The effect of air pollution on GDP: evidence from a natural experiment in India

Anshuman Tiwari*

03 May, 2025

Abstract

Estimating the effect of air pollution on aggregate economic output is challenging because pollution increases with GDP, biasing estimates upward. Standard approaches like differencing with district and year fixed effects reduce but do not eliminate this bias, necessitating a credible instrument. I leverage a natural experiment in India created by groundwater-conservation mandates in two northern states, which shifted crop-residue fires from October into November, when cool air and calm winds trap particulate matter. Satellite fire detections and wind trajectories show that a 10% increase in exposure to upwind November fires raises annual district PM2.5 by 0.3%, while October fires have no impact. Using this exogenous variation as an instrument in first-differenced district panels, I estimate a 1% rise in PM2.5 reduces real GDP by 0.18%, highlighting substantial economic damage from pollution in developing countries.

^{*} Energy Policy Institute at Chicago - India (email: atiwari2@uchicago.edu). All errors are mine. Financial support from the Environmental Defense Fund, the Grantham Research Foundation and LSE STICERD is gratefully acknowlegded.

1 Introduction

The literature documents wide-ranging impacts of particulate matter less than 2.5 microns in diameter (PM2.5), including on human health and mortality (Schlenker and Walker 2016; Deryugina et al. 2019), worker and firm productivity (Graff Zivin and Neidell 2012; Chang et al. 2019; Fu et al. 2021; Adhvaryu et al. 2022; Borgschulte et al. 2022a; Leroutier and Ollivier 2025), among others. But evidence for the effect of pollution on aggregate economic output as measured through the Gross Domestic Product (GDP) is surprisingly rare, with Dechezleprêtre et al. (2019) being an exception. One reason for this lack of evidence may be that solving the endogeneity problem of positive correlation between growth in GDP and pollution is difficult. This paper develops a novel instrument that leverages variation in PM2.5 driven by exogenous policy changes in upwind crop burning and mediated by meteorological conditions to quantify the consequences for economic growth.

India has the highest average PM2.5 concentrations in the world at 7 times the WHO standards (Greenstone 2021). It is also home to almost 20% of the world's population. Therefore, it is an important context to study this problem. It is also a rapidly growing economy, and while PM2.5 concentrations are a consequence of that growth, they may also pose a threat to it. I analyze the effect of PM2.5 levels on GDP using newly available panel data for 530 Indian districts between 2007 and 2013. In order to account for the non-stationary nature of the GDP data¹, I utilize district-specific time trends as well as a first differences approach. The latter performs better with strongly non-stationary data series and is commonly used in macroeconomic analysis of GDP data (Wooldridge 2010). Identification of the causal effect of PM2.5 on district GDP relies on yearly deviations in PM2.5 being plausibly exogenous, controlling for year and district fixed effects as well as time trends. However, causality may yet run from GDP to PM2.5, with larger yearly deviations in pollution systematically being a result of higher economic growth in that district in the given year. I develop a novel solution to this endogeneity concern by leveraging an exogenous groundwater policy change that inadvertently increased downwind PM2.5 without directly affecting GDP growth in those downwind districts.

Groundwater depletion due to over-exploitation of aquifers is a well-documented and pressing problem in India, particularly in the two northern states of Punjab and Haryana (World Bank 2021). These states had long encouraged farmers to extract groundwater at close to zero marginal price for the irrigation of the water-intensive rice crop.² Farmers in these states had

¹ Average yearly district GDP growth rate in this period was 7%.

 $^{^2}$ Rice is synonymous with paddy, with interchangeable references to paddy or rice fields.

been able to grow sufficient foodgrains, and the country had avoided regular famine events as a result (Pingali 2012). But, by the early 2000s, these state governments had realized the need to protect groundwater resources in the interest of long-term local growth. Rather than incentivizing a shift away from the rice crop or instituting a marginal price for groundwater³, these states decided to force farmers to push back the dates of rice transplantation⁴ from mid-May to mid-June, hoping that the arrival of the monsoon by late June would reduce dependence on groundwater to grow rice.

Farmers in Punjab and Haryana had followed the practice of setting fire to their fields after the rice harvest in October in order to prepare the same fields for planting the staple wheat crop. These fires clear out leftover residue after rice harvest that come in the way of planting wheat seeds, and had become popular as the cheapest method to get rid of this residue. Although agricultural fires are nominally illegal, enforcement is rare, with an average expected fine in Haryana of 0.75 USD while the average marginal cost to clear the field without fires is 50 USD (Behrer 2019; Lohan et al. 2018). By forcing a shift in the transplantation dates to early June in order to conserve groundwater, these states also shifted rice cultivation dates and the incidence of burning from October into November. Any delays in the planting of the wheat crop can reduce yields (McDonald et al. 2022), so many farmers also had a stronger incentive to utilize fires to clear fields.

I utilize a two-way fixed effects design with information on the timing of the groundwater laws to document that the laws increased November fire count by 54% and the measure of biomass burnt by 72% in districts of Punjab and Haryana. Small anticipation effects imply that these may be slight underestimates. At the same time, October fire count and measure of biomass burnt decreased by 58% and 57% respectively. Since winter fire activity is predominantly concentrated in these two months, these results document the shift in monthly fire patterns toward early winter. Therefore this policy change had the unintended consequence of shifting the peak of agricultural fires from October into November, when the onset of winter brings lower wind speeds and temperatures that slow the dispersion of particulate matter (Vallero 2014).

Next, I develop a novel method to capture the effect of this increase in November fire activity on annual PM2.5 levels. I construct an origin-destination-by-day specific measure of how exposed any given district is to fires in upwind districts on any given day. I calculate this

 $^{^{3}}$ The marginal price of groundwater extraction continues to be close to zero, since electricity for pumps is subsidized with flat tariffs rather than marginal pricing, and any outstanding farm electricity bills are rarely paid back to state distribution companies.

⁴ This process of moving seedlings grown in nurseries into fields reduces weed removal and produces higher yields. More details at http://www.knowledgebank.irri.org/training/fact-sheets/crop-establishment/manual-transplanting

metric by weighting the proxy of biomass burnt in an origin district by the fraction of time during the day when wind was blowing from that origin to a destination district, penalized by the distance between these districts. Then I calculate monthly "fire exposure" of a district by summing across all possible origins separately for each month of the year. I show that November fire exposure strongly affects annual PM2.5 levels, with changes in upwind fire exposure explaining 4.2% of annual deviations in within-district PM2.5, compared to 16% explained by local weather. In contrast, October fire exposure does not have a strong effect on annual PM2.5 levels. This result can be explained by the calmer winds and cooler air of the early winter in November that trap particulate matter, in contrast to stronger winds and higher precipitation that characterize the late monsoon month of October and result in rapid removal of particulate matter from the atmosphere.

Finally, I use the November fire exposure as an instrument to tackle the endogeneity concern that shocks to GDP growth are correlated with increase in local pollution even after accounting for fixed but unobservable determinants and time trends. With this instrument, I show that a 1% increase in PM2.5 levels reduces GDP by 0.18% in Indian districts. In comparison, Dechezleprêtre et al. (2019) find an elasticity of -0.08 for NUTS-3 level regions in Europe using thermal inversions as the instrument. Thermal inversions are likely to occur in some places more than others, leaving some residual correlation with fixed determinants of GDP such as the presence of ports, network effects etc. The use of location fixed effects can reduce this problem; but there may not be large changes in the distribution of thermal inversions over a short period of time, so this may be a weak instrument with fixed effects. A strength of my approach is to rely on an exogenous policy change that drives clear and large changes in pollution across Indian districts during the time period of study. This provides an especially credible method to estimate the effect of air pollution on economic output.

This paper contributes to two main literatures. The relationship between economic growth and the environment at various levels of development is poorly understood. The Environmental Kuznets Curve hypothesized an inverted-U shape, with economic growth worsening environmental outcomes at low levels of GDP per capita, but improving these outcomes at high levels (Grossman and Krueger 1995; Stern 2017).⁵ I focus on uncovering a causal mechanism behind how air pollution affect GDP per capita, documenting that the protection of groundwater resources in India increased air pollution, contributing to a small but growing literature that is focused on uncovering such causal mechanisms (Jayachandran 2022). The

 $^{^{5}}$ For a sample of urban areas across the world, Jayachandran (2022) documents that greenhouse gas emissions keep increasing with GDP per capita, lead pollution displays the EKC inverted-U pattern, air pollution (particulate matter concentrations) displays a linear and correlation while Ozone does not display any correlation at all.

resulting cost of air pollution to economic growth is larger in India compared to evidence from Europe (elasticity of -0.18% vs -0.08% in Dechezleprêtre et al. (2019))

This paper also contributes to the literature on how institutions affect environmental outcomes in developing countries. More stringent regulation (Burgess et al. 2019), use of technology (Assunção et al. 2022) and higher resource allocation to monitoring and enforcement (Duflo et al. 2018) can improve environmental outcomes in developing countries. Weak state capacity impedes the implementation of regulations on the books that prohibit the use of fires in agriculture in Punjab and Haryana, leading to large economic costs. Another explanation may be that while the groundwater externality is localized to the two states, the air pollution externality is an inter-state phenomenon. Lipscomb and Mobarak (2017) show that decentralization of regulatory authority in Brazilian municipalities leads to larger water pollution externalities across border. Kahn et al. (2015) document that providing promotion incentives to reduce some water pollutants to local officials reduces their externality on downstream neighbors. While the states of Punjab and Haryana are able to successfully implement one set of laws intended to conserve local groundwater, they are unable (or unwilling) to implement regulations on fires, which cause downwind externalities on top of local ones. This result suggests that designing regulatory institutions for environmental protection at the appropriate level is important in determining the outcomes of regulation, including for economic growth.

The rest of the paper is structured as follows. Section 2 describes the data while section 3 presents the context, section 4 presents the research design, section 5 describes the results, and section 6 concludes.

2 Data

2.1 Air quality

An important consideration for air quality data is complete geographical coverage. Whereas ground-level monitoring station coverage in India is extremely sparse (Greenstone and Hanna 2014), satellite imagery-based products provide complete coverage. Secondly, ground-level observations may be susceptible to manipulation (Greenstone et al. 2022; Ghanem and Zhang 2014). Therefore, the main source of data on air quality is Hammer et al. (2020), a gridded reanalysis product of global surface PM2.5 concentrations at a resolution of 0.01° that should be much less susceptible to such manipulation. This product combines satellite imagery data on Aerosol Optical Depth with state-of-the-art chemical transport models, and

calibrates the output to global ground-based observations. It is easy to aggregate the gridded product to the necessary resolution for analysis at pixel, city or district level. Forthcoming sections will detail the aggregation procedure for each analysis.

2.2 GDP

To estimate the impacts of PM2.5 on GDP, we would like data at the most granular subnational level possible. While satellite-based data on PM2.5 levels are available at a 1x1 km grid, output data are rarely available at such sub-national scales. Fortunately, GDP measures at the district level in India between 2007-2013 have recently been compiled by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in their District Level Database (DLD).⁶ I clean and combine these data with other district-level data using district identifiers from ICRISAT and the Census of India, 2011.

2.3 Agricultural fires

The burning of residue from crop harvest is referred to as agricultural fires. To analyse the impact of the groundwater conservation laws on the monthly pattern of fires, the ideal data would include precise location of each individual fire set for the purpose of burning crop residue. But there are no representative ground-level observations of this phenomenon. To overcome this challenge, I utilize the Fire Information for Resource Management System (FIRMS) product from the National Aeronautics and Space Administration (NASA) agency of the United States that is widely used to identify terrestrial fires. This product provides information on daily fires detected at latitude/longitude level across the world and has been recently used to analyze agricultural fires in the economics literature (Behrer 2019).

FIRMS provides a few related products: Near-Real Time (NRT) fires using the MODIS instrument aboard Terra and Aqua satellites, standard product from the same instrument but with a 2-3 month lag and another NRT product using the VIIRS instrument from the Suomi-NPP and NOAA-20 satellites. The main difference between the first two and the third is the resolution of the data. MODIS products are at 1 km resolution and are available from 2000 (more reliable from 2002 when Aqua satellite was launched) whereas VIIRS products are at 375 m but only available from 2012. The primary analysis utilizes the MODIS standard product which differs from the NRT data in that corrections are made to the imprecise location of the Aqua satellite in the NRT data. Imagery data from Aqua and Terra satellites

⁶ http://data.icrisat.org/dld/src/crops.html

is available at least four times daily for each pixel on Earth and is processed using a NASA algorithm isolate a ground-level fire signal from other signals such as solar flares.

I combine this data with information on land use from the European Space Agency Climate Change Initiative's land cover map (version 2.07)⁷. This allows the subset of fires that is found on agricultural land to be separated from natural forest fires since this paper is interested in agricultural fires. I aggregate and resample the land cover data which is at a resolution of 300 m to the fire data grid (at 1 km resolution), with an indicator for agricultural land use as the main output from this process. All fires are then masked based on this indicator variable to find the subset of agricultural fires.

2.4 Meteorology

Hourly wind data are used to construct exposure to agricultural fires for every origindestination pixel pair. Details of the methodology follow in the section 3 below. The source of these wind data is the European Center for Medium Range Weather Forecasting (ECMWF) ERA5 family of global gridded reanalysis datasets.⁸. Reanalysis data combine ground-level observations and satellite data with Chemical Transport Models that represent physical and chemical processes in the atmosphere to produce reliable and complete coverage for the world. Since ground-level observations are particularly sparse in developing countries these reanalysis data are widely used in the literature on climate and air pollution in Economics (Auffhammer et al. 2013) Hourly wind speed and direction data are taken from the ERA5-Land hourly dataset which is available at a resolution of 0.1° . These are combined with daily agricultural fires at the pixel level to construct the fire exposure variable. Apart from being used to quantify the contribution of distant residue burning on local air pollution, fire exposure also is an instrument for pollution at the city and district level in estimation of certain elasticities. Finally, I also construct temporal averages for weather variables including rainfall, temperature and relative humidity from this dataset to be used as controls in the regression analysis.

⁷ Data is available at https://cds.climate.copernicus.eu/cdsapp/#!/dataset/satellite-land-cover

⁸ Data available at https://cds.climate.copernicus.eu

3 Background

3.1 Air pollution in India

Almost all of India's population in 2011 experienced pollution levels substantially higher than the World Health Organization (WHO) limit for particulate matter below 2.5 microns in size.⁹ Exposure to such pollution can affect the physical productivity of workers (Graff Zivin and Neidell 2012; Chang et al. 2016, 2019), reduce their labor supply and earnings (Hanna and Oliva 2015; Borgschulte et al. 2022b; Hoffmann and Rud 2024), and affect property prices (Bayer et al. 2009; Freeman et al. 2017). Yet, there is lack of good evidence on how such high pollution levels affect aggregate economic output in a large country like India. This is the gap I seek to fill in this paper.

3.2 Groundwater conservation and air pollution

I utilize exogenous variation in air pollution that is driven by the groundwater conservation policy. Before discussing the construction and use of this instrument in the research design section, I provide some background on how groundwater conservation could have caused exogenous variation in air pollution.

India is the largest user of groundwater in the world; but with almost 20% of the world's population, it only has about 4% of the world's freshwater resources (World Bank 2021). The resulting overuse of groundwater to meet population needs has caused rapid aquifer depletion and led to an urgent environmental crisis, particularly affecting the alluvial plains of North-Western India. I discuss the factors behind this depletion in the North-Western states of Punjab and Haryana, leading to the passage of a set of groundwater conservation laws in 2009. I then discuss how these laws may have unintentionally pushed agricultural fires into early winter, when their impact on air pollution is exacerbated due to meteorological conditions.

⁹ This form of pollution is commonly knows as PM2.5 and is known to cause serious health effects. Both short-term and prolonged exposure to PM2.5 can lead to heart attacks, asthma, decreased lung function or cancer, stroke and a variety of other conditions, and cause premature mortality in people with heart or lung diseases (Greenstone 2021). The WHO annual average limit is 5 μ g/m3. According to their PM2.5 database, Delhi's annual average PM2.5 level was 153 μ g/m3 in 2013, in comparison with New York city's average of 14 and London's at 16.

3.2.1 The need for groundwater conservation

Until the advent of the so-called Green Revolution of the 1960s that raised agricultural productivity dramatically across India and much of the poor world (Pingali 2012), North-Western India was a primarily wheat-growing region with little rice consumption or production locally. One of the institutional innovations of the Green Revolution in these states was the provision of large subsidies for tubewells and borewells. Individual farmers could now access shallow groundwater to irrigate fields even if they did not have access to the large, pre-existing canals systems built by the colonial British empire. Over time, modern pumps running on electricity were combined with practically zero tariffs to farmers so that they could irrigate their fields at minimum cost.

This newfound access to groundwater allowed farmers to diversify their crop portfolio by allowing the cultivation of the highly water-intensive rice crop during the "Kharif" or monsoon season (June-October). The wheat crop is cultivated during the "Rabi" or winter season, when the lower temperatures and plenty of sunshine provide perfect weather conditions for growth (Kataki et al. 2001); planting happens in early winter and harvest in early spring.

The state of Punjab contributed less than 1% of India's rice in 1961; by the late 1990s this figure was up to 10%; absolute rice output across India rose from 11 million tonnes to 75 million tonnes in this period, underlining the massive increase in rice cultivation in Punjab (Subramanian 2017). Similar trends in rice cultivation were seen in Haryana. This fundamental change in the cropping patterns of the region exacerbated the depletion of groundwater resources, since the paddy fields were flooded primarily using groundwater, pumped out before the annual monsoon reached Punjab and Haryana. Taken together, the unregulated exploitation of groundwater had led to an acute water crisis by the early 2000s, although concerns about excessive extraction almost 1.5 times the natural recharge rate had been expressed by agricultural scientists and government committees going back to the 1980s (Singh 2009).

Despite the alarm expressed by various stakeholders, the state governments largely ignored the problem until the early 2000s. When asked about these concerns, the then-Chief minister of Punjab, Prakash Singh Badal, is quoted in the media as saying, "The problem is not as acute as is being projected. It is a theoretical evaluation and there is no truth in it" (Down To Earth 1999). The political economy of both states, but particularly of Punjab, centers around medium and large sized farmers who receive a range of state subsidies that incentivizes rice and wheat cultivation. Apart from the Green Revolution era technological subsidies for higher-yielding seeds, fertilizers and pesticides, tubewells and electric pumps, provision of cheap electricity is also important in explaining groundwater levels (Ryan and Sudarshan 2020). Often these dues are not paid to state distribution companies at all, resulting in lack of investment in the power grid (World Bank 2018). But, most importantly, the procurement of wheat and rice crop by the state governments of Punjab and Haryana at so-called Minimum Support Prices that distort market signals (Parikh et al. 2003) precludes farmers from switching to other crops with higher price and yield risks.

3.2.2 What was the groundwater conservation mandate?

The practice of transplantation of rice before mid-June was thought to be particularly cause too much reliance on groundwater (Singh 2009). In response, sections of the state bureaucracy had made efforts starting in the early 2000s to shift the transplantation of rice closer to the monsoon, since this was thought to ease the strain on groundwater use. The two governments took executive action through ordinances in 2008¹⁰ to extend the practice of delaying rice transplantation state-wide. Given the generally favorable response to this ordinance, the legislatures of Punjab and Haryana separately ratified the Preservation of Subsoil Water Acts of 2009 ("laws" from now on) in an effort to conserve groundwater.

These laws prohibited early transplanting of rice before the monsoon in an attempt to reduce groundwater usage for irrigation. Much of the rice transplantation would occur in the peak of summer during May when evapotranspiration (water loss from plants as well as soils and water bodies) is high. These laws specified June 10 as the earliest transplantation date, and it was shifted further to June 20 later¹¹. When planting rice in May, farmers were solely dependent on groundwater reserves for rice growth; moving transplantation to June allowed rice growth to depend more on monsoon rainfall. This was expected to lead to a lower rate of groundwater extraction.¹²

3.2.3 Potential displacement of fires into November due to policy

The primary use of fires in Indian agriculture today is to clear the field of leftover residue from harvesting a crop, before sowing and planting the next crop (Shyamsundar et al. 2019); this differs from slash-and-burn agriculture that is practiced in parts of Africa and Indonesia

¹⁰ These do not have the same power in Indian law as a statute and cannot be renewed beyond a few months.

 $^{^{11}{\}rm The}$ Indian Met Department (IMD) sets out July 1 as the expected date of onset of the monsoon in North-Western India. Details here

¹² Groundwater recharge is typically a slow-moving process that takes place over a longer period than the period of study here. I plan to conduct an assessment of the change in groundwater levels to the present day due the policy in the future

(Andini et al. 2018). Figure 1 shows that agricultural fires are concentrated in the states of Punjab and Haryana, which are also characterized by a Rice-Wheat crop system.

In the Rice-Wheat system of Punjab and Haryana, fires are used to clear rice residue before the planting of the wheat crop on the same land, since rice residue comes in the way of planting wheat. This practice dates back at least to the 1990s. The earliest observations of fires from the NASA FIRMS database (described in the next section) starting in 2002 clearly demonstrate that North-western India already had a disproportionate share of fires in Indian agriculture.

The delay in rice transplantation due to the laws also pushed back harvest dates. The resulting delay in rice harvest from mid October to late October and November meant that farmers had fewer days between rice harvest and wheat plantation. Any delays in wheat plantation beyond the first two weeks of winter reduces yields substantially (McDonald et al. 2022). Therefore, the law may have had the unintended consequence of increasing the intensity of agricultural fires in November, when slower winds and lower temperatures tend to worsen downwind air quality.

4 Research Design

4.1 Effect of PM2.5 on GDP

This section describes the estimation strategy for the causal impact of higher PM2.5 levels on district GDP in India. I build up to an instrumental variables strategy for PM2.5 that allows the quantification of impact of the groundwater laws on downwind GDP. Before describing this IV strategy, equation (1) presents an OLS regression model of the effect of PM2.5 on district GDP.

$$log(GDP_{du}) = \beta log(PM_{du}) + \gamma Weather_{du} + g_d * t + \alpha_d + Y_u + \epsilon_{du}$$
(1)

$$Weather_{dy} = \{Temp_{dy}, Temp_squared_{dy}, Rain_{dy}, Rain_squared_{dy}, Relative_Humidity_{dy}, Surface_Pressure_{dy}, Windspeed_{dy}\}$$

The quantity of interest β is the percentage reduction in GDP for a 1% increase in PM2.5 levels. This model contains district and year fixed effects D_d and Y_y respectively, which control for fixed factors that raise productivity or increase pollution as well as account for any common macroeconomic shocks. Weather variables such as temperature, rainfall, humidity and wind speed are known to affect PM2.5 (Dechezleprêtre et al. 2019; Bondy et al. 2020). Therefore, I control for yearly average weather that could determine the level of pollution from given emissions.

The main residual concern with identification of β in this model is that deviations of GDP and PM can be jointly determined. Higher economic activity in a given year can itself cause an increase in PM that year by increasing emissions. At the same time, higher deviation in PM can stunt GDP growth that year through channels such as increased worker morbidity and lower labor productivity. We would like to identify the second channel and avoid the former.

I adopt three approaches to tackle this endogeneity issue. Figure 4 shows that GDP exhibits strong growth in this period and therefore is not stationary; between 2007-2013, average Indian GDP growth rate was 7%. First, I fit a district-specific linear time trend $g_d * t$ in GDP. The time trend will capture district-specific factors that cause constant GDP growth, leaving only deviations from the trend line in the outcome. This approach can also help reduce omitted variables bias (OVB) that jointly determines both GDP and PM2.5 (eg. demand shocks that affect certain districts). Such OVB can cause the causal chain to run from GDP to PM2.5, leading to reverse causality that biases the estimate upwards, since an increase in economic activity increases PM2.5 levels.

Secondly, I also conduct analysis using first differences (FD) that is the preferred over fixed effects to deal with non-stationary, autocorrelated data series in both outcome and explanatory variables. An FD specification that also includes a fixed effect is the same as allowing for a district-specific linear growth rate g_d . The FD approach is commonly used in the macroeconomic literature to deal with serial correlation in aggregated GDP data. Equation (2) specifies the regression framework for the FD model.

$$\Delta log(GDP_{dy}) = \beta \,\Delta log(PM_{dy}) + \gamma \,\Delta Weather_{dy} + g_d + \Delta Y_y + \Delta \epsilon_{dy} \tag{2}$$

This specification examines how the growth rate of PM2.5 affects the growth rate of GDP, controlling for year-on-year changes in weather and common macroeconomic conditions. The district fixed effect g_d captures the constant growth rate of GDP in these districts. But even with the FD design, there may still be some OVB in the leftover variation, leading to reverse causality that biases the results upwards.

The third approach utilizes the fact that the stock of pollution in a district is partially due to sources outside the district, notably agricultural fires in this instance. Using an instrument

for PM2.5 defined in the next section with both the panel and FD specifications allows us to tackle the reverse causality challenge.

Inference : Recent methodological guidance on inference from Abadie et al. (2023) cautions against leaning towards conservatism. They distinguish between sampling-based and design-based uncertainty. Sampling-based uncertainty arises when there is within-unit correlation in a sample, such as worker incomes in a given location. Aggregated GDP across a large spatial unit such as a district naturally mitigates against this within-unit correlation, eliminating any sampling-based uncertainty. Design-based uncertainty arises from uncertainty about what would have happened under different treatment assignments to any given district. According to Abadie et al. (2023), this is the main source of uncertainty for studies that use administrative data on the entire population, such as district-level GDP. They show that standard inference tools such as clustered or Conley errors are too conservative in these settings since they were developed to correct for sampling-based uncertainty.

In my setting, the fire exposure instrument leverages district-specific variation in treatment (weighted upwind burning across origins), creating localized exogenous shocks that significantly reduce residual spatial correlation. Despite the warnings of Abadie et al. (2023), I tend toward the conservative by clustering standard errors at the region-year level, where regions are groups of contiguous districts that share similarities in economic and geographic fundamentals such as level of development, soil types, weather, and fire exposure etc.¹³

4.2 Effect of groundwater policy on the timing of fires

Before describing the construction of an instrument for air pollution using policy-driven variation in the timing of fires, I show that this policy indeed shifted the monthly pattern of fires. I utilize a difference-in-differences research design with fixed effects to conduct this analysis. The outcome variables in each district-month-year period from 2002 to 2020 are the count of fires and the total strength of these fires as measured by the fire radiative power. These are aggregated to the district-level to reflect the administrative unit at which state policy is implemented in India. I estimate a Poisson fixed effects model to recover the coefficient of interest, assuming the standard exponential link function (Behrer 2019; Ranson 2014) for the count or measure of biomass burnt F_{dmy} in district d, month m and year y. The conditional expectation function given regressors \mathbf{X}_{dmy} is as follows

 $^{^{13}}$ There are 530 districts and 96 regions in the sample, so that there are 5.5 districts on average in each region. Each district had an area of approximately 100 sq km, on average.

$$\mathbf{E}[F_{dmy}|\mathbf{X}_{dmy}] = exp(\sum_{m \in [1,12]} \delta_m D_{dmy} + \alpha_d + \tau_{sm} * Y_y) \tag{3}$$

where the RHS inside the exponential function contains \mathbf{X}_{dmy} . Since the laws came into force at the state-level in 2009, the treatment indicator D_{dmy} turns on for district-months in Punjab and Haryana in and after 2009. District fixed effects D_d control for unobserved determinants of fires that do not change over time. Comparison of fires within state-bymonth cells (τ_{sy}) flexibly controls for other within-state determinants of fire seasonality such as different crop calendars, crop mixes etc. that do not change over time. Year fixed effect Y_y controls for any common trends across the country (such as the country-wide increase in fires driven by the Mahatma Gandhi National Employment Guarantee Scheme or NREGS documented by Behrer (2019)).

The count nature of the data and the nontrivial presence of zeros in the count data motivate the use of a Poisson model. A log transform of F_{dmy} would create bias in a linear model estimation whereas an inverse hyperbolic sine transform makes the interpretation of the elasticity slightly more complicated (Bellemare and Wichman 2020). Further, the Poisson FE model only requires that the conditional expectation function be specified correctly for consistent estimation of the parameters (Wooldridge 2010). It produces unbiased estimates of the coefficients even if the fire data do not match the Poisson distributional assumptions (Wooldridge 1999a, 1999b; Lin and Wooldridge 2019). This is not true of other models that are used to handle count data such as negative binomial (Blackburn 2015). I estimate this model using quasi-maximum likelihood method through the *fixest* package in R (Berge et al. 2022).¹⁴

Taking log of (3) yields the following

$$log(\mathbf{E}[F_{dmy}|\mathbf{X}_d my]) = \sum_{m \in [1,12]} \delta_m D_{dmy} + \alpha_d + \tau_{my}$$
(4)

Therefore the coefficients of interest δ_m give the monthly elasticity of fire count to the policy. As with any difference-in-differences design, the main identifying assumption for the δ_m s is that trends in monthly fires would be similar between treatment and control districts in the absence of the policy change. I discuss this assumption in more details in the results section. Standard errors are clustered two ways at the district and state-by-year level to account for

¹⁴The Poisson model can be used with non-integer data such as the measure of biomass burnt as well, and the strengths of the Poisson over other model when the data have nontrivial presence of zeros also holds (Wooldridge 2010)

the district-level autocorrelation as well as implementation at the state-by-year level.

4.3 Constructing instrument for air quality using agricultural fires

Now I describe the construction of the fire exposure metric. This metric can be constructed for every month of the year; but our instrument will be November fire exposure so as to leverage the exogenous displacement of fires from October to November.

I capture the contribution of daily agricultural fires F_{ot} from 1x1 degree origin pixel o on air quality in destination pixel d for month M, as follows

$$\omega_{od,M} = \left(\sum_{t \in M} \frac{windfrac_{odt} * F_{ot}}{dist_{od}}\right)$$
(5)

wind $frac_{odt}$ is the daily average fraction of time that the wind at o blows towards d on day t, whereas $dist_{od}$ represents the distance between the centroids of o and d. In order to calculate $wind frac_{odt}$, I start by assigning each hourly wind observation in o on day t into one of 36 bins of 10 degree span each, based on the wind direction that hour (true north is 0 degree as in the figure). I then construct the wind speed-weighted fraction of time the wind was blowing in each of these 36 bins by aggregating hourly observations for day t. $wind frac_{odt}$ is then calculated by summing up wind fraction for the bins which are in the direction of d from o as shown in figure 3.

I construct this instrument for various distances between origin and destination districts. These are increased sequentially so that the distance that maximizes power in predicting PM2.5 can be selected (Details in the next section of this estimation). Variation in the instrument is driven by two factors: (i) changes in the temporal distribution of fires at origin, and (ii) changes in the daily wind patterns at origin during the given month.

This instrument substantially differs from the wind-based instrument in Deryugina et al. (2019). They utilize daily *local* variation in wind patterns that can change where pollution comes from on that given day. I account for daily but *upwind* variation that is *specifically* affecting the destination of interest, through the weighting of $windfrac_{odt}$, penalized by $dist_{od}$. This instrument allows me to explicitly leverage groundwater policy-driven variation in air pollution.

A hypothesis from the scientific literature is that fires in the winter are much worse for downwind PM due to meteorological conditions that favor longer suspension and entrapment of particulate matter in the lower atmosphere of downwind districts. I test this hypothesis that fires in the winter are worse for PM2.5 in fire-exposed downwind districts by testing their effect on annual PM2.5. Before discussing the results on the effect of PM2.5 on district GDP in section 5.3, I discuss the results of this first stage in section 5.2, based on specifications in equations 1 and 2.

5 Results

Table 1 summarizes the main variables used in the analysis.

5.1 Effect of groundwater laws on monthly fire patterns

To begin the results section, I refer to figure 2 that shows some growth in the fire count and fire strength for November occurring just before the laws were passed in 2009, with a stronger trend upwards after the passage of the law, before stabilizing by 2015 or so. This suggests some anticipation effects in November, since the count of fires is trending up 2-3 years before the policy came into effect. These anticipation effects can be attributed to the informal implementation of the policy before 2009 that is discussed in section 3.2.2. This may have driven the shifts in fire patterns by slightly delaying the cultivation dates before 2009, with formal implementation inducing a larger shift. The lack of pre-trends on October fires combined with a downward shift after 2009 supports this view. Therefore, the effect for November my be an underestimate.¹⁵

Table 2 presents estimates of the causal effect of these laws on monthly fires, based on 3. Columns 1 and 3 provide the mean number of fires and measure of biomass burnt in Punjab and Haryana, before the passage of these laws. Those columns show peaks of fires during the months of April, October and November. Fires in the latter two months are used to clear the monsoon season rice residue, as described earlier. Fires in April are used to clear the wheat crop residue after the harvest is done. The time pressure of needing to be rid of the rice residue before wheat plantation that leads to the fires after monsoon rice harvest does not arise after the wheat harvest. Yet we see substantial fire activity in April. This wheat residue burning practice may have come about due to habit formation from setting fires to

¹⁵ Given the recent literature on the bias of TWFE, I plan to test for conditional parallel trends with anticipation effects as well as the treatment effect of interest using the framework of Callaway and Sant'Anna (2021). Their approach would work well in this setting since they rely on never-treated units to estimate treatment effects. Therefore, I plan to utilize their R *did* package to estimate these effects in the future, better accounting for anticipation effects.

the rice residue. However, it is less troublesome for downwind pollution than are fires during early winter, since meteorological conditions in April do not favor suspension of particulate matter over the plains of North India.

Turning to the results in columns 1 and 3 of table 2 now, the main result is that the laws increase the log of expected fire count in November by 0.43, and log of expected fire strength by 0.54. Estimates for October are negative, providing evidence that the laws probably succeeded in pushing rice cultivation by a few weeks to a month, and therefore peak fire activity into November. Estimates for the other months except June and July are negative. For the months from December to May, this would suggest a domino effect of the later rice cultivation on other crop burning, since the entire crop calendar gets pushed back. The spring season fire peak (from the wheat harvest) that used to happen in April and May seems to shift slightly toward June and July, generating the positive estimates for those two months. The negative estimates for August and September probably also come from the enforced delay in rice plantation that would have affected some farmers who would plant rice in early May otherwise. Finally, since there are very few fires to begin with in July, and since July happens to be the rainy season, the shifting of the wheat fire season perhaps does not have the same consequences for downwind pollution that the shifting of the rice season does.

I present robustness results to alternative specifications and sample selection in table A1. These include the following: OLS estimation rather than Poisson, including fires data from 2000 and 2001,¹⁶ and limiting the analysis to the sample for which GDP data is available.¹⁷ The results are consistent with table 2 in all these robustness checks, with only the fire strength when limiting to the GDP sample becoming insignificant. This lack of power could be due to the effects of policy not having had enough time to accumulate by 2013 or to anticipation effects just before 2009. It should certainly not be taken as an indication that the laws did not increase November fire activity.

5.2 Effect of November fires on annual downwind pollution

Before turning to the causal effect of annual PM2.5 on district GDP, I discuss the effect of fire exposure on annual district PM2.5 levels. These results are equivalent to the first stage for the 2SLS results on GDP in the next section. As noted in the previous sections, fires in the winter are particularly harmful for PM2.5 levels due to prevailing meteorological conditions over

 $^{^{16}}$ The NASA Aqua satellite was launched in 2002 and drastically improved estimates of fire activity in the FIRMS database

 $^{^{17}}$ 530 district between 2007-2013

North India that favor slower dispersion of the particulate matter over space. Further, the groundwater laws pushed agricultural fires in Punjab and Haryana toward November (early Winter). Therefore, I focus on the effect of November fire exposure on PM2.5. Certain fires can be stronger because more organic material is burnt, thereby producing higher amounts of particulate matter. Therefore I use FRP to maximize signal in the instrument relative to using count of fires.

I implement various distance cut-offs on the exposure measure: origin districts at a distance larger than the cut-off are not used to construct FRP exposure for destination district. This is done for two reasons. Firstly, while wind fraction *times* inverse-distance weighting captures some of the pollution decay over distance, it could miss out on some important features that govern decay, such as (i) rainfall along the path, which can cause the "wet deposition" of particulate matter (Vallero 2014) (ii) meteorological conditions along the path such as wind speed, temperature and relative humidity that could also alter the trajectory or cause further deposition out of the atmosphere and (iii) geographical features such as mountains along the way. For this reason, I hypothesize that larger cut-offs could add more noise to the instrument. Therefore, I test which distance cut-off maximizes the within-R2, in order to quantify the trade-off between signal and noise when increasing the distance cut-offs.

Table 3 shows results for cut-offs between 500 and 1000 km. In panel A, I present results from a fixed effects model that includes a district-specific time trend, equivalent to the first stage for equation 1. Panel B presents results from the first difference model with district fixed effects in equation 2, therefore assuming a district-specific trend in growth of PM2.5. Both these sets of results show strong and robust elasticities of PM2.5 to November FRP exposure, peaking at a cut-off of 900 km (for both the coefficient size and within-R2). The main result here is that a 1% increase in November FRP exposure increases annual PM2.5 levels by 0.029% (0.032%) with the FE (FD) model. It further illustrates the trade-off between signal and noise when increasing distance to origin in constructing the instrument.¹⁸ Globally, 900 km maximizes within-R2 when explaining PM2.5 using November FRP exposure. I therefore use that as the preferred distance to construct the instrument for PM2.5 in the next section.

In table 4, I confirm that higher FRP exposure only from fires during winter months affects annual PM2.5 levels. This can be explained by unfavorable meteorological conditions during winter that cause the particulate matter emissions from agricultural fires to stay suspended for longer. However, fires in the winter months other than November are not affected by the groundwater laws in Punjab and Haryana. Therefore, in order to quantify the effect of

 $^{^{18}}$ Results for regressions with a 100 km cut-off to no distance cut-off at all show an increasing within R2 until 900 km when they start dropping of monotonically.

increased November fires due to the laws later, I use only November-based FRP exposure instrument in the analysis of the effect of PM2.5 on GDP in the next section.

Now, I address a concern that distance may be correlated with geographic determinants of PM2.5 (and GDP later). Controlling for district fixed effects in these regressions helps address that concern. But, distance also enters the instrument itself non-linearly; it may be that the district fixed effect does not fully address the issue. Therefore, I construct the instrument for each of these cut-offs by adding up wind fraction-weighted FRP from qualifying origins *without* inverse-distance weighting. Thus distance directly does not enter this instrument. Results for these regressions are presented in appendix table A2. They do not suggest any cause for concern that distance entering the instrument non-linearly causes any bias in the first stage.

5.3 Effect of PM2.5 on GDP

In this section, I turn to the impact of annual PM2.5 levels on annual GDP in Indian districts in panel A of table 5. I present results with the fixed effects in columns 1-3, and with the first difference specification in columns 4-5. Column 1 presents the OLS estimate controlling for weather and including district and year fixed effects, but without district-specific linear time trends. The coefficient is positive and strongly significant. The causal effect of higher PM2.5 on GDP should be negative, given the harmful effects on human health and productivity, and potential effects on agriculture and machinery. The positive coefficient suggests that this specification is not sufficient to address the concern about omitted variables that jointly determine GDP and PM2.5, such as yearly demand shocks that cause higher GDP growth due to certain districts being more trade-exposed, for example. Higher economic activity in that year would increase PM2.5 levels in that district, and district fixed effects are insufficient to capture the co-movement of these variables. The estimate is biased upwards since the causal chain runs from GDP to PM2.5 in such cases. The sample period witnessed very strong GDP growth in Indian districts, making this a particular concern in this setting.

Column 2 presents results with the addition of a district-specific linear time trend to reduce this concern. The coefficient turns negative now, although it is imprecise, suggesting that this time trend is able to reduce the upward bias from the reverse causality of GDP to pollution. It also suggests the importance of including such time trends for non-stationary GDP data when focusing on the effect of jointly determined variables such as air pollution, as opposed to plausibly exogenous variables such as temperature deviations (Dell et al. 2012).

Before discussing the 2SLS estimates in columns 3 and 5, I focus on column 4 which presents

the first difference estimate along with a district fixed effect, in effect assuming a districtspecific trend in the growth rate of GDP. The coefficient is -0.03 and significant at the 5% level. The FD specification works much better with non-stationary data, and therefore this coefficient is less biased and also more precisely measured than the fixed effects regression with time trends in column 2.

Both these approaches solve some of the omitted variable problem plaguing estimation of the effect of PM2.5 on district GDP. However, any joint residual variation from the trend still causes upward bias in the estimates. I turn to the instrumental variable strategy to address this residual concern. In column 3, I present 2SLS results from the fixed effects model with district-specific time trends, instrumenting for PM2.5 using November FRP exposure with a 900 km distance cut-off. The estimate is now much larger, although the IV also increases standard errors as expected.

Panel B reproduces relevant first stage estimates from table 3. To test for weak instruments, I also present two statistics below the first stage estimates. Stock and Yogo (2005) suggest the use of the Cragg-Donald F-stat in a multivariate setting to test for weak instruments, with a rule of thumb that a value less than 10 indicates a potentially weak instrument. The Cragg-Donald F-stat is about 101.4; but this relies on iid assumptions for the errors. Therefore, I also report the Kleibergen-Paap (KP) F-stat which is equivalent to the robust F-stat with one endogenous regressor, as in this setting. The F-stat of 25.3 is comfortably above 10, and therefore concerns about weak instruments do not arise here.¹⁹

Column 5 presents the 2SLS results from the first difference model. The point estimate is slightly larger than column 3, and is estimated much more precisely. The KP F-stat is 26.4, again comfortably larger than 10. I consider the specification in column 5 as the preferred specification. These estimates suggest that increasing PM2.5 levels by 1% in a given year has a large negative causal effect of 0.18% on district GDP.

6 Conclusion

This paper estimates the causal impact of air pollution on economic output using a natural experiment provided by groundwater conservation mandates in Punjab and Haryana.

¹⁹Andrews et al. (2019) recommend the use of the effective F-statistic (MOP F-stat) of Olea and Pflueger (2013) in the case of a single endogenous regressor. This statistic is not easily calculated in any R or Stata package that implements IV with panel data. However, Andrews et al. (2019) also note that with one single endogenous regressor, the MOP F-stat is equivalent to the KP and robust F-stats. Therefore, the provided F-stat is the correct one to test for weak instruments.

To address endogeneity concerns inherent in measuring pollution's impact on GDP—since higher economic output typically generates more pollution—I exploit the exogenous shift in agricultural fires induced by these mandates as an instrumental variable. Specifically, the mandates shifted biomass burning from October to November, when meteorological conditions significantly enhance pollution exposure for downwind districts.

I construct a novel measure of exposure to upwind November fires, demonstrating that it substantially explains annual variations in PM2.5 concentrations across districts. I find that 1% higher upwind November fire exposure increase annual PM2.5 by 0.03%. Using first-differenced models to control for district-specific trends and instrumenting pollution levels with this policy-induced variation, I find that a 1% increase in annual PM2.5 concentrations reduces district GDP by 0.18%.

There are also some limitations to this approach. First, the estimate relies on the exposure instrument affecting downwind districts in accordance with its structure. While a chemical transport model could do better in modeling this relationship, it is much more resource-intensive to operate and may not do especially well for seasonal sources such as agricultural fires. I also leave it to future research to further clarify the economic mechanisms behind this relationship—whether the observed GDP reduction primarily occurs through declines in industrial productivity, agricultural output, or labor health and productivity impacts. Investigating firm adaptation strategies, such as relocating or adjusting production schedules, will also help deepen our understanding of pollution's economic costs.

In conclusion, this study contributes to the literature by providing a credible estimate of pollution's economic impact. By leveraging a novel instrument for air pollution that arises due to exogenous policy changes, it also underscores the need for integrated environmental and economic policy frameworks that consider cross-sectoral and spatial spillovers.

7 References

- Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2023. "When Should You Adjust Standard Errors for Clustering?"." The Quarterly Journal of Economics 138 (1): 1–35.
- Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2022. "Management and Shocks to Worker Productivity." *Journal of Political Economy* 130 (1): 1–47.
- Andini, Ade, Sébastien Bonnet, Patrick Rousset, and Udin Hasanudin. 2018. "Impact of Open Burning of Crop Residues on Air Pollution and Climate Change in Indonesia." *Current Science* 115 (12): 2259–66.
- Andrews, Isaiah, James H. Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11 (1): 727–53.
- Assunção, Juliano, Clarissa Gandour, and Romero Rocha. 2022. "DETERring Deforestation in the Amazon: Environmental Monitoring and Law Enforcement." *American Economic Journal: Applied Economics*.
- Auffhammer, Maximilian, Solomon M. Hsiang, Wolfram Schlenker, and Adam Sobel. 2013. "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change." *Review of Environmental Economics and Policy* 7 (2): 181–98.
- Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins. 2009. "Migration and Hedonic Valuation: The Case of Air Quality." Journal of Environmental Economics and Management 58 (1): 1–14.
- Behrer, A. Patrick. 2019. "The Impact of Anti-Poverty Efforts on Indian Agriculture and Air Pollution." Unpublished manuscript.
- Bellemare, Marc F., and Casey J. Wichman. 2020. "Elasticities and the Inverse Hyperbolic Sine Transformation." Oxford Bulletin of Economics and Statistics 82 (1): 50–61.
- Berge, Laurent, Sebastian Krantz, and Grant McDermott. 2022. "Fixest: Fast Fixed-Effects Estimations."
- Blackburn, Mckinley L. 2015. "The Relative Performance of Poisson and Negative Binomial Regression Estimators." Oxford Bulletin of Economics and Statistics 77 (4): 605–16.
- Bondy, Malvina, Sefi Roth, and Lutz Sager. 2020. "Crime Is in the Air: The Contemporaneous Relationship Between Air Pollution and Crime." Journal of the Association of Environmental and Resource Economists 7 (3): 555–85.
- Borgschulte, Mark, David Molitor, and Eric Zou. 2022b. "Air Pollution and the Labor Market: Evidence from Wildfire Smoke." Working {Paper}. Working Paper Series. National Bureau of Economic Research.
 - ——. 2022a. "Air Pollution and the Labor Market: Evidence from Wildfire Smoke."

{NBER} working paper 29952. Working Paper Series. National Bureau of Economic Research.

- Burgess, Robin, Francisco J M Costa, and Ben Olken. 2019. "The Brazilian Amazon's Double Reversal of Fortune." Preprint. SocArXiv.
- Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230.
- Chang, Tom Y., Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2019. "The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China." *American Economic Journal: Applied Economics* 11 (1): 151–72.
- Chang, Tom Y., Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2016. "Particulate Pollution and the Productivity of Pear Packers." *American Economic Journal: Economic Policy* 8 (3): 141–69.
- Dechezleprêtre, Antoine, Nicholas Rivers, and Balazs Stadler. 2019. "The Economic Cost of Air Pollution: Evidence from Europe." {OECD} {Working} paper 1584. Paris.
- Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." American Economic Journal: Macroeconomics 4 (3): 66–95.
- 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.
- Down To Earth. 1999. "Parched Punjab." *Parched Punjab.* https://www.downtoearth.org. in/coverage/parched-punjab-20535.
- Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2018. "The Value of Regulatory Discretion: Estimates From Environmental Inspections in India." *Econometrica* 86 (6): 2123–60.
- Freeman, Richard, Wenquan Liang, Ran Song, and Christopher Timmins. 2017. "Willingness to Pay for Clean Air in China." Working paper 24157. National Bureau of Economic Research.
- Fu, Shihe, V Brian Viard, and Peng Zhang. 2021. "Air Pollution and Manufacturing Firm Productivity: Nationwide Estimates for China." *The Economic Journal* 131 (640): 3241– 73.
- Ghanem, Dalia, and Junjie Zhang. 2014. "Effortless Perfection: Do Chinese Cities Manipulate Air Pollution Data?" Journal of Environmental Economics and Management 68 (2): 203–25.
- Graff Zivin, Joshua, and Matthew Neidell. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review* 102 (7): 3652–73.

Greenstone, Michael. 2021. "AQLI India Fact Sheet."

- Greenstone, Michael, and Rema Hanna. 2014. "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India." *American Economic Review* 104 (10): 3038–72.
- Greenstone, Michael, Guojun He, Ruixue Jia, and Tong Liu. 2022. "Can Technology Solve the Principal-Agent Problem? Evidence from China's War on Air Pollution." American Economic Review: Insights, April.
- Grossman, Gene M., and Alan B. Krueger. 1995. "Economic Growth and the Environment*." The Quarterly Journal of Economics 110 (2): 353–77.
- Hammer, Melanie S., Aaron van Donkelaar, Chi Li, Alexei Lyapustin, Andrew M. Sayer,
 N. Christina Hsu, Robert C. Levy, et al. 2020. "Global Estimates and Long-Term
 Trends of Fine Particulate Matter Concentrations (1998–2018)." *Environmental Science*& Technology 54 (13): 7879–90.
- Hanna, Rema, and Paulina Oliva. 2015. "The Effect of Pollution on Labor Supply: Evidence from a Natural Experiment in Mexico City." *Journal of Public Economics* 122 (February): 68–79.
- Hoffmann, Bridget, and Juan Pablo Rud. 2024. "The Unequal Effects of Pollution on Labor Supply." *Econometrica* 92 (4): 1063–96.
- Jayachandran, Seema. 2022. "How Economic Development Influences the Environment." Annual Review of Economics 14 (September): 1–30.
- Kahn, Matthew E., Pei Li, and Daxuan Zhao. 2015. "Water Pollution Progress at Borders: The Role of Changes in China's Political Promotion Incentives." American Economic Journal: Economic Policy 7 (4): 223–42.
- Kataki, P. K., P. Hobbs, and B. Adhikary. 2001. "The Rice-Wheat Cropping System of South Asia." *Journal of Crop Production* 3 (2): 1–26.
- Leroutier, Marion, and Hélène Ollivier. 2025. "The Cost of Air Pollution for Workers and Firms." {SSRN} {Scholarly} {Paper}. Rochester, NY: Social Science Research Network.
- Lin, Wei, and Jeffrey M. Wooldridge. 2019. "Chapter 2 Testing and Correcting for Endogeneity in Nonlinear Unobserved Effects Models." In *Panel Data Econometrics*, edited by Mike Tsionas, 21–43. Academic Press.
- Lipscomb, Molly, and Ahmed Mushfiq Mobarak. 2017. "Decentralization and Pollution Spillovers: Evidence from the Re-Drawing of County Borders in Brazil*." The Review of Economic Studies 84 (1): 464–502.
- Lohan, Shiv Kumar, H. S. Jat, Arvind Kumar Yadav, H. S. Sidhu, M. L. Jat, Madhu Choudhary, Jyotsna Kiran Peter, and P. C. Sharma. 2018. "Burning Issues of Paddy Residue Management in North-West States of India." *Renewable and Sustainable Energy Reviews* 81 (January): 693–706.

- McDonald, Andrew J., Balwinder-Singh, Alwin Keil, Amit Srivastava, Peter Craufurd, Avinash Kishore, Virender Kumar, et al. 2022. "Time Management Governs Climate Resilience and Productivity in the Coupled Rice–Wheat Cropping Systems of Eastern India." Nature Food 3 (7): 542–51.
- Olea, José Luis Montiel, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." Journal of Business & Economic Statistics 31 (3): 358–69.
- Parikh, Kirit S., A. Ganesh-Kumar, and Gangadhar Darbha. 2003. "Growth and Welfare Consequences of Rise in MSP." *Economic and Political Weekly* 38 (9): 891–95.
- Pingali, Prabhu L. 2012. "Green Revolution: Impacts, Limits, and the Path Ahead." Proceedings of the National Academy of Sciences 109 (31): 12302–8.
- Ranson, Matthew. 2014. "Crime, Weather, and Climate Change." Journal of Environmental Economics and Management 67 (3): 274–302.
- Ryan, Nick, and Anant Sudarshan. 2020. "Do India's Farmers Use Too Much Water?" {VoxDev}.
- Schlenker, Wolfram, and W. Reed Walker. 2016. "Airports, Air Pollution, and Contemporaneous Health." *The Review of Economic Studies* 83 (2): 768–809.
- Shyamsundar, P., N. P. Springer, H. Tallis, S. Polasky, M. L. Jat, H. S. Sidhu, P. P. Krishnapriya, et al. 2019. "Fields on Fire: Alternatives to Crop Residue Burning in India." *Science (New York, N.Y.)*, August.
- Singh, Karam. 2009. "Act to Save Groundwater in Punjab: Its Impact on Water Table, Electricity Subsidy and Environment." Agricultural Economics Research Review 22 (January).
- Stern, David. 2017. "The Environmental Kuznets Curve After 25 Years." Journal of Bioeconomics 19 (1): 7–28.
- Stock, James H., and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor* of Thomas Rothenberg, edited by Donald W. K. Andrews and James H. Stock, 80–108. Cambridge: Cambridge University Press.
- Subramanian, Kapil. 2017. "How Delhi's Air Pollution Crisis May Have Origins in the Green Revolution." *The Hindu*, November.
- Vallero, Daniel. 2014. Fundamentals of Air Pollution 5th Edition. 5th ed. Elsevier.
- Wooldridge, Jeffrey M. 1999a. "Quasi-Likelihood Methods for Count Data." In Handbook of Applied Econometrics Volume 2: Microeconomics, 321–68. John Wiley & Sons, Ltd.
 - ———. 1999b. "Distribution-Free Estimation of Some Nonlinear Panel Data Models." *Jour*nal of Econometrics 90 (1): 77–97.
 - —. 2010. Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge,

MA, USA: MIT Press.

World Bank. 2018. "Power Sector Distortions Cost India Billions." World Bank.

———. 2021. "Addressing Groundwater Depletion: Lessons from India, the World's Largest User of Groundwater."

8 Figures and Tables



Figure 1: Strength of fires in Indian districts $(2010) \leftrightarrow$

Note: The states of Punjab and Haryana are outlined in light blue.



Figure 2: Trends in fire count and fire radiative power $(2002-2020) \leftrightarrow$

(a) **Fire count**



Figure 3: Construction of the fire exposure instrument \hookleftarrow

(b) Direction from origin to destination

Note: The pink lines on top are fractions of time during the day when the wind was blowing in that bin.



Figure 4: Trend in fire exposure, PM and GDP (2007-2013) \hookleftarrow

Note: Growth from the 2007 baseline value of each variable is plotted

Variable	Ν	Mean	SD	Min	Max			
Panel A: Monthly Fire Measures and Gro	undwater	Law (2002	2-2020)					
Count of fires	143640	5.340	32.720	0	1148			
Total Fire Radiative Power (mw)	143640	102.75	670.34	0	45044			
Groundwater Law Dummy	143640	0.065	0.247	0	1			
Panel B: Exposure to Upwind November Fires (FRP-based) with distance cut-off (2007-2013)								
Nov FRP exposure, $cut-off = 500$	3731	35.167	86.607	0.020	675.725			
Nov FRP exposure, $cut-off = 600$	3731	38.924	87.728	0.053	675.902			
Nov FRP exposure, $cut-off = 700$	3731	42.431	88.123	0.063	675.920			
Nov FRP exposure, $cut-off = 800$	3731	45.904	88.126	0.085	675.985			
Nov FRP exposure, $cut-off = 900$	3731	49.451	87.799	0.087	676.289			
Nov FRP exposure, $cut-off = 1000$	3731	52.759	87.268	0.164	676.300			
Panel C: Annual Particulate Matter and	GDP (200	7-2013)						
Mean PM2.5 (micrograms/m3)	3731	62.517	27.678	17.828	147.946			
GDP (Billions of Rupees, Constant 2004)	3731	81.301	164.07	2.414	3728			
Panel D: Annual Weather (2007-2013)								
Mean Temperature (°C)	3731	25.011	3.767	-10.369	29.847			
Total Rainfall (mm)	3731	2165.9	430.47	0	2809			
Mean Relative Humidity (Ratio)	3731	0.640	0.081	0.388	0.852			
Mean Surface Pressure (kilo-pascal)	3731	96.85	4.946	56.460	100.83			
Mean Windspeed (m/s)	3731	1.437	0.598	0.329	3.831			

Table 1: Summary Statistics

Notes: All data is at the district level. The sample consists of 530 districts, except for Panel A which consists of 630 districts (out of 640 census 2011 districts). The reduction is due to ICRISAT GDP data only being available between 2007-2013 for a subset of districts. \leftarrow

	Fire (Count	Fire Radat	tive Power
	Pre-2009 Mean [SD]	(1)	Pre-2009 Mean [SD]	(2)
 Ianuary	1 881	-0 749***	18 459	-0 746***
January	[3 037]	(0.143)	[38, 224]	(0.154)
February	2.384	-0.659**	25.36	-0.809***
rebruary	[3, 759]	(0.000)	[55, 396]	(0.218)
March	[0.100] 9.11	-0.529***	$\begin{bmatrix} 00.000 \end{bmatrix}$	-0 771***
waren	[4 482]	(0.145)	[<u>81</u> 6]	(0.154)
April	[4.402] 20.527	1 00***	[01.0] 440.866	(0.104) 0.780***
Арт	[27, 007]	(0.260)	[601 855]	-0.109 (0.266)
May	[21.301] 62 546	(0.200) 0.430***	[001.000] 1330.016	(0.200) 0.286*
May	[72,012]	(0.118)	[1659, 531]	(0.157)
Juno	$\begin{bmatrix} 12.912 \end{bmatrix}$	(0.110) 0.253	13.74	0.040
June	0.494 [1 306]	(0.233)	15.74 [55.981]	(0.158)
Tultz	$\begin{bmatrix} 1.300 \end{bmatrix}$	(0.101) 0.542***	$\begin{bmatrix} 0.0.201 \end{bmatrix}$	(0.130) 0.726***
July	[0.149]	(0.106)	2.401	(0.265)
August	$\begin{bmatrix} 0.324 \end{bmatrix}$	(0.190) 1 10***	[9.059] 6.021	(0.200) 1.98***
August	[1, 077]	(0.280)	[10.750]	-1.20
Contombon	[1.077]	(0.209) 1.00***	[19.709] 59.210	(0.249) 1.04***
September	4.020	-1.83	38.319	-1.94
Ostalian	[10.900]	(0.162)	[145.041]	(0.169) 0.955***
October	192.287	-0.85(2940.382	-0.855
N	[208.394]	(0.118)		(0.138)
November	49.840	(0.11c)	(88.759 [9965 141]	(0.150)
	[130.000]	(0.110)	[2205.141]	(0.159)
December	3.084	-0.600	26.838	-0.526
	[3.997]	(0.162)	[40.39]	(0.175)
Observations	4018	$140,\!372$	4018	$140,\!372$
Pseudo R2		0.784		0.797
Years	2002-2018	2002-2018	2002-2018	2002-2018
Districts	41	630	41	630

Table 2: Poisson Estimates of Impact of Groundwater Laws on Monthly Fires

continued

State x Month FE	Х	Х
Year FE	Х	Х
District FE	Х	Х

Notes: Years 2002-2018. Columns 1 and 3 provide mean and SD of fire count and fire strength before 2009 in Punjab and Haryana. Columns labeled (1) and (2) provide Poisson estimates. Standard errors are clustered at district and State x Year. *p<0.1; **p<0.05; ***p<0.01. \leftarrow

		Dependent Variable: log(PM)						
	(1)	(2)	(3)	(4)	(5)	(6)		
Panel A: Fixed Effects M	Iodel							
log(Nov FRP Exposure)	0.011^{***} (0.004)	0.016^{***} (0.004)	0.022^{***} (0.005)	0.027^{***} (0.006)	0.029^{***} (0.006)	0.028^{***} (0.006)		
Observations Within R2	$3,731 \\ 0.539$	$3,731 \\ 0.542$	$3,731 \\ 0.546$	$3,731 \\ 0.550$	$3,731 \\ 0.551$	$3,731 \\ 0.549$		
Panel B: First Difference	es Model							
log(Nov FRP Exposure)	0.008^{***} (0.003)	0.009^{***} (0.003)	0.010^{***} (0.003)	0.031^{***} (0.006)	0.032^{***} (0.006)	0.031^{***} (0.006)		
Observations Within R2	$3,178 \\ 0.171$	$3,178 \\ 0.172$	$3,178 \\ 0.173$	$3,201 \\ 0.197$	$3,201 \\ 0.197$	3,201 0.193		
Distance Cutoff	[500 km]	[600 km]	[700 km]	[800 km]	[900 km]	[1000 km]		
Weather Controls District and Year FE	X X	X X	X X	X X	X X	X X		
District x Time Trend	Х	Х	Х	Х	Х	Х		

Table 3: Impact of distance-weighted November fire exposure on annual PM2.5

Notes: Years 2007-2013. The sample is limited to districts for which GDP data is available. Each column of Panel A and B provides estimates from the same regression specification but with a different distance cut-off when constructing the FRP exposure instrument. Estimates in each panel are equivalent to the first stage for columns 3 and 5 in table 5. Standard errors are clustered at district and Region x Year. *p<0.1; **p<0.05; ***p<0.01. \leftarrow

	Dependent Variable: $\log(PM)$					
Exposure Month	Jan	Feb	Mar	Apr	May	Jun
Panel A: Estimates for Janua	ary to June					
log(Monthly FRP Exposure)	0.007	0.017^{***}	0.011^{**}	-0.002	0.007	-0.006***
	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)	(0.002)
Observations	3,731	3,731	3,731	3,731	3,731	3,718
Within R2	0.535	0.540	0.537	0.534	0.535	0.537
	Dependent Variable: log(PM)					
Exposure Month	Jul	Aug	Sep	Oct	Nov	Dec
Panel B: Estimates for July t	to December					
log(Monthly FRP Exposure)	-0.004	-0.002	0.003	-0.010*	0.029***	0.012^{*}
	(0.003)	(0.003)	(0.003)	(0.005)	(0.006)	(0.007)
Observations	3,697	3,726	3,730	3,731	3,731	3,731
Within R2	0.534	0.534	0.535	0.536	0.551	0.536
Distance Cutoff	[900 km]	[900 km]	[900]	[900 km]	[900 km]	[900 km]
Weather Controls	Х	Х	Х	Х	Х	Х
District and Year FE	Х	Х	Х	Х	Х	Х
District x Time Trend	Х	Х	Х	Х	Х	Х

Table 4: Impact of distance-weighted Monthly fire exposure on annual PM2.5

Notes: Years 2007-2013. The sample is limited to districts for which GDP data is available. Each single column in Panel A and B displays estimates for the regression of annual PM2.5 on exposure to fires during that month of the year only, using the same specification as in table 3. Standard errors are clustered at district and Region x Year. *p<0.1; **p<0.05; ***p<0.01. \leftarrow

	Dependent Variable							
		$\log(\text{GDP})$		$\Delta \log(\text{GDP})$				
	(1)	(2)	(3)	(4)	(5)			
Panel A: OLS and 2SLS	' Results							
$\log(PM2.5)$	0.147^{***} (0.035)	-0.008 (0.016)	-0.159 (0.097)	-0.030^{**} (0.014)	-0.179^{***} (0.069)			
Observations R2	$3,731 \\ 0.996$	3,731 0.999	$3,731 \\ 0.999$	$3,201 \\ 0.379$	$3,201 \\ 0.326$			
Weather Controls District and Year FE District x Time Trend	X X	X X X	X X X	X X	X X			
First Differences 2SLS Estimate			Х	Х	X X			
Panel B: First Stage Res	sults							
log(Nov FRP Exposure)			0.029^{***} (0.006)		0.032^{***} (0.006)			
Cragg-Donald F-stat Kleibergen-Paap F-stat			101.4 25.3		116.5 26.4			

Table 5: Impact of Air Pollution (PM2.5) on GDP

Notes: Years 2007-2013. The sample is limited to districts for which GDP data is available. Panel A, columns 1-3, show estimates for both OLS and 2SLS regressions of log GDP level on log PM2.5, starting without a time trend, then controlling for a time trend and finally conducting 2SLS with time trend. Columns 4 of panel A shows an OLS estimate using first differences while column 5 instruments for first difference of log PM with first difference of log Nov Exposure (900 km cut-off). Standard errors are clustered at district and Region x Year. *p<0.1; **p<0.05; ***p<0.01. \leftrightarrow

9 Appendix

		Fire Count			Radative P	ower
	(1)	(2)	(3)	(4)	(5)	(6)
January	-0.143**	-0.747***	-0.746***	-0.164***	-0.765***	-0.918***
	(0.059)	(0.133)	(0.078)	(0.054)	(0.156)	(0.080)
February	-0.368***	-0.737***	-0.786***	-0.407***	-0.867***	-0.949***
	(0.080)	(0.252)	(0.205)	(0.077)	(0.202)	(0.229)
March	-0.398***	-0.668***	-0.914***	-0.500***	-0.957***	-1.27***
	(0.049)	(0.142)	(0.080)	(0.057)	(0.156)	(0.137)
April	-0.790***	-1.06***	0.198	-0.735**	-0.766***	0.028
	(0.289)	(0.254)	(0.186)	(0.335)	(0.263)	(0.183)
May	-0.118*	-0.534***	-0.097	-0.072	-0.410***	-0.086
	(0.059)	(0.118)	(0.076)	(0.072)	(0.158)	(0.080)
June	-0.078	0.243	0.591^{***}	-0.233***	0.054	0.577^{***}
	(0.052)	(0.165)	(0.170)	(0.077)	(0.157)	(0.118)
July	-0.152*	0.517^{***}	1.20***	-0.159***	0.698***	1.55^{***}
	(0.076)	(0.150)	(0.401)	(0.055)	(0.220)	(0.224)
August	-0.495***	-1.05***	-0.035	-0.699***	-1.25***	0.333^{*}
	(0.082)	(0.295)	(0.070)	(0.065)	(0.254)	(0.182)
September	-1.09***	-1.95***	-1.42***	-1.28***	-2.07***	-1.52***
	(0.258)	(0.193)	(0.144)	(0.250)	(0.195)	(0.141)
October	-0.223***	-0.817^{***}	-0.380***	-0.190*	-0.821***	-0.535***
	(0.074)	(0.119)	(0.070)	(0.107)	(0.159)	(0.079)
November	1.05^{***}	0.509^{***}	0.238***	1.15^{**}	0.613***	0.124
	(0.382)	(0.120)	(0.081)	(0.430)	(0.161)	(0.091)
December	-0.370**	-0.677***	-0.323**	-0.359*	-0.640***	-0.355***
	(0.154)	(0.157)	(0.132)	(0.178)	(0.166)	(0.103)
Observations	56,082	149,257	43,904	56,082	$149,\!257$	43,904
Specification	OLS	Poisson	Poisson	OLS	Poisson	Poisson

Table A1: Impact of Groundwater Laws on Monthly Fires in Punjab and Haryana - Robustness

continued

Years	2002-2018	2000-2018	2007 - 2013	2002-2018	2000-2018	2007-2013
Districts	630	630	630	630	630	630
State x Month FE	Х	Х	Х	Х	Х	Х
Year FE	Х	Х	Х	Х	Х	Х
District FE	Х	Х	Х	Х	Х	Х

Notes: Provides robustness checks for table 2. Columns 1 and 4 conduct OLS estimation with log(fire count) and log(FRP) as the dependent variables. Columns 2 and 5 conduct the Poisson estimation with fires data from 2000 and 2001, when the fires are less reliable. Columns 3 and 6 conduct Poisson estimation by restricting sample to data from the 530 districts over 2007-2013 which have GDP data available. Standard errors are clustered at district and State x Year. *p<0.1; **p<0.05; ***p<0.01. \leftrightarrow

		Dependent Variable: log(PM)						
	(1)	(2)	(3)	(4)	(5)	(6)		
Panel A: Fixed Effects M	Iodel							
$\log(Nov FRP Exposure)$	0.014^{***} (0.004)	0.021^{***} (0.005)	0.028^{***} (0.006)	0.032^{***} (0.006)	0.032^{***} (0.006)	0.028^{***} (0.006)		
Observations Within R2	$3,731 \\ 0.542$	$3,731 \\ 0.545$	$3,731 \\ 0.551$	$3,731 \\ 0.554$	$3,731 \\ 0.553$	$3,731 \\ 0.549$		
Panel B: First Differences Model								
$\log(Nov FRP Exposure)$	0.010^{***} (0.003)	0.011^{***} (0.003)	0.011^{***} (0.003)	0.034^{***} (0.006)	0.033^{***} (0.006)	0.030^{***} (0.006)		
Observations Within R2	$3,178 \\ 0.173$	$3,178 \\ 0.174$	$3,178 \\ 0.174$	$3,201 \\ 0.201$	$3,201 \\ 0.198$	3,201 0.191		
Distance Cutoff	[500 km]	[600 km]	[700 km]	[800 km]	[900 km]	[1000 km]		
Weather Controls	Х	Х	Х	Х	Х	Х		
District and Year FE	Х	Х	Х	Х	Х	Х		
District x Time Trend	Х	Х	Х	Х	Х	Х		

Table A2: Impact of November fire exposure without distance weighting on annual PM2.5

Notes: Years 2007-2013. Robustness to dropping distance from construction of exposure instrument in table 3. The sample is limited to districts for which GDP data is available. Each column of Panel A and B provides estimates from the same regression specification but with a different distance cut-off when constructing the FRP exposure instrument. Standard errors are clustered at district and Region x Year. *p<0.1; **p<0.05; ***p<0.01. \leftarrow