The Economic Benefits versus Environmental Costs of India's Coal-Fired Power Plants

Geoffrey Barrows, Teevrat Garg, and Akshaya Jha*†

January 22, 2018

Preliminary and incomplete; please do not cite without permission.

Abstract

Developing countries characterized by increasing electricity demand face a dilemma: fossil-fuel fired generation is cheap and reliable yet has substantial environmental consequences. Using a difference-in-differences approach comparing locations close to versus far away from coal-fired power plants in India, we show that increases in coal-fired capacity result in sizable increases in local air pollution levels and infant mortality. In contrast, using data on firm-level outcomes, district-level agricultural outcomes, and district-level "night-lights", we find that coal-fired capacity increases have relatively small (but precisely estimated) impacts on <u>local</u> economic benefits. Combined, our results indicate that the environmental costs of coal-fired power plants vary substantially over space while the economic benefits from these plants are distributed across the state or region; this suggests that coal-fired power plants should be sited based primarily on their environmental costs.

*Geoffrey Barrows: Ecole Polytechnique, Email: geoffrey-masters.barrows@polytechnique.edu. Teevrat Garg: School of Global Policy and Strategy, UC San Diego, Email: teevrat@ucsd.edu. Akshaya Jha: H. John Heinz III College, Carnegie Mellon University, 4800 Forbes Avenue, Pittsburgh, PA 15213. Email: akshayaj@andrew.cmu.edu.

†We thank Lucas Davis, Tim Fitzgerald, Lynne Kiesling, Dean Lueck, Leslie Martin, Edson Severnini, and Frank Wolak, as well as seminar and conference participants at the 2017 AERE Summer Conference, the 2017 Annual Conference for the Society for Institutional and Organizational Economics (SIOE), Berkeley Energy Camp, and the University of Illinois at Urbana-Champaign. We thank Kyle Navis and William Honaker for excellent research assistance. Garg acknowledges funding from the Center for Global Transformation at UCSD. Any remaining errors are our own.

1 Introduction

Despite substantial growth in clean/renewable energy production in the developed world, most poor countries still rely heavily on coal-fired power plants for electricity.[1] While cheap and reliable electricity is critical for the growth of developing countries (Wolfram, Shelef and Gertler, 2012; Gertler et al., 2016), coal-fired power plants dramatically impair local environmental quality, which has been shown to adversely affect health, productivity, and a range of other outcomes.[2] How do the social costs of coal-fired power generation – born by individuals that live and work near the power plants – compare to the benefits of electricity, which potentially are shared much more equally across space? Despite a large literature linking pollution to health, we still know very little about the costs and benefits of electricity suppliers, especially in the developing world.[3]

In this paper, we exploit temporal and geographic variation in the expansion of coal-fired power plants in India together with a newly constructed database of district-level infant mortality rates, wages, and productivity to derive the first ex-post estimates of the costs and benefits of coal-fired power plants in the developing world. The centerpiece of our analysis is a unique panel of district-level infant mortality rates (IMR) broken down by urban vs rural and gender over the period 1996 -2014 that we assembled from administrative records from the Indian Vital Statistics Civil Registry. These data allows us to compare changes in IMR over time across districts that saw new plant construction and districts that did not. This setting is particularly well-suited to estimate causal impacts because India significantly expanded its coal-fired capacity in recent years, and because migration rates in India are unusually low (Munshi and Rosenzweig, 2009), which means endogenous sorting plays a limited role in our context (Moretti and Neidell, 2011).

We address two main questions. First, we ask: what are the health effects of coal-fired power plants for people living and working near the plants? In theory, coal-fired power plants increase ambient pollution levels, which negatively impact health. But coal-fired power plants also generate benefits for end-use consumers of electricity that could contribute to *better* health outcomes. For example, increased electricity supply lowers price of heating and cooling, which could benefit health outcomes. Also, lower energy

[1]For example, coal-fired power plants generate 75% of grid-based electricity in India today, and this number is expect to increase to over 90% by 2030 (Shearer, Fofrich and Davis, 2017).

[2]See Zivin and Neidell (2013*a*) and Currie et al. (2014) for reviews on the effects of air pollution on human health in general.

[3]We know of only one other paper that estimates the health effects from power plant openings – Clay, Lewis and Severnini (2015), who estimate impacts in the US. We review the rest of the related literature below

costs to firms could feed through to higher wages and lower output prices, which could both benefit health outcomes. To the extent that these benefits accrue differentially to people living and working near the power plants, the reduced-form impact captures the net effect.

Second, we investigate local economic benefits directly. If electricity travels costlessly throughout the country, then local economic benefits should be minimal.[4] However, if there are transportation costs associated with electricity, either in the form of nominal costs or transmission loss, then firms or individuals located nearer to the plant may benefit more from the construction of a new plant compared to people living further away. We are interested in estimating these *local* benefits of coal-fired power plants both so that we can decompose the net health effects into direct effects (via pollution) and indirect effects (via consumption of electricity, employment, wages, final good prices, etc), and so that we can compare the local benefits against the local costs.

Our main specification exploits geographic and temporal variation across districts in coal-fired electricity production capacity. To measure exposure, we compute both total in-district capacity as the sum of plant-level capacity for all plants in the district, as well as area-based measures of the share of a district covered by disks of different radii around all plants (whether the plants are sited in the district or not). We then relate these two measures either to district-level or firm-level annual outcomes. The identifying assumption is that outcomes in districts that saw exposure increases would have trended the same way as districts that did not see exposure increases, absent the realized coal-fired plant expansion. In support of this assumption, we provide evidence that none of the outcomes measures we consider were trending differently for high exposure vs low exposure districts before the construction of plants.

Controlling for district fixed effects, state-year fixed effects, district level temperature and precipitation and weighting by live births, we find on net that coal-fired power plants dramatically increase infant mortality rates. While infant mortality rates were falling generally over the period, we find that an extra GW of installed capacity increased infant mortality rates (IMR) by 1.9 - 2.4 deaths per 1000 live births, depending on the specification. On a base rate of approximately 13 deaths per 1000 live births, this represents an 14-18% increase in IMR. Additionally, we find that these impacts were entirely driven by urban IMR, but do not see any statistically significant difference between male vs female

[4] Even if electricity travels near costlessly within states in India, there might be supply-side effects on health through employment at the plant itself. Though if the only supply-side effect stemmed from employment at the plant itself, it is unlikely that we would detect any difference at the district level.

infants. These results are robust to a variety of different specification checks.[5]

We next test for local economic benefits from coal-fired power plants on a array of outcome variables. First, we test for local benefits to manufacturing firms using the nationally representative Annual Survey of Industries (ASI). The ASI contains detailed information on electricity use, wages, material inputs, gross sales, and product level output prices by firm. The (relatively) newly available panel version of the ASI allows us to track firms over time, so we are able to control for firm fixed effects in the estimation. However, the panel version of the ASI omits district identifiers. In order to exploit the geographic variation in plant rollout, we designed our own procedure for merging district identifiers from the cross-section version.[6] With our geographically linked panel data, we find that increases in coal-fired capacity cause manufacturing firms to substitute away from own-generator production of electricity. This result implies that distance to the plant matters, at least for manufacturing firms. However, despite the robust result on electricity use, we fail to reject the null of no impact on wages, employment, labor productivity, total factor productivity, and product-level output prices. The null result on wages and prices suggest that economic benefits do not accrue differentially to worker/consumers living near power plants via supply-side channels. Additionally, the null result on sales and productivity imply that manufacturing firms do not benefit directly as a result of having coal-fired plants nearby, even though they substitute towards grid-based electricity.

Next, we test for benefits to agricultural wages and yields. Electricity is an important input into the production of home goods and agricultural outputs. Increased electricity supply could push up labor demand via input complementarities, which would raise wages and yields. To test for impacts in the agricultural sector, we examine data from the Village Dynamics in South Asia Meso data set, which is compiled y researchers at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT 2015). The data set provides district-level information on annual agricultural production, prices, acreage, and yields, by crop. We generate aggregate price-weighted district level measures of total yield in each

[5]In the appendix, we also provide evidence that coal-fired power plants differentially increased satellite based measures of local air pollution. However, since our input variable potentially affects health through many different pollutants, estimating the IV impacts of any one pollutant on IMR is invalid (as in Greenstone and Hanna (2014)). Furthermore, while the impacts on NO_2 are precisely estimated, we fail to reject a null result on $PM_{2.5}$. The conclusion that coal-fired power plants fail to increase $PM_{2.5}$ concentrations seems highly suspect, and suggests that even though satellite-based measures of pollution provide better coverage than ground monitors, researchers should take caution in relying exclusively on satellite-based measures to estimate pollution effects.

[6]Martin, Nataraj and Harrison (2017) was the first to merge district identifyers from the cross section, though we design our own procedure for this.

district for the six major crops (rice, wheat, sugarcane, groundnut, sorghum, and maize) as well as the five major monsoon crops (read: excluding wheat). ICRISAT also provides data on district level averages of rural wages and employment separately for men and women. With these data, we fail to reject the null of no impact on either wages or yields, further supporting the lack of local economic benefits.

Lastly on the benefits, we estimate the local impact on aggregate economic activity using satellite-based measures of light intensity at nights, "nightlights" (Henderson, Storeygard and Weil, 2012). We fail to reject the null of no impact.

Taken together, our paper shows that coal-fired power plants generate large local health costs to people living near the coal plants, while yielding minor (if any) local benefits. From a policy perspective, our findings highlight that building new coal-fired power plants sufficiently far away from population centers minimizes environmental costs without significantly affecting the economic benefits from these plants; a similar logic applies to shutting down existing coal-fired plants. Importantly, though we do not comprehensively quantify either the full local environmental costs or economic benefits from coal-fired power plants, the magnitude of our effect linking coal-fired capacity to infant mortality is strikingly large. This suggests that policymakers would have to apply a very low value of statistical life to the total number of deaths due to coal-fired electricity generation in order to justify the existing level and spatial distribution of coal-fired generation in India (Greenstone and Jack, 2015).

Our paper contributes to three strands of literature. First our paper contributes to a small group of papers that estimates the health impacts of power/industrial plants. Luechinger (2014) and Tanaka (2015) study the impacts of regulations aimed at power plants in Germany and China, respectively, finding that regulation-induced reductions in air pollution lower IMR. Cesur, Tekin and Ulker (2017) finds that natural gas infrastructure expansion lowered IMR in Turkey through fuel-switching away from coal use by households. Beach and Hanlon (2016) finds that cities in the England in the 1850s with more industrial coal use had higher IMRs. And Lavaine and Neidell (2017) finds that strikes in France that shut down oil refineries contributed to improved health outcomes. Relative to these papers, we study plant openings, rather than regulation or changes in infrastructure or industrial composition. While the epidemiological effects of coal should be the same (conditional on disease environment and individual health), the attendant economic benefits are likely different from plant openings compared to regulatory effects. Thus, the net relationship between pollution and health in our context could be quite different from previous work.

Additionally, opening new coal-fired plants generates considerably larger pollution changes than regulations on the intensive margin, so which might produce larger impacts if the dose-response function is non-linear. Additionally, the estimates of the elasticity between health and pollution are mostly based on developed-world studies, which might not be valid in developing countries (Arceo-Gomez, Hanna and Oliva, n.d.).

Also within this group, there are three closely related papers that *do* study the impact of plant openings. First, Currie et al. (2015) studies the health effects from 1600 toxic plant opening and closings in the US. Compared to Currie et al. (2015), we focus just on coal-fired power plants (where Currie et al. (2015) includes lots of different industrial plants), consider different pollutants, and look at a larger radius around the plant such that treatment could include some local economic benefits as well. Clay, Lewis and Severnini (2015) studies the benefits and costs of coal plant openings in the U.S. in the mid 20th century. With low levels of environmental regulation and electrification rates, mid 20th century U.S. resembles our development setting to some degree. Finally, Gupta and Spears (2017) also estimate the health impacts of coal-fired power plant openings in India, though Gupta and Spears (2017) only studies plants that opened between 2005 and 2012 and do not estimate impacts on infant mortality or firm-level outcomes.

Second, we contribute to the literature on electrification/electricity supply. This literature tends to focus on impacts of grid expansion (Burlig and Preonas, 2016; Barron and Torero, 2017; Dinkelman, 2011; Lipscomb, Mobarak and Barham, 2013), which generates benefits from electrical services for households, though not much costs in terms of pollution from power plant output.[7] To the extent that benefits from electricity services from plant openings disproportionately affect households nearer to the power plants, our estimates include these benefits. Also on the supply side, Allcott, Collard-Wexler and O'Connell (2016) estimates firm-level impacts of supply shocks to hydroelectric power plants in India.

Finally, our paper contributes to the broader literature on environment and health in the developing world. While this literature is small relative to it's developed-world counterpart (Greenstone and Jack, 2015), a handful of studies have estimated the health impacts of air pollution from sources such as wild fires (Jayachandran, 2009), vehicles (Greenstone and Hanna, 2014), exporting (Bombardini and Li, 2016), and thermal inversions (Arceo-Gomez, Hanna and Oliva, n.d.). By relating health outcomes directly to proximity to polluting activity, we bypass the thorny issue of measuring pollution with sparsely distributed ground monitors.

[7] though increased electricity consumption obviously also means increased supply, the attendant effects of this supply increase is not a focus of this literature

2 Conceptual Framework

In this section, we present a conceptual framework to explain how the construction of new coal-fired power plants impacts health and wages. In general, health and wages are jointly determined, so reduced-form estimates of coal capacity on health and wages fail to identify the individual channels. However, under certain assumptions or in certain circumstances, reduced-form estimates can shed light on the magnitudes of local costs and benefits.

A representative worker-consumer supplies 1 unit of labor inelastically and derives utility from health H, and a composite good X, with

$$U = U(X, H) \tag{1}$$

Health depends on the incidence of becoming sick ϕ and ex-post medical expenses, M. The incidence to become sick depends on ambient outdoor air pollution Z, which results from coal-fired electricity plants connected to the grid. Indexing coal-fired electricity plants by θ, we have

$$Z \quad = Z(\theta) \tag{2}$$

and

$$H \quad = \quad H\left(M, \phi(Z(\theta))\right) \tag{3}$$

where $H()$ is twice continuously differentiable.

Worker wage w is determined in equilibrium by the local demand for labor. Labor demand in turn is determined in part by supply of electricity as in Allcott, Collard-Wexler and O'Connell (2016). Wages also depend on work productivity, which depends on H. Thus, we have

$$w \quad = w(H, \theta) \tag{4}$$

Worker-consumers solve the problem

$$\underset{X,M}{\text{Max}} \quad \mathcal{L} = U(X, H) - \lambda\left[p_X * X + p_M * M - w(H, \theta)\right] \tag{5}$$

This set up is similar to the one from Zivin and Neidell (2013*b*), though we omitted medical expenditures that mitigate the impacts of pollution on ϕ and add a direct income effect of coal capacity through the supply side.

First order conditions with respect to X and M implicitly define optimal consumption plans

$$X^* = X^*\left(p_X, p_M, \phi(\theta), \theta\right) \tag{6}$$
$$M^* = M^*\left(p_X, p_M, \phi(\theta), \theta\right) \tag{7}$$

In a context with low migration costs, welfare costs of θ can be estimated by the change in housing prices. Optimal consumption plans X^* and M^* determine the indirect utility function, which can then be interpreted as the value of living in a given region.

However, in a context with high migration costs, it cannot be assumed that residential sorting equilibrates the housing market. In this case, there is no convenient sufficient statistic for welfare. To get a sense of the costs and benefits, we investigate reduced-form impacts to health and wages. Substituting M^* into 3 and 4 and totally differentiating, we have

$$\frac{dH}{d\theta} = \frac{\partial H}{\partial M}\left[\frac{\partial M}{\partial \phi}\frac{\partial \phi}{\partial Z}\frac{\partial Z}{\partial \theta} + \frac{\partial M}{\partial \theta}\right] + \frac{\partial H}{\partial \phi}\frac{\partial \phi}{\partial Z}\frac{\partial Z}{\partial \theta} \tag{8}$$

which can be rearranged as

$$\frac{dH}{d\theta} = \underbrace{\left[\frac{\partial H}{\partial M}\frac{\partial M}{\partial \phi} + \frac{\partial H}{\partial \phi}\right]\frac{\partial \phi}{\partial Z}\frac{\partial Z}{\partial \theta}}_{\text{Net Pollution Effect}} + \underbrace{\frac{\partial H}{\partial M}\frac{\partial M}{\partial \theta}}_{\text{Supply Side Health Effect}} \tag{9}$$

and

$$\frac{dw}{d\theta} = \underbrace{\frac{\partial w}{\partial H}\frac{dH}{d\theta}}_{\text{Health Productivity Effect}} + \underbrace{\frac{\partial w}{\partial \theta}}_{\text{Supply Side Income Effect}} \tag{10}$$

The reduced-form health effect $\frac{dH}{d\theta}$ is the sum of a Net Pollution Effect (NPE) and a Supply Side Health Effect (SSHE). The former results from both increased incidence of sickness due to coal-fired power plants and from increased ex-post medical expenditures induced by coal-related illnesses. Hence, the NPE is net of mitigating behaviors (hence the term “Net”). The SSHE results from increased medical expenditures induced by higher wages. As long as medical expenses are a normal good, the SSHE is weakly greater than zero. The reduced-form wage effect $\frac{dw}{d\theta}$ is the sum of a Health Productivity Effect (HPE) and a Supply Side Income Effect (SSIE). The former results from lower productivity due to weaker health status, while the latter represents the pure income effect from higher labor demand.

The NPE can be thought of as a pure environmental cost of the coal-fired plant. The SSIE is a pure benefit of the plant. The SSHE is also a benefit of the coal-fired plant in the sense that it leads to better health, while the HPE is also a cost of the plant, in that it leads to lower wages.

What can we say about the local costs vs benefits of the coal fired power plant? First, suppose we only consider the impact on health. In this case, we only identify the net effect on health. We cannot say how large is the NPE nor the SSHE. The magnitude of $\frac{dH}{d\theta}$ merely reflects the difference between the two.

Now assume we can estimate $\frac{dH}{d\theta}$ and $\frac{dw}{d\theta}$. If we assume that SSHE=HPE=0, then $\frac{dH}{d\theta}$ and $\frac{dw}{d\theta}$ reduce to just the NPE and the SSIE, and we can quantify the local environmental costs vs local environmental benefits. Without this assumption, it is not clear what $\frac{dH}{d\theta}$ and $\frac{dw}{d\theta}$ say about local costs vs local benefits.

In the event that we estimate $\frac{dH}{d\theta} < 0$ and $\frac{dw}{d\theta} = 0$, we have two possible cases. Either, $\frac{\partial w}{\partial H} = \frac{\partial w}{\partial \theta} = 0$, in which case there are no local economic benefits and $\frac{dw}{d\theta}$ represents just the NPE. Alternatively, we could have $\frac{\partial w}{\partial \theta} = -\frac{\partial w}{\partial H}\frac{dH}{d\theta}$. I.e., the SSIE exactly offsets the HPE, yielding no reduced-form effect on wages. In this case, $\frac{dH}{d\theta}$ understates the NPE and the SSIE is proportional to the elasticity of wages with respect to health

$$\frac{\partial w}{\partial \theta} \propto \frac{\partial w}{\partial H} \tag{11}$$

3 Background and Empirical Strategy

India is an ideal place to study the impact of coal-fired power plants because (i) India has dramatically increased it's coal-fired capacity in recent years (ii) regional health data can be linked to coal plants at high spatial resolution (iii) migration rates are very low, so endogenous sorting should not be an issue. In this section, we briefly review key features of the electricity production sector in India, present our data on coal-fired power plants, and then discuss our empirical strategy.

3.1 Electricity Production in India

Today, roughly 75% (73%) of India's (China's) electricity generation came from coal-fired sources in 2015 (based on World Bank statistics). Moreover, coal-fired power plants in India generated roughly 950TWh of electricity in 2016, making India one of the largest consumers of coal in the world. Unfortunately, India does not currently have cost-effective, reliable alternatives to coal for generating electricity. For example, unlike the United States, India does not have plentiful reserves of natural gas; only 5% of India's electricity generation came from natural gas fired sources in 2014 (World Bank statistics). Even as of the third quarter of 2017, Bloomberg estimates that the levelized cost of generating electricity from coal-fired sources in India is 52.03 USD/MWh compared with 94.66 USD/MWH for natural gas fired sources. Finally, though the share of renewable resources is rapidly growing, these sources only provided 5% of India's electricity generation in 2015 and remain relatively costly to install.

Relative to coal mined in the United States, Indian coal typically has high ash content (ranging from 35-50%), high moisture content (4-20%), low sulfur content (0.2-0.7%), and low calorific values (between 2500-5000 kcal/kg, which is much less than the normal range of 5000 to 8000 kcal/kg) (Mittal, Sharma and Singh, 2012). The low sulfur content results in a relatively low level of SO_2 emissions from burning Indian coal, but the high moisture and ash contents along with the low heat content makes Indian coal particularly environmental unfriendly in terms of CO_2, $PM_{2.5}$, and NO_2 emissions. CO_2 emissions are a global stock pollutant that contribute to climate change rather than a local pollutant such as $PM_{2.5}$ that has negative health impacts (including increased mortality risk). However, most plants in India have installed electrostatic precipitators (ESPs) designed to mitigate fine particulate ($PM_{2.5}$); we do not observe the installation date for these ESPs in our data nor whether a given power plant even has an ESP.

The location of coal-fired power plants in India has been influenced by regulation beginning with the Third Five Year Plan (1961-66). Currently, new power plants are typically built further than 25km of the outer periphery of a city, 500m away from the flood plain of the river system, and based on availability of land, water, and coal (either transportation infrastructure or a mine). The current guidelines certainly reflect some of the environmental costs of placing coal-fired power plants in cities as well as the relatively low costs of transmitting electricity long distances. However, older coal-fired power plants still in operation today were sometimes built in or near cities in order to easily serve electricity demand in these cities; two particularly salient examples are the recently shut down plant in Delhi (Badarpur Thermal Power Station) and the still operating plant located near Mumbai (Trombay Thermal Power Station).

Finally, India now has a single electricity transmission grid that covers the entire country; this was not always the case. For example, India connected its southern power grid with the national grid only in late 2013.[8] Understanding the extent to which India's transmission grid can effectively move electricity from coal-fired power plants to electricity demand centers is crucial for identifying whether the economic benefits from these coal-fired power plants are primarily local or regional.[9] In particular, our paper sets out to answer: are the districts in India with coal-fired power plants compensated via economic growth for the local environmental costs they face from their plants?

3.2 Empirical Strategy

We exploit geographic and temporal variation in the expansion of coal-fired production capacity in India in recent years to estimate causal impacts. To measure exposure, we assembled data on the precise geographic location of each coal-fired power plant in India (latitude and longitude) as well as opening date (and shut down date, if applicable), and the date of any additional capacity installations to the plant from administrative records of the Central Electricity Authority (CEA).

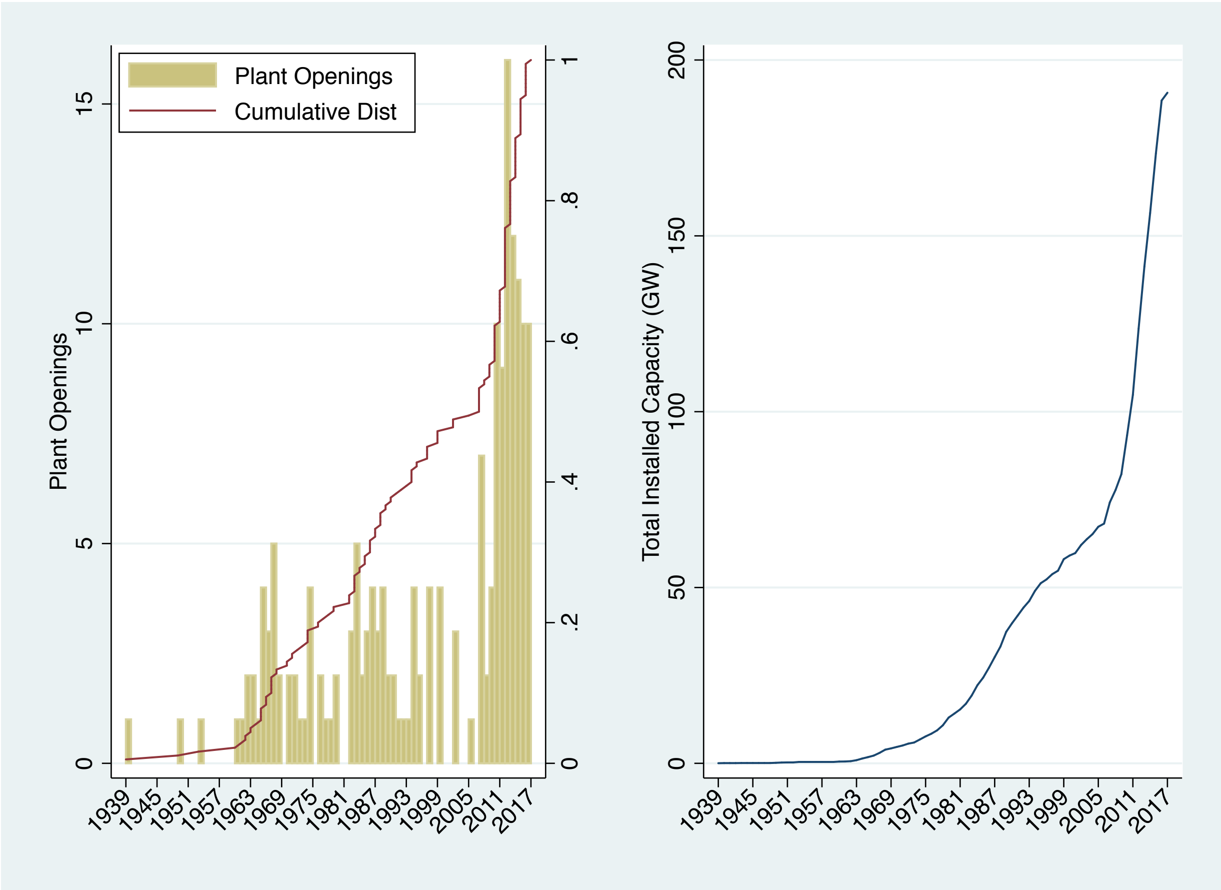
Figure 1 plots the rollout of coal-fired power plants in India by year between 1939 and 2017. The left panel depicts plant openings (left axis) and cumulative plant openings (right axis) by year. In total, we count 180 coal-fired power plant openings over the period. From the red line in the left panel (which displays the cumulative distribution function of plant openings), we see that 50% of plant openings occurred after 2006, so there has clearly been a large acceleration in the construction of coal-fired power plants. In the right panel of Figure 1, we plot total installed capacity over time in Gigawatts (GW) and find a similar acceleration in the level of installed capacity after 2006. By the final year of our sample (2017), we observe 191 GW of coal-fired capacity, with approximately 50% of this total having been constructed since 2009.[10]

[8] For more information, see this news article titled "India interconnects southern power grid with 765 kV transmission line": `http://www.elp.com/articles/2014/01/india-interconnects-southern-power-grid-with-765-kv-transmission-line.html`.

[9] Ryan (2017) quantifies the economic benefits from increased electricity transmission in India inclusive of the reduction in the exercise of market power due to these increases in transmission.

[10] Total installed capacity is measured in GW. To convert GW to the maximum amount of possible electricity generated in a year in Gigawatt-hours (GWh), we multiply by 8760 hours. For the final year in the sample, this gives us $191 \times \frac{8760}{1000} = 1670.5$ TWh of potential electricity generated in a year.

Figure 1: Roll-out of Coal-Fired Power Plants over Time

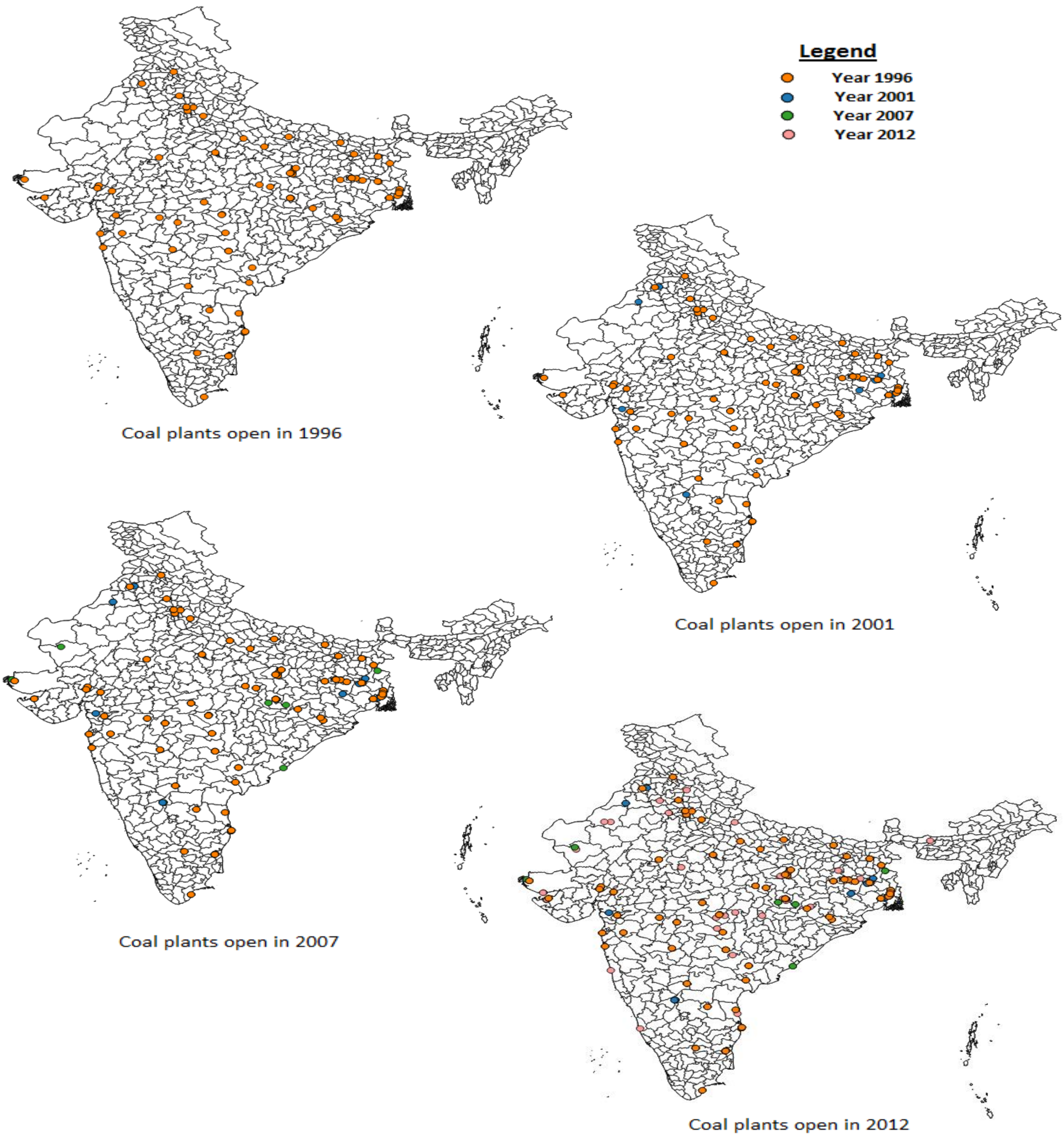


Notes: On the left, left axis counts number of plant openings per year. Right axis plots cumulative share of plants opened by year (red line).

Figure 2 presents the location of online plants for 4 years in the sample. Here, we see the geographic variation in the rollout. In Figure 2, we can see that coal-fired plants are distributed throughout India, though there are some clusters of plants in the North and Northeast. It is also clear from Figure 2 that plants are often built near district boundaries. Given that air pollution can travel fairly large distances (up to 100km), proximity to plants outside the district borders should also be taken into account in computing exposure.

The main results in the paper relate variation in coal-fired electricity capacity to district-level outcome metrics. At the district-level, we compute two measures of exposure. First,

Figure 2: Map of Coal-Fired Power Plants



Notes: Sub-figures show the location of active plants in the year indicated. Colors correspond to the year of plant opening.

we sum up total installed capacity in each district in each year

$$DistCap_{d,t} = \sum_{p=1}^{P} PlantCap_{p,t} \tag{12}$$

where d indexes districts, p plants, and t years. Second, to take account of air pollution from neighboring districts, we compute the share of each district that intersects with circles of different radius around each plants. That is, we construct a weighted average of installed capacity where the weights correspond to the share of each district that intersects a circle of radius k around each plant $sh_{p,d}^{radius=k}$:

$$Exposure_{d,t}^{radius=k} = \sum_{p=1}^{P} PlantCap_{p,t} * sh_{p,d}^{radius=k} \tag{13}$$

For small radii, only plants within the district borders will receive positive weights. However, when we increase the radius, the installed capacity from neighboring plants will enter into the sum in (13).

With these exposure measures, we estimate district-level impacts in a difference-in-difference-like framework

$$Y_{d,t} = \alpha_d + \theta_{s,t} + \beta * Coal_{d,t} + X_{d,t}\gamma + \epsilon_{d,t} \tag{14}$$

where $Y_{d,t}$ represents a district level outcome , $Coal_{d,t}$ is either $DistCap_{d,t}$ or area-weighted exposure $Exposure_{d,t}^{radius=k}$, α_d is a district fixed effect, $X_{d,t}$ are time-varying district-level controls and $\epsilon_{d,t}$ an idiosyncratic error term. Baseline specifications also control for flexible state-year effects $\theta_{s,t}$, which absorb any time-varying trends at the state-level.

The identifying assumption in equation (14) is that outcomes for districts that saw increased exposure to coal-fired power plants would have trended in a parallel fashion to districts that did not see increases, absent the observed rollout. While this assumption is inherently untestable, we can test for whether outcomes were trending together in the years prior to capacity expansions. Following SS, we estimate a flexible "event study" specification that allows for multiple "events" per district. Let j indicate an "event," which corresponds to an increase in district-level coal-fired capacity. We estimate:

$$Y_{d,t} = \alpha_d + \theta_{s,t} + \sum_{j}^{J_d} \sum_{l=-L}^{L} \beta^l * 1(t - e_d^j = l) + X_{d,t}\gamma + \epsilon_{d,t} \tag{15}$$

where t index years in calendar time, l indexes event time, and e_d^j is the year of event j occurring in district d. The β^l's capture the impact of an event on outcome Y l years in the past or future. If the timing of coal plant openings are exogenous to unobserved determinants of outcome Y, we should have $\beta^l = 0$ for all $l \leq 0$.

For each outcome variable, we will present estimates of (14) and (15). In the case of (14), we estimate taking by turns $DistCap_{d,t}$ and $Exposure_{d,t}^{radius=k}$, for different radii k.

4 Local Health Costs of Coal-fired Power Plants

This section estimates the net cost to infant mortality of coal-fired capacity. We first discuss our data for district-level infant mortality rates from 1996-2014, and then present our empirical results.

4.1 IMR Data

We obtain data on infant mortality and number of live births for 1996-2014 recorded in the Civil Registration System (CRS) and aggregated at the district level as part of the annual report on the Vital Statistics of India. Each state is responsible for reporting district level measures on live births, infant deaths and total deaths disaggregated by urban and rural localities and by gender. While the estimates from CRS are typically lower than other estimates of infant mortality, the district level variation is similar to other data sources such as the Sample Registration System (Greenstone and Hanna, 2014).

Table 1 presents descriptive statistics on infant deaths per 1000 live births. Columns 1-4 (5-8) present statistics for districts that never (ever) site a coal-fired power plant. An observation corresponds to a district-year.

Figure 3 plots average IMR over time. The left panel plots the overall average as well as the average for rural vs urban locations, while the right panel breaks out rural and urban each by gender. We can see that average IMR has been falling in India, both overall and within each sub-population.

4.2 IMR Results

Tables 2 and 3 presents our main findings with respect to infant health. In Table 2, Panels A and B estimate equation (14) taking $DistCap_{d,t}$ and $Exposure_{d,t}^{radius=100}$, respectively as the measure of exposure. Table 3 takes $Exposure_{d,t}^{radius=k}$ as exposure for different k. In all

Table 1: Infant Death per 1000 Live Births

	No Coal Plants *567 Districts*				Coal Plants *108 Districts*			
	(1) mean	(2) min	(3) max	(4) obs	(5) mean	(6) min	(7) max	(8) obs
Total	13.6	0.01	800	6489	11.1	0.01	99	1521
Urban	15.6	0.02	2406	5999	12.4	0.01	204	1489
Male	13.7	0.06	983	3811	12.3	0.02	101	902
Female	11.7	0.05	855	3825	10.1	0.06	98	907
Rural	14.6	0.01	597	6100	10.1	0.04	92	1426
Male	12.6	0.03	250	3781	7.9	0.05	120	844
Female	12.3	0.04	142	3786	8.0	0.05	106	849

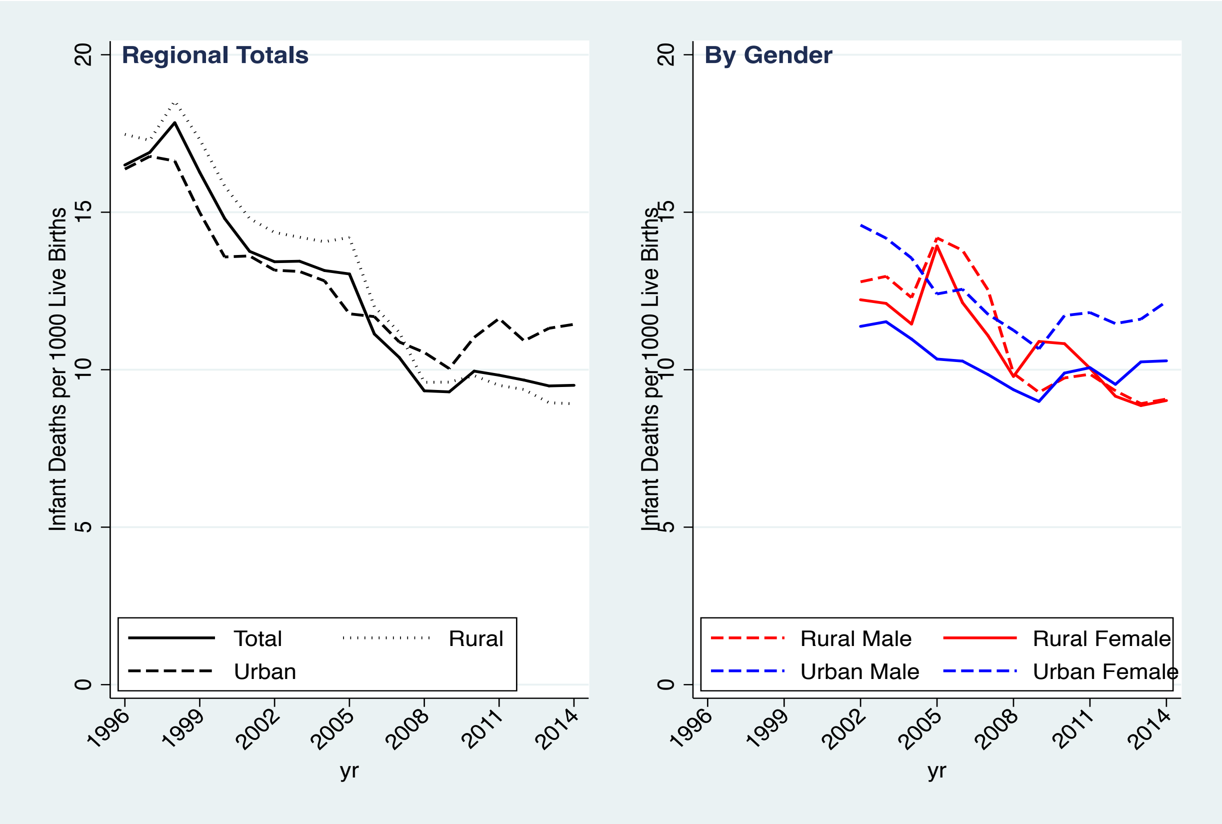
Notes: Observation corresponds to a district-year. Columns 1-4 (5-8) present statistics for districts that never (ever) site a coal-fired power plant.

cases, the exposure measure has been transformed so that a unit increase corresponds to 1 standard deviation (1σ). Finally, Panel C presents estimates of equation (15), where we group event time into 5 periods. Event time $\in [-4, -2]$ indicates years that are between 2 and 4 years before the capacity increase. Event time $\in [-1, 1]$ indicates years that are within 1 year of the capacity increase on either side of the construction date. Event time $\in [2, 4]$ and $\in [5, \infty]$ indicate years that are 2 to 4 years and 5 to infinity years after the event, respectively. The excluded category is Event time $\in [-\infty, -5]$, so all coefficients in Panel C are relative to the years 5 or more before the event. All regressions control for district and state-yr fixed effects, as well as district-level temperature, precipitation, and number of live births. All regressions are weighted by the number of live births and standard errors are clustered at the district level.

Column 1-3 in Tables 2 and 3 present the results for the full sample. We have 5841 district-years observations across 457 districts for which we have non-missing values for total, urban, and rural IMR. In Panel A, we find that total and urban IMR increase with in-district capacity, though the point estimates are not statistically distinguishable from zero. Point estimates in Panel B of Table 2 and throughout Table 3 , where we account for proximity to coal-fired plants beyond the district borders are very similar. In Panel C of Table 2, we find fail that trends in IMR are related to coal-fired capacity prior to the capacity increase (Event time $\in [-4, -2]$), while IMR sometimes increases after the event.

The weak statistical evidence of causal impacts in column 1-3 are perhaps surprising,

Figure 3: IMR over time



Notes: .

given the amount of toxic pollution generated by burning coal. However, in columns 4-6, we find that the weak correlations are driven entirely by two outlier districts that built enormous coal plants at the end of the sample. When we throw out these two district from the estimation, we find large and statistically significant impacts on total IMR and Urban IMR. In panel A of Table 2, we find that a 1σ increase in coal-fired capacity exposure leads to 1.27 more infant deaths per 1000 live births overall, and 2.05 more deaths per 1000 in urban areas. These estimates are statistically significant at the 5% and 1% levels, respectively. In column 6, we find no corresponding impact on rural IMR. We find the same qualitative result in Table 3, though the point estimates are attenuated somewhat. Finally, we find no evidence of differential trends prior to capacity increases (panel C), which bolsters support for the parallel trends assumption. These results are robust to the inclusion of year fixed effects instead of state-year effects, as well as omitting analytical weights. We can also vary the definition of $Exposure_{d,t}^{radius=k}$ to reflect the share of total area of the district covered, and results are qualitatively unchanged.

In terms of magnitudes, the point estimates in panel A imply that a 1GW increase in coal-fired capacity increases total IMR 1.27*.52 =2.43 deaths, with a 95% confidence interval of (0.47, 4.40). On a base rate of 13.15, this level increase implies an 18.5% increase in IMR overall, with a 95% confidence interval of (3.6%, 33.4%). Looking just at Urban IMR, the point estimate in panel A implies a 1GW increase in IMR causes 2.05*0.52=3.94 more deaths (95% confidence interval of (1.15, 6.73)), or an increase of 26.4% (95% confidence interval of (7.7%, 45.1%)). The annual district-level costs of this average increase in overall infant mortality rates, using a value of statistical life of 1.2 million dollars (USD 2000) (Viscusi and Aldy, 2003), is roughly 93 million dollars (USD 2000).[11]

Next, in Tables 4 and 5, we present results separately by gender. IMR rates by gender are available only for the more recent years (2002 - 2014). In columns 1 and 4 of Tables 4 and 5, we present results for the full sample aggregated to urban (column 1) and rural (column 5). Sample in columns 1 and 5 are slightly larger than in Tables 2 and 3 because we do not impose here that all three aggregate measures (total, urban, and rural) have non-missing values. Columns 2 and 6 restrict to observations for which regional totals as well as male and female breakdowns are all non-missing. Hence, the estimates in columns 2-4 and 6-8 derive from the same set of observations. We have again excluded the two outlier districts. Specifications are identical to Tables 2 and 3.

In Tables 2 and 3, we find again that overall impact are driven entirely by urban areas. Neither the overall levels nor the gender breakdowns yield results that are statistically distinguishable form zero for the rural populations. In columns 5-8, we find that impacts on urban IMR are large and statistically significant at conventional levels. Additionally, we find that the impact on male IMR is stronger than the impacts on female IMR.

There are two possible explanations for why coal impacts are higher for urban than for rural IMR. First, it could be that coal plants are systematically placed closer to urban areas. Hence, urban parts of districts experience higher exposure to coal plants. Second, even if exposure were equal across urban vs rural areas, there could be non-linearities in the does-response function. Urban areas tend to have higher pollution rates. If IMR is particularly sensitive to pollution at higher levels, then we would expect to see the type of heterogeneity in Tables 2 and 3.

[11] The full calculation is 2.43 deaths per 1000 * 33,094 live births * 1.2 million USD = 92.93 million USD

Table 2: IMR Results

	Full Sample			Restricted Sample		
	Total	Urban	Rural	Total	Urban	Rural
	(1)	(2)	(3)	(4)	(5)	(6)
<u>*Panel A*</u>						
Cap (1σ)	0.54	0.88	0.05	1.27**	2.05***	0.04
	(0.43)	(0.60)	(0.40)	(0.52)	(0.74)	(0.88)
R squared	0.78	0.58	0.75	0.78	0.59	0.75
<u>*Panel B*</u>						
ACap (1σ)	0.53	0.90	0.09	1.00*	1.69**	0.12
	(0.39)	(0.57)	(0.38)	(0.52)	(0.78)	(0.67)
R squared	0.78	0.58	0.75	0.78	0.59	0.75
<u>*Panel C*</u>						
Event time $\in [-4, -2]$	0.39	0.54	0.06	0.32	0.52	-0.03
	(0.36)	(0.59)	(0.42)	(0.37)	(0.63)	(0.43)
Event time $\in [-1, 1]$	0.61	2.01***	-0.48	0.70	2.28***	-0.60
	(0.53)	(0.71)	(0.54)	(0.58)	(0.74)	(0.59)
Event time $\in [2, 4]$	0.18	0.88	-0.42	0.19	0.98	-0.53
	(0.54)	(0.94)	(0.46)	(0.58)	(0.99)	(0.51)
Event time $\in [5, \infty]$	1.30**	1.63	0.22	1.34*	1.79*	0.10
	(0.65)	(0.99)	(0.68)	(0.68)	(1.03)	(0.73)
R squared	0.78	0.59	0.75	0.78	0.59	0.75
mdv	10.6	12.2	9.6	10.6	12.2	9.7
# Obs	5841	5841	5841	5817	5817	5817
# Districts	457	457	457	455	455	455

Notes: All regressions include district and state-yr fixed effects, as well as district-level controls for temperature, precipitation, and number of live births. All regressions are weighted by the number of live births and standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

To investigate the mechanism behind the heterogeneity in urban vs rural, we estimate heterogeneity in density of population around the coal plant sites. If it is true that coal plants are placed closer to urban areas, then we should see higher population density closer to the plants. We first compute average population density in concentric circles around

Table 3: IMR Results ACap

	Full Sample			Restricted Sample		
	Total (1)	Urban (2)	Rural (3)	Total (4)	Urban (5)	Rural (6)
Panel A : Radius 1						
ACap (1σ)	0.52	0.85	0.05	1.25**	2.00***	0.03
	(0.42)	(0.59)	(0.40)	(0.52)	(0.74)	(0.88)
R squared	0.78	0.58	0.75	0.78	0.59	0.75
Panel B : Radius 10						
ACap (1σ)	0.69	0.98	0.13	1.33***	1.89**	0.18
	(0.47)	(0.63)	(0.49)	(0.48)	(0.74)	(0.89)
R squared	0.78	0.58	0.75	0.78	0.59	0.75
Panel C : Radius 50						
ACap (1σ)	0.74	0.99	0.09	1.27**	1.73**	0.11
	(0.50)	(0.67)	(0.59)	(0.55)	(0.87)	(0.95)
R squared	0.78	0.58	0.75	0.78	0.59	0.75
Panel D : Radius 100						
ACap (1σ)	0.53	0.90	0.09	1.00*	1.69**	0.12
	(0.39)	(0.57)	(0.38)	(0.52)	(0.78)	(0.67)
R squared	0.78	0.58	0.75	0.78	0.59	0.75
mdv	10.6	12.2	9.6	10.6	12.2	9.7
# Obs	5841	5841	5841	5817	5817	5817
# Districts	457	457	457	455	455	455

Notes: All regressions include district and state-yr fixed effects, as well as district-level controls for temperature, precipitation, and number of live births. All regressions are weighted by the number of live births and standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

each of our 180 power plants for radius of 1km, 5km, 10km, 20km, 100km, 200k, and 500km from census data for years 2000, 2005, 2010, and 2015. We then estimate:

$$log(D_{p,t,b}) = \alpha_{p,t} + \sum_{l=1}^{B} \beta^l * 1(b = l) + \epsilon_{p,t,b} \tag{16}$$

where $log(D_{p,t,b})$ is the log of population density in year t around plant p in a buffer of radius

Table 4: IMR Results Urban vs Rural Breakdowns

	Rural				Urban			
	Total	Total	Male	Female	Total	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Cap (1σ)	0.07	0.58	0.62	0.52	2.26***	1.81**	2.11**	1.56*
	(0.85)	(0.89)	(0.92)	(0.87)	(0.77)	(0.90)	(1.03)	(0.82)
R squared	0.70	0.81	0.79	0.80	0.45	0.74	0.74	0.70
Panel B								
ACap (1σ)	0.04	0.25	0.22	0.27	1.94**	1.20	1.29	1.08
	(0.67)	(0.70)	(0.73)	(0.68)	(0.85)	(0.82)	(0.90)	(0.80)
R squared	0.70	0.81	0.79	0.80	0.45	0.74	0.74	0.70
Panel C								
Event time $\in [-4,-2]$	-0.03	0.25	0.20	0.30	0.54	0.97	0.43	1.57**
	(0.41)	(0.39)	(0.42)	(0.41)	(0.60)	(0.62)	(0.73)	(0.69)
Event time $\in [-1,1]$	-0.50	0.24	0.29	0.18	2.50***	2.08***	2.20**	2.01***
	(0.56)	(0.54)	(0.56)	(0.56)	(0.81)	(0.78)	(0.88)	(0.76)
Event time $\in [2,4]$	-0.46	-0.13	-0.04	-0.23	1.74	1.47	1.97	0.97
	(0.49)	(0.54)	(0.56)	(0.58)	(1.10)	(1.21)	(1.52)	(1.01)
Event time $\in [5,\infty]$	0.20	0.09	0.20	-0.04	2.24**	2.06	2.02	2.26
	(0.73)	(0.91)	(0.94)	(0.91)	(1.06)	(1.76)	(2.09)	(1.52)
R squared	0.70	0.81	0.79	0.80	0.45	0.74	0.74	0.70
mdv	9.6	8.2	8.4	8.0	12.5	11.6	12.7	10.4
# Obs	6192	3280	3280	3280	6203	3280	3280	3280
# Districts	476	358	358	358	471	358	358	358

Notes: All regressions include district and state-yr fixed effects, as well as district-level controls for temperature, precipitation, and number of live births. All regressions are weighted by the number of live births and standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

b, $\alpha_{p,t}$ is a plant year fixed effect. The coefficients of interest are the β^l's, which measure the average population density by distance from the plant, controlling for plant-year effects.

Estimates of equation 16 are presented in Figure 4. We omit the dummy variable for buffer of 100km, so all point estimates are relative to a distance of 100km. In Figure 16, we find that population density is nonlinear in distance from the plant. The density within 5km of the plant is no greater than the density between 50 and 100 kms. However, the area

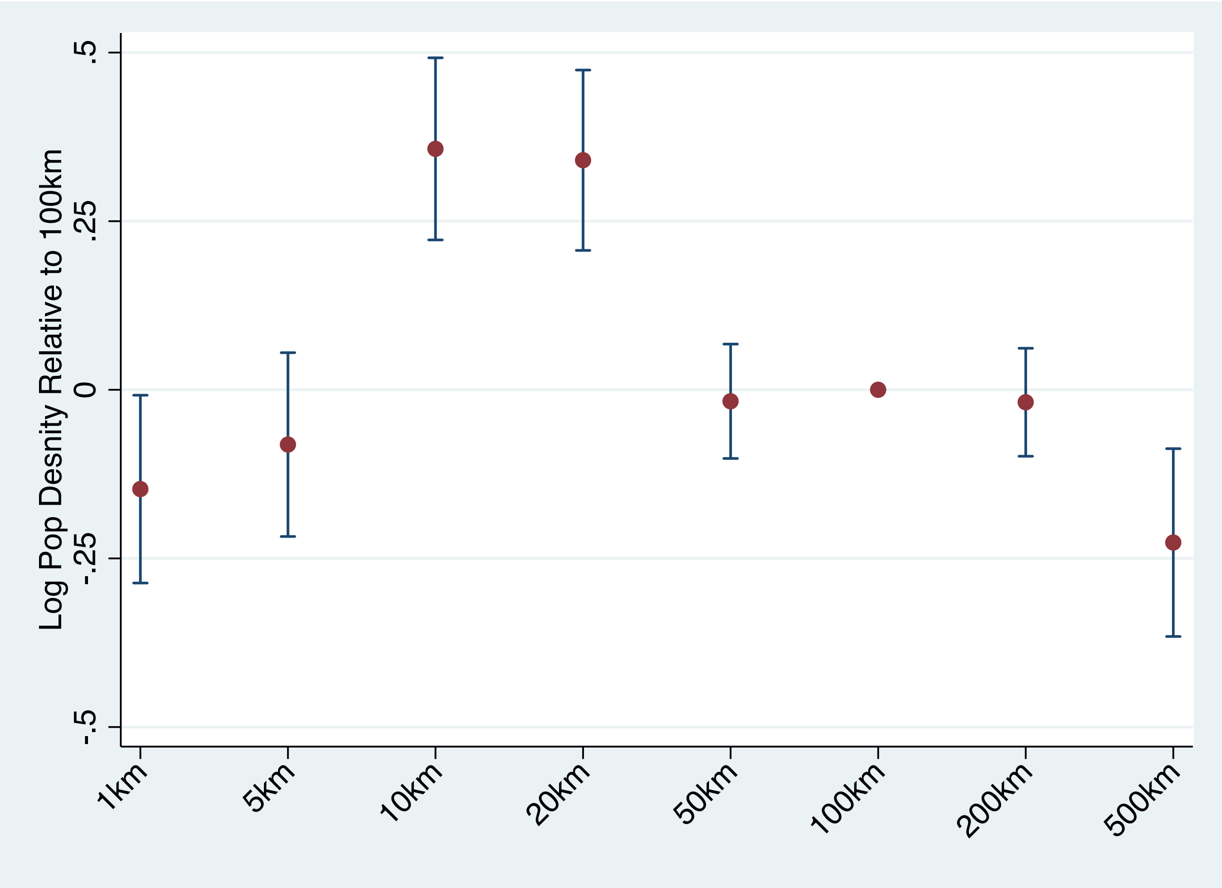
Table 5: IMR Results Urban vs Rural Breakdowns ACAP

	Rural				Urban			
	Total	Total	Male	Female	Total	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A : Radius 1								
ACap (1σ)	0.02	0.58	0.62	0.52	2.22***	1.81**	2.11**	1.56*
	(0.86)	(0.89)	(0.92)	(0.87)	(0.77)	(0.90)	(1.03)	(0.82)
R squared	0.70	0.81	0.79	0.80	0.45	0.74	0.74	0.70
Panel B : Radius 10								
ACap (1σ)	-0.03	0.69	0.76	0.61	2.13***	1.92**	2.22**	1.66*
	(0.92)	(0.91)	(0.94)	(0.90)	(0.78)	(0.98)	(1.09)	(0.89)
R squared	0.70	0.81	0.79	0.80	0.45	0.74	0.74	0.70
Panel C : Radius 50								
ACap (1σ)	-0.02	0.45	0.42	0.47	1.97**	1.49	1.55	1.46
	(0.95)	(1.03)	(1.07)	(0.99)	(0.90)	(1.09)	(1.21)	(1.01)
R squared	0.70	0.81	0.79	0.80	0.45	0.74	0.74	0.70
Panel D : Radius 100								
ACap (1σ)	0.04	0.25	0.22	0.27	1.94**	1.20	1.29	1.08
	(0.67)	(0.70)	(0.73)	(0.68)	(0.85)	(0.82)	(0.90)	(0.80)
R squared	0.70	0.81	0.79	0.80	0.45	0.74	0.74	0.70
mdv	9.6	8.2	8.4	8.0	12.5	11.6	12.7	10.4
# Obs	6192	3280	3280	3280	6203	3280	3280	3280
# Districts	476	358	358	358	471	358	358	358

Notes: All regressions include district and state-yr fixed effects, as well as district-level controls for temperature, precipitation, and number of live births. All regressions are weighted by the number of live births and standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

between 5 and 10km and 10 and 20km are about 35% more dense than the area between 50km and 100km. The results indicate that coal plants tend to be placed between 5-20km from high density areas, i.e. urban areas, which indicates that endogenous placement plays a role in the heterogeneous IMR effects.

Figure 4: Plant Placement



Notes: Figure presents point estimates and 95% confidence intervals from estimates of equation 16.

5 The (Lack of) Local Benefits of Coal Capacity

This section provides statistical evidence arguing that the vast majority of the economic benefits from a coal-fired power plant are dispersed equally to people at the state-level or national-level; there are little additional economic benefits to having a coal-fired power plant located either nearby or within your district relative to being in the same state as the power plant. To make this claim, we focus on the impact of district-level coal-fired capacity on tow sectors of the economy – manufacturing and agriculture, and one summary variable of economic activity – "night-lights" (the average luminosity of different areas at night).

5.1 The Effect of Coal-Fired Capacity on Manufacturing Firms

In this section, we test for differential impacts of increases in generating capacity on manufacturing firm production outcomes. Coal-fired plant construction shifts out the aggregate supply of electricity, which lowers the price for manufacturing firms. If electricity flows across regions without cost, then the price effects will be shared equally across all firms, regardless of where the plants are built. However, if there are nontrivial transportation costs (for example, transmission constraints or line losses) associated with electricity supply, then new supply should disproportionately benefit manufacturing firms nearer to the new plant. To estimate local economic benefits, we merge our two measures of district capacity to annual firm-level outcomes in the Indian Annual Survey of Industries (ASI).

Allcott, Collard-Wexler and O'Connell (2016) show that temporary electricity shortages ("blackouts") are a significant problem for manufacturing firms in India. Allcott, Collard-Wexler and O'Connell (2016) also show that blackouts lead to lower material input expenditures and lower sales. In a theoretical section, Allcott, Collard-Wexler and O'Connell (2016) also show that blackouts can lead to lower labor demand as well as lower productivity. The construction of coal-fired power plants should alleviate pressure on the grid and lower the incidence of blackout, which should trigger all the mechanisms indicated by Allcott, Collard-Wexler and O'Connell (2016). In particular, if local coal-fired capacity construction leads to higher wages or more employment, then workers living near the plants might experience the supply-side health benefits discussed in the conceptual framework. The reduced form impacts in section 4 indicate that however big these supply-side effects are, the direct effect from pollution is larger, since the reduce form impact on health is negative. Still, we can assess the supply-side effects directly by testing for impact on labor.

5.1.1 ASI Data

The ASI is a plant-level survey of formal manufacturing units conducted every year in India, which we aggregate up to the firm level. All registered plants with more than 10 workers are included in the sampling frame. Large firms are surveyed every year, while smaller firms are surveyed randomly only in some years. Sampling is conducted to achieve representation at the state-by-industry level, but researchers have generally taken the survey to be approximately representative at the district level.

The ASI reports annual information on a range of outcomes, including outputs in quantities and values, labor inputs broken down by worker type, material inputs, capital stocks, and detailed energy use. In terms of energy, we observe nominal price reported for electricity purchased from the grid, quantity of electricity purchased from the grid (from which we compute nominal unit values of electricity purchased from the grid), as well as quantity of electricity consumed from on-site production. Nominal outputs and inputs are deflated using industry-specific price deflators, and labor, capital, and electricity prices are deflated by country-level inflation rates.

Descriptive statistics are reported in Table 13. In total, we observe 107,782 firms operating between the years 1999 and 2010, yielding 254,186 firm-year observations.[12] The average firm generates 141 million Rs (in year 2000 Rs) in sales per year, or roughly 3 million USD. The average firms consumes 0.918 GWh of electricity each year in total, with 0.770 GWh coming from the grid. The rest is generated on site.[13] Firms report average prices of 4.672 rs/KWH over the period, with implied price per unit falling slightly below this level. Average man-days worked and effective wage (inclusive of benefits) are 27,000 and 196,000 RS/Day, respectively. Converting to annual levels, these figures imply about 85 full-time workers (working 6-day weeks), each working for an annual salary of 61,000 Rs, or about 1,350 USD.

From these raw data, we compute total revenue factor productivity (TRFP) following Allcott, Collard-Wexler and O'Connell (2016). We assume deflated firm-level revenue is the result of Cobb-Douglas production function in deflated expenditures on materials and labor and deflated value of capital stock. With this assumption, the factor elasticities for variable inputs labor and materials are equal to the input factor share in total revenue.

[12] Our algorithm for merging in district identifiers results in dropping the year 1998.

[13] Summing over all firms, we compute total electricity consumption of about 40 TWh per year. Over the same period, we calculate that average coal-fired capacity equaled 69 GW, or $69 \times \frac{8760}{1000} = 604$ TWh of production. Thus, the firms in the sample only account for around 7% of the total electricity production from coal-fired plants.

Table 6: Descriptive Statistics ASI

	Mean	Std. Dev.	# of Obs.
Panel A : Firm level			
Electricity Consumption			
Total Electricity Consumption (GWH)	0.918	2.565	223832
Electricity Purchased from Grid (GWH)	0.770	2.088	223834
Electricity Generated on site (GWH)	0.067	0.306	229028
Share of Electricity Generated on site (0-1)	0.060	0.141	224036
real electricity price implied (val/qty) rs/KWH	4.672	0.902	221642
real electricity price stated (val/qty) rs/KWH	4.244	1.392	209078
Inputs			
Man-Days Worked (millions)	0.027	0.044	243074
Effective Wage (1000 Rs per day)	0.196	0.134	239074
Capital Share in Total Sales	0.507	0.929	238808
Labor Share in Total Sales	0.122	0.138	244437
Material Share in Total Sales	0.676	0.207	219857
Energy Share in Total Sales	0.058	0.088	222029
Output and Productivity			
Real Gross Sales (billions Rs)	0.141	0.336	243775
Log TRFP – GMM	1.201	0.504	206210
Panel B : Firm-product level			
Sales (Bill Rs)	0.069	0.187	348909
Qty (various units)	0.413	1.654	348909
Unit Value (Bill Rs/unit)	0.026	0.081	348908

Notes: Top and bottom 1% of dependent-variable values have been excluded.

Capital stocks are assumed to adjust with a 1-year lag. Thus, following the procedure from Allcott, Collard-Wexler and O'Connell (2016), we compute production coefficients for each industry-year assuming Cobb-Douglas technology, and then subtract fitted values of labor and material from deflated gross sales. Next, we estimate the capital elasticity by regressing this residual on capital share, instrumenting the capital share with the lag of its value. We then take the difference between log sales and fitted sales as productivity (Log

TRFP – GMM). Descriptive statistics for Log TRFP – GMM are reported in Table 13.

Finally, in the ASI, firms also list quantity and value of sales at the product-line level along with units of quantity. Products are classified according to the ASI commodity classification (ASICC) code. Firms list up to 10 individual product lines along with an "other" category. With these product-level data, we can estimate impacts on output prices by computing unit values, as long as units are constant over time within firm-product line. Descriptive statistics at the firm-product level are reported in panel B of Table 13. The average firm-product line generates 69 million Rs in revenue.

In order to relate outcomes in the ASI to our district-level exposure measures, we need to associate each firm in the ASI to a specific district. The panel version of the ASI allows us to track firms over time, but the panel version omits the district identifiers. District codes are included in the cross-section version of the ASI (without firm-level identifiers). We follow Martin, Nataraj and Harrison (2017) and merge district codes from the cross-sectional data in order to exploit the spatial variation in coal-fired plant rollout.[14]

We then estimate firm-and firm-product versions of (14) and (15)

$$Y_{i,d,t} = \alpha_i + \theta_{s,t} + \beta * Coal_{d,t} + X_{d,t}\gamma + \epsilon_{i,d,t} \tag{17}$$

and

$$Y_{i,d,t} = \alpha_i + \theta_{s,t} + \sum_{j}^{J_d} \sum_{l=-L}^{L} \beta^l * 1(t - e_d^j = l) + X_{d,t}\gamma + \epsilon_{i,d,t} \tag{18}$$

where i indexes either firm or firm-product. At the firm-product level, we also include product category-by-year fixed effects to absorb product-specific time trends.

5.1.2 ASI Results

We begin by estimating impacts of local coal-fired electricity capacity on the electricity consumption of firms in the ASI. Tables 7 and 8 report results for total electricity consumption (GWH), electricity purchased from the grid (GWH), unit value of electricity (rs/KWH), and indicator for generating any electricity on site (0,1), and the quantity of electricity generated on site (GWH). All variables except the own generation indicator (column 4) are logged and the top and bottom 1% of values have been excluded. As before, coal exposure is converted so that a 1-unit increase corresponds to 1 standard deviation.

[14]We develop our own algorithm (departing from Martin, Nataraj and Harrison (2017)) to merge district identifiers from the cross-section files that minimizes information loss.

Table 7: ASI Electricity

	etotal (1)	epurchase (2)	unit value (3)	1(owngen) (4)	eowngen (5)
Panel A : In-District					
Cap (1σ)	0.014	0.036	0.017***	-0.022***	-0.138**
	(0.021)	(0.023)	(0.005)	(0.007)	(0.070)
R squared	0.93	0.93	0.82	0.64	0.79
Panel B: Area-Weighted					
ACap (1σ)	0.039**	0.049**	0.010*	-0.011	-0.131
	(0.017)	(0.021)	(0.006)	(0.009)	(0.102)
R squared	0.93	0.93	0.82	0.64	0.79
Panel C: Event Study					
ET $\in [-4,-2]$	-0.003	0.007	0.004	0.015	-0.021
	(0.020)	(0.027)	(0.006)	(0.014)	(0.045)
ET $\in [-1,1]$	-0.029	0.007	0.026***	-0.006	-0.137**
	(0.026)	(0.026)	(0.007)	(0.015)	(0.061)
ET $\in [2,4]$	-0.001	0.005	0.008	-0.017	-0.029
	(0.025)	(0.030)	(0.006)	(0.013)	(0.054)
ET $\in [5,\infty]$	-0.053	-0.011	0.015**	-0.033*	-0.300*
	(0.052)	(0.056)	(0.007)	(0.019)	(0.175)
R squared	0.93	0.93	0.82	0.64	0.79
mdv	-1.8	-1.9	1.9	0.4	-2.9
# Obs	152693	152564	152501	179564	49709
# Firms	46334	46295	46415	52413	15092

Notes: All regressions include firm and state-yr fixed effects, as well as district-level controls for temperature and precipitation. Top and bottom 1% of dependent-variable values have been excluded. Standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

In panel A of Tables 7, we find that total electricity consumption and electricity purchased from the grid increase with coal-fired capacity, but the point estimates are small and statistically indistinguishable from 0. In Panel B, where exposure is measured by the weighted share of a district covered by a radius of 100km around a coal plant, estimates are larger and statistically significant. A 1σ increase in coal exposure leads to 3.9% and 4.9%

more total electricity and prid-based electricity, respectively. However, we see in Table 8 that these results are sensitive to the specification. Point estimates are always positive, and in the range of 1% - 5%, though only statistically significant in some cases. Additionally, in panel C, we find no evidence of differential trends in electricity consumption prior to increases in coal-fired capacity.

In columns 4-5 of Tables 7 and 8, we estimate impacts to the propensity to produce any electricity on-site, and the quantity of electricity produced on-site. In panel A of Table 7, we find that firms respond to construction of coal plants in their own districts by reducing the propensity to produce generation on-site, and reduce the quantity, conditional on having a generator. On the extensive margin, a 1σ increase in coal capacity reduces the propensity of firms to generate electricity on-site by 2.2 percentage points on a base rate of 41%, or about 5%. On the intensive margin, we find that a 1σ increase in capacity lower the quantity of electricity generated on-site by 13.8%. These results are statisticaly significant at the 1% and 5% levels, respectively. However, in Table 8, we find that these results are also somewhat sensitive to the specification. For exposure measures based on radii up to 50km, we find that capacity increases lowers the propensity to use a generator. But for the exposure measure based on 100km radius, the point estimate is statistically indistinguishable from 0. Together, these results indicate that the construction of new coal-fired capacity differentially impact the electricity consumption patterns of firms nearer vs farther away from the power plants, but the impacts are modest in magnitude.

Next, in Tables 9 and 10, we test for impacts to labor, output, and productivity. Tables 7 and 8 indicate some evidence of local impacts to electricity purchases of firms. As shown in Allcott, Collard-Wexler and O'Connell (2016), these impacts to electricity impacts may pass through to labor, sales, and productivity.

In columns 1-2 of Tables 9 and 10, we find no evidence that firms pay higher wages or hire more labor. Using both measures of exposure, point estimates are small and statistically indistinguishable from 0. For days worked, we can reject at the 5% level any impact of a one standard deviation increase of coal-fired capacity outside the interval (-2.8% , 1.4%); similarly, for effective wages, we can reject any impacts outside the interval (-0.8% , 0.8%).[15] Additionally, in column 4, we cannot reject the null of no impact to firm-level sales. It seems that to whatever extent firms purchase more electricity due to local con-

[15] The full calculation is $(-0.032 - (1.96 \times 0.047)) \times 0.226 = -0.028$ for the lower bound of the 95% confidence interval and $(-0.032 + (1.96 \times 0.047)) \times 0.226 = 0.014$ for the upper bound for days worked; similarly, for effective wages, we calculate: $(-0.001 - (1.96 \times 0.018)) \times 0.226 = -0.008$ for the lower bound of the 95% confidence interval and $(-0.001 + (1.96 \times 0.018)) \times 0.226 = 0.008$ for the upper bound of the 95% confidence interval.

Table 8: ASI Electricity Acap

	etotal	epurchase	unit value	1(owngen)	eowngen
	(1)	(2)	(3)	(4)	(5)
Panel A : Radius 1					
ACap (1σ)	0.014	0.035	0.016***	-0.021***	-0.135*
	(0.021)	(0.023)	(0.005)	(0.007)	(0.071)
R squared	0.93	0.93	0.82	0.64	0.79
Panel B: Radius 10					
ACap (1σ)	0.012	0.028	0.018***	-0.018**	-0.076
	(0.021)	(0.025)	(0.006)	(0.007)	(0.085)
R squared	0.93	0.93	0.82	0.64	0.79
Panel C: Radius 50					
ACap (1σ)	0.031	0.049**	0.015**	-0.025***	-0.104
	(0.022)	(0.025)	(0.007)	(0.009)	(0.093)
R squared	0.93	0.93	0.82	0.64	0.79
Panel D: Radius 100					
ACap (1σ)	0.039**	0.049**	0.010*	-0.011	-0.131
	(0.017)	(0.021)	(0.006)	(0.009)	(0.102)
R squared	0.93	0.93	0.82	0.64	0.79
mdv	-1.8	-1.9	1.9	0.4	-2.9
# Obs	152693	152564	152501	179564	49709
# Firms	46334	46295	46415	52413	15092

Notes: All regressions include firm and state-yr fixed effects, as well as district-level controls for temperature and precipitation. Top and bottom 1% of dependent-variable values have been excluded. Standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

struction of coal-fired power plants, these input expenditures do not translate into more labor expenditures or more output. In terms of the conceptual framework, the null result on wages and employment indicate that there are no local supply-side benefits to health of worker-consumers living near the plants.

Next, in columns 3 and 5, we test for impacts on labor productivity and TFP, respectively. In panel A of Table 9, we find that capacity increases tend to *lower* labor productivity and TFP, though the point estimates are small and indistinguishable from 0. These results are echoed through Table 10, except for a slightly positive impact on TFP using the exposure measure with radius 100. Again, while there is some mixed evidence of benefits to electricity inputs, there do not appear to be any robust impacts to productivity.

Table 9: ASI Economic

	Days Worked	Wages	LP	Sales	TFP
	(1)	(2)	(3)	(4)	(5)
Panel A : In-District					
Cap (1σ)	-0.022	-0.006	-0.003	-0.014	-0.005
	(0.014)	(0.005)	(0.012)	(0.014)	(0.008)
R squared	0.92	0.88	0.84	0.92	0.81
Panel B: Area-Weighted					
ACap (1σ)	-0.006	0.003	0.019	-0.012	0.016*
	(0.013)	(0.006)	(0.016)	(0.015)	(0.010)
R squared	0.92	0.88	0.84	0.92	0.81
Panel C: Event Study					
ET $\in [-4, -2]$	0.016	0.004	-0.003	0.013	-0.004
	(0.014)	(0.012)	(0.012)	(0.027)	(0.007)
ET $\in [-1, 1]$	-0.022	-0.001	-0.010	-0.026	-0.007
	(0.021)	(0.011)	(0.012)	(0.025)	(0.012)
ET $\in [2, 4]$	-0.011	0.006	0.011	-0.029	0.001
	(0.015)	(0.007)	(0.010)	(0.026)	(0.007)
ET $\in [5, \infty]$	-0.058**	-0.025**	-0.020	-0.051	-0.014
	(0.027)	(0.012)	(0.017)	(0.032)	(0.023)
R squared	0.92	0.88	0.84	0.92	0.81
mdv	9.5	5.6	-2.8	3.4	1.2
# Obs	169767	167228	168171	171655	140863
# Firms	50542	49897	49237	50420	43152

Notes: All regressions include firm and state-yr fixed effects, as well as district-level controls for temperature and precipitation. Top and bottom 1% of dependent-variable values have been excluded. Standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Lastly, we test for impacts to firm-product outputs. One way that worker-consumers might benefit from local construction of coal-fired power plants might be through lower output prices. Lower prices would effect welfare directly, but they might also impact health through the budget constraint. Lower prices means more disposable income to

Table 10: ASI Economic Acap

	Days Worked (1)	Wages (2)	LP (3)	Sales (4)	TFP (5)
Panel A : Radius 1					
ACap (1σ)	-0.019	-0.006	-0.002	-0.012	-0.005
	(0.014)	(0.005)	(0.012)	(0.014)	(0.008)
R squared	0.92	0.88	0.84	0.92	0.81
Panel B: Radius 10					
ACap (1σ)	-0.011	-0.009	-0.002	-0.005	-0.003
	(0.015)	(0.006)	(0.013)	(0.016)	(0.007)
R squared	0.92	0.88	0.84	0.92	0.81
Panel C: Radius 50					
ACap (1σ)	-0.009	-0.005	0.010	-0.016	0.014
	(0.015)	(0.006)	(0.016)	(0.016)	(0.010)
R squared	0.92	0.88	0.84	0.92	0.81
Panel D: Radius 100					
ACap (1σ)	-0.006	0.003	0.019	-0.012	0.016*
	(0.013)	(0.006)	(0.016)	(0.015)	(0.010)
R squared	0.92	0.88	0.84	0.92	0.81
mdv	9.5	5.6	-2.8	3.4	1.2
# Obs	169767	167228	168171	171655	140863
# Firms	50542	49897	49237	50420	43152

Notes: All regressions include firm and state-yr fixed effects, as well as district-level controls for temperature and precipitation. Top and bottom 1% of dependent-variable values have been excluded. Standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

spend on medical services, which improve health. Again, to any extent that these benefits materialize, the reduced-form impacts in section 4 indicate that the direct pollution effects dominate. Still, by estimating the price effects directly, we learn how the pollution effect relates to the reduced-form estimates.

In Tables 11 and 12, we find no evidence of impacts to firm-level revenues (column 1) real sales (column 2) or unit values (column 3). The point estimates on unit values are negative, but we in none of our specification can we reject a null of no impact.

Table 11: ASI Product-level

	Sales (1)	Qty (2)	UV (3)
Panel A : In-District			
Cap (1σ)	-0.046	-0.006	-0.025
	(0.045)	(0.056)	(0.049)
R squared	0.93	0.93	0.93
Panel B: Area-Weighted			
ACap (1σ)	-0.033	0.019	-0.049
	(0.052)	(0.062)	(0.047)
R squared	0.93	0.93	0.93
Panel C: Event Study			
ET $\in [-4, -2]$	-0.013	-0.011	0.003
	(0.034)	(0.047)	(0.029)
ET $\in [-1, 1]$	-0.030	-0.016	-0.022
	(0.028)	(0.056)	(0.041)
ET $\in [2, 4]$	-0.004	-0.069*	0.043
	(0.024)	(0.037)	(0.035)
ET $\in [5, \infty]$	-0.049	0.003	-0.115
	(0.036)	(0.079)	(0.075)
R squared	0.93	0.93	0.93
mdv	2.3	8.6	-6.4
# Obs	127327	127333	127642
# Firm-products	46256	46219	46230

Notes: All regressions include firm-product, state-yr, and product-category-year fixed effects, as well as district-level controls for temperature and precipitation. Top and bottom 1% of dependent-variable values have been excluded. Standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 12: ASI Product-level Acap

	Sales (1)	Qty (2)	UV (3)
Panel A : Radius 1			
ACap (1σ)	-0.044	-0.005	-0.025
	(0.044)	(0.054)	(0.047)
R squared	0.93	0.93	0.93
Panel B: Radius 10			
ACap (1σ)	-0.051	0.016	-0.046
	(0.041)	(0.050)	(0.044)
R squared	0.93	0.93	0.93
Panel C: Radius 50			
ACap (1σ)	-0.059	0.021	-0.060
	(0.046)	(0.058)	(0.043)
R squared	0.93	0.93	0.93
Panel D: Radius 100			
ACap (1σ)	-0.033	0.019	-0.049
	(0.052)	(0.062)	(0.047)
R squared	0.93	0.93	0.93
mdv	2.3	8.6	-6.4
# Obs	127327	127333	127642
# Firm-products	46256	46219	46230

Notes: All regressions include firm and state-yr fixed effects, as well as district-level controls for temperature and precipitation. Top and bottom 1% of dependent-variable values have been excluded. Standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

5.2 The Effect of Coal-Fired Capacity on Agriculture and Night-lights

We next test for impacts on agricultural outcomes and night-time luminosity as measured by satellites ("Night-lights"). In the conceptual framework, we showed how local benefits to industry in terms of wages or employment may lead to better health outcomes through increased medical expenditures. We might expect to see better these supply-side benefits in the manufacturing sector, which has a higher input share of electricity. Still, electricity is important for agricultural production as well. In addition, local coal-plant construction could lower the effective price of heating and cooling for households, which may make them more productive. We test for these channels in the reduced form with district-level agricultural yield and wages.

Next, as a test of overall economic benefits, we estimate impacts to "Night-lights". "Night-lights" have been used both as a measure of economic growth (Henderson, Storeygard and Weil, 2012) and a measure of rural electrification (Burlig and Preonas, 2016).

5.2.1 Agriculture and Night-lights Data

Agricultural data come from the Village Dynamics in South Asia Meso data set, which is compiled by researchers at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT 2015). The data set provides district-level information on annual agricultural production, prices, acreage, and yields, by crop. We generate aggregate price-weighted district level measures of total yield in each district for the six major crops (rice, wheat, sugarcane, groundnut, sorghum, and maize) as well as the five major monsoon crops (read: excluding wheat). ICRISAT also provides data on district level averages of rural wages and employment separately for men and women. We report summary statistics in Table 13 panel A.

Night lights data

Table 13: Agricultural and Nightlights Data

	No Coal Plants *567 Districts*				Coal Plants *108 Districts*			
	(1) mean	(2) min	(3) max	(4) obs	(5) mean	(6) min	(7) max	(8) obs
Panel A : Agricultural Data								
Yield	5.1	0.0	6.5	7274	5.1	1.3	6.6	2205
Yieldm	4.9	-1.1	7.0	7270	5.0	-3.2	6.6	2205
Male Wages	46.2	0.0	540.0	4764	42.0	0.0	283.3	1571
Female Wages	32.4	0.0	341.7	4370	28.6	0.0	215.6	1400
Panel B : Night Lights								
Night Lights SA	0.35	-2.24	16.87	9555	1.16	-1.86	54.78	2226
Night Lights WA	0.37	-2.25	16.87	9555	1.09	-1.55	49.71	2226

Notes: Top and bottom 1% of dependent-variable values have been excluded.

5.2.2 Agriculture and Night-lights Results

Tables 14 and 15 estimate impacts on agricultural outcomes and night lights. All specifications include district-level fixed effects and state-year effects as before, as well as controls for temperature and precipitation. Dependent variables are logged, so point estimates are directly interperable as semi elasticities with respect to a 1σ increase in coal capacity. In panel A of Table 14, we find that yields decline mildly with coal fired capacity, though the point estimates are statistically indistinguishable from 0. Looking at Table 14, we are never able to reject a null of no impact on yields.

In columns 3-4, we estimate impacts on male and female wages, respectively. Here, the point estimates indicate that wages increase with coal-fired capacity; though again, we can never reject the null of no impact.

Finally, columns 5 and 6 demonstrate that changes in district-level coal-fired generating capacity do not have a statistically significant impact on night-lights. Moreover, the economic magnitude of these coefficients is small; a 1 GW increase in district-level capacity (akin to adding a very large coal-fired power plant) only results in a roughly 2.5% average increase in night-lights if we consider the in-district measure of coal-fired capacity (see Column 1). Henderson, Storeygard and Weil (2012) uses night-lights as a measure of economic growth, while Burlig and Preonas (2016) uses night-lights as a measure of rural electrification; based on either interpretation of night-lights, our results indicate that a district with a coal-fired power plant receives no *additional* economic benefits from this plant relative to other districts within the state. If we use the conversion factor between night-lights and economic output mentioned in Henderson, Storeygard and Weil (2012), *the upper 95% confidence bound* for our coefficient estimate in Column 1 implies an annual increase of 12.68 million dollars (in 2000 USD) in district-level gross domestic product (GDP) from a one standard deviation increase in district-level coal-fired capacity.[16]

[16] From Henderson, Storeygard and Weil (2012) (pg. 996): "Third, we obtain an estimate of the structural elasticity of growth in night lights with respect to true GDP growth; the point estimate we obtain is just over one.".

Table 14: Agricultural and Nightlight Results

	Yield (1)	YieldM (2)	Wage M (3)	Wage F (4)	NL (SA) (5)	NL (WA) (6)
Panel A						
Cap (1σ)	-0.017	-0.020	0.012	0.040	-0.003	-0.002
	(0.021)	(0.026)	(0.025)	(0.027)	(0.060)	(0.059)
R squared	0.92	0.92	0.91	0.92	0.91	0.92
Panel B						
ACap (1σ)	0.006	0.008	0.005	-0.000	-0.060	-0.065
	(0.031)	(0.035)	(0.031)	(0.024)	(0.055)	(0.056)
R squared	0.92	0.92	0.91	0.92	0.91	0.92
Panel C						
Event time $\in [-4, -2]$	-0.015	-0.007	-0.038**	-0.030	0.047	0.033
	(0.016)	(0.020)	(0.017)	(0.024)	(0.039)	(0.035)
Event time $\in [-1, 1]$	0.003	0.013	-0.023	0.014	0.032	0.039
	(0.016)	(0.024)	(0.019)	(0.020)	(0.078)	(0.082)
Event time $\in [2, 4]$	0.000	-0.003	-0.030*	-0.006	-0.006	0.003
	(0.022)	(0.030)	(0.017)	(0.020)	(0.062)	(0.063)
Event time $\in [5, \infty]$	-0.038	-0.020	-0.005	0.025	0.065	0.056
	(0.029)	(0.028)	(0.024)	(0.042)	(0.077)	(0.079)
R squared	0.92	0.92	0.91	0.92	0.91	0.92
mdv	5.285	5.168	4.279	3.981	1.001	1.050
# Obs	3720	3720	2493	2079	7420	7420
# Districts	280	280	247	224	489	489

Notes: All regressions include district and state-yr fixed effects, as well as district-level controls for temperature and precipitation. District-yer observations are weighted by number of live births. Standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 15: Agricultural Results Acap

	Yield (1)	YieldM (2)	Wage M (3)	Wage F (4)	NL (SA) (5)	NL (WA) (6)
Panel A : Radius 1						
ACap (1σ)	-0.017	-0.021	0.012	0.040	0.003	0.003
	(0.021)	(0.026)	(0.025)	(0.027)	(0.060)	(0.059)
R squared	0.92	0.92	0.91	0.92	0.91	0.92
Panel B : Radius 10						
ACap (1σ)	-0.015	-0.018	0.020	0.043	0.009	0.007
	(0.019)	(0.024)	(0.026)	(0.027)	(0.059)	(0.058)
R squared	0.92	0.92	0.91	0.92	0.91	0.92
Panel C : Radius 50						
ACap (1σ)	-0.011	-0.010	0.026	0.037	-0.051	-0.059
	(0.024)	(0.029)	(0.029)	(0.025)	(0.064)	(0.064)
R squared	0.92	0.92	0.91	0.92	0.91	0.92
Panel D : Radius 100						
ACap (1σ)	0.006	0.008	0.005	-0.000	-0.060	-0.065
	(0.031)	(0.035)	(0.031)	(0.024)	(0.055)	(0.056)
R squared	0.92	0.92	0.91	0.92	0.91	0.92
mdv	5.285	5.168	4.279	3.981	1.001	1.050
# Obs	3720	3720	2493	2079	7420	7420
# Districts	280	280	247	224	489	489

Notes: All regressions include district and state-yr fixed effects, as well as district-level controls for temperature and precipitation. District-yer observations are weighted by number of live births. Standard errors are clustered at the district level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

6 Conclusion

The trajectory of economic development in developing countries is characterized by increased demand for electricity. This demand is often met through relatively inexpensive coal-fired electricity generation. Against the backdrop of India's rapid build-out of coal-fired power plants, we quantify the local infant mortality costs versus the local economic benefits arising from coal-fired capacity expansions over the last two decades using an adaptation of the standard difference-in-differences framework. We find that capacity increases generate sizable local health costs on net but yield only modest local economic benefits. Moreover, the local economic benefits we do find seem to accrue to manufacturing firms rather than workers or residents living nearby these coal-fired power plants. To the extent that these plants generate significant economic benefits, these benefits are likely distributed at the state or regional level.

Our results have several important implications for policymakers in India and other developing countries considering similar expansions of coal-fired power generation. First, our paper provides suggestive evidence that only a very low implied value of statistical life would need to be considered by policymakers in order to justify how much electricity generation in India currently comes from coal-fired generation sources. Of course, a more thorough accounting of both the environmental costs and economic benefits of coal-fired electricity generation is required to definitely assert that the *level* of electricity generation coming from coal-fired sources is either too high or too low.

However, given that most of the benefits are likely regional but costs are local, it is clear that the *placement* of new coal-fired power plants and planned capacity increases should ideally be located away from population centers. Yet, a large number of coal-fired power plants in India are currently located near urban centers. This also raises important questions on the political economy of power generation; who should bear the costs of the electricity generation that everybody benefits from? We leave it to future work to quantify the extent to which economic, environmental, and political factors drive the placement of coal-fired power plants in India in practice.

References

Allcott, Hunt, Allan Collard-Wexler, and Stephen D O'Connell. 2016. "How do electricity shortages affect industry? Evidence from India." *The American Economic Review*, 106(3): 587–624.

Arceo-Gomez, Eva O, Rema Hanna, and Paulina Oliva. n.d.. "Forthcoming."Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City."." *Economic Journal*.

Barron, Manuel, and Maximo Torero. 2017. "Household electrification and indoor air pollution." *Journal of Environmental Economics and Management*, 86: 81–92.

Beach, Brian, and W Walker Hanlon. 2016. "Coal smoke and mortality in an early industrial economy." *The Economic Journal*.

Bombardini, Matilde, and Bingjing Li. 2016. "Trade, pollution and mortality in china." National Bureau of Economic Research.

Burlig, Fiona, and Louis Preonas. 2016. "Out of the Darkness and Into the Light? Development Effects of Rural Electrification." Working Paper,,(October).

Cesur, Resul, Erdal Tekin, and Aydogan Ulker. 2017. "Air pollution and infant mortality: evidence from the expansion of natural gas infrastructure." *The Economic Journal*, 127(600): 330–362.

Clay, Karen, Joshua Lewis, and Edson Severnini. 2015. "Canary in a Coal Mine: Impact of Mid-20th Century Air Pollution Induced by Coal-Fired Power Generation on Infant Mortality and Property Values." Working paper.

Currie, Janet, Joshua Graff Zivin, Jamie Mullins, and Matthew Neidell. 2014. "What do we know about short-and long-term effects of early-life exposure to pollution?" *Annu. Rev. Resour. Econ.*, 6(1): 217–247.

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. "Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings." *The American economic review*, 105(2): 678–709.

Dinkelman, Taryn. 2011. "The effects of rural electrification on employment: New evidence from South Africa." *The American Economic Review*, 101(7): 3078–3108.

Gertler, Paul J., Orie Shelef, Catherine D. Wolfram, and Alan Fuchs. 2016. "The Demand for Energy-Using Assets among the World's Rising Middle Classes." *American Economic Review*, 106(6): 1366–1401.

Greenstone, Michael, and B Kelsey Jack. 2015. “Envirodevonomics: A research agenda for an emerging field.” *Journal of Economic Literature*, 53(1): 5–42.

Greenstone, Michael, and Rema Hanna. 2014. “Environmental regulations, air and water pollution, and infant mortality in India.” *The American Economic Review*, 104(10): 3038–3072.

Gupta, Aashish, and Dean 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.*

Henderson, J Vernon, Adam Storeygard, and David N Weil. 2012. “Measuring economic growth from outer space.” *The American Economic Review*, 102(2): 994–1028.

Jayachandran, Seema. 2009. “Air quality and early-life mortality evidence from Indonesia’s wildfires.” *Journal of Human resources*, 44(4): 916–954.

Lavaine, Emmanuelle, and Matthew Neidell. 2017. “Energy production and health externalities: Evidence from oil refinery strikes in france.” *Journal of the Association of Environmental and Resource Economists*, 4(2): 447–477.

Lipscomb, Molly, Mushfiq A Mobarak, and Tania Barham. 2013. “Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil.” *American Economic Journal: Applied Economics*, 5(2): 200–231.

Luechinger, Simon. 2014. “Air pollution and infant mortality: A natural experiment from power plant desulfurization.” *Journal of health economics*, 37: 219–231.

Martin, Leslie A, Shanthi Nataraj, and Ann E Harrison. 2017. “In with the big, out with the small: Removing small-scale reservations in India.” *The American Economic Review*, 107(2): 354–386.

Mittal, Moti L, Chhemendra Sharma, and Richa Singh. 2012. “Estimates of emissions from coal fired thermal power plants in India.” 13–16.

Moretti, Enrico, and Matthew Neidell. 2011. “Pollution, health, and avoidance behavior evidence from the ports of Los Angeles.” *Journal of human Resources*, 46(1): 154–175.

Munshi, Kaivan, and Mark Rosenzweig. 2009. “Why is mobility in India so low? Social insurance, inequality, and growth.” National Bureau of Economic Research.

Ryan, Nicholas. 2017. "The competitive effects of transmission infrastructure in the indian electricity market." National Bureau of Economic Research.

Shearer, Christine, Robert Fofrich, and Steven J Davis. 2017. "Future CO2 emissions and electricity generation from proposed coal-fired power plants in India." *Earth's Future*, 5(4): 408–416.

Tanaka, Shinsuke. 2015. "Environmental regulations on air pollution in China and their impact on infant mortality." *Journal of Health Economics*, 42: 90–103.

Van Donkelaar, Aaron, Randall V Martin, Michael Brauer, N Christina Hsu, Ralph A Kahn, Robert C Levy, Alexei Lyapustin, Andrew M Sayer, and David M Winker. 2016. "Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors." *Environmental science & technology*, 50(7): 3762–3772.

Viscusi, W Kip, and Joseph E Aldy. 2003. "The value of a statistical life: a critical review of market estimates throughout the world." *Journal of risk and uncertainty*, 27(1): 5–76.

Wolfram, Catherine, Orie Shelef, and Paul Gertler. 2012. "How Will Energy Demand Develop in the Developing World?" *Journal of Economic Perspectives*, 26(1): 119–38.

Zivin, Joshua Graff, and Matthew Neidell. 2013*a*. "Environment, health, and human capital." *Journal of Economic Literature*, 51(3): 689–730.

Zivin, Joshua Graff, and Matthew Neidell. 2013*b*. "Environment, health, and human capital." *Journal of Economic Literature*, 51(3): 689–730.

A Additional Tables and Figures

B Coal-Fired Capacity and Local Air Pollution

In this section, we test for impacts to local air pollution from increase in coal-fired capacity. We present two sets of results. First, we measure average air pollution levels in concentric circles of increasing radius around each power plant site for each year in the pollution data. With these data, we estimate a difference-in-difference model where we compare the change in pollution before and after a plant either comes online or increases capacity for areas near the plant vs areas far from the plant. Next, we take average pollution measures at the district level and relate to our two measures of district exposure.

B.1 Data Sources

Van Donkelaar et al. (2016) constructs annual, globally gridded data at the $0.1° \times 0.1°$ resolution on ambient NO_2 concentration levels for the sample period 1996-2012; we use these data to construct average NO_2 levels around different distance buffers (ex: 1km, 5km, 100km) from each coal-fired power plant site. Van Donkelaar et al. (2016) also provides annual $PM_{2.5}$ data gridded at the $0.01° \times 0.01°$ resolution for 1998-2015, and the Modern-Era Retrospective analysis for Research and Applications (MERRA) database lists monthly $PM_{2.5}$ and SO_2 for the sample period 1980-2016 gridded at the $0.5° \times 0.625°$ resolution; these data are used in supplementary analyses. Finally, we present results differentiating local air pollution effects upwind versus downwind from the plant site; the wind direction for these specifications is calculated using MERRA data.

B.2 Empirical Methodology: Coal-Fired capacity and Local Air Pollution

We estimate the following ordinary least squares (OLS) regression specification linking coal-fired capacity to the (log of) local air pollution levels around each plant site:

$$log(Y_{p,t,b}) = \alpha_{p,b} + \theta_{s,y} + \beta_0 Capacity_{p,t} + \beta_1 Capacity_{p,t} 1(b = \text{Not 500km Buffer}) + \epsilon_{p,t,b} \quad (19)$$

where y indexes year-of-sample. $Capacity_{p,y}$ is the annual electricity generating capacity (in GW) at plant site p, while $1(b = \text{Not 500km Buffer})$ is an indicator that takes on 1 if the observation corresponds to an average of the outcome variable taken using a distance bandwidth less than 500km rather than 500km (we consider $b = (1, 5, 10, 20, 50, 100, 200)km$ versus $b = 500km$). We include plant-buffer fixed effects ($\alpha_{p,b}$) as well as state/year-of-sample fixed effects ($\theta_{s,y}$).

We also consider the following specification where an indicator for whether the plant has opened by time t is the independent variable of interest:

$$log(Y_{p,t,b}) = \alpha_{p,b} + \theta_{s,y} + 1(\text{Plant p has opened})_{p,t}(\beta_0 + \beta_1 1(b = \text{Not 500km Buffer})) + \epsilon_{p,t,b} \quad (20)$$

As before, we include plant-buffer fixed effects ($\alpha_{p,b}$) and state/year-of-sample fixed effects ($\theta_{s,y}$).

Finally, for both of these specifications, we take the average NO_2 levels at different distance bandwidths separately for upwind versus downwind locations; taking the coal capacity regression as an example, we estimate:

$$log(NO_2)_{p,y,b,w} = \alpha_{p,b,w} + \theta_{s,y} + \epsilon_{p,y,b}$$

$$+Capacity_{p,y}(\beta_0 + \beta_1 1(b = \text{Not 500km Buffer}) + \beta_2 1(w = \text{Downwind})$$

$$+\beta_3 Capacity_{p,y} 1(b = \text{Not 500km Buffer}) 1(w = \text{Downwind}))$$

where $w \in \{\text{downwind}, \text{upwind}\}$ indexes an average taken over upwind versus downwind locations.

B.3 Empirical Results: Coal-Fired Capacity Local Air Pollution

Tables 16 and 17 present results at the buffer level. In each table, the top panel presents the specifications where coal-fired capacity is the independent variable of interest, while the bottom panel presents the specifications where an indicator for whether a plant opened on or before the year-of-sample is the independent variable of interest. In both cases, the independent variable of interest is interacted with an indicator that's one if the observation corresponds to an average taken over the bandwidth less than 500km rather than 500km (1km in Column (1), 5km in Column (2), ..., and 200km in Column (7)). All specifications include both plant-buffer fixed effects ("Area"), state-year fixed effects, and plant-buffer controls for temperature and precipitation. Standard errors are clustered at the district level.

In Table 16, we find that both capacity increases (panel A) and plant openings (panel B) increase NO_2 concentrations. In panel A, we find that a 1GW average increase in electricity generating capacity at a plant-site results in a 7.7% (3.2%) average increase in

Table 16: Buffer Results NO2

Dependent Variable:	Log NO2					
	(1) 1km	(2) 5km	(3) 10km	(4) 50km	(5) 100km	(6) 200km
Panel A						
CapacityXclose	0.077***	0.077***	0.078***	0.060***	0.032**	0.006
	(0.021)	(0.021)	(0.022)	(0.016)	(0.014)	(0.013)
R squared	0.974	0.974	0.974	0.980	0.980	0.985
Panel B						
OnlineXclose	0.080***	0.080***	0.079***	0.067***	0.038***	0.013
	(0.019)	(0.019)	(0.019)	(0.012)	(0.009)	(0.011)
R squared	0.974	0.974	0.974	0.980	0.980	0.985
Area FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X
State X Year FE	X	X	X	X	X	X
# Obs	5962	5962	5962	5962	5962	5962
# Plants	180	180	180	180	180	180
# Plant-Years	2981	2981	2981	2981	2981	2981

Notes: All regressions control for annual temperature and precipitation. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

NO_2 levels around 1km (100km) from this plant site relative to 500km away. In panel B, the average plant opening (not including capacity additions at existing plant sites) results in an average increase of 8.0% (3.8%) in NO_2 levels 1km (100km) versus 500km away from the plant site. In both panels, the point estimates are statistically significant up to 100km, and then become statistically indistinguishable from 0.

By contrast, in Table 17 we find no evidence that either capacity increases or plant openings increase $PM_{2.5}$ concentrations. Across all specifications, point estimates are small and statistically indistinguishable from 0.

Next, in Tables 18 and 19, we present estimates at the district level. In both tables, columns 1-4 present results for NO_2, while columns 5-8 present results for $PM_{2.5}$. Columns 1 and 5 present specifications with year fixed effects and no analytical weights. Subsequent columns add state-by-year affects and analytical weights for live births. All specifications include district fixed effects and cluster on the district level. Coal exposure has been transformed so that a 1 unit increase corresponds to an increase in 1 standard deviation

Table 17: Buffer Results PM2.5

Dependent Variable:	Log PM2.5					
	(1) 1km	(2) 5km	(3) 10km	(4) 50km	(5) 100km	(6) 200km
Panel A						
CapacityXclose	0.000	0.001	0.001	0.001	0.002	0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
R squared	0.986	0.986	0.986	0.987	0.988	0.992
Panel B						
OnlineXclose	0.001	0.001	0.002	0.002	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
R squared	0.986	0.986	0.986	0.987	0.988	0.992
Area FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X
State X Year FE	X	X	X	X	X	X
# Obs	6480	6480	6480	6480	6480	6480
# Plants	180	180	180	180	180	180
# Plant-Years	3240	3240	3240	3240	3240	3240

Notes: All regressions control for annual temperature and precipitation. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

for comparability.

In panel A of Table 18, we find that a 1σ increase in coal-fired capacity increases ambient NO_2 levels between 3.2% and 4.2%, depending on the specification. In panel B, where we allow for impact from coal plants in neighboring districts, 1σ increase in exposure generates between 8.3% and 10.3% more NO_2. By contrast, in columns 5-8, we find no evidence that district-level $PM_{2.5}$ is related to coal exposure. Point estimates are in the neighborhood of 1% and never statistically significant. Additionally, in panel C, we cannot reject the null of no impact in the years prior to coal capacity expansions. These results are robust to varying the radius of exposure around the plant, as documented in Table 19.

Table 18: Pollution Results

	Log NO2				Log PM25			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Cap (1σ)	0.042**	0.036***	0.038**	0.032**	0.018	0.003	0.025	0.001
	(0.017)	(0.013)	(0.016)	(0.013)	(0.013)	(0.006)	(0.017)	(0.006)
R squared	0.92	0.97	0.94	0.97	0.97	0.99	0.98	0.99
Panel B								
ACap (1σ)	0.083***	0.103***	0.093***	0.088***	-0.018	-0.002	-0.006	-0.004
	(0.031)	(0.022)	(0.021)	(0.022)	(0.022)	(0.006)	(0.025)	(0.007)
R squared	0.92	0.97	0.95	0.97	0.97	0.99	0.98	0.99
Panel C								
Event time $\in [-4, -2]$	0.011	0.010	0.002	0.003	0.002	0.004	0.002	0.002
	(0.015)	(0.011)	(0.011)	(0.009)	(0.007)	(0.005)	(0.006)	(0.006)
Event time $\in [-1, -1]$	0.018	0.009	0.014	0.007	0.013	-0.000	0.023*	0.000
	(0.014)	(0.011)	(0.010)	(0.009)	(0.010)	(0.005)	(0.012)	(0.005)
Event time $\in [2, 4]$	0.021*	0.009	0.017*	0.006	0.017**	0.007	0.017	0.004
	(0.012)	(0.010)	(0.010)	(0.009)	(0.008)	(0.006)	(0.011)	(0.007)
Event time $\in [5, \infty]$	0.029*	0.032**	0.024	0.026*	0.012	0.014	0.009	0.008
	(0.017)	(0.014)	(0.018)	(0.015)	(0.012)	(0.009)	(0.012)	(0.009)
R squared	0.92	0.97	0.94	0.97	0.97	0.99	0.98	0.99
District FE	X	X	X	X	X	X	X	X
Controls	X	X	X	X	X	X	X	X
Year FE	X		X		X		X	
State X Year FE		X		X		X		X
Analytical Weights			X	X			X	X
# Obs	4551	4527	4551	4527	4106	4085	4106	4085
# Districts	428	425	428	425	422	419	422	419
Mean Dep Var	-1.05	-1.05	-0.95	-0.95	2.93	2.93	3.00	3.00

Notes: All regressions control for annual temperature and precipitation. Analytical weights reflect number of live births. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 19: Pollution Results ACap

	Log NO2				Log PM25			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A : Radius 1								
ACap (1σ)	0.043**	0.036***	0.038**	0.032**	0.019	0.003	0.026	0.001
	(0.017)	(0.013)	(0.016)	(0.013)	(0.013)	(0.006)	(0.017)	(0.006)
R squared	0.92	0.97	0.94	0.97	0.97	0.99	0.98	0.99
Panel B : Radius 10								
ACap (1σ)	0.053***	0.042***	0.046***	0.037***	0.022*	0.003	0.027*	-0.000
	(0.018)	(0.013)	(0.015)	(0.013)	(0.012)	(0.006)	(0.015)	(0.006)
R squared	0.92	0.97	0.94	0.97	0.97	0.99	0.98	0.99
Panel C : Radius 50								
ACap (1σ)	0.110***	0.087***	0.097***	0.071***	0.003	0.000	0.012	-0.002
	(0.036)	(0.023)	(0.031)	(0.019)	(0.026)	(0.007)	(0.025)	(0.007)
R squared	0.92	0.97	0.94	0.97	0.97	0.99	0.98	0.99
Panel D : Radius 100								
ACap (1σ)	0.083***	0.103***	0.093***	0.088***	-0.018	-0.002	-0.006	-0.004
	(0.031)	(0.022)	(0.021)	(0.022)	(0.022)	(0.006)	(0.025)	(0.007)
R squared	0.92	0.97	0.95	0.97	0.97	0.99	0.98	0.99
District FE	X	X	X	X	X	X	X	X
Controls	X	X	X	X	X	X	X	X
Year FE	X		X		X		X	
State X Year FE		X		X		X		X
Analytical Weights			X	X			X	X
# Obs	4551	4527	4551	4527	4106	4085	4106	4085
# Districts	428	425	428	425	422	419	422	419
mdv	-1.05	-1.05	-0.95	-0.95	2.93	2.93	3.00	3.00

Notes: All regressions control for annual temperature and precipitation. Analytical weights reflect number of live births. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

C Geographic Sorting

Using our population data, we test for aggregate sorting towards or away from the coal plants. We estimate

$$log(D_{p,t,b}) = \alpha_{p,b} + \theta_{s,t} + 1(\text{Plant p has opened})_{p,t}(\beta_0 + \beta_1 1(b = \text{Not 500km Buffer})) + \epsilon_{p,t,b} \tag{21}$$

where $log(D_{p,t,b})$ is the log of population density in year t around plant p in a buffer of radius b, $\alpha_{p,b}$ is a plant-buffer fixed effect, and $\theta_{s,t}$ is a state-year fixed effect. The coefficient of interest is β_1, the interaction effect from increases in capacity on log population density closer as opposed to further from the plant. If the increase in coal capacity leads to outward migration, we would expect $\beta_1 < 0$.

Estimates of equation 21 are presented in Figure 5. We find no evidence of sorting in response to the capacity increases.

Figure 5: Migration

Notes: Figure presents point estimates and 95% confidence intervals from estimates of equation 21.