Identifying the Effect of Coal Plants on Air Pollution

E. Somanathan*, Sagnik Dey†, Kanishka Kacker* and Rishabh Choudhary‡

October 8, 2020

Abstract

In the past 15 years, India has witnessed a dramatic rise in the construction of coal fired power plants. Recent work has shown severe negative health impacts of this rise, but these focus on short-term effects alone while long term effects are of deeper policy interest. Using a panel of Indian power plants spanning nine years from 2011 to 2019, we estimate the marginal contribution of an additional unit of coal to particulate matter emissions of a diameter less than 2.5 micrometers (PM2.5), a particularly harmful pollutant, measured from satellite gridded data. We use plant starts, unscheduled stoppages and closures as instruments for coal burning. Our identification assumption is that at a high frequency – our data is at the plant-month level – stoppages and starts will affect pollution levels only through coal burning. On average, increases in coal burnt raises pollution levels up to 900 km away from a power plant. With exposure response functions already documented, our estimates can be used to provide long term impacts on health.

*Indian Statistical Institute, Delhi; †: Indian Institute of Technology, Delhi; ‡ World Bank, Delhi
Corresponding author email: kkacker@isid.ac.in, kanishka.kacker@gmail.com

1 Introduction

Most of the developing world is on a path of rapidly increasing demand for electricity [Wolfram et al., 2012]. Of this world, India and China constitute the two biggest countries. Coal plant construction has rapidly risen in both countries. Coal fired power plants run far below capacity in India and not all households have access to electricity as yet. For the forseeable future, therefore, much of the increased demand for electricity is likely to come through further increases in coal burning. Increased burning of coal in power plants has enormous implications for local health outcomes and global climate emissions targets, as environmental externalities from coal fired power plants appear to be large [Gupta and Spears [2017] Barrows et al. [2019]].

India is witnessing some of the worst air pollution ever recorded in the known history of the world: 9 of the top 10 most polluted cities are in India, with air quality indexes an order of magnitude over what are considered safe limits [World Health Organization, 2018]. While coal burning emits other pollutants - oxides of sulfur and nitrogen as well as carbon - knowing its effect on PM2.5 is vital since PM2.5 is particularly harmful amongst this suite of pollutants, responsible for 4 million deaths in 2016 from heart disease, stroke, lung cancer, chronic lung disease and respiratory infections [Health Effects Institute, 2018].

Evidence on the health impacts from burning coal, however, comes mostly from developed countries. There are good reasons to believe these estimates might be substantially different in developing countries: exposure is larger while regulatory capacity is weaker in developing countries, making impacts possibly larger than in developed countries [Gupta and Spears, 2017]. There is a small but growing literature on the health impacts from coal burning in India: Barrows et al. [2019] and Gupta and Spears [2017] shown severe negative health impacts of this rise, as a consequence of harmful emissions from coal plants. These studies focus, however, on short term impacts, while long term impacts are of deeper policy interest ([Pope-III and Dockery, 2006], [Ebenstein et al., 2017], [Burnett et al., 2018], and [Saraswat and Bansal, 2019]).

We aim to fill in this gap by estimating the marginal increase in exposure to harmful emissions arising from an additional unit of coal burnt. Our main focus is on emissions of particulate matter (PM) of a diameter less than 2.5 micrometers (PM 2.5). Source apportionment studies [Venkatraman et al., 2018] suggest coal burning in power plants are a significant source of PM 2.5, however this methodology is not designed to generate causal estimates of the marginal impact of increased coal burning which is what is needed going forward. With our estimates of the increase in exposure and data on local population densities, the change in mortality as a consequence of increased exposure to PM2.5 emissions can be easily calculated using exposure-response functions.

Using a panel of Indian power plants spanning five years from 2011 to 2019, we estimate the marginal contribution of an additional unit of coal burnt to exposure to PM2.5, measured from gridded satellite data. Importantly, as we intend to measure the total exposure to coal burning from any plant, we account for all areas downwind of coal plants that can be affected by emissions from the plant. To this end, we average all PM2.5 emissions that are downwind of a coal plant.

In the regression of PM2.5 on coal burning, the amount of coal burnt will be endogenously determined. Changes in coal burning will come from changes in electricity generated, which will be related to economic activity. And economic activity can raise particulate emissions through channels other than coal burning. To resolve this problem, we use shocks to coal burning arising from plant openings, unscheduled stoppages and closings as instruments for actual coal burnt. Our central identifying assumption is that shocks to coal burning will affect the level of pollution solely through the amount of coal burnt. Crucial to the validity of this assumption is the frequency at which we observe coal burning: at high frequency levels – our data is at the monthly level - shocks are unlikely to be correlated with any unobserved background economic variables but will influence pollution through the channel just described.

Our estimates imply that on average, increases in coal burnt by one standard deviation can raise downwind pollution levels by 0.10 to 0.15 standard deviations. These effects are felt up to 1000 kilometers (620 miles) downwind from the plant. In the next section, we give a brief background to the electricity sector in India. We follow this up with our methodology and a description of our data. Results from our econometric model are then presented. A discussion of these then concludes the paper.

2 Background to the Electricity Sector in India

The number of coal plants in India has steadily increased over the past twenty years, with a significant acceleration during the ten year period between 2007 and 2017 that doubles total capacity available from coal based plants. Figure 1 shows installed capacity by fuel type. Here we can see coal based capacity expands strongly particularly during the period 2007 to 2017 while most other fuel types rise only slightly. We can also see renewables steadily increasing their contribution and currently is the second largest contributor to capacity behind coal. Capacity expansion slows down for coal post 2017, although it continues to rise for renewables.

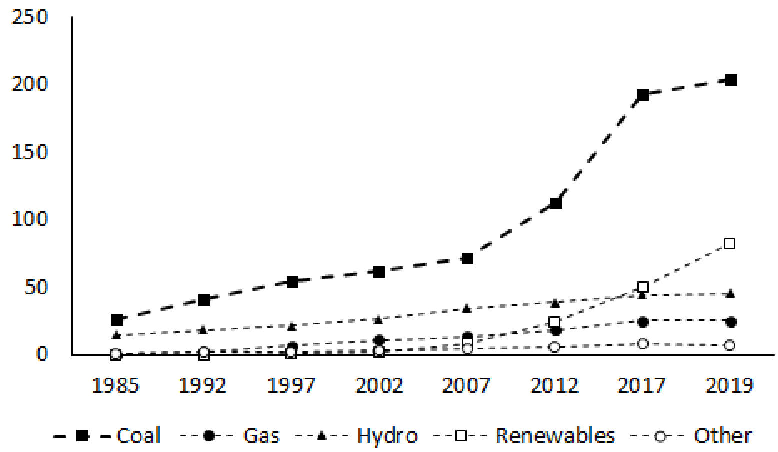

Figure 1: Capacity Expansion in Power Generation ('000 MW), by Fuel Type (Source: CEA Executive Summary on Power Sector Oct 2019)

The rapid expansion in capacity has come through new coal plant construction rather than additions to capacity in existing plants. Figure 2 shows the capacity additions arising from new coal plants alone from 2005 onward.

We can see a significant increase in coal plant construction from 2010 to 2015, and the numbers indicated here suggest most of the capacity expansion from 2007 to 2017 comes from new plants being built.

Figure 2: New Coal Plant Capacity ('000 MW)
(Source: EndCoal plant tracker http://endcoal.org/global-coal-plant-tracker/summary-statistics/)

Although capacity significantly expands, most Indian power plants do not operate at efficient levels. For instance, Figure 3 below plots average plant load factors for thermal plants - we can see a steady fall by 16% in the twelve year period from 2006-07 to 2018-19.[1] The reasons for this are unclear, although coal quality and a lack of incentives might explain some part of the inefficiency [Chan et al., 2014]. What is clear, though, is that the inefficient operation of such a large number of plants - our sample contains 132 plants - is costly as expensive fixed high capital investments are less utilized than they should be. It would also imply that future demand increases are likely to be met by further use of existing coal plants, at least to some degree.

[1] A plant load factor is the ratio of the actual output of a plant to the output if it operates at full capacity, for the same time interval.

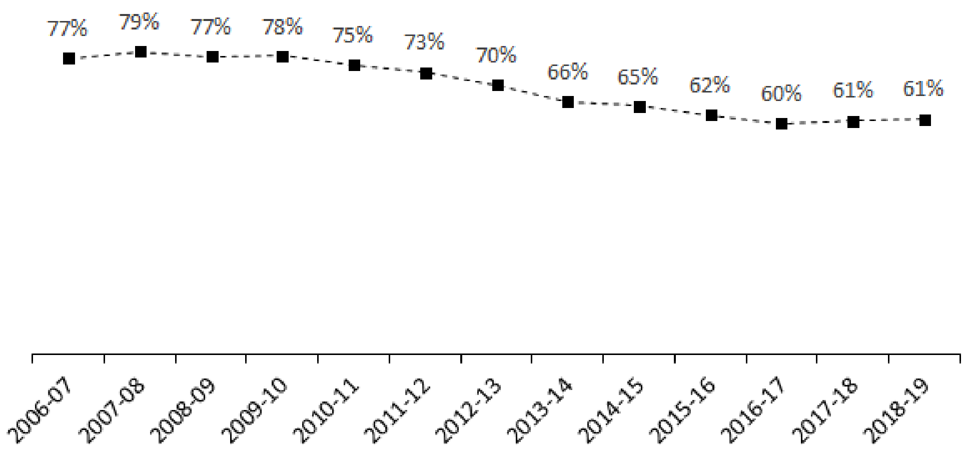

Figure 3: Plant Load Factors
(Source: CEA Annual Report 2019)

It is also likely that demand for electricity is set to rise. Despite the vast increase in capacity, and the inefficient scale of operation, the demand for electricity is still not fully met. Outages are still common across the country: Min et al. [2017] report that by 2012 even though electricity supply expanded 6 times, shortages increased to 8.5%. More recent reports from the Central Electricity Authority (CEA) reveals shortages in 11 states and union territories [Central Electricity Authority, 2019]. [2] Indeed a common finding across Gupta and Spears [2017] and Barrows et al. [2019] is that people living close to a power plant tend not to benefit from increased electricity access. The implication is that coal burning is likely to rise or at least not to fall over the coming years.

The impact of this expansion in coal plant building can be at least casually seen by observing trends in PM2.5 concentrations. Figure 4 shows lowess

[2] India has a total of 37 states and union territories. States reported in 2019 to have electricity shortages include Jammu & Kashmir, Uttarakhand, Maharashtra, Kerala, Puducherry, Bihar, Jharkhand, Orissa, West Bengal, Arunachal Pradesh, and Assam.

estimates of the trend in ambient PM2.5 concentrations, constructed from gridded satellite data.[3] These estimates are calculated from the entire sample for each month from the years 2011 to 2019. We thus have a total of 108 observations, with the first observation in January 2011 which is given a value of zero on the horizontal axis.

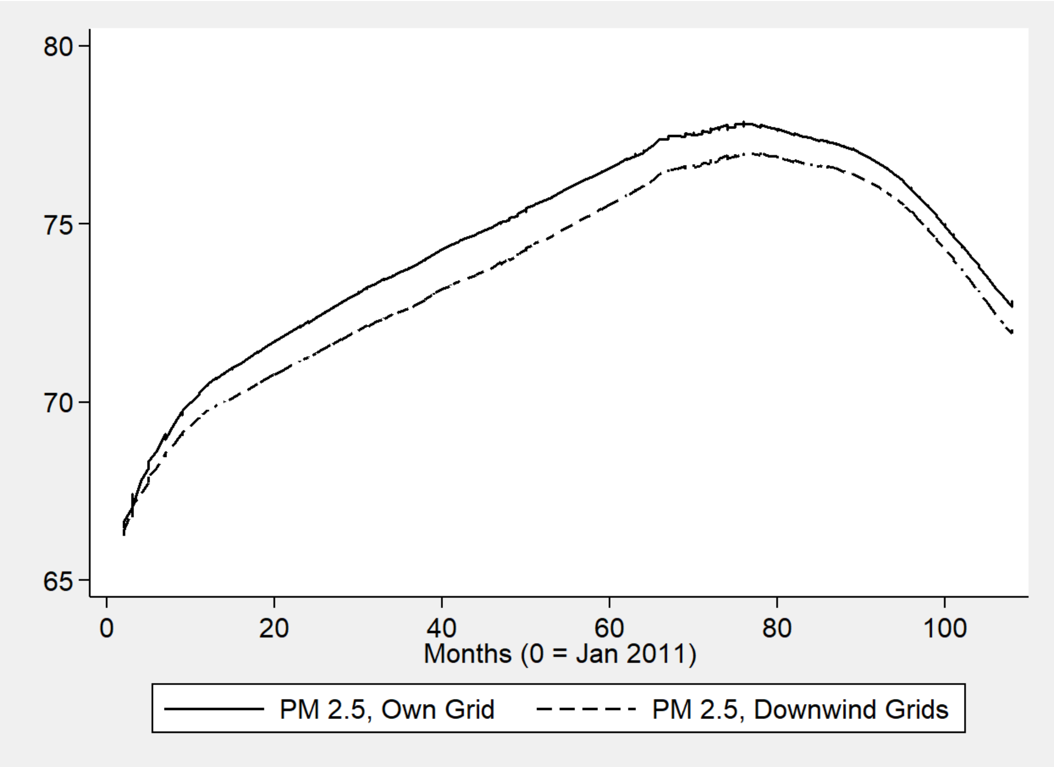

Figure 4: Lowess estimates of particulate concentration. Time is on the horizontal axis, with 0 equaling January 2011. PM 2.5, both own grid and downwind, is plotted on the vertical axis. Data taken from gridded satellite based analysis, with each grid covering a 50 km X 50 km area.
(Source: Author Calculations)

We can see coal burning tracks upward up until about 2014 after which

[3]The country is split into a number of grid cells, each sized 50 km by 50 km. For every day for each of these grid cells for the years 2011 to 2019, we have information on PM2.5 concentrations except when cloud cover is strong enough to render the data missing. We then average these observations for each month.

it starts to fall gradually. This fall really accelerates 2017 onward. Once this acceleration takes place, the PM levels also start to decline. For comparison, and to understand the seriousness of this issue, the 2005 WHO guideline level of PM2.5 considered safe was set at 10 micrograms per meter cube. The lowest PM 2.5 level in our data (16 micrograms per meter cube) is higher than this level.

PM2.5 concentrations also show a distinct spatial pattern, as figure 5 makes clear. The left panel shows the location of coal plants in our sample, with red dots representing a grid where a coal plant exists at any point in time in our sample. The right panel shows concentrations of PM2.5. We can see the northern belt of the country just below the Himalayas experiences the highest concentration of PM2.5 levels, and as we move south the concentration levels drop. Many coal plants are also located in the northern part of the country, and we can see an approximate positive relationship between coal plant location and PM2.5 concentrations. This is not the only factor of course: some of the emissions from coal burning in power plants will be carried to other grid cells by wind and there are other sources of PM2.5.

Figures 4 and 5 therefore indicate some evidence that coal power plants raise pollution concentrations. Over time more coal plants are built, raising PM2.5 emissions. Across the country, where more coal plants exist, that neighborhood has higher PM2.5 emissions. We now move to more complete evidence regarding the link between coal plants and PM2.5 levels.

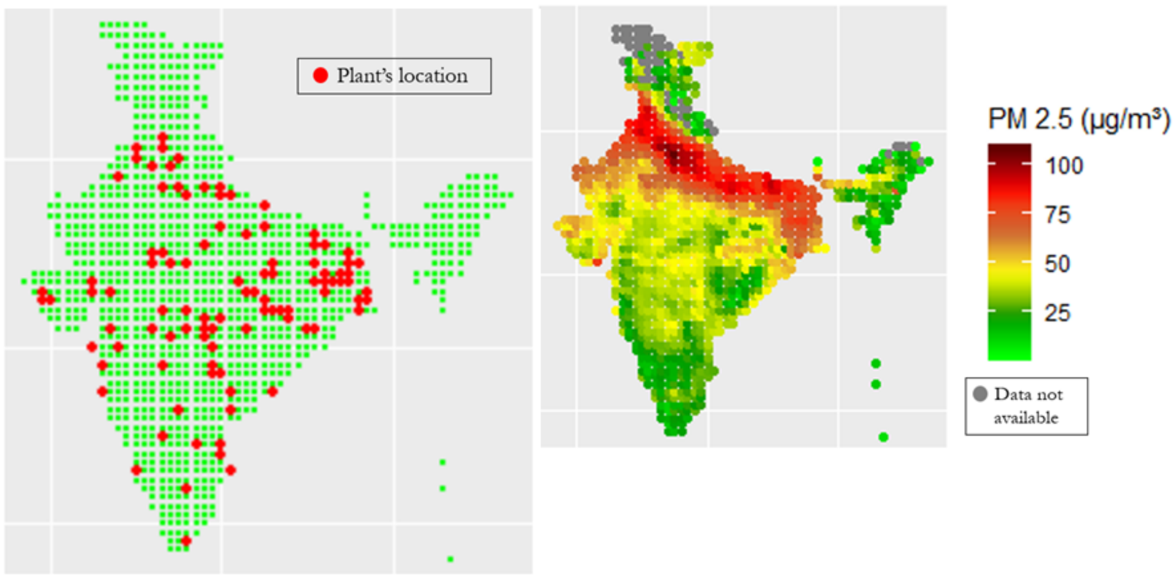

Figure 5: The spatial spread of coal plants and PM2.5 emissions. Data for PM2.5 averaged across the years 2011 to 2015.
(Source: Author Calculations)

3 Coal Burning and Pollution: Describing the Data

We take data on coal burning from the Central Electricity Authority. The CEA releases monthly reports on the amount of coal burnt and electricity generated in each power plant; we take an imputed coal amount burnt based on the generation data. The way we do this is as follows: we know monthly generation and can sum this across a sequence of 12 months to get at an aggregate amount of electricity generated. Coal burning is summed up similarly, and then dividing electricity generated by coal burnt results in an multiplicative factor for each plant. Then we use this along with the monthly electricity generated to impute coal burnt.

The reason for doing this is because the electricity generation data are more reliable. Each plant typically reports coal burnt with a lag of between 1 to 2 months on average, compared to the electricity generated data. Electricity generation and coal burning is self-reported by each plant. The CEA cross-checks these reports against regional load center meter readings for the former and Coal India Limited statements for the latter. When Coal India Limited statements do not resolve the discrepancy, the CEA uses generation data to impute coal burning.

Data on PM2.5 levels comes from satellite based analysis, and is reported at the level of a grid that spans 50 square kilometers. These data are available at a daily frequency, we average it by month. The years covered ranges from 2011 to 2019, thus capturing the most significant build up in coal plant construction.

We supplement these sources of information with three important covariates - wind speed, precipation and temperature. Larger wind speeds will blow particulate emissions out from a grid but can also blow particulates into a grid. Rainfall washes some of the particulates out of the air, while temperature can affect concentration of PM levels through thermal inversions. Information on these covariates comes from the ERA5 dataset [Copernicus Climate Change Service, 2017]. [4]

Our data is a panel of power plants observed at a monthly frequency. The total number of power plants in our data equals 193. We collapse the data down to the two-digit geographic coordinates to give use 161 unique

[4]Thermal inversions occur when a layer of warm air moves over colder air, trapping the colder air and not allowing it to move vertically. In India, this happens close to winter - late October, early November - over the Indo-Gangetic plain when the sharpest rise in PM2.5 levels in the country occurs. See https://en.wikipedia.org/wiki/Inversion_(meteorology).

location identifiers; this collapsing is necessary to give sensible estimates and is described in detail below. Since our data spans 9 years, at a monthly level, this gives us a maximum possible number of observations of 17,388. Data for some plants is missing in some months. For pollutant concentration, wind speed, precipitation and temperature the data comes from gridded satellite data: there a total of 1156 grid cells. Data is also missing for pollutant concentrations if the satellite images aren't of sufficient clarity.

Each plant is assigned to a grid based on its coordinates and the grid coordinates. If a plant happens to fall within a grid's coordinates, it is assigned that grid cells' pollutant concentration, wind speed, precipitation and temperature. To account for the effects of wind, we construct a cone of influence around each plant. This cone is essentially an arc with a predefined angle and length: we choose an angle of 90 degrees and varying lengths. Using information on wind direction at the grid the plant is located in, the 90 degree angle and various lengths we can draw cones of varying influence. Any grid whose centroid falls within this cone is labelled as downwind from the plant. Figures 6 and 7 show grid cells that are downwind of plants in January 2015 and July 2015, with two specifications of cone length: 110km and 220km. We also see that depending on month, downwind grid cells change - holding angle and length constant - as wind directions change.

Rather than use the full cone of influence for each plant, we will use portions of the cone to investigate the effects of coal burning on downwind areas. That is, rather that use all areas up to a distance from the plant, we will use areas between a minimum and maximum distance from the plant. This means we use *rings* of the cone. The motivation for doing so comes

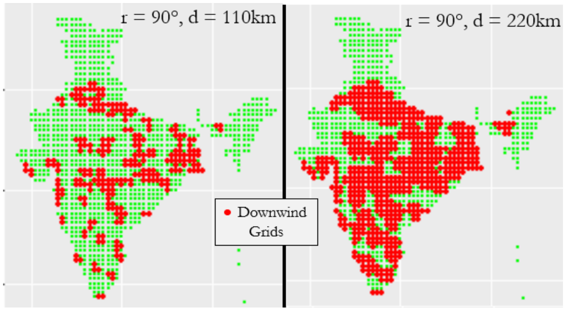

Figure 6: Downwind grid cells, January 2015

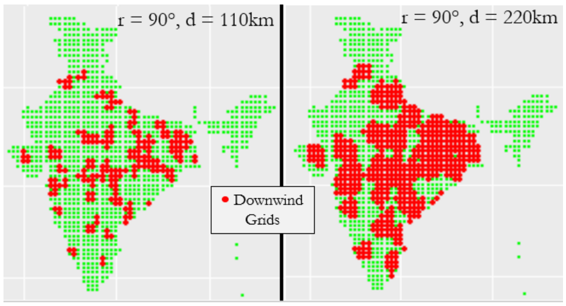

Figure 7: Downwind grid cells, July 2015

from the fact that smoke from coal burning is released at a significant height through smoke stacks, and winds are stronger at higher heights. We might expect therefore that in the immediate areas downwind of the plant, the marginal contribution of the plant to pollution might be lower than areas further away, with the effect eventually decaying.[5]

Accounting for wind will lead us to more informative estimates, but exposes a tradeoff in the decision of angle and length. The further away we go from a plant, the more we can possibly track where the particulates blow leading to an estimate closer to the actual effect. At the same time, we also raise the possibility of introducing more noise as we may be including irrelevant grid cells. There is no clear way to resolve this tradeoff, so we run our specification for various choices of length.

Panel A in Table 1 describes our main variables. We can see average PM2.5 levels are much higher than WHO recommendations. In Panel B of

[5] An additional constraint we impose is to ensure we only use grid cells that can be compared over time. If wind directions change over time, then because of our first-differenced specification we might include a grid cell that is currently downwind but wasn't so in the previous time period. In this case, the increase in PM 2.5 comes purely from the construction of the dataset. Therefore, we restrict our analysis to grid cells that are downwind for both time periods.

the same table, we show how coal burnt and PM 2.5 levels change as we move further away from the plant. At larger distances from the plant, we can see PM 2.5 levels fall while coal burning remains roughly the same. The changes in coal burnt come from the fact that sometimes plants drop out as there are not enough downwind grids with PM 2.5 data - either because we are close to political boundaries (our data stops at the political boundary of India) or the data are missing. While the PM 2.5 levels fall by a large amount, the changes in coal burnt are slight.

Table 1: Summary Description of the Data

Panel A	0 - 100 km ring				
	Observations	Mean	Standard Deviation	Min	Max
Coal Burnt ('000 Tons)	8,537	380.59	451.81	0	3165
Grid where Plant is Located					
Precipitation (mm)	11,083	2.90	4.65	0	43.63
Temperature (Celcius)	11,083	25.19	5.32	10.46	37.52
Wind Speed (metres/second)	11,083	1.66	1.16	0.02	8.63
Downwind Grids					
Precipitation (mm)	11,083	2.95	4.69	0	43.76
Temperature (Celcius)	11,083	25.34	5.18	10.51	37.66
Wind Speed (metres/second)	11,083	1.52	1.03	0.04	7.89
PM 2.5 (micrograms/metre cube)	11,083	74.10	38.09	16.70	223.10
Instruments					
open	9,601	0.02		0	1
close	9,581	0.02		0	1

Panel B

	0 - 100 km		100 - 200 km		200 - 300 km		300 - 400 km		400 - 500 km	
	Obs	Mean	Obs	Mean	Obs	Mean	Obs	Mean	Obs	Mean
Coal Burnt ('000 tons)	8,537	380.59	8,660	386.01	8,582	384.36	8,365	381.44	8,022	386.50
PM 2.5	11,083	74.10	10,999	73.86	10,896	73.81	10,645	73.67	10,197	72.84

	500 - 600 km		600 - 700 km		700 -800 km		800 - 900 km		900 - 1000 km	
	Obs	Mean	Obs	Mean	Obs	Mean	Obs	Mean	Obs	Mean
Coal Burnt ('000 tons)	7,710	386.09	7,549	383.66	7,068	356.95	6,748	364.22	6,577	373.34
PM 2.5	9,816	71.44	9,574	71.49	9,016	70.60	8,545	70.15	8,354	69.59

All data are at the plant-month level. Data on PM 2.5, temperature, precipitation and wind speed is from gridded satellite data. Each grid is 0.5-degree, covering an area of 50 square kilometers. Coal information from Central Electricity Authority reports. Temperature, precipitation and wind speed data from Copernicus Climate Change Service [2017].

4 Plant Openings and Closings as Shocks to Identify the impact of Coal Burning

The equation we wish to estimate is:

$$p_{i,t} = \alpha + \theta_i + \beta * coal_{i,t} + B * X_{i,t} + \epsilon_{it}$$

Here i indexes plant, and t indexes time which is a year-month combination. Each observation is identified by the triple of plant, year and month. α is an intercept; θ_i represents a plant specific effect; $coal_{i,t}$ is the coal burnt at time t in the grid that plant i is located in; X is a vector of controls, wind speed, temperature and precipitation in our case with a vector of coefficients B and ϵ is an error term. At time t, these controls take on the value for the grid plant i is located in. A plant specific effect captures certain properties that would otherwise conflate with the estimated β. Plants may vary in size implying a larger amount of coal burnt. Plants located in the Indo-Gangetic belt will be associated with a larger ambient PM2.5 concentration than other plants purely due to geography: winds carrying particulates blow from south to north and get trapped in the plain as the Himalayas form a natural northern boundary.

The outcome p indicates pollutant concentrations. Here we have defined it to include pollutant concentrations only in the grid plant i is allocated to. In order to account for the effects of wind, however, we would want to average all of the pollutant concentrations in grid cells downwind from plant i. Suppose there are a total of J grid cells. We would then estimate the following:

$$\overline{p_{j,t}} = \alpha + \theta_i + \beta * coal_{i,t} + B * X_{i,t} + \epsilon_{it}$$

$\overline{p_{j,t}}$ is the average pollutant concentration in all J grid cells. [6] In this specification, we allow for coal burning in plant i to affect pollutant concentration in all downwind grid cells, not just the grid the plant is located in. Note we are still only allowing for the controls in the grid the plant is located in to affect pollutant concentrations but this will be relaxed later.

Let $\overline{p_{j,t}} = P_{i,t}$. That is, P represents the average pollutant concentration in grid cells downwind from plant i and in the grid i is located in. To account for the plant specific unobservable θ_i, we take a difference from a lagged value. As θ_i is invariant with time, it will drop out, as will the intercept term α.[7]

$$P_{i,t} - P_{i,t-1} = \beta * coal_{i,t} - \beta * coal_{i,t-1} + B * X_{i,t} - B * X_{i,t-1} + \epsilon_{it} - \epsilon_{i,t-1}$$

That is, we estimate a first-difference model, which we shall refer to as

$$\Delta P_{i,t} = \beta * \Delta coal_{i,t} + B * \Delta X_{i,t} + \Delta \epsilon_{it} \tag{1}$$

We use a first-difference rather than a fixed effect approach, although the latter is also possible, owing to some features of the data. Accounting for the plant specific factor θ_i through a fixed effect approach rather than a first-differenced technique renders the coefficient on *open* problematic for the

[6] That is: $\overline{p_{j,t}} = \frac{\sum_{j=1}^{J} p_{j,t}}{J}$

[7] It is possible to also use a fixed effect approach. As we discuss later, such a technique delivers problematic estimates.

following reason. Typically we observe plants that open to slowly ramp up coal burning, although the first month when it opens is nevertheless a discrete jump up from zero. Since the coefficient on open reflects the difference from the average coal burnt in that plant in a fixed effect specification, this turns out to be negative, which is not sensible.

The estimate of β will be biased, however, because of the amount of coal burnt will be endogenously determined.[8] Changes in coal burning will come from changes in electricity generated, which will be related to economic activity. And economic activity can raise particulate emissions through channels other than coal burning.

To resolve this problem, we use coal plant openings and closings as instruments for $\Delta coal_{i,t}$:

$$\Delta coal_{i,t} = \gamma_1 * open_{i,t} + \gamma_2 * close_{i,t} + \eta_{i,t}$$

where $open_{i,t}$ is defined as follows:

$$open_{i,t} = \begin{cases} 1, & \text{if } coal_{i,t} > 0 \ and \ coal_{i,t-1} = 0 \\ 0, & \text{otherwise} \end{cases}$$

and $close_{i,t}$ is defined similarly:

$$close_{i,t} = \begin{cases} 1, & \text{if } coal_{i,t} = 0 \ and \ coal_{i,t-1} > 0 \\ 0, & \text{otherwise} \end{cases}$$

[8] $\Delta\ \alpha_t$ captures any month to month changes in PM2.5 concentrations due to seasonal changes.

We allow for the effects of *open* and *close* to vary by month. Planned maintenance of power plants is likely to take place in periods when demand isn't at its peak: for India, this would be around winter. Closures during the summer are likely unplanned, and we would expect these to be both shorter and possibly imply larger coal burning, as power has to be restored immediately.

$$\Delta coal_{i,t} = \sum_{m=1}^{12} \gamma_{om} * \lambda_m \times open_{i,t} + \sum_{m=1}^{12} \gamma_{cm} * \lambda_m \times close_{i,t} + \eta_{i,t} \qquad (2)$$

The identifying assumption is that plant openings and closings affect pollution levels only through their impact on coal burning. At a high enough frequency, plant openings and closings will create discrete changes in coal burning but other covariates, such as economic growth, are unlikely to change discretely in a short time frame. With the instrumental variables technique, we use only the predicted changes in coal burning coming from openings and closings, thus ensuring the estimated β is free from endogeneity bias.

With our definition of opening and closing dummies, we present evidence below that shows we have sufficient variation in these variables. Figure 8 shows the number of plant openings and closings over time. We change our definition of *open* and *close* slightly for ease of exposition, with $open_{i,t}$ coded as 1 if the plant opens in month-year t and $close_{i,t}$ coded as -1 if the plant closes in month-year t. In the regression results, however, we will use the earlier definition - that is, $close_{i,t}$ coded as 1 if the plant closes in month-year t. We can see a general trend toward increasing openings and closings, but closings are more likely toward the end of the sample period. Panel A of Table 1 shows summary statistics on plant openings and closings.

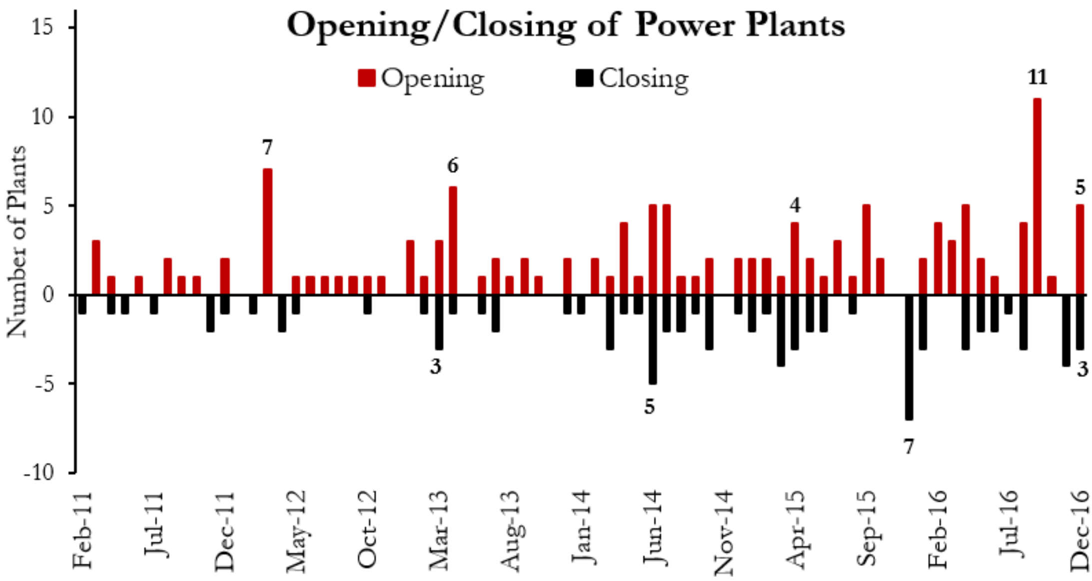

Figure 8: Plant Openings and Closings, a trend over time
(Source: Author Calculations)

We modify the second stage equation 1 in two ways. First, we relax the assumption that aggregate pollutant concentrations are unaffected by atmospheric conditions in downwind grid cells. We include a first differenced variable $\Delta EX_{i,t}$: $EX_{i,t}$ is a vector of controls that include the average wind speed, average temperature and average precipitation across all grid cells downwind of plant i, and D is a vector of coefficients.

Second, seasonal variation in pollutant concentration is captured by λ_m, where m indexes months as in the first stage. Particulate matter concentration varies widely within rather than across years, thus indicator variables for each month will capture such seasonality. We also include year indicator variables to capture any changes in concentration across years.

Our final econometric specification is given below. Driscoll-Kraay standard errors are used to account for heteroskedasticity, serial and spatial correlation.

$$\Delta P_{i,t} = \beta * \Delta coal_{i,t} + B * \Delta X_{i,t} + D * \Delta EX_{i,t} + \sum_{y=1}^{9} \lambda_y + \sum_{m=1}^{12} \lambda_m + \Delta \epsilon_{i,t} \quad (3)$$

$$\Delta coal_{i,t} = \sum_{m=1}^{12} \gamma_{om} * \lambda_m \times open_{i,t} + \sum_{m=1}^{12} \gamma_{cm} * \lambda_m \times close_{i,t} + \eta_{i,t} \quad (4)$$

When we estimate this set of equations off the data, we collapse the data by the following process: We round off the plants' coordinates to two decimal places, and collapse the data down to these latitude-longitude pairs. However, the definition of *open* and *close* are taken from plant-level data. The reason for carrying this additional step is that when we make our final simulations of morbidity impacts, we need to be sure we aren't double counting. Say two plants are situated very close by, and so their emissions will fall into the same set of grids downwind. Collapsing the data as we do ensures we are estimating the impact on downwind particulate matter from coal burnt by these two plants taken together and we don't treat the two plants as separate observations, as they are in fact contributing together to the pollution downwind.

5 Estimates of the Marginal Impact of Coal Burning on PM2.5 levels

Figures 9 shows the estimates of β, together with 95% confidence intervals, from equation 3 above. As we can see, they are all of the expected sign - more coal burning increases downwind pollution - and statistically significantly different from zero. The size of the coefficient is large: a standard deviation in coal burning raises downwind pollution by between 0.10 to 0.15 standard deviations.

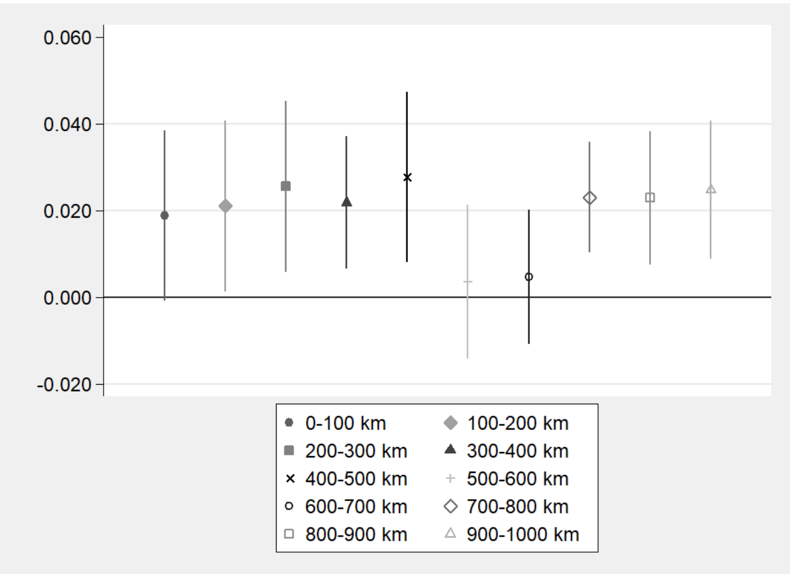

Figure 9: Estimates of β when coal consumption taken from electricity generation

F-statistics for the various rings are shown in Figure 10. We can see the instruments are highly correlated with the endogenous coal burning variable.

Finally, figure 11 show the first stage coefficients for each cone. For every

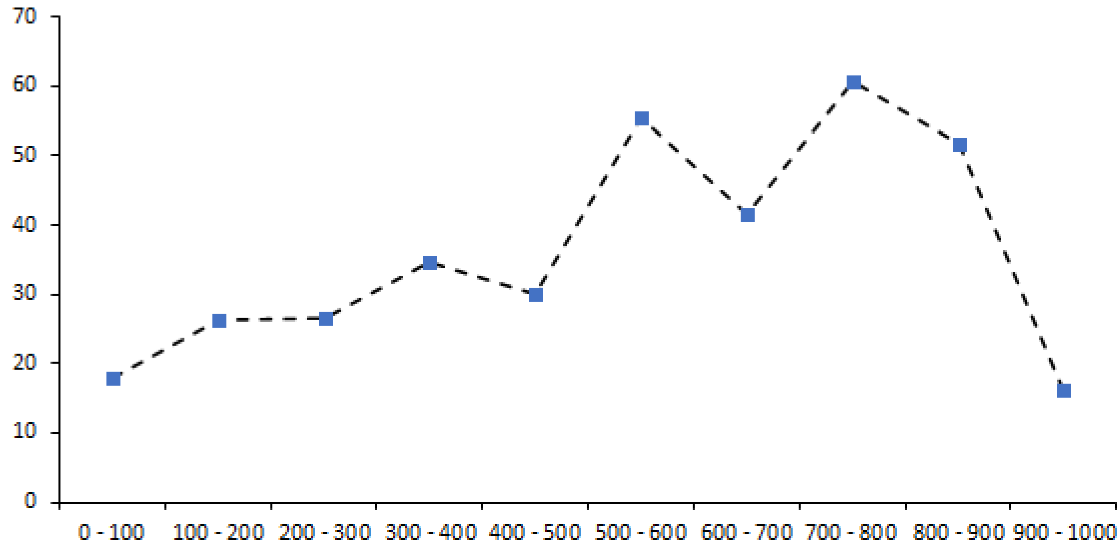

Figure 10: First Stage Kleibergen-Paap F-Statistics

month, there is a separate *open* and *close* coefficient. The sign on these is as per expectations - openings increase the amount of coal burnt and closings reduce it. The values of zero represent when the coefficient is dropped due to multicollinearity, presumably due to a lack of variation. When the values are small, the coefficients are statistically insignificant at the 5% level of significance, but when they are large, they are statistically significant at the 5% level of significance.

Diagnostics for the instruments therefore suggest we have valid and relevant instruments. Individual coefficients on most of the instruments are statistically significant, and of the correct sign with meaningful estimates when statistically significant. The F-statistics for the first stage indicate we have strong instruments. Since the errors are clustered by grid, we cannot use the Stock-Yogo critical values to evaluate whether we have a high enough F-statistic as these bounds are derived from the assumption of homoskedastic

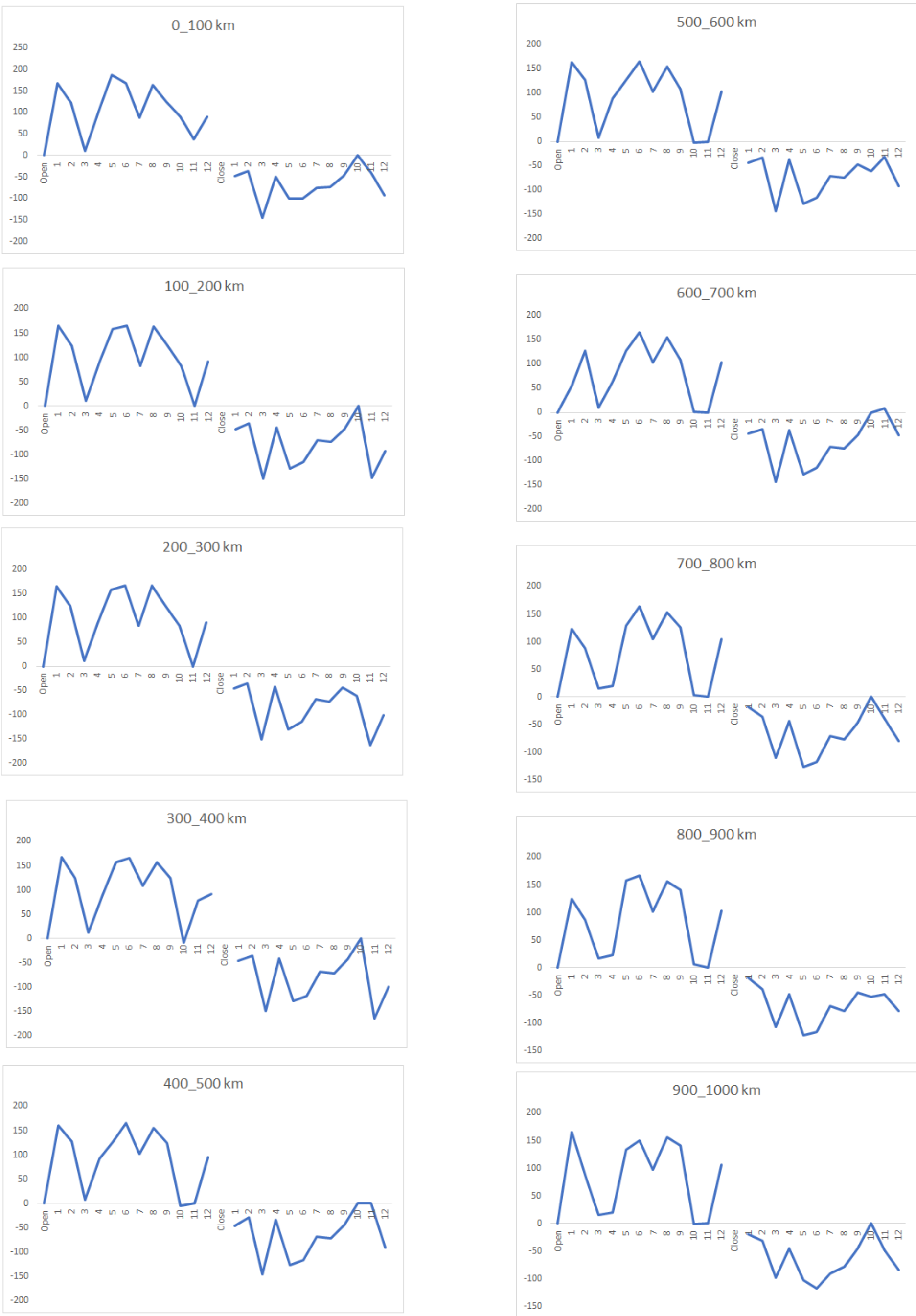

Figure 11: First Stage Coefficients for *open*, *close*, for each month coal consumption as from electricity generation reports

error terms. Critical values for the F-statistic in the case of heteroskedastic error terms are as yet unknown. The Hansen J test of overidentifying restrictions shows our instruments are valid.

6 Conclusion

In this paper, we account for the influence of coal burning on PM2.5 concentrations, using shocks to coal burning from plant openings and closings as instruments for coal burning. We find coal burning in one plant affects pollution positively and strongly in locations downwind from the plant.

Ultimately our goal is to provide policy makers with an estimate of the marginal impact of coal burning on health over the long term. As mentioned in the introduction, coal plants in India run at low capacity, electricity demand is likely to rise over time, and there are still households that do not have access to electricity ([Gupta and Spears, 2017], [Barrows et al., 2019]). All of these imply coal burning is likely to rise, and at the very least, not fall over time in order to meet this rise in demand.

In order to use these estimates to give a sense of the exposure resulting from coal burning, we will need to weight the results by population densities. With this information, we can calculate the implied long term health cost from coal burning using exposure-response functions already documented in the literature ([Pope-III and Dockery, 2006], [Ebenstein et al., 2017], [Burnett et al., 2018] and [Saraswat and Bansal, 2019]).

References

Barrows, G., T. Garg, and A. Jha (2019, May). Spatial disperson of costs and benefits of Power Plants in India.

Burnett, R., H. Chen, M. Szyszkowicz, N. Fann, B. Hubbell, C. A. P. III, J. S. Apte, M. Brauer, A. Cohen, S. Weichenthal, J. Coggins, Q. Di, B. Brunekreef, J. Frostad, S. S. Lim, H. Kan, K. D. Walker, G. D. Thurston, R. B. Hayes, C. C. Lim, M. C. Turner, M. Jerrett, D. Krewski, S. M. Gapstur, W. R. Diver, B. Ostro, D. Goldberg, D. L. Crouse, R. V. Martin, P. Peters, L. Pinault, M. Tjepkema, A. van Donkelaar, P. J. Villeneuve, A. B. Miller, P. Yin, M. Zhou, L. Wang, N. A. H. Janssen, M. Marra, R. W. Atkinson, H. Tsang, T. Q. Thach, J. B. Cannon, R. T. Allen, J. E. Hart, F. Laden, G. Cesaroni, F. Forastiere, G. Weinmayr, A. Jaensch, G. Nagel, H. Concin, and J. V. Spadaro (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences 115*(38), 9592–9597.

Central Electricity Authority (2019). Load Generation Balance Report.

Chan, H. S., M. L. Cropper, and K. Malik (2014). Why are power plants in India less efficient than power plants in the United States? *American Economic Review 104*(5), 586–590.

Copernicus Climate Change Service (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS).

Ebenstein, A., M. Fan, M. Greenstone, G. He, and M. Zhou (2017). New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences 114*(39), 10384–10389.

Gupta, A. and D. Spears (2017). Health externalities of India's expansion of coal plants: Evidence from a national panel of 40,000 households. *Journal of Environmental Economics and Management 86*, 262–276.

Health Effects Institute (2018). State of Global Air, Special Report. Boston, MA:Health Effects Institute.

Min, B., Z. O'Keefe, and F. Zhang (2017). Whose power gets cut? Using high-frequency satellite images to measure power supply irregularity. The World Bank Policy Research Working Paper 8131.

Pope-III, C. A. and D. W. Dockery (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the air & waste management association 56*(6), 709–742.

Saraswat, Y. and S. Bansal (2019, December). Health Effects of Sustained Exposure to Fine Particulate Matter: Evidence from India.

Venkatraman, C., M. Brauer, K. Tibrewal, P. Sadavarte, Q. Ma, A. Cohen, S. Chaliyakunnel, J. Frostad, Z. Klimont, R. V. Martin, D. B. Millet, S. Philip, K. Walker, and S. Wang (2018). Source influence on emission pathways and ambient PM 2.5 pollution over India (2015–2050). *Atmospheric Chemistry and Physics 18*(11), 8017–8039.

Wolfram, C., O. Shelef, and P. Gertler (2012). How will energy demand develop in the developing world? *Journal of Economic Perspectives 26*(1), 119–138.

World Health Organization (2018). Ambient Air Quality Database.