

Air pollution and agricultural productivity in a developing country*

Joshua D. Merfeld[†]

2023-12-02

Abstract

I document negative externalities of air pollution in the Indian agricultural sector. Using variation in pollution induced by changes in wind across years, I show that higher levels of pollution lead to decreased agricultural productivity. The negative effects of pollution are larger in areas growing more labor-intensive crops, indicating that the pollution works partly through direct effects on labor productivity. I also show that pollution from coal plants has a larger effect on agricultural productivity than other types of pollution. A back-of-the-envelope calculation indicates that aggregate output losses from pollution are substantial; 2011 total output would have been between 2.3 and 4.5 percent higher if pollution had remained at 2001 levels, depending on the crop.

Keywords: pollution, productivity, agriculture, labor, India

JEL Codes: H40, I15, J22, O13, Q52, Q53

*I thank Ariel Ortiz-Bobea, Marc Bellemare, and David Sungho Park for comments and suggestions that greatly improved the quality of the paper, as well as conference/seminar attendees at the 2023 KER International Conference, Seoul National, the University of Reading, and the EEA. The usual disclaimer applies.

[†]KDI School of Public Policy and Management and IZA; merfeld@kdis.ac.kr

1 Introduction

In 2021, energy generation from coal reached an all-time high,¹ driven in part by rapid population growth in developing countries, including India. While natural gas has become a popular alternative in much of the world, the Indian government met much of the increased demand through the construction of coal power plants. From 1990 to 2010, India’s population grew by more than 40 percent, leading to a large increase in demand for power.² During those same two decades, power generation from all coal units of at least 30 megawatts in the country more than doubled, from 42.4 gigawatts to more than 100 gigawatts.³

While electricity may lead to economic and productivity growth (Dinkelman 2011; Kline and Moretti 2014; Rud 2012; Van de Walle et al. 2017), though electrification alone may not be a sufficient condition (K. Lee, Miguel, and Wolfram 2020), these improvements do not come without downsides, especially when it comes to power derived from coal. Emissions from the production and burning of coal – pollutants like sulfur dioxide, nitrogen oxides, methane, and mercury, for example – are also directly harmful to human health and the environment. Although the use of coal had been declining, recent events – including Russia’s invasion of Ukraine – have led to an increase in the demand for coal.⁴

Pollution more generally is also increasing in much of the world. Carbon dioxide equivalent per capita has almost doubled over the last 50 years in India, more than doubled in Indonesia, and almost tripled in China.⁵ Much of this increase is due to the development process itself – as countries develop, their economies transform, with more people moving out of agriculture and into manufacturing and services (Herrendorf, Rogerson, and Valentinyi 2014). Despite this development, agriculture remains a key sector in many developing countries, employing a large proportion of the poor (Dercon 2009), and agriculture is a key sector for poverty reduction (Christiaensen, Demery, and Kuhl 2011). This raises an important question: do rising pollution levels in developing countries threaten agricultural productivity?

In this paper, I study the effects of pollution on agricultural productivity in India. Previous work has suggested that there may be harmful direct effects of air pollution (Heck et al. 1982; Marshall et al. 1997) and water pollution (Reddy and Behera 2006) on agricultural productivity, including work showing that gold mines – which release large amounts of pollutants into the surrounding environment – lead to lower agricultural productivity in Ghana (Aragón and Rud 2016). These effects can work through the well-known effects on human health and productivity, but also directly through land productivity. Black carbon in soot,

¹www.iea.org/reports/coal-fired-electricity

²data.worldbank.org

³globalenergymonitor.org/

⁴www.npr.org/2022/08/15/1117560560/a-rising-demand-for-coal-amidst-war-in-ukraine

⁵<https://edgar.jrc.ec.europa.eu>

for example, can directly affect solar radiation reaching plants on the surface (Ramanathan and Carmichael 2008; Burney and Ramanathan 2014), while, on the other hand, sulfur in some pollution sources can actually be productivity enhancing (Sanders and Barreca 2022). Relative to previous work in economics, this paper tackles pollution more broadly, before turning to a specific source of pollution: coal plants. I show that air pollution leads to lower agricultural productivity in India and that pollution from coal plants is particularly harmful to agriculture.

Identifying the causal effects of pollution is difficult. For starters, the construction of coal plants is likely to be endogenous, though the relationship between construction and economic conditions is not a priori clear. On the one hand, governments may locate new plants in fast-growing areas, while on the other hand, they may decide the exact location of plants based on the political power of local citizens (Kopas et al. 2020). In the case of the data I use here, coal plants are constructed in areas with higher population and higher levels of pollution. In terms of agriculture, higher productivity predicts more coal plants in a given year as well as the construction of a coal plant over the next decade, indicating possible bias when estimating effects of pollution on agricultural productivity.

To identify the effects of air pollution on agricultural productivity, I identify areas of high pollution emission using the location of coal plants. While coal plants may add emissions, I show that they are located in areas with relatively higher levels of pollution, in general. Using these locations as point sources of pollution, I use plausibly exogenous changes in wind direction – in a similar spirit to Deryugina et al. (2019) – to measure the effects of pollution on key agricultural outcomes. I first plot the location of coal plants in India by year. Using a village-level shapefile, I identify all villages that are located within 30km of a coal plant at any time during the sample period, which comprises the 2000s to the 2010s. I then calculate the direction from each coal plant to all villages within 30km. I include all areas, regardless of whether the coal plant has opened yet, to allow me to disaggregate the effects of pollution from coal plants with other kinds of pollution later in the paper.

Using daily wind data and these spatial variables, I create an identifier separately for every village for whether a given village is located downwind from a coal plant on any given day. I then aggregate this exposure variable based on the desired analysis; for agricultural productivity, for example, I create an exposure variable measuring the total number of exposure days in the five-month agricultural season, focusing only on the monsoon (Kharif) season. More days being located downwind from a high-pollution location leads to substantially higher average concentrations of particulate matter 2.5 (PM 2.5), which is defined as particulate matter with a diameter of less than 2.5 micrometers.

Using wind direction as an instrument for particulate matter, an increase of PM 2.5 by one percent leads to a large decrease in agricultural productivity, with an elasticity of approximately -0.63 in the most conservative specification. The IV estimates indicate that the effect of pollution is more negative than naïve OLS estimates, which is consistent with the evidence on the construction of coal plants; higher levels of pollution are correlated with higher levels of agricultural productivity, a relationship that the IV strategy adjusts downwards. Robustness checks confirm these main results. Randomization inference shows that the estimates are robust to alternative measures of uncertainty, while levels of pollution do not affect current agricultural productivity.

Heterogeneity analysis shows that the negative effects of pollution are relatively similar based on median exposure to pollution but larger in villages with higher initial yields. Finally, given how many poor farmers rely on rainfall and how weather shocks are on the rise across the globe, I also look at possible multiplicative effects of shocks by interacting the pollution variables with a rainfall variable. Higher exposure along with less rainfall increases the negative effect of either shock individually, meaning these shocks are compounding. Given concerns regarding agricultural productivity and climate change, the rise of pollution in India may be particularly problematic for poor farmers.

A key question is the mechanism: does pollution reduce labor productivity, land productivity, or both? While previous work from the United States indicates that pollution from coal plants, specifically, can lead to higher *land* productivity (Sanders and Barreca 2022), agriculture in India differs greatly from agriculture in the United States. To test for possible effects on labor productivity, I look at differential effects based on labor intensity of the predominant crops in different villages. Specifically, I separate effects in predominantly rice-growing areas from other areas; since rice is very labor intensive, we would expect to see larger negative effects of pollution in rice-growing regions relative to other regions if one mechanism through which pollution affects productivity is direct effects on labor. I show that pollution negatively affects agricultural productivity in non-rice areas but that the effects in rice-growing areas are around 50 percent larger in magnitude. These results indicate that pollution likely directly affects labor productivity, though I am not able to rule out direct effects on land, either, due to a lack of data.

The final set of analyses disentangles the effects of pollution from the effects of coal pollution, specifically, by using wind direction and the rollout of coal plants across the country. Combining instrumental variables with a differences-in-differences strategy, I show that agricultural productivity has an elasticity of around -0.22 with respect to non-coal pollution. The opening of a coal plant, however, doubles the elasticity, to more than -0.44, indicating that coal plant emissions are particularly harmful to agricultural productivity, at least in India. This is consistent with coal emissions being quite different from other forms of pollution; coal

emissions contain different types and quantities of harmful substances, meaning effects may differ greatly.

I carry out a simple back-of-the-envelope calculation to determine how much pollution affected agricultural output in the country as a whole in 2011. Using the most conservative estimate of the effect of pollution on agricultural productivity, I find that rice output would have been 2.3 percent higher, maize output would have been 4.5 percent higher, and wheat output would have been 3.8 percent higher in 2011 if pollution had remained at 2001 levels.

An important caveat to these results is that I cannot disentangle exactly which pollutants drive results. While I use PM 2.5 as a measure of overall exposure to pollution, effects could be driven by any pollutants carried by the wind. This leaves open an important avenue of future research regarding the effects of different types of pollutants on agricultural productivity.

The results are consistent with evidence of effects of pollution on different forms of labor activity. For example, general levels of pollution decrease productivity of call center works in China (Chang et al. 2019), lead people to perform worse on cognitive functions (Ebenstein, Lavy, and Roth 2016; Wen and Burke 2022), lower labor market earnings (Borgschulte, Molitor, and Zou 2022), and also lead to lower levels of farm labor productivity in California, driven by changes in ozone concentration (Graff Zivin and Neidell 2012). Also in China, He, Liu, and Salvo (2019) find small but significant negative effects of pollution on manufacturing productivity, while Chen, Oliva, and Zhang (2022) show pollution leads to labor out migration. Relevant to my results, Hanna and Oliva (2015) show that decreases in pollution lead to increases in labor allocation in Mexico.

Several papers are closely related to this one. Aragón and Rud (2016) show that gold mines in Ghana have substantial negative impacts on agricultural productivity in the surrounding areas. Burney and Ramanathan (2014) model the effects of climate change and pollutants on agricultural productivity in India. Finally, Sanders and Barreca (2022) actually find that coal emissions can be good for agriculture in the United States, due to the presence of sulfur, which can increase agricultural productivity. This paper differs from the previous literature in three important ways. First, I explore effects of pollution more generally as well as from coal plants, which are much more ubiquitous than gold mines across the world. Second, I use a different identification strategy from most other papers, relying on wind blowing pollution from coal plants. Finally, agriculture in India and the United States can be quite different, with the latter being much more mechanized. This means effects of pollution could differ substantially depending on the context, a conclusion that the current results support.

2 Data

The main goal of this paper is to examine whether exposure to pollution affects agricultural productivity, as well as to provide evidence about the effects of pollution from coal plants, in particular. I use several sources of data, which I describe here and list in Table A1 of the appendix.

The first set of data lists the location of coal plant across the globe. This data comes from Global Energy Monitor⁶ and lists all units generating at least 30 megawatts of electricity. The data on coal plants includes the year of opening (and, if applicable, the year of retirement), the GPS (latitude/longitude) location of each plant, and the amount of power produced by the units at each plant. For this paper, I do not use the information on the capacity of the plant.

Figure A1 in the appendix displays the location of coal plants in India for two specific years: 1990 and 2010. There is a clear increase in the prevalence of coal plants across the country over the two decades. Additionally, the overall capacity from coal plants increased from 42.4 gigawatts to 100.4 gigawatts, an increase of 136.6 percent in just 20 years. This increase is driven both by the roll out of new plants as well as the construction of new units at existing plants, the latter of which is not shown in Figure A1.

The second dataset includes agricultural productivity for both the monsoon and winter seasons, from 2002 to 2018. To match other data, I only use the data up to 2013. This data comes from Gangopadhyay et al. (2022) and estimates land productivity (i.e. yield, in tons per hectare) for the major crops in India, using satellite-derived data. While there is a wealth of literature on measurement error in land (Desiere and Jolliffe 2018; Abay et al. 2022), the authors use a static cropland mask for the productivity estimates. This means that any measurement error will hopefully be constant across time and subsumed by the village fixed effects. While the measure of output is also noisy, the use of satellite-derived data and a static cropland mask reduces concerns about measurement error being responsible for the results below, especially since common sources of measurement error are unlikely to be correlated with wind direction. The authors define the monsoon season as June to October and the winter season as November to March. Due to large differences in the crops grown and the production technologies used across seasons, I focus on only the monsoon season and keep the authors' definition when matching across data below. The data on agricultural productivity is publicly available from Nature's data-sharing website.⁷ Since the main regressions below use village fixed effects, any differences in cropping patterns across areas should not bias estimates on changes, as long as these patterns do not change markedly in response to annual pollution levels.

The agricultural productivity data is available as raster files with a resolution of 500m. To aggregate this

⁶globalenergymonitor.org/projects/global-coal-plant-tracker/

⁷springernature.figshare.com

data up to a useful administrative unit, I use village-level shapefiles provided by Asher et al. (2021) and publicly available on the SHRUG platform.⁸ I aggregate the agricultural productivity data to the village level by extracting mean productivity for each feature in the shapefile. I do this separately for the monsoon season each year, from 2002 to 2013.

To measure exposure to pollution from point sources around coal plants, I first locate all village centroids located within 30km from a coal plant in a given year. I choose 30km due to previous research on the effects of (air) pollution (Aragón and Rud 2016; Bencko and Symon 1977; Li and Gibson 2014). I then calculate the direction from these point sources to all village centroids within that 30km radius. To define exposure, I pull daily wind direction data from the National Center for Atmospheric Research.⁹ For each day, I document whether the wind is blowing towards each village centroid¹⁰ and then temporally aggregate this daily data to match the data on productivity. Specifically, agricultural productivity is defined across five months (i.e. the monsoon season is from June to October) so I count the total days a given village is exposed to wind blowing from *any* coal plant within 30km during those five months.

The top panel in Figure A2 in the appendix is an example wind direction raster. The raster shows wind direction for the entirety of India on January 1st, 2010, as well as the location of coal power plants on that date. The prevailing winds on the date differ across the country and, though not shown in the figure, across days. This means that the overall exposure for a given area to coal plant emissions changes across time.

For pollution data, I use estimated particulate matter (PM). Specifically, I use estimates of PM 2.5, which includes particles with a diameter of 2.5 micrometers or less. This data comes from Hammer et al. (2020) and is available as a raster with a resolution of 0.01 degrees. I aggregate this data to the village level by extracting the mean PM for each feature in the shapefile. I do this separately for each month from June 2002 to October 2013. The bottom panel in Figure A2 in the appendix is an example of the PM data. The figure shows the estimated PM 2.5 from for each village in Guna district of Madhya Pradesh on January 1st, 2010.

While I use PM 2.5 as the estimated level of pollution, it is important to clarify that I cannot explicitly say that PM 2.5, per se, is responsible for any effects. Pollution includes a variety of different components, including PM 2.5, PM 10, sulfur dioxide, and nitrogen dioxide, to name but a few. While I use PM 2.5 as a proxy for pollution, it is possible that other components are responsible for the effects. Additionally, the PM 2.5 data is estimated using satellite data and a machine learning algorithm. While this is a common approach in the literature (see Hammer et al. (2020) for more details), it is important to note that the

⁸www.devdatalab.org/shrug_download/

⁹climatedataguide.ucar.edu/

¹⁰I define “towards” as within five degrees to help take into account that I am using village centroids. The wind data includes both the x and y components of wind speed, which I can use to calculate angles from point sources to village centroids.

estimates are not direct measurements of pollution. However, a valid instrumental variable can help correct for any measurement error present in the data (Glewwe 2012; Gillen, Snowberg, and Yariv 2019).

To help understand some of the mechanisms driving the relationship between exposure to emissions and agricultural productivity, I also look at the differential effects by the labor intensity of the predominant crop in each area. For this, I use data from Monfreda, Ramankutty, and Foley (2008), who estimate area harvested for different crops across the global around the year 2000. I identify areas where rice (paddy) is the most commonly grown crop, under the assumption that rice is more labor intensive than other crops (Vollrath 2011; Michler 2020). Michler (2020), for example, shows that rice, on average, uses almost 1,800 hours per hectare, while the second-highest crop in his data, cotton, uses just 850. He also estimates that wheat, another common crop in India, uses just 600 hours per hectare.

3 Methods

I am primarily interested in the effects of pollution on agricultural productivity. We might consider a regression of the following form:

$$y_{it} = \alpha_i + \gamma_t + \beta PM_{it} + \delta rain_{it} + \psi temp_{it} + \varepsilon_{it}, \quad (1)$$

where y_{it} is outcome y for unit i at time t , α_i is a geographic fixed effects, γ_t is time fixed effects, PM_{it} is the (log) concentration of particulate matter 2.5, $rain_{it}$ is total rainfall during the season, and $temp_{it}$ is average temperature during the season.

However, even with the fixed effects, this regression likely suffers from endogeneity; places with more pollution may be very different from places with less pollution. It is difficult to sign the direction of the bias a priori, because it depends on the local context and how the coal plants are placed. For example, if coal plants are placed in places in need of power – for example, due to rapid growth – then the estimates might be upwards biased. On the other hand, if officials purposefully place coal plants in poorer areas for political reasons, then the estimates might be downwards biased.

What are the characteristics of areas near the construction of new coal plants? Table 1 presents four separate regressions, using state-level fixed effects due to the spatial variation in coal plants. The first two columns use 1991 census data – downloaded from SHRUG – and the second two columns use 2001 census data, all at the village level. The first column presents results from regressing a dummy for whether there is a coal plant within the village polygon in 1991 on several census variables. It seems that the exact location of a coal plant

Table 1: Local characteristics and the construction of coal plants

	1991 census		2001 census	
	1991	2001	2001	2011
Population (log)	-0.107 (0.097)	0.291*** (0.083)	0.241*** (0.079)	0.128 (0.080)
Area (log)	0.240*** (0.065)	-0.110 (0.088)	0.133* (0.079)	0.119 (0.085)
Ag. productivity			0.092 (0.070)	0.121 (0.097)
Pollution (log PM 2.5)			0.612*** (0.201)	0.530** (0.256)
Has plant?		14.842*** (0.568)		13.574*** (0.158)
Observations	209,110	209,110	408,913	415,334

Note: State fixed effects are included in regressions. Robust standard errors are in parentheses. The outcome in the first column is whether the village is within 30km of a coal plant in 1991. The second column is whether a village in 1991 will have a coal plant in 2001, conditional on not having one in 1991. The last two columns are similarly defined, except using 2001 and 2011 as the years. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

does not correlate with overall population of the village, but it does correlate with area; in other words, coal plants are more likely to be located in larger villages but, conditional on area, not in places with more or fewer people.

The second column keeps the 1991 census data but changes the outcome variable to a dummy for whether there *will be* a coal plant constructed within 30km before 2001. This column restricts the sample to only villages that do not already have a coal plant within 30km in 1991. We see that the pattern of coefficients changes from column one to column two; higher populations predict the construction of coal plant in the village over the next ten years. Note that I do not have information on productivity and pollution from 1991.

The last two columns are particularly relevant to the current results. They present similar results to those in the first two columns, but use the 2001 census data and a dummy for the presence of a coal plant in 2001 and 2011 in columns three and four, respectively. Importantly, I include agricultural productivity in the monsoon season of 2002 – the earliest year available – as well as average pollution during the monsoon season for 1998 through 2000 as additional predictors. Pollution is a strong predictor for the presence of a coal plant in 2001 as well as for whether a coal plant will be constructed before 2011. While the coefficient is small, it is actually relatively large given that relatively few villages will get a new plant in the next decade. Agricultural productivity is positively correlated with the presence of a coal plant, but not significantly so.

Since pollution is likely to be endogenous, a naïve OLS regression will not return unbiased estimates. The experimental ideal would be to randomly assign pollution to villages, but this is obviously not possible.

Instead, I use an instrumental variables (IV) approach to identify the causal effects of pollution on agricultural productivity. The key identifying assumption is that wind direction is exogenous from year to year, conditional on village fixed effects and weather. While the location of pollution sources might be endogenous, the amount of pollution received by a specific village in a given year can vary randomly based on wind direction. By counting the number of days in a given year that a village is downwind from a polluted area and conditioning on village fixed effects and weather (rainfall and temperature), variation across years is plausibly exogenous. Since I include village fixed effects, identification comes from within-village changes in wind direction across years.

Consider the reduced form effect of wind exposure on agricultural productivity:

$$y_{it} = \alpha_i + \gamma_t + \beta wind_{it} + \delta rain_{it} + \psi temp_{it} + \varepsilon_{it}, \quad (2)$$

where variables are defined identically to above but with wind as the primary variable of interest instead of pollution. Combining these into a two-stage procedure, we can instrument for pollution using wind direction; small changes in wind direction can lead to large changes in pollution exposure for villages located near point sources of pollution, allowing me to identify the causal effects of pollution by isolating the random component driven by wind direction. This is similar in spirit to Deryugina et al. (2019).

Identification relies on changes in wind direction across time for the same geographic units. In other words, conditional on the fixed effects and weather, I assume changes in wind direction are as good as random and uncorrelated with the outcome – primarily agricultural productivity in this paper – except through exposure to pollution.

The inclusion of rainfall and temperature is meant to reduce concerns that wind direction may not affect agricultural productivity but, instead, be correlated with it through changes in weather. For example, prevailing winds from one direction may be associated with dry weather, while winds from another direction may be more likely to carry moisture and rain. It is not a priori clear how this would bias the estimates with respect to pollution and coal plants, however. The differences in weather patterns is very context specific and can differ even for villages located near one another.

I believe there are two reasons this is unlikely to be the case here. First, villages located near one another have very similar weather patterns, but wind direction leads to large differences in pollution exposure even for villages right next to one another. Second, in some specifications, I include very flexible specifications for weather. These inclusions do not lead to very large changes in the estimated relationship between pollution

and agricultural productivity, suggesting that weather is unlikely to be responsible for a spurious correlation between productivity and pollution in the main results.

It is worth taking the time to validate the use of wind direction as a measure of exposure to pollution. Consider the data used for the regressions presented in Table A3 of the appendix, which is the first stage of the 2SLS procedure. The outcome variable is particulate matter – specifically, PM2.5, which is particulate matter no larger than 2.5 micrometers in diameter – which comes from Hammer et al. (2020). PM2.5 is one of the harmful byproducts from many industrial processes, including power generation. Particulate matter is a major component of the pollution emitted by coal plants, along with sulfur dioxide, different types of nitrogen oxides, and mercury.¹¹ This data is available at the monthly level, but I aggregate variables to the season level since that is the level of analysis in the main results.

Both the particulate matter data and the wind data is at the village level, which means the regressions in Table A3 are at the village level. As such, the geographic fixed effects are village fixed effects and the temporal fixed effects are year fixed effects. Since the exposure variable is at the village level and we follow villages over time, the standard errors are also clustered at the village level. I add more and more flexible specifications for weather, with the most flexible specification in the fourth column. Note that the relationship between wind direction and particulate matter is very similar across all four columns, suggesting that the inclusion of weather does not change the relationship between wind direction and pollution. This is important because it suggests that the use of wind direction as a proxy for pollution exposure is not biased by weather patterns.

3.1 Identifying mechanisms

A key question is how pollution affects agricultural productivity. Specifically, does pollution directly affect labor productivity, land productivity, or both? Unfortunately, there is a lack of data that would allow me to directly test these mechanisms, since I do not have data on the amount of labor or land used in each village. Instead, I look at heterogeneity in the effects of pollution based on the predominant crop grown in each village.

Michler (2020) presents estimates of labor intensity for several major crops in India. The most labor-intensive crop is rice, which uses almost 1,800 hours per hectare. The second-most labor-intensive crop is cotton, which uses just 850 hours per hectare. Wheat, another common crop in India, uses just 600 hours per hectare. This suggests that pollution should be more harmful for rice than for other crops if there are direct impacts on labor productivity.

¹¹www.epa.gov/airmarkets/power-plants-and-neighborhood-communities

To identify this possible mechanism, I identify areas that grow predominantly rice. Specifically, I estimate the following regression:

$$y_{it} = \alpha_i + \gamma_t + \omega PM_{it} + \phi rice_{it} + \beta PM_{it} \times rice_{it} + \delta rain_{it} + \psi temp_{it} + \varepsilon_{it}, \quad (3)$$

where PM_{it} denotes the effect of pollution in non-rice areas and $PM_{it} \times rice_{it}$ is the differential effect in rice-growing areas. Here, I create an indicator variable for rice areas that is time invariant, so $rice_{it}$ drops out of the estimation. I instrument for PM_{it} and for $PM_{it} \times rice_{it}$ using wind and wind interacted with the indicator for rice.

While this is a useful first step, it is not a perfect identification strategy. First, there could be other differences in the types of areas that grow rice, and these differences could actually be responsible for any heterogeneity. Second, this strategy does not allow me to say anything about the effects of pollution on land productivity.

3.2 Identifying effects of pollution from coal plants

I use the location of coal plants to identify locations of high pollution emission, but there can be other pollution sources located near coal plants, as well. I nonetheless also want to identify the effects of pollution coming specifically from coal plants, in addition to the overall effects of pollution in the other parts of the paper. I do this by combining the two-stage procedure above with a differences-in-differences framework. Specifically, I create a new variable identifying the exact year the coal plant opens in a given area and calculate the change in the effects of pollution on agricultural productivity.

Consider the following regression of interest:

$$y_{it} = \alpha_i + \gamma_t + \omega PM_{it} + \phi open_{it} + \beta PM_{it} \times open_{it} + \delta rain_{it} + \psi temp_{it} + \varepsilon_{it}, \quad (4)$$

where the interaction term $PM_{it} \times open_{it}$ indicates the change in the effect of pollution following the opening of the coal plant. Note that this is not the same as saying that pollution increased; instead, it is focusing on the change in the effect of a change in pollution of one unit on the outcome of interest, agricultural yield. I instrument for both PM_{it} and $PM_{it} \times open_{it}$ using wind and the interaction between wind and the open indicator, or $wind_{it} \times open_{it}$.

Table 2: Wind direction and agricultural productivity

	(1)	(2)	(3)	(4)
wind days (10s)	-0.0078*** (0.0007)	-0.0065*** (0.0007)	-0.0051*** (0.0007)	-0.0053*** (0.0007)
weather	No	Yes	Yes	No
weather (expanded)	No	No	Yes	No
weather (expanded, bins)	No	No	No	Yes
fixed effects:				
village	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
observations	1,266,922	1,266,922	1,266,922	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated (log) agricultural yield, defined as tons per hectare. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

4 Results

I first present the reduced form results for agricultural productivity. The outcome for all regressions in this section is the log of agricultural land productivity, which is defined as tons per hectare. The first set of results are in Table 2. The unit of analysis is the village-season – from 2002 to 2013 – focusing only on the monsoon (kharif) season. The data used to construct the yield measure is calculated using five months of data, so the wind variable is defined for the same five months and is the sum of days in which the wind is pointed towards the village centroid from a coal plant.

The first column presents the most simple results, with only the wind exposure variable and the fixed effects. The second column adds a rainfall and average temperature variable. While I leave them out of the table for simplicity, the coefficients are in the expected direction: rainfall is positive and average temperature is negative. The third column adds higher-order weather variables for both rainfall and temperature, as well as the interaction between them. This attenuates the relationship between wind and productivity somewhat, but the coefficient remains negative and statistically significant.

The final column includes the most flexible form for the weather variables. The z-score for rainfall is turned into bins – with cutoffs at $z = -1$ and $z = 1$ – based on rainfall in each month of the season. In other words, over the five months, I create 15 different indicator variables for rainfall based on the z-score in each month. One of these indicators drops out for each month of the year, leaving 10. I also include interactions between all of these and temperature, as well as between the bins and temperature squared. The coefficient on wind remains negative and statistically significant, and is actually slightly larger in magnitude than the results in column three, which is a much less flexible specification for weather. This leads me to conclude that the results are not being driven by weather patterns.

To put the size of the coefficient in context, it is worth digging a little deeper into the exposure variable. Wind is scaled by 10 for presentation, meaning that an increase in *exposure* for a village from its mean leads to a 0.5 percent drop in agricultural productivity. The standard deviation of the within-village change in exposure to wind from polluted areas is 4.35, meaning a one-standard-deviation increase from the mean leads to a 0.22 percent decrease in agricultural productivity due to nothing but being located downwind from a pollution source.

Of course, it is not wind, per se, that we are interested in. Instead, we are actually interested in better understanding the effects of exposure to pollution on agricultural productivity. Table A2 in the appendix presents the same regressions as Table 2, with a key difference: instead of wind exposure, it uses log average concentration of PM2.5 – from Hammer et al. (2020) – during the season as the key explanatory variable. There is again a consistent, negative relationship between particulate matter and agricultural productivity. Since both variables are logged, the coefficients are similar to elasticities, indicating that a one-percent increase in pollution leads to an approximately 0.04 percent decrease in agricultural productivity in three of the columns.

Recall, however, the results from above: coal plants were placed in places with higher average agricultural productivity. This means that the simple regressions in Table A2 might actually be biased upwards; although pollution might negatively affect productivity, pollution is higher in more productive areas due to the endogenous placement of pollution sources in the study areas. One possible way to recover the actual effect of pollution on agricultural productivity is with an instrument. If, conditional on the fixed effects and weather variables, wind direction is exogenous and only affects agricultural productivity through pollution, then it is a valid instrument for pollution.

I present these IV results in Table 3, with the first-stage F-statistic at the bottom. The first-stage results – the full results are in Table A3 of the appendix – make clear how strong of an instrument wind is; all three F-statistics are above 1,000, meeting even the higher bar for the first-stage F-test argued for by D. Lee et al. (2022). The coefficient on wind in the first stage is stable at around 0.0085 across specifications.

The second-stage results first indicate that PM2.5 is a very strong predictor of agricultural productivity. Using the results in column three, a one percent increase in particulate matter leads to a 0.62 percent decrease in agricultural productivity. To put this in further perspective, the mean particulate matter concentration is just short of 40 (39.68) in the monsoon sample, while the within-village absolute deviation in particulate matter is 3.98, meaning year-to-year swings of several percent in agricultural productivity due to pollution exposure alone are common.

Table 3: Pollution and agricultural productivity, IV estimates

	(1)	(2)	(3)	(4)
Particulate matter (log PM 2.5)	-0.941*** (0.078)	-0.769*** (0.075)	-0.621*** (0.080)	-0.626*** (0.077)
Weather	No	Yes	Yes	No
Weather (expanded)	No	No	Yes	No
Weather (expanded, bins)	No	No	No	Yes
Fixed effects:				
Village	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
F (first stage)	1,236	1,266	1,205	1,246
Observations	1,266,922	1,266,922	1,266,922	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated (log) agricultural yield, defined as tons per hectare. * p<0.10 ** p<0.05 *** p<0.01

Column four presents one last specification; it includes a much more flexible specification for weather. This is the same binned specification as I describe above. The coefficient in column four shows almost not change from those in column three, again relieving some concerns that weather may not be correctly specified, leading to a spurious correlation between particulate matter and agricultural productivity.

The comparison of the results to those in the previous table also confirm the concerns around endogenous placement of pollution sources. The coefficients in Table 3 are substantially more negative than those in Table A2. In other words, the simple regression of agricultural productivity on pollution results in coefficients on pollution that are upwards biased, since higher levels of pollution occur in more productive areas, on average. The IV strategy corrects this bias and leads to a robust, negative relationship between pollution and agricultural productivity.

4.1 Robustness checks

In this section, I report the results of two separate robustness checks for the main results presented above. In the previous results, standard errors are clustered at the level of the village, which is also the level at which treatment (exposure to pollution) is assigned as well as the level at which agricultural productivity is extracted. Past work has shown that standard errors might be underestimated in such contexts (Bertrand, Duflo, and Mullainathan 2004). This is a particular concern in the present case, since agricultural productivity is extracted at the village level but estimated using data that has high spatial correlation (Gangopadhyay et al. 2022).

To help remedy concerns, I perform randomization inference (e.g. Barrios et al. (2012)) by randomly assigning treatment (wind and pollution) across villages and repeating this process 1,000 times. I do this separately by

Table 4: Agricultural productivity and pollution leads

	(1)	(2)
particulate matter (one-year lead)	-0.033 (0.067)	
particulate matter (two-year lead)		-0.070 (0.052)
weather (expanded)	No	No
fixed effects:		
village	Yes	Yes
year	Yes	Yes
F	592	783
observations	1,161,265	1,055,562

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated (log) agricultural yield (tons per hectare). * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

selecting from within district-years only, to help take into account district-year correlation in outcomes and as a way to mitigate possible concerns that traditional clustered standard errors may actually be much too conservative (Abadie et al. 2023). Since this is an instrumental variables procedure, I randomly select the wind and pollution variable *together*, from the same observation.

I present these results in Figure A3 of the appendix. The within-district correlation is apparent when selecting villages from within the same district in the same year. The distribution is approximately normal but has a mean and median of around -0.4. Nonetheless, the actual treatment effect is larger in magnitude than almost all of the bootstrapped replications; only three of the 1,000 replications results in a large coefficient than the true coefficient, leading to a p-value of well less than 0.01. The results are apparently robust to alternative specifications of uncertainty.

The second robustness check includes leads of pollution and wind direction. I instrument for future pollution using future wind direction, which mirrors the strategy employed in the main results. I present these results in Table 4. Neither a one-year nor a two-year lead predicts current agricultural productivity, which supports the identification assumptions.

4.2 Heterogeneity

Having established that exposure to pollution has a negative effect on agricultural productivity, I now move to the heterogeneity in these effects to better understand whether certain areas are driving the effects. Table 5 presents three different analyses: heterogeneity by median exposure, heterogeneity by initial yield, and possible multiplicative effects of shocks. All regressions use the IV strategy detailed above and I include the F

Table 5: Heterogeneity in the effects of pollution on productivity

	Wind		Yield		
	>p(50)	<=p(50)	>p(50)	<=p(50)	
Particulate matter (log PM 2.5)	-0.415*** (0.114)	-0.336*** (0.099)	-1.026*** (0.060)	-0.568*** (0.166)	-0.306*** (0.077)
PM times Rain					0.119*** (0.002)
Weather	Yes	Yes	Yes	Yes	Yes
Fixed effects:					
Village	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
F (1st stage, PM)	597	482	675	428	674
F (1st stage, PM times rain)					6,393
Observations	617,804	649,118	634,342	632,580	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated (log) agricultural yield, defined as tons per hectare. The first two columns split the sample based on the median exposure (by village). The second two columns split the sample based on initial – the first year available – agricultural productivity. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

statistic from the first stage regressions (of which there are two in the last column) at the bottom of the table.

The first two columns split the sample based on median exposure. In other words, for each village, I calculate the median exposure (based on wind) and then use this single value for each village to split the sample into villages above and below the median. The coefficient in the first column is only slightly larger than the coefficient in the second column, indicating that pollution does not have large differential effects based on exposure. One possible explanation is that exposure is not cumulative, but continues to be harmful regardless of the baseline level.

Columns three and four split the sample based on median yield at the very beginning of the sample (the first year available, 2002). Column three presents results for villages above the median while column four presents results for villages below the median. The largest effects are apparently concentrated in the upper half of the distribution, meaning the most productive areas are also most susceptible to the effects of pollution. This could have important implications for the overall effects of the placement of any point sources of pollution.

Given ongoing changes in the climate, a key question is how welfare will change as shocks become more and more common. The last column looks at possible multiplicative effects of shocks by interacting pollution with rainfall and seeing whether the effects compound one another. Since there are now two endogenous regressors, I instrument for both using wind exposure and an interaction between wind exposure and rainfall.

The coefficient on the level term for pollution is negative. Since rainfall is defined as a z score, this represents the effect of pollution only when rainfall is at its mean value for a given season. The interaction term, on the

other hand, is positive. Since higher values of rainfall are generally good, this positive coefficient represents a compounding effect of pollution and rainfall; when rainfall is low (bad), the effects of particulate matter on agricultural productivity is larger (more negative). Another way to think about this, however, is that higher values of rainfall dampen the effects of pollution shocks. This could have very profound effects on welfare, since this essentially increases the variance of agricultural productivity across years (which is already increasing due to climate change), at least as it relates to pollution exposure. Given the importance of variability to welfare (Dercon 2002; Merfeld and Morduch 2023), this could have important implications for the effects of pollution on welfare.

4.3 Mechanisms

A key question about the results is the mechanism. While previous results indicate that pollution may actually increase land productivity (Sanders and Barreca 2022), the results in that paper come from the United States, where agricultural practices vary greatly from those in India. On the other hand, we also know that pollution can affect labor productivity in other contexts (Chang et al. 2019; Ebenstein, Lavy, and Roth 2016; Graff Zivin and Neidell 2012; Wen and Burke 2022). Unfortunately, there is no available labor data that would allow me to directly test the effects of pollution on agricultural labor productivity. While the National Sample Survey is nationally representative, it does not allow one to back out productivity, only total labor allocation in the preceding week. Additional data options – like the ARIS-REDS or IHDS – are not collected frequently enough to enable the kind of analysis employed here.

Instead, I look at heterogeneity by the modal crop grown in each area. I use data from Monfreda, Ramankutty, and Foley (2008), who estimate area harvested by crop at a resolution of approximately 10km. These estimates come from around the year 2000, at the beginning of my sample period. I identify areas where rice is the most commonly grown crop. Rice (paddy) is commonly regarded as a labor-intensive crop in India (Vollrath 2011; Michler 2020), meaning that it may be more susceptible to pollution if some of the effects of pollution are through direct effects on labor productivity.

Figure A4 in the appendix shows the distribution of the most common crop across the villages in my sample. Though there are large areas where rice is the most commonly grown crop, there is also quite a bit of overlap, especially in the middle parts of the country. However, given the geographic concentration of rice production, I will implement a robustness check with state-by-year fixed effects below.

Table 6 presents the main results in this section. I continue to employ an IV strategy, interacting pollution with an indicator for rice and instrumenting this with an interaction between wind and rice. The first row

Table 6: Agricultural productivity by most common crop

	(1)	(2)
Particulate matter (log PM 2.5)	-0.287** (0.133)	-0.394*** (0.134)
PM 2.5 times rice	-0.338*** (0.061)	-0.205*** (0.060)
Weather (expanded)	Yes	No
Weather (expanded, bins)	No	Yes
Fixed effects:		
Village	Yes	Yes
Year	Yes	Yes
F (PM 2.5)	1,369	1,168
F (PM 2.5 times rice)	2,107	2,201
Observations	1,266,922	1,266,922

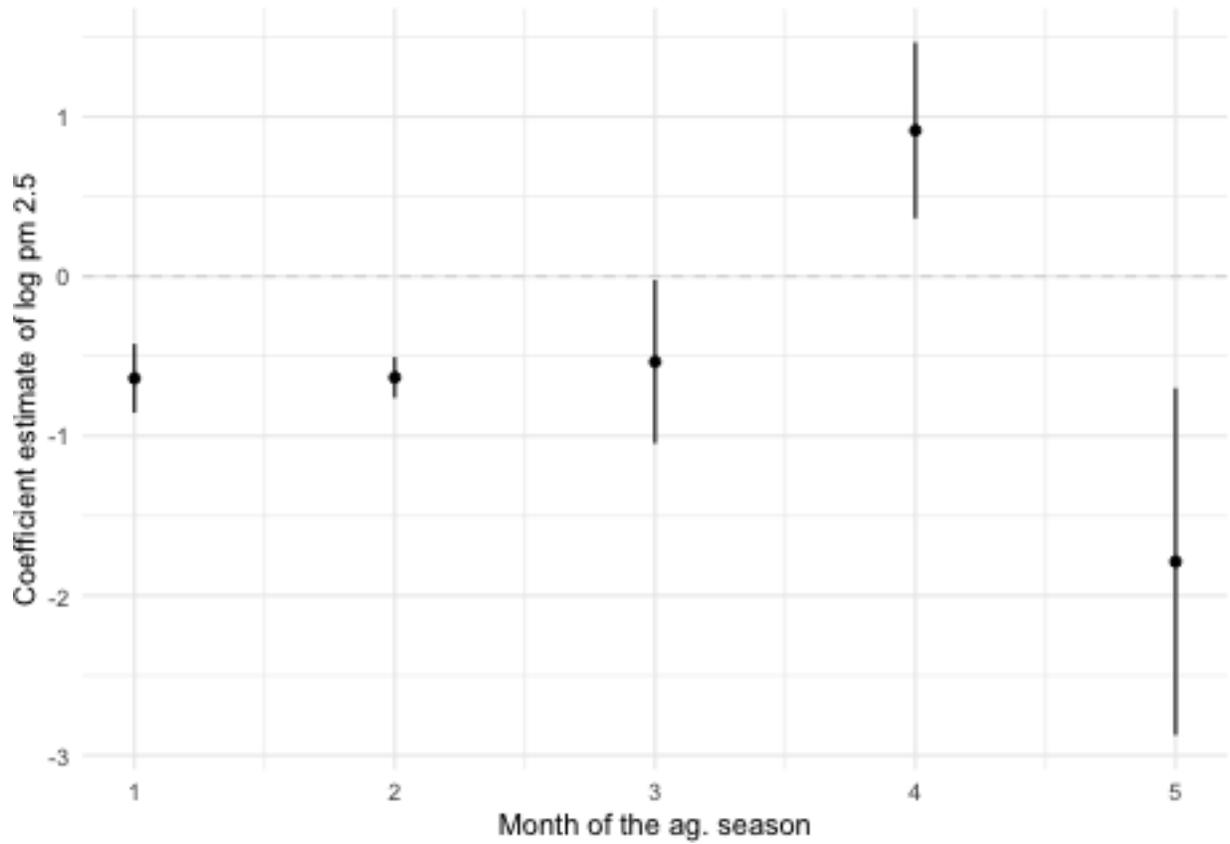
Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated (log) agricultural yield (tons per hectare). The rice dummy equals one in villages where rice is the most commonly harvested crop. * $p < 0.10$
** $p < 0.05$ *** $p < 0.01$

shows the effects of overall pollution on agricultural productivity in villages where rice is not the most commonly grown crop. Column two, which has the most flexible weather specification, indicates that an increase of particulate matter by one percent decreases productivity in these areas by around 0.4 percent. The second row is an interaction term between rice and PM 2.5, meaning the linear combination of the two coefficients gives the overall effect of pollution on productivity in rice-predominant villages. The coefficient of -0.2 shows that pollution decreases productivity by around 50 percent more in rice-growing villages than in non-rice growing villages. Since rice is more labor intensive, this is evidence that the effects of pollution work at least partly through the effects of pollution on labor productivity. This does not, however, rule out any direct effects on land productivity.

Given the geographic concentration of crops, Table A4 in the appendix presents the same results but with state-by-year fixed effects, instead of just year fixed effects. This will compare the effects of pollution within states in the same year, hopefully decreasing concerns that something other than labor intensity is the main driver of the heterogeneity documented in Table 6. The results are consistent with those presented here, and show that pollution decreases productivity more in rice-growing villages than in other villages.

Another way to tease out differential effects on labor productivity, specifically, is to look at the effects of pollution by month. If pollution is working through labor, then we would expect to see larger effects when labor is relatively more important. The harvest season, the fifth month of the agricultural season in the data, stands out as the most likely candidate for two reasons. First, labor allocation is generally quite high on plots

Figure 1: Effects of pollution on productivity by month



Coefficients are from a single regression, where the outcome is (log) agricultural yield and pollution in each of the five months of the season is instrumented with wind in each month. Standard errors are indicated by the bars and are clustered at the village level.

during the harvest season. Second, pollution at the very end of the season is unlikely to have large effects on overall land productivity, as opposed to pollution in prior months, since growth is generally complete by the end of the season.

Figure 1 presents the coefficients of the IV regression separately for each month of the agricultural season, with the very beginning of the season on the left side of the graph and the harvest season on the right side of the graph. Interestingly, the coefficient for the fourth month of the season is actually positive. It is hard to know exactly why this is the case, but it could be an interaction between the positive effects of coal pollution on agricultural productivity found by Sanders and Barreca (2022) and the fact that very little labor is used just prior to harvest. However, the overall explanation is beyond the scope of the current paper.

The other four months all see negative effects of pollution on agricultural productivity. However, the effect during the harvest season is the largest, at more than twice the magnitude of any of the other three months. In other words, it seems that pollution has the largest effects during harvest, which is consistent with the hypothesis that pollution affects labor productivity. This is also consistent with the results in Table 6, which show that pollution has larger effects in rice-growing areas, which are more labor intensive.

4.4 Coal plants or overall pollution?

Finally, while I identify high-pollution areas using the location of coal plants, an important question is how much of the overall effect is driven by coal plants and how much is driven by other polluting industries located in areas near coal plants. I try to answer this question in Table 7, where I implement an instrumental variables differences-in-differences, essentially instrumenting for pollution prior to the opening of a coal plant with wind and instrumenting for the interaction term (pollution times plant opening) with wind times plant opening.

Columns one and two of Table 7 present specifications with slightly different weather controls. Column two presents the most extensive set of controls – the flexible specification with binned weather variables – and I focus on that column here. First, there are now two first stages but the F-tests indicate that the instruments are strong for both, including the interaction term ($F = 205$ in the second column). Second, the first row shows the effect of pollution on agricultural productivity prior to the opening of a coal plant; the coefficient is negative, indicating negative effects of non-coal pollution. Third, the interaction term is strongly negative and approximately the same in magnitude as the level term, indicating a very clear effect specific to coal plants, in addition to the overall negative effect of pollution.

Importantly, since these results come from an IV, both coefficients denote the change in agricultural yield

Table 7: Agricultural productivity and pollution source

	(1)	(2)
Particulate matter (log PM 2.5)	-0.228*** (0.066)	-0.223*** (0.065)
PM 2.5 times Coal	-0.250*** (0.024)	-0.220*** (0.025)
Weather (expanded)	Yes	No
Weather (expanded, bins)	No	Yes
Fixed effects:		
Year	Yes	Yes
Village	Yes	Yes
F (PM)	769	772
F (PM times Coal)	231	205
Observations	1,266,922	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated (log) agricultural yield, defined as tons per hectare. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

due to an increase in pollution by the same units. In other words, this is not identifying differences in overall pollution, per se, but rather differences in the effects of an increase of pollution by the same amount. Since I did not find heterogeneity by exposure, it does not seem that this is driven by non-linear effects of pollution on productivity. Instead, these different effects could be due to, for example, differences in emission patterns for different kinds of plants and factories; coal plants, for example, are responsible for much more mercury than other kinds of polluting industries.¹²

5 Overall effects of pollution on agricultural productivity

Pollution has increased markedly in India over the last few decades. We can use the results above to quantify the overall effects of pollution on agricultural productivity. However, there are two reasons just using the sample of villages from the main results above might not be indicative of overall changes in India. First, the sample of villages have higher levels of pollution than other villages, as shown in Table 1. Second, the sample of villages are also located near coal plants, which appear to have larger negative effects on agricultural productivity than other sources of pollution.

For this reason, I instead turn to data from ICRISAT,¹³ which has district-level total production data from different years. To track with the timing of the main data used above, I pull total agricultural production from 2001 and 2011 for three of the most common crops in India: rice, wheat, and maize.¹⁴ To simplify

¹²https://www.epa.gov/sites/default/files/2016-12/documents/nei2014v1_tsd.pdf

¹³<http://data.icrisat.org/dld>

¹⁴Maize might not seem like the first choice for India. However, maize is one of the most common crops grown in other developing

the calculation, I assume total area devoted to each crop is unchanged across the decade. I then merge this with the pollution data from Hammer et al. (2020) using a 2011 district-level census shapefile, which has identifiers that match the ICRISAT data. Figure A5 in the appendix shows the distribution of total output for the three crops in 2001. The data covers much, but not all, of the country.

In the output data, I observe values for 2001 and 2011. I can calculate the overall change in log output across the two years, as well as the overall change in pollution levels. I use the most conservative point estimate *without* coal plants in Table 7, which is -0.223 . For each district, I calculate the counterfactual for each crop as follows:

$$\Delta \log(\text{counterfactual})_d - 0.223 \times \Delta \log(PM2.5)_d = \Delta \log(\text{actual})_d, \quad (5)$$

where counterfactual is the value I want to calculate, while actual is the actual observed change in output. I invert the equation to find the predicted counterfactual without the change in pollution. In other words, I calculate what the change would have been if pollution had not changed at all for any given district. Since this relies on log changes, I restrict the sample only to districts that have output data for a given crop in 2001. Due to a few very large changes in pollution in a few areas, I winsorize log counterfactual changes at the 95th percentile for all crops. After calculating the counterfactual for each district separately, I sum total output to the country level in order to judge the overall effects of pollution on agricultural productivity.

The overall change depends on the distribution of changes in pollution as well as the distribution of output across the three crops. I first present the actual and predicted changes in (log) output in Figures A6, A7, and A8 in the appendix. Since pollution increased almost everywhere, the counterfactual change is (almost) always higher than the actual change.

I present the main results of the back-of-the-envelope calculation in Table 8. I align the three crops across the rows, with the 2001 and 2011 total output (summed across all districts, in '000s of tons) in the first two columns. Total actual output increased for all three crops, with the largest absolute increase being for wheat – which increased by around 37 percent in relative terms – and the largest relative increase being for maize – which increased by around 60 percent in total output.

The key set of results is the comparison between column two and column three. Column two gives the actual total output by crop in 2011, while column three presents the counterfactual output, calculated as what would have happened if pollution had not changed at all. There is quite a bit of heterogeneity across crops. For rice, total output would have been around 2.3 percent higher if pollution had remained at 2001 levels.

countries – especially in Sub-Saharan Africa – so I use it here as a benchmark.

Table 8: Change in output and counterfactual, all districts

	2001 actual output	2011 actual output	2011 counter- factual	Difference (pct.)
Rice	79,962	93,238	95,405	2.3
Wheat	58,780	93,975	97,520	3.8
Maize	11,935	20,897	21,832	4.5

Note: Output is the total sum across all districts, in '000s of tons. The counterfactual is calculated as described in Equation 5.

Maize, on the other hand, would have seen a 4.2 percent increase and wheat would have seen a 3.8 percent increase. These results underscore the point that the effects of pollution are not uniform across crops, even if we assume equal effects of pollution, due to differences in the distribution of pollution changes across the country.

6 Conclusion

In this paper, I show some of the negative effects associated with pollution, in general, and coal pollution, in particular. Using exogenous variation in pollution driven by changes in wind direction, results here show that pollution has a large, negative effect on agricultural productivity in India. These negative effects are at least partly due to changes in labor productivity; areas growing more labor-intensive crops see larger decreases in agricultural productivity due to pollution. Nonetheless, I am unable to rule out direct effects on land productivity with the data available.

Using instrumental variables in a differences-in-differences set up, I also show that the effects of pollution from coal, in particular, are particularly harmful to agricultural productivity. Changes of an equivalent increase in pollution lead to three times larger negative effects on agricultural productivity following the opening of a coal plant, compared to just before.

These results document relatively large, local externalities in the agricultural sector in a developing country. Since the sector is a refuge for some of the poorest citizens in the world, the results raise important questions about the current speed of the energy transition as it relates to sustainable development and poverty reduction, more generally. While we often frame the debate around greenhouse gasses in terms of climate change, there are important local reasons to reduce pollution emissions, as well. The technology for greener energy exists; at this point, it is simply a question of policies (Markard 2018; Blazquez, Fuentes, and Manzano 2020).

These results, of course, do not address possible positive effects of the building of additional power plants, which can lead to increased electricity access, or shifting more of the population into non-agricultural sectors,

which can raise standards of living. As such, the main conclusion is that the location of pollution emissions – especially from coal plants – may be particularly important for multiple aspects of economic activity, not just health and mortality.

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.
- Abay, Kibrom A, Christopher B Barrett, Talip Kilic, Heather Moylan, John Ilukor, and Wilbert Drazi Vundru. 2022. “Nonclassical Measurement Error and Farmers’ Response to Information Reveal Behavioral Anomalies.”
- Aragón, Fernando M, and Juan Pablo Rud. 2016. “Polluting industries and agricultural productivity: Evidence from mining in Ghana.” *The Economic Journal* 126 (597): 1980–2011.
- Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad. 2021. “Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India using the SHRUG Open Data Platform.” *The World Bank Economic Review*.
- Barrios, Thomas, Rebecca Diamond, Guido W Imbens, and Michal Kolesár. 2012. “Clustering, Spatial Correlations, and Randomization Inference.” *Journal of the American Statistical Association* 107 (498): 578–91.
- Bencko, Vladimir, and Karel Symon. 1977. “Health Aspects of Burning Coal with a High Arsenic Content.” *Environmental Research* 13 (3): 378–85.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. “How Much Should We Trust Differences-in-Differences Estimates?” *The Quarterly Journal of Economics* 119 (1): 249–75.
- Blazquez, Jorge, Rolando Fuentes, and Baltasar Manzano. 2020. “On Some Economic Principles of the Energy Transition.” *Energy Policy* 147: 111807.
- Borgschulte, Mark, David Molitor, and Eric Yongchen Zou. 2022. “Air pollution and the labor market: Evidence from wildfire smoke.” *Review of Economics and Statistics*, 1–46.
- Burney, Jennifer, and V Ramanathan. 2014. “Recent climate and air pollution impacts on Indian agriculture.” *Proceedings of the National Academy of Sciences* 111 (46): 16319–24.
- 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.
- Chen, Shuai, Paulina Oliva, and Peng Zhang. 2022. “The effect of air pollution on migration: evidence from China.” *Journal of Development Economics* 156: 102833.
- Christiaensen, Luc, Lionel Demery, and Jesper Kuhl. 2011. “The (evolving) role of agriculture in poverty reduction — An empirical perspective.” *Journal of Development Economics* 96 (2): 239–54.
- Dercon, Stefan. 2002. “Income Risk, Coping Strategies, and Safety Nets.” *The World Bank Research Observer* 17 (2): 141–66.
- . 2009. “Rural poverty: Old challenges in new contexts.” *The World Bank Research Observer* 24 (1): 1–28.
- 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.
- Desiere, Sam, and Dean Jolliffe. 2018. “Land productivity and plot size: Is measurement error driving the inverse relationship?” *Journal of Development Economics* 130: 84–98.
- Dinkelman, Taryn. 2011. “The effects of rural electrification on employment: New evidence from South Africa.” *American Economic Review* 101 (7): 3078–108.
- Ebenstein, Avraham, Victor Lavy, and Sefi Roth. 2016. “The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution.” *American Economic Journal: Applied Economics* 8 (4): 36–65.
- Gangopadhyay, Prasun K, Paresh B Shirsath, Vinay K Dadhwal, and Pramod K Aggarwal. 2022. “A new two-decade (2001–2019) high-resolution agricultural primary productivity dataset for India.” *Scientific Data* 9 (1): 1–12.
- Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. “Experimenting with measurement error: Techniques with applications to the Caltech cohort study.” *Journal of Political Economy* 127 (4): 1826–63.
- Glewwe, Paul. 2012. “How much of observed economic mobility is measurement error? IV methods to reduce measurement error bias, with an application to Vietnam.” *The World Bank Economic Review* 26 (2): 236–64.
- Graff Zivin, Joshua, and Matthew Neidell. 2012. “The impact of pollution on worker productivity.” *American*

- Economic Review* 102 (7): 3652–73.
- 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: 68–79.
- He, Jiaxiu, Haoming Liu, and Alberto Salvo. 2019. “Severe air pollution and labor productivity: Evidence from industrial towns in China.” *American Economic Journal: Applied Economics* 11 (1): 173–201.
- Heck, Walter W, OC Taylor, Richard Adams, Gail Bingham, Joseph Miller, Eric Preston, and Leonard Weinstein. 1982. “Assessment of crop loss from ozone.” *Journal of the Air Pollution Control Association* 32 (4): 353–61.
- Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi. 2014. “Growth and Structural Transformation.” *Handbook of Economic Growth* 2: 855–941.
- Kline, Patrick, and Enrico Moretti. 2014. “Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority.” *The Quarterly Journal of Economics* 129 (1): 275–331.
- Kopas, Jacob, Erin York, Xiaomeng Jin, SP Harish, Ryan Kennedy, Shiran Victoria Shen, and Johannes Urpelainen. 2020. “Environmental justice in India: incidence of air pollution from coal-fired power plants.” *Ecological Economics* 176: 106711.
- Lee, David, Justin McCrary, Marcelo J Moreira, and Jack Porter. 2022. “Valid t-ratio Inference for IV.” *American Economic Review* 112 (10): 3260–90.
- Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2020. “Does Household Electrification Supercharge Economic Development?” *Journal of Economic Perspectives* 34 (1): 122–44.
- Li, Ya-Ru, and Jacqueline MacDonald Gibson. 2014. “Health and air quality benefits of policies to reduce coal-fired power plant emissions: a case study in North Carolina.” *Environmental Science & Technology* 48 (17): 10019–27.
- Markard, Jochen. 2018. “The Next Phase of the Energy Transition and Its Implications for Research and Policy.” *Nature Energy* 3 (8): 628–33.
- Marshall, Fiona, Mike Ashmore, Fiona Hinchcliffe, et al. 1997. *A hidden threat to food production: Air pollution and agriculture in the developing world*. International Institute for Environment and Development.
- Merfeld, Joshua D, and Jonathan Morduch. 2023. “Poverty at Higher Frequency.” *KDI School of Pub Policy & Management Paper No. DP23-03*.
- Michler, Jeffrey D. 2020. “Agriculture in the process of development: A micro-perspective.” *World Development* 129: 104888.
- Monfreda, Chad, Navin Ramankutty, and Jonathan A Foley. 2008. “Farming the planet: Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000.” *Global Biogeochemical Cycles* 22 (1).
- Ramanathan, Veerabhadran, and Gregory Carmichael. 2008. “Global and regional climate changes due to black carbon.” *Nature Geoscience* 1 (4): 221–27.
- Reddy, V Ratna, and Bhagirath Behera. 2006. “Impact of water pollution on rural communities: An economic analysis.” *Ecological Economics* 58 (3): 520–37.
- Rud, Juan Pablo. 2012. “Electricity provision and industrial development: Evidence from India.” *Journal of Development Economics* 97 (2): 352–67.
- Sanders, Nicholas J, and Alan I Barreca. 2022. “Adaptation to Environmental Change: Agriculture and the Unexpected Incidence of the Acid Rain Program.” *American Economic Journal: Economic Policy* 14 (1): 373–401.
- Van de Walle, Dominique, Martin Ravallion, Vibhuti Mendiratta, and Gayatri Koolwal. 2017. “Long-term gains from electrification in rural India.” *The World Bank Economic Review* 31 (2): 385–411.
- Vollrath, Dietrich. 2011. “The Agricultural Basis of Comparative Development.” *Journal of Economic Growth* 16: 343–70.
- Wen, Jeff, and Marshall Burke. 2022. “Lower test scores from wildfire smoke exposure.” *Nature Sustainability* 5 (11): 947–55.

Appendix A

Table A1: Data sources

	source	geographic coverage	temporal coverage
shapefile	Asher et al. (2021)	India	
coal plants	Global Energy Monitor	global	yearly
wind	NCAR	global	daily
pollution	Hammer et al. (2020)	global	monthly
agriculture	Gangopadhya et al. (2022)	India	two seasons/year
weather	TerraClimate	global	monthly
crops	Monfreda et al. (2008)	global	2000 (year)

Global Energy Monitor: globalenergymonitor.org/projects/global-coal-plant-tracker. NCAR: climatedataguide.ucar.edu/. TerraClimate: www.climatologylab.org/terraclimate.html

Figure A1: Coal plants in India from 1990 to 2010

A. 1990



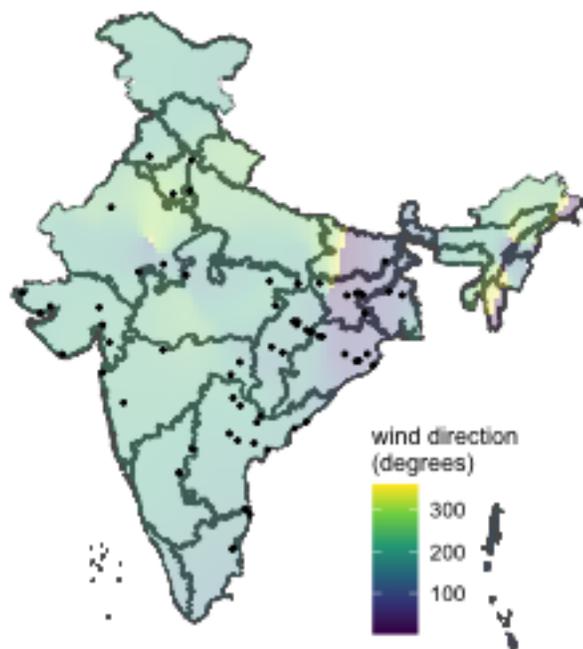
B. 2010



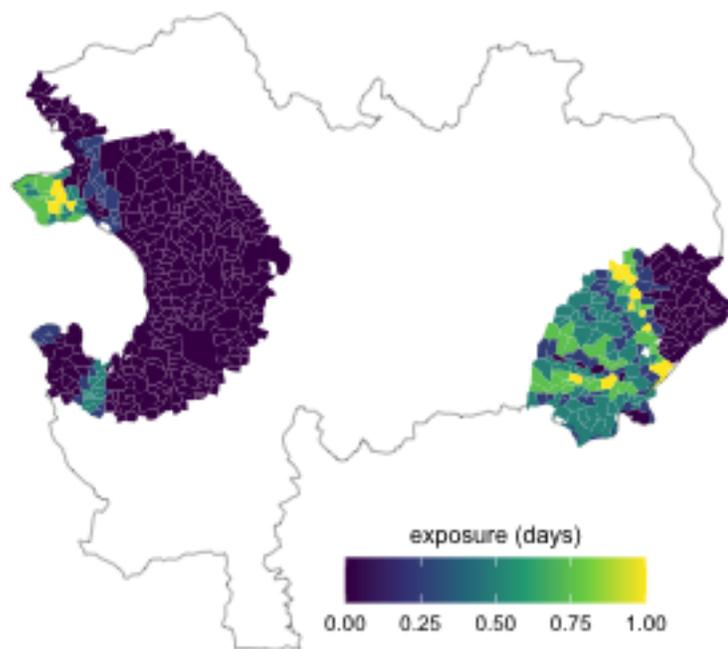
Note: The left figure shows the location of coal plants of at least 30MW in 1990. The right figure shows the location of such coal plants in 2010.

Figure A2: Wind direction and aggregation examples (2010-01-01)

A. Wind direction



B. Aggregation to district



Note: The top figure shows the average wind direction on January 1st, 2010, with zero degrees indicating north. The points are the location of coal plants on that date. The bottom figure shows the distribution of pollution exposure in a specific district – Guna district in Madhya Pradesh – on the same date.

Table A2: Particulate matter (PM 2.5) and agricultural productivity

	(1)	(2)	(3)	(4)
Particulate matter (log PM 2.5)	-0.042*** (0.003)	-0.045*** (0.003)	-0.008** (0.003)	-0.041*** (0.003)
Weather	No	Yes	Yes	No
Weather (expanded)	No	No	Yes	No
weather (expanded, bins)	No	No	No	Yes
Fixed effects:				
Village	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Observations	1,266,922	1,266,922	1,266,922	1,266,922

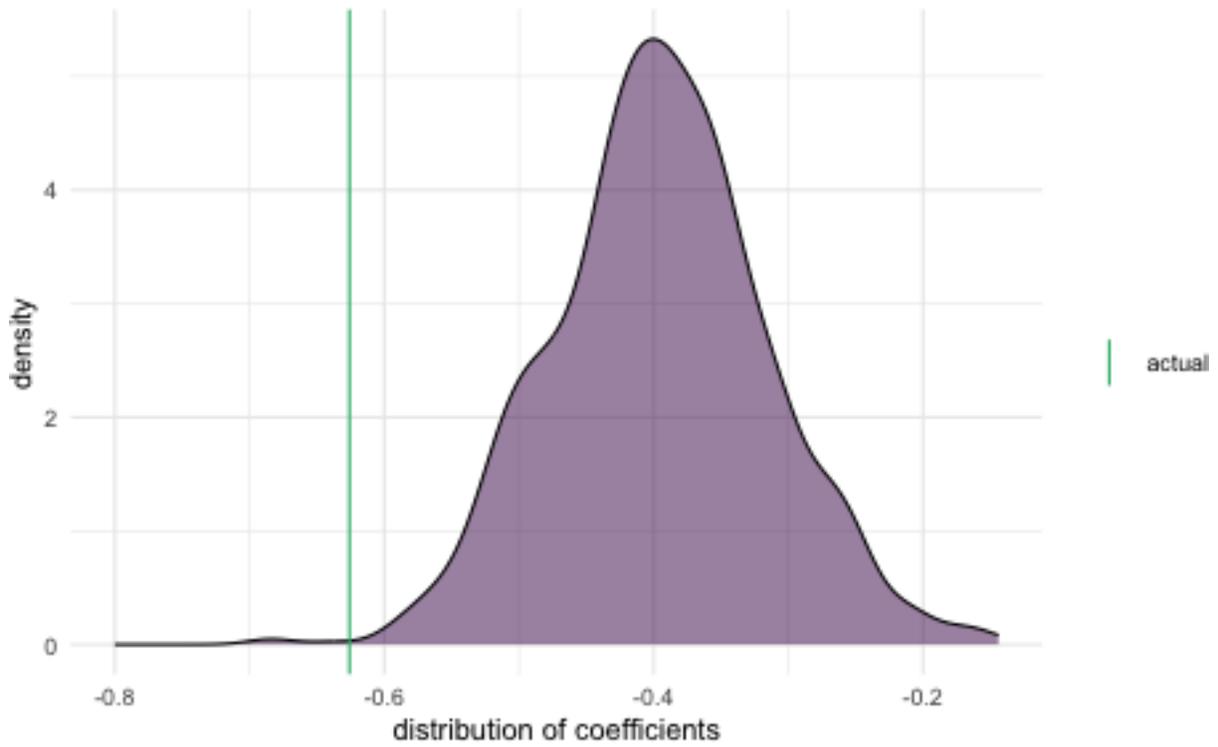
Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated (log) agricultural yield, defined as tons per hectare. * p<0.10 ** p<0.05 *** p<0.01

Table A3: Wind direction and particulate matter

	(1)	(2)	(3)	(4)
wind	0.0083*** (0.0002)	0.0085*** (0.0002)	0.0083*** (0.0002)	0.0085*** (0.0002)
weather	No	Yes	Yes	No
weather (expanded)	No	No	Yes	No
weather (expanded, bins)	No	No	No	Yes
fixed effects:				
village	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
F	1,236	1,266	1,205	1,246
observations	1,266,922	1,266,922	1,266,922	1,266,922

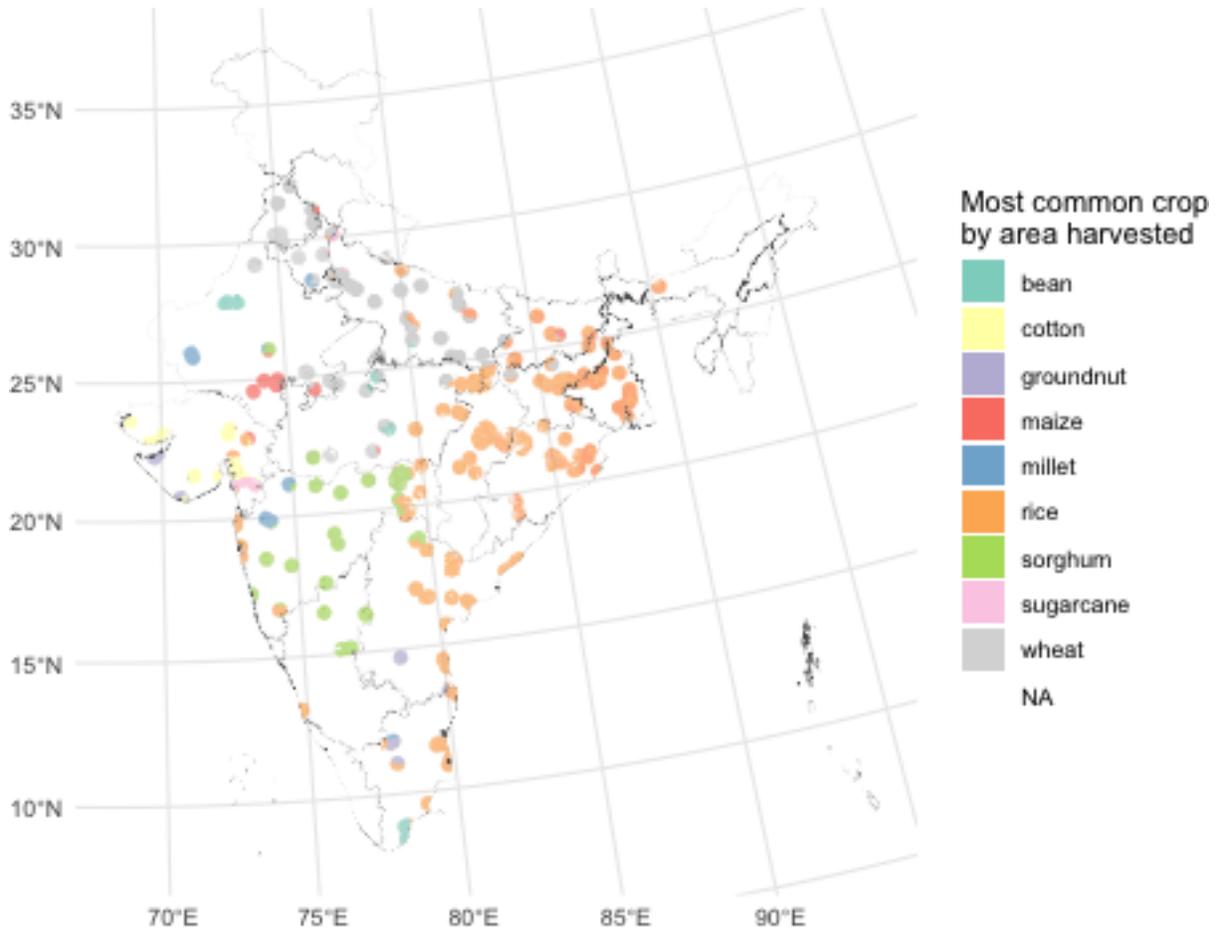
Note: Standard errors are in parentheses and are clustered at the village level. * p<0.10
 ** p<0.05 *** p<0.01

Figure A3: Randomization inference



Note: The figure presents the distribution of coefficients by randomly assigning wind direction and pollution exposure across villages within the same year. Sampling is done within year and district. The vertical line is the actual coefficient from column (4) of Table 3.

Figure A4: Most common crop by village



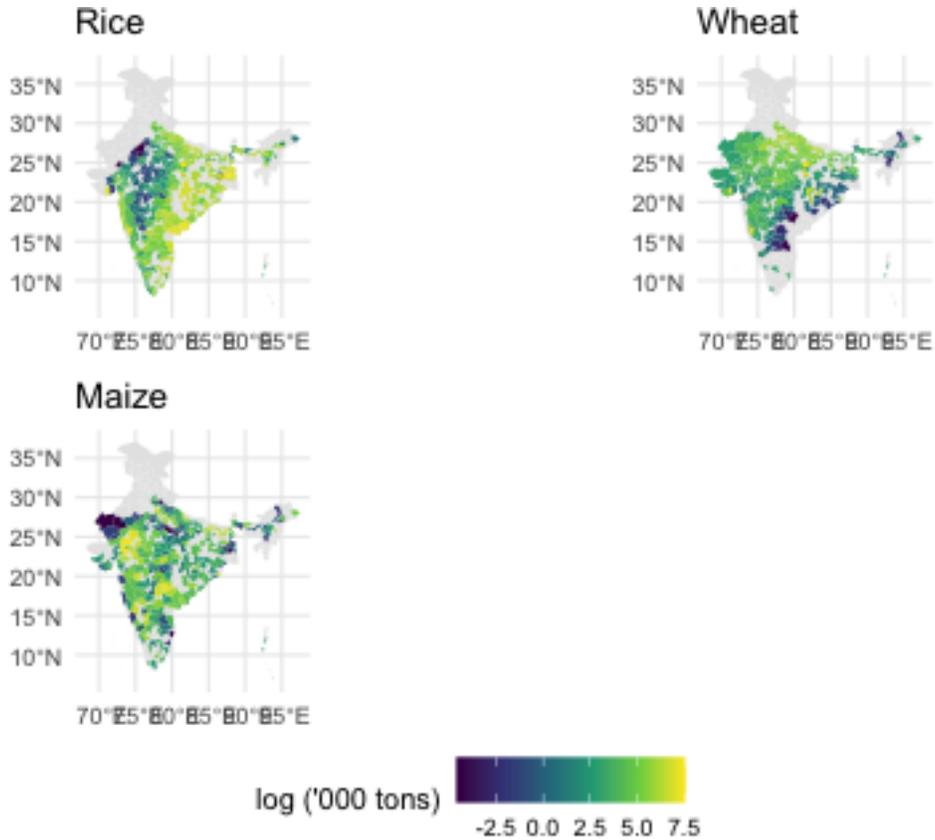
Note: The figure shows the most common crop grown in each village, by area harvested.

Table A4: Agricultural productivity by most common crop

	(1)	(2)
Particulate matter (log PM 2.5)	-0.811*** (0.158)	-1.034*** (0.144)
PM 2.5 times rice	-1.204*** (0.095)	-0.966*** (0.092)
Weather (expanded)	Yes	No
Weather (expanded, bins)	No	Yes
Fixed effects:		
Village	Yes	Yes
Year	Yes	Yes
F (PM 2.5)	883	941
F (PM 2.5 times rice)	1,355	1,395
Observations	1,266,922	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated (log) agricultural yield (tons per hectare). The rice dummy equals one in villages where rice is the most commonly harvested crop. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Figure A5: District-level yields in 2001



Yield data comes from ICRISAT.

Figure A6: Distribution of actual and counterfactual log changes in output, rice

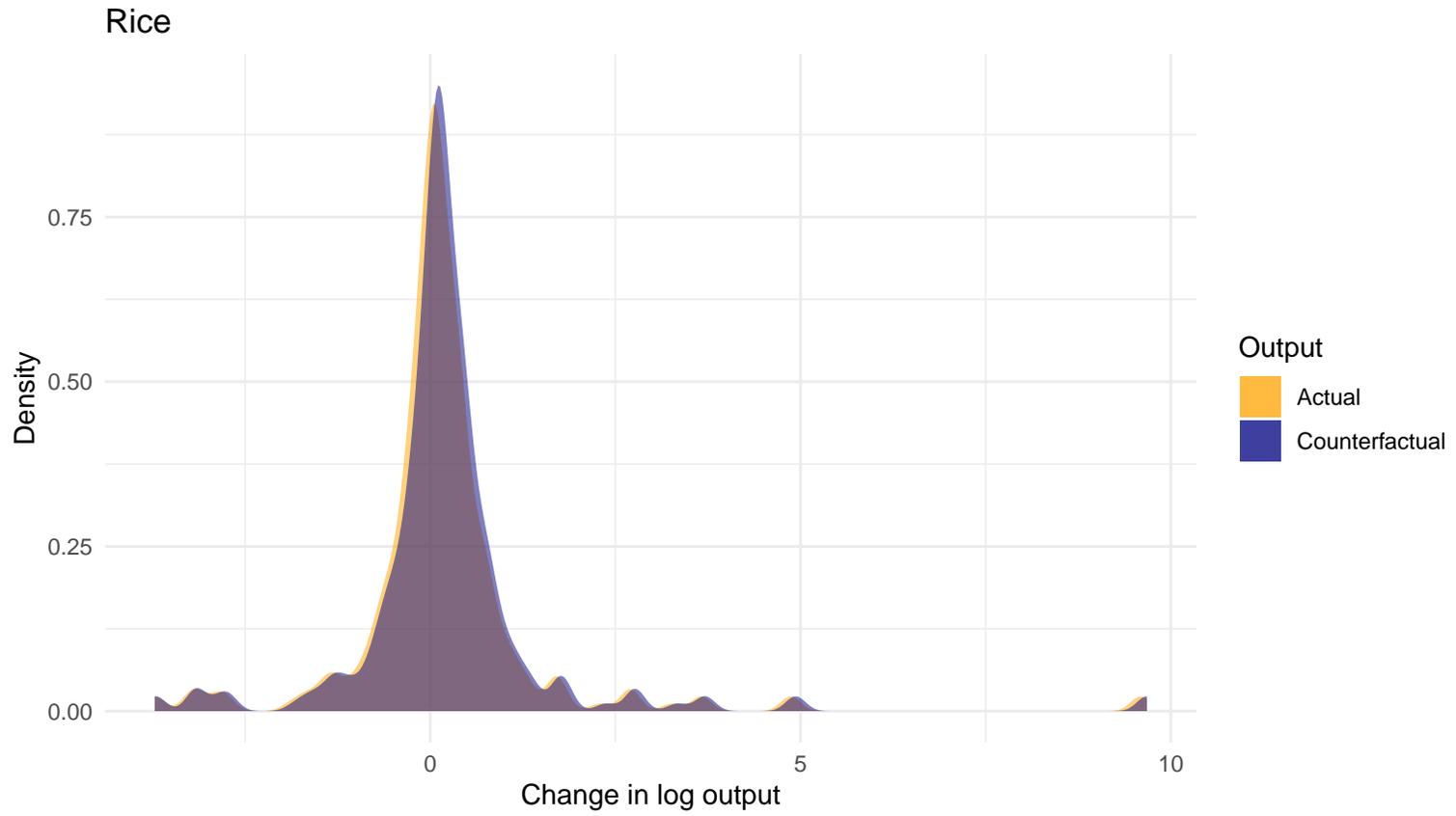


Figure A7: Distribution of actual and counterfactual log changes in output, wheat

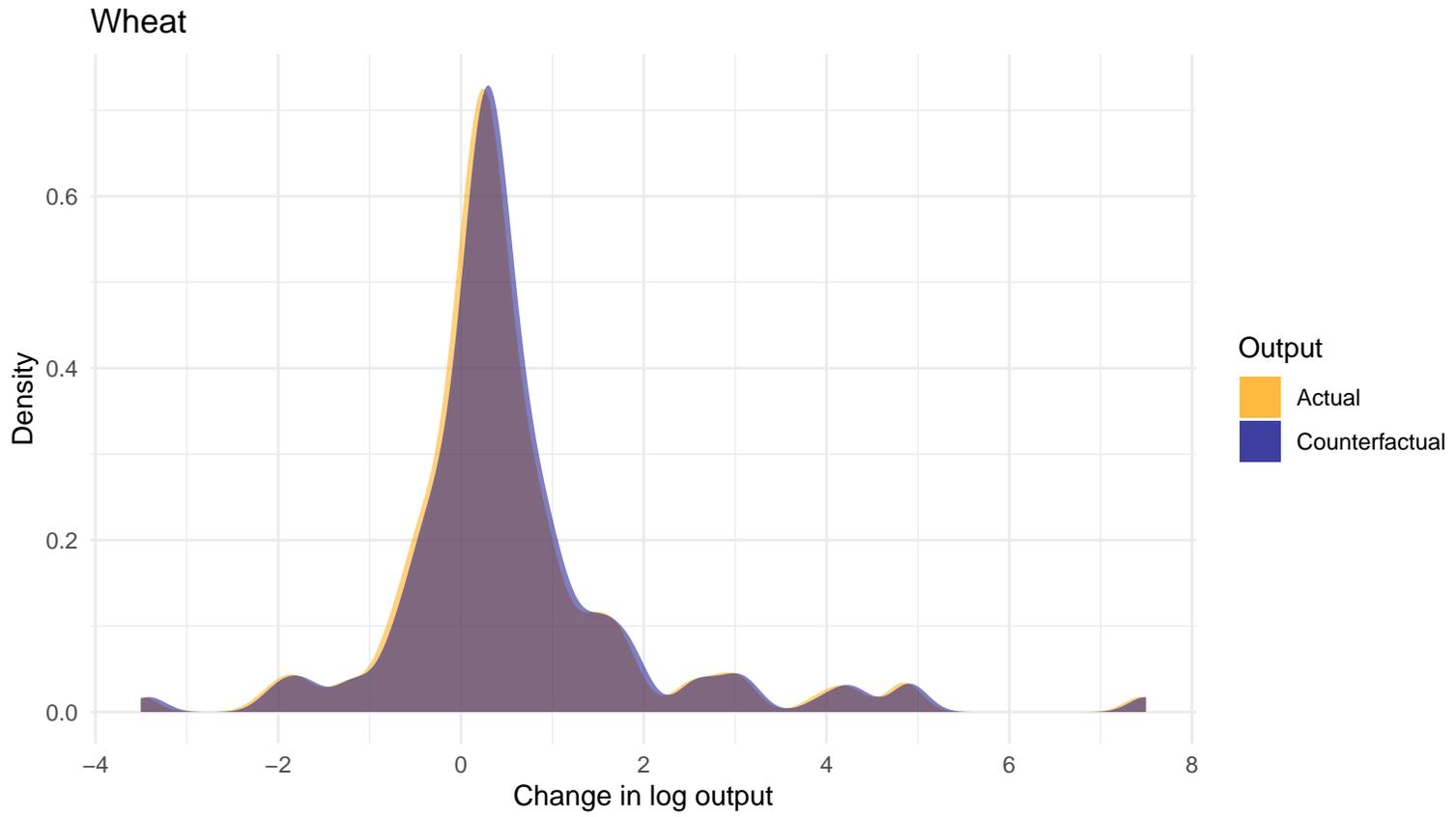


Figure A8: Distribution of actual and counterfactual log changes in output, maize

