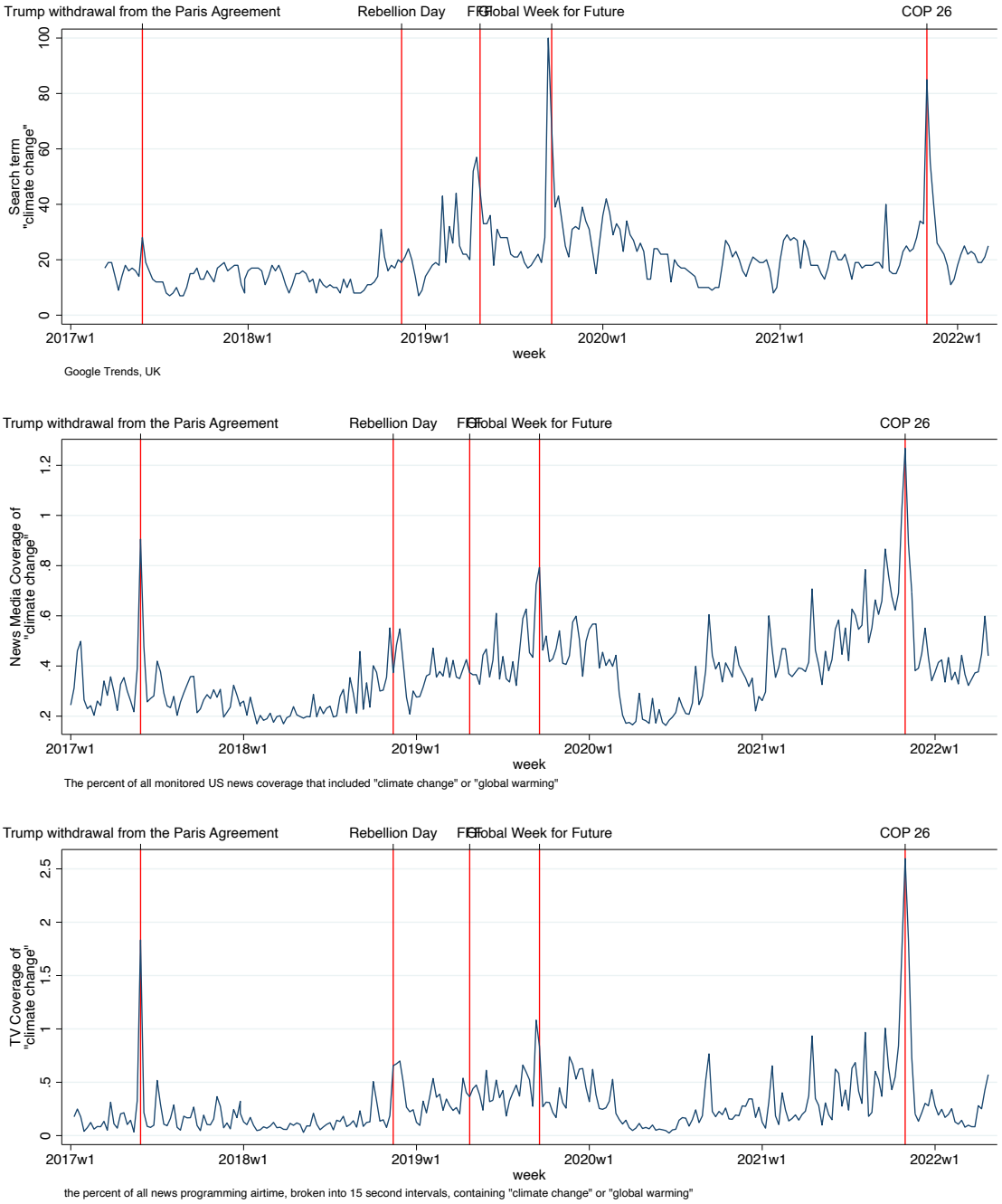
Main Figures and Tables

Figure 3.1: Monthly Count of Protests in Europe and North America



Notes: This graph, derived from ACLED dataset, displays the count of protests per month. Each point represents the total number of protests in a given month, with those exceeding 300 protests omitted for clarity. The scatter plot points depict monthly protest frequencies, while the line illustrates the linear fit, indicating the overall trend.

Figure 3.2: Climate Change Engagement in Google Trends, Print Media, and TV



Notes: Panel A illustrates trends in Google search intensity for the term 'climate change' in the UK, extracted from Google Trends. Key events, such as Trump's withdrawal from the Paris Agreement and COP 26, are marked, indicating their potential influence on public interest. Panel B focuses on news media coverage in the US, showing the proportion of news items featuring 'climate change' or 'global warming', with data sourced from the GDELT project. Panel C explores television news coverage, showing the percentage of airtime allocated to discussing these issues, also based on the GDELT data.

Table 3.1: US national-level weekly analysis

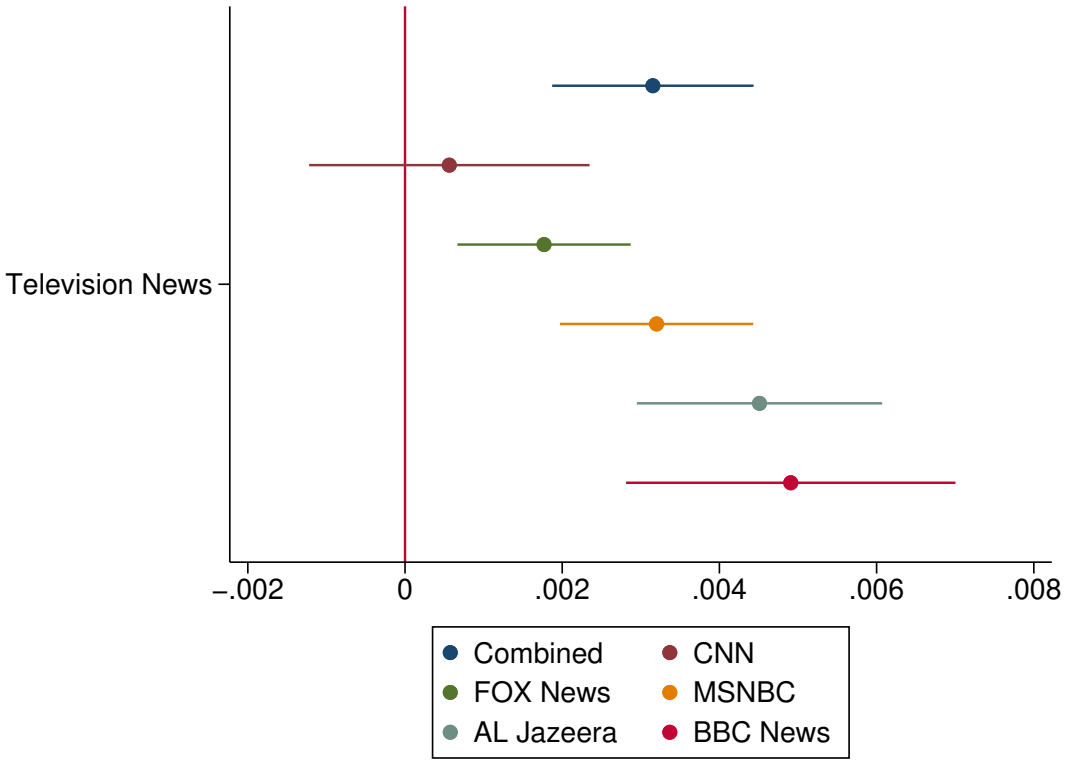
	Search Intensity		News Climate		TV Climate		News Climate Exc. Protest		TV Climate Exc. Protest	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
No. Protests	.022***		.013***		.015***		.0083***		.013***	
	(.0025)		(.003)		(.003)		(.0031)		(.0031)	
No. Attendees		.16***		.076***		.094***		.04		.08***
		(.023)		(.027)		(.027)		(.027)		(.027)
N	199	199	212	212	212	212	212	212	212	212
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	.487	.423	.236	.196	.23	.189	.212	.193	.215	.178

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents findings from eleven regression analyses utilizing weekly data from CountLove to examine the impact of protests on various indicators of public engagement with climate change. The regressions correlate national-level metrics for a given week with the quantity of protests or participants during that week. Models 1-2 analyze Google Trends search intensity for 'climate change', Models 3-4 and 7-8 examine the percentage of news coverage on 'climate change' and 'climate change excluding protests' respectively, while Models 5-6 and 9-10 focus on the same metrics in TV coverage. The dependent variables in columns 3 to 10 are sourced from GDELT. All dependent variables are standardized to have a mean of zero and a variance of one. The primary independent variables are the number of protests and attendees. The first column sources data from the ACLED and encompasses the period from January 1, 2020, to May 1, 2022, while subsequent columns draw on data from CountLove and cover from April 1, 2017, to January 31, 2021. Each model controls for linear and quadratic time trends, and seasonal effects captured by month fixed effects.

Figure 3.3: Protest coverage across different TV stations



Notes: This graph illustrates the estimated coefficients for protest coverage on various TV news networks, each represented by a distinct regression model. These are Combined (aggregate of all stations), CNN, FOX News, MSNBC, Al Jazeera, and BBC News. The coefficient for each network, depicted on the Y-axis, measures the extent to which protests influenced TV coverage of climate change.

Table 3.2: US DMA-level daily analysis

	(1) Search Intensity	(2) Search Intensity	(3) Search Intensity	(4) Search Intensity
Protest × Precip	-4.5081***	-2.6866***	-2.3467***	-0.9595*
	(1.6089)	(0.5286)	(0.4322)	(0.5600)
Protest	26.3063***	1.8394	0.8064	0.5046
	(1.0830)	(1.6634)	(1.4304)	(1.6707)
DMA FE	No	Yes	Yes	Yes
Date FE	No	No	Yes	Yes
Linear Time Trend	No	No	No	Yes
Observations	163876	163876	163876	163876

Robust standard errors clustered at the DMA level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table displays regression results on the influence of protests and weather on Google Trends search intensity, using data from Google Trends combined with ACLED post-January 1, 2020. It introduces an interaction term for protests in DMAs and rainfall, differentiating the impact of the protest between rainy and dry weather conditions. We compute the average rainfall between 12PM and 4PM on the day of the protest to construct the precipitation variable. In cases of multiple protests in the same DMA on the same day, the aggregate number of protests is used. The analysis is conducted at a daily frequency.

Table 3.3: Protest and vote shares in EP elections

	Green Party	Conservative	Socialist	Agrarian/Centre	Christian Democrats
Protest × Precip	-0.357***	0.502	-0.144	1.274	-0.145
	(0.125)	(0.413)	(0.134)	(1.032)	(0.292)
Protest	5.208***	-2.128**	1.164	-1.028	-3.209**
	(0.635)	(1.080)	(0.805)	(2.168)	(1.427)
Observations	957	906	1314	113	1067
Mean	6.100	8.854	14.25	8.135	14.92
Standard Deviation	6.989	9.468	10.14	8.652	15.64
	Liberal	Radical Left	Radical Right	Regionalist	Voter Turnout
Protest × Precip	0.161	0.155**	-0.881***	0.0433	-0.744***
	(0.133)	(0.0631)	(0.165)	(0.147)	(0.189)
Protest	2.605***	-0.654	6.243***	-0.738	4.366***
	(0.755)	(0.423)	(1.024)	(1.048)	(1.001)
Observations	1186	1508	1279	680	1064
Mean	9.352	4.676	14.97	3.396	52.57
Standard Deviation	8.906	4.866	14.18	7.438	11.51

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents OLS regression results from Europe at the NUTS3 level. The first explanatory variable is an interaction between protest occurrence and average precipitation between 12 noon and 4 p.m. on the day of the protest, i.e., March 15, 2019. Protest is an indicator variable that takes a value of 1 if a Fridays for Future protest was held in the NUTS3 region prior to the EP elections. Different political parties in each country are categorized into party families on the basis of their ideology using data from the Chapel Hill Expert Survey. Controls include the share of population having tertiary education (at the NUTS2 level) and the long-run average precipitation in the month of March, calculated using precipitation data in the years 2005-2018 (at the NUTS3 level).

Table 3.4: Hansard textual analysis

	Discussion Density				Valence Measure			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Protests	0.0111**	0.0204*	0.0035	0.0262	0.0116***	0.0146**	0.0106**	0.0188**
	(0.0028)	(0.0064)	(0.0026)	(0.0192)	(0.0009)	(0.0042)	(0.0016)	(0.0047)
Constituencies	All	Labour	Conservative	Other	All	Labour	Conservative	Other
Observations	753	253	368	132	753	253	368	132

Robust standard errors in parentheses

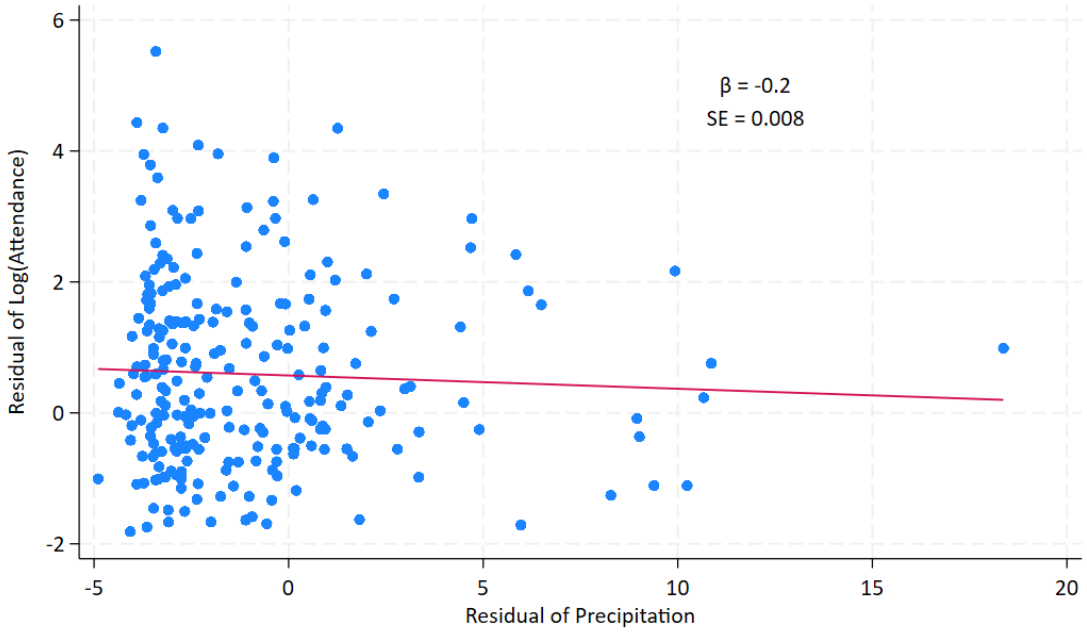
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table was generated using data from the Hansard dataset and the ACLED protest data. The Hansard dataset provides counts of Member of Parliament (MP) mentions of "climate change" or similar words in parliamentary records, while the ACLED dataset offers information on protest events in the UK. The table presents regression results examining the relationship between the number of protests in a constituency and the frequency of MP mentions post-2019 in parliamentary records. The analysis considers different models, including controls for pre-2019 mentions and separate analyses for Conservative and Labor MPs. All regressions control for the level of frequency of mentions of "climate change" and related keywords in the constituency before 2019. Robust standard errors are clustered at the country level.

Appendix

3.7 Appendix: Figures and Tables

Figure 3.4: Attendance and precipitation on the day of protest



Notes: The residuals on y-axis are generated by regressing log of attendance at the protests on March 15, 2019 (as reported on the *Fridays for Future* website) on the long-run average precipitation in each NUTS3 region. Similarly, the residuals on the x-axis are generated by regressing precipitation between 12PM and 4PM on the day of the protest on the long-run average precipitation in each NUTS3 region (both obtained from the ERA5-Land dataset). Since the reporting on attendance is incomplete and unreliable, we also restrict the specification to include protests attended by at least 75 people, i.e., protests large enough to gather some attention in the media. Results are qualitatively robust to changing this threshold up and down.

Table 3.5: Protest and vote shares in EP elections with rainfall percentiles

	Green Party	Conservative	Socialist	Agrarian/Centre	Christian Democrats
Protest × Precip (percentile)	0.0457**	-0.0736**	0.0175	-0.138	0.0155
	(0.0148)	(0.0274)	(0.0205)	(0.0901)	(0.0398)
Protest	1.455*	3.018	-0.438	9.570	-4.293*
	(0.593)	(1.603)	(0.946)	(6.389)	(1.903)
Observations	957	906	1314	113	1067
Mean	6.100	8.854	14.25	8.135	14.92
Standard Deviation	6.989	9.468	10.14	8.652	15.64
	Liberal	Radical Left	Radical Right	Regionalist	Voter Turnout
Protest × Precip (percentile)	-0.0218	-0.0300**	0.178***	-0.00747	0.112***
	(0.0176)	(0.00973)	(0.0256)	(0.0208)	(0.0238)
Protest	4.098***	0.949**	5.069***	-0.232	3.004**
	(0.780)	(0.349)	(1.440)	(0.565)	(1.149)
Observations	1186	1508	1279	680	1064
Mean	9.352	4.676	14.97	3.396	52.57
Standard Deviation	8.906	4.866	14.18	7.438	11.51

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table is similar to Table 3.3 but uses inverse percentiles of precipitation instead of continuous values. The first explanatory variable is an interaction between protest occurrence and the inverse of precipitation percentiles between 12 noon and 4 p.m. on the day of the protest, i.e., March 15, 2019. Protest is an indicator variable that takes a value of 1 if an FFF protest was held in the NUTS3 region prior to the EP elections. Different political parties in each country are categorized into party families on the basis of their ideology using data from the Chapel Hill Expert Survey. Controls include the share of population having tertiary education (at the NUTS2 level) and the long-run average precipitation in the month of March, calculated using precipitation data in the years 2005-2018 (at the NUTS3 level).

Table 3.6: Protest and vote shares in EP elections with same party composition

	Green Party	Radical Left	Radical Right	Turnout
Protest × Precip	-0.389***	0.270***	-0.428**	-0.700***
	(0.106)	(0.0625)	(0.179)	(0.183)
Protest	4.972***	-1.765***	5.006***	3.499***
	(0.642)	(0.437)	(1.130)	(1.008)
Observations	677	1312	929	677
Mean	4.872	4.490	12.65	55.82
Standard Deviation	5.723	4.815	10.96	8.633

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table is similar to Table 3.3 but looks at the subset of NUTS3 regions that have a Green party, a Radical Left party, and a Radical Right party. The first explanatory variable is an interaction between protest occurrence and the inverse of precipitation percentiles between 12 noon and 4 p.m. on the day of the protest, i.e., March 15, 2019. Protest is an indicator variable that takes a value of 1 if an FFF protest was held in the NUTS3 region prior to the EP elections. Different political parties in each country are categorized into party families on the basis of their ideology using data from the Chapel Hill Expert Survey. Controls include the share of population having tertiary education (at the NUTS2 level) and the long-run average precipitation in the month of March, calculated using precipitation data in the years 2005-2018 (at the NUTS3 level).

Table 3.7: Protest and vote shares in EP elections with full balanced sample

	Green Party	Radical Left	Radical Right	Turnout
Protest × Precip	-0.389***	0.260***	-0.567***	-0.700***
	(0.106)	(0.0530)	(0.131)	(0.183)
Protest	4.972***	-1.366***	5.083***	3.499***
	(0.642)	(0.316)	(0.947)	(1.008)
Observations	677	677	677	677
Mean	4.872	4.237	12.82	55.82
Standard Deviation	5.723	2.535	7.619	8.633

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table is similar to Table 3.6 but takes the average of the vote shares of Radical Left and Radical Right parties if more than one contested the election from that NUTS3 region. The first explanatory variable is an interaction between protest occurrence and the inverse of precipitation percentiles between 12 noon and 4 p.m. on the day of the protest, i.e., March 15, 2019. Protest is an indicator variable that takes a value of 1 if an FFF protest was held in the local authority prior to the EP elections. Different political parties in each country are categorized into party families on the basis of their ideology using data from the Chapel Hill Expert Survey. Controls include the share of population having tertiary education (at the NUTS2 level) and the long-run average precipitation in the month of March, calculated using precipitation data in the years 2005-2018 (at the NUTS3 level).