Opening the Brown Box: Production Responses to Environmental Regulation[†]

Rebecca De Simone[‡] S. Lakshmi Naaraayanan[§] Kunal Sachdeva[||]

December 2023

Abstract

We study production responses to emission capping regulation on manufacturing firms. We find that firms lowered their pollution as they transitioned from self-generated to externally sourced electricity, shifted towards producing less coal-intensive products, and increased their abatement expenditures. Firms preserved profitability by increasing their production towards higher-margin products. However, firms in highly polluting industries produced fewer products, and in the aggregate, leading to lower product variety, higher markups, an altered firm-size distribution, and lower business formation. Our findings highlight both the mechanisms behind how mandated pollution reduction can be effective and its costs, suggesting a loss in agglomeration externalities.

Keywords: Abatement, Firm Production, Profitability, Environmental Regulation
JEL Classification: G18, G31, G38, Q50, Q58

[†]We thank João Cocco, Lauren Cohen, Tony Cookson, Julian Franks, Kasper Nielsen, Henri Servaes, Antoinette Schoar, Pablo Slutzky, Daniel Wolfenzon, and conference participants at the London Adam Smith Juniors Conference, Economic and Financial Barriers and Opportunities to the Green Transition in Emerging Economies (Warwick and CAFRAL), and Workshop in Finance and Development for helpful comments. De Simone and Naaraayanan thank the London Business School's Research and Materials Development Grant and Wheeler Institute for Business and Development for supporting this research. All errors are our own.

[‡]London Business School. Email: rdesimone@london.edu.
[§]London Business School. Email: lnaaraayanan@london.edu.
[||]Jesse H. Jones Graduate School of Business, Rice University. Email: kunal.sachdeva@rice.edu.

Policymakers are directing national efforts towards combating climate change, with regulations that target firm emissions (IPCC, 2023; World Economic Forum, 2023). At the same time, they care how emissions reductions are achieved, seeking firms to reduce their energy use by making abatement investments and changing their production processes. Yet these policymakers face capacity constraints to enforce rules and limited information to effectively target emissions reduction (Duflo, Greenstone, Pande, and Ryan, 2013, 2018).

While there is robust evidence on the effectiveness of environmental regulations in curbing emissions, much remains unknown about their impact on production decisions and the economic mechanisms behind firms' trade-offs when balancing emission reduction with economic impacts.[1] Firms often respond by shifting emissions to less-regulated jurisdictions and within their supply chains (Bartram, Hou, and Kim, 2022; Ben-David, Jang, Kleimeier, and Viehs, 2021; Duchin, Gao, and Xu, 2022; Schiller, 2018). At the same time, there is mixed evidence on whether such regulations lower firm output and productivity, and increase consumer costs (Bertrand, Djankov, Hanna, and Mullainathan, 2007; Fowlie, 2010; Greenstone, List, and Syverson, 2012). Alternatively, there is a view that these costs might be temporary and the regulations could eventually boost productivity.[2] Beyond these firm-level outcomes, the primary challenge in empirically assessing the impacts of environmental regulation centers on the opacity of responses within firms, encompassing changes in operational strategies, product inputs and outputs, and energy management.

In this paper, we combine detailed data on firm production and abatement expenditures with variation in environmental regulation emanating from their implementation. We study *within-firm* responses in production decisions, both on the input and output sides, allowing us to uncover the economic forces driving emissions reductions. We begin by showing that the regulation meaningfully decreased emissions using hand-collected regulatory data and satellite emission readings. At the firm level, we exploit detailed product-level data on inputs to show that firms operating in the highest-polluting industries optimize their energy use and shift from in-house to external procurement of electricity. Moreover, we find that the average firm makes

[1] See, for example, Greenstone (2002); Greenstone and Hanna (2014); He, Wang, and Zhang (2020) and cites therein.

[2] Proposed mechanisms for this include R&D spillovers, first-mover advantages (Harrison, Martin, and Nataraj, 2017; Jaffe and Palmer, 1997; Lanjouw and Mody, 1996; Porter and Linde, 1995), and by encouraging firms to optimize energy use and adopt green technologies (Fan, Zivin, Kou, Liu, and Wang, 2019; Newell, Jaffe, and Stavins, 1999; Wu, Yu, Jiaxing, and Zhou, 2023).

substantial investments in pollution abatement, both on extensive and intensive margins. On the output side, we find that firms adjust by reallocating production towards their highest-margin products while maintaining overall profitability. In terms of aggregate dynamics, we observe that firms in high-pollution industries are significantly impacted as they transition to lower-emission products. Meanwhile, firms in less polluting industries experience increased profit margins. Overall, there is a noticeable decline in business dynamism, evidenced by reduced new firm entry and product variety.

Our setting is an emissions capping regulation targeting industrial clusters—co-located dynamic concentrations of related businesses—imposed by the Central Pollution Control Board (CPCB) in India. In 2009, the CPCB introduced the Comprehensive Environmental Pollution Index (CEPI) to quantify pollution levels of industrial clusters and their impact on local populations. They used this index to enforce emission reductions for firms located in these clusters based on whether the cluster CEPI values exceeded pre-defined thresholds. We exploit the resulting discontinuities in enforcement intensity, both within and across industrial clusters, in a difference-in-discontinuities (DiRD) design. Combined with detailed firm- and product-level data, this allows us to identify *within-firm* changes in production decisions on both the input and output sides, and cleanly quantify the costs and benefits of such regulations.

The analyses proceed in four parts. First, we show that the regulation reduced aggregate emissions in industrial clusters. To do so, we hand-collect data from follow-up monitoring studies conducted by the CPCB and find improvements in environmental impact, on average, among industrial clusters with the highest enforcement intensity. We complement these analyses with satellite readings on industrial emissions and document a significant and persistent reduction. To rule out concerns regarding unobserved time-varying trends driving these results, we conduct a placebo test on emissions from energy producers, which are subject to similar seasonality and economic fluctuations but were not targeted by the regulation. Reassuringly, we do not find any change, both in terms of economic magnitude and statistical significance, in these emissions.

Second, we use the granular product-level data to understand the drivers behind the aggregate reduction in emissions. We leverage detailed product-level information on manufacturing firms, unique to India, due to mandatory disclosure requirements (Bau and Matray, 2023; De Loecker, Goldberg, Khandelwal, and Pavcnik, 2016). Specifically, on the input side, we map product-level energy consumption to carbon emissions to estimate product-level emissions (Lyubich, Shapiro, and Walker, 2018). We find that firms respond to the regulation by decreasing the amount of energy inputs per product and the estimated CO_2 emissions per unit produced. Further, we find that, on average, firms shift away from coal use and from producing to purchasing electricity. Notably, we find that firms increase their abatement expenditures, on average, both on extensive and intensive margins. Additionally, we exploit cross-sectional differences in monitoring, emanating from stricter standards imposed by the CPCB and the central government, on high-polluting industries (Harrison, Martin, and Nataraj, 2017). Our analyses suggest that firms in high-polluting industries within treated industrial clusters relative to others, primarily change their input mix while firms in other industries increase their abatement expenditures.

Third, on the output side, we examine changes to product mix and product-level pricing. On average, we find that firms do not change the quantity or the number of products produced. However, there is a significant reduction in product variety driven by a lower probability of adding a new product in a given year. These average treatment effects mask significant heterogeneity within cluster differences in firm response. Specifically, firms in high-polluting industries increase the quantity of products produced while they produce fewer products compared to other firms in the same industrial cluster and product markets. Moreover, they shift their product portfolio *away* from their highest margin and coal-intensive products. In contrast, firms in other industries shift *towards* highest-margin and coal-intensive products. Together, these results suggest that firms in high-polluting industries within industrial clusters, relative to firms in other industries, drive the average reduction in emissions.

Fourth, at the firm-level, on average, these product-level changes result in an increase in the efficiency at converting inputs to revenue (revenue productivity) while preserving their profitability. Additionally, firms do not pass on the increased costs through product price adjustments. However, notably, we observe an increase in product margins and revenue productivity among firms in other industries in addition to a concomitant decrease in raw material expenditures. The reverse is true for firms in high-polluting industries within these clusters. These results suggest a potential impact on the competitiveness of clusters, with the disproportionate cost of the regulation borne by firms in the highest-polluting industries.

A natural question is whether and how these regulations impact industrial clusters in the aggregate. Development strategies in emerging economies often emphasize and rely on industrial clusters to catalyze growth and innovation primarily through agglomeration externalities (Juhász, Lane, Oehlsen, and Pérez, 2022). Given its importance, we examine changes to firm entry and exit into these clusters around the regulation. We document a significant decrease in firm entry across the firm size distribution among both large and small firms. This effect is strongest in industrial clusters with the greatest enforcement intensity. These findings suggest a dampening competitive pressure within the cluster with lower potential agglomeration benefits. Thus, while we find that regulated firms improve their efficiency by lowering their energy intensity in the production process, it is unclear whether this compensates for the loss in business dynamism resulting from lower firm entry.

Lastly, we examine other margins of adjustments through which firms may reduce emissions. One potential margin is that firms could relocate their production outside the industrial cluster or shift their emissions by expanding capacity elsewhere. We present two pieces of evidence ruling out these possibilities. First, we do not find an change in the average probability of a merger and acquisition for firms located in the industrial cluster and as a function of enforcement intensity. Second, we do not find a change in the average probability of announcing a new plant or abandoning the expansion of existing plants. Together, these results suggest a reduction in emissions from production changes instead of firms shifting their emissions elsewhere.

Our results on opening up the 'brown box' of how firms change their inputs and outputs are important in designing effective environmental regulations for industrial clusters. A key insight is that these regulations prompted a fixed-cost shift away from high-emission energy sources alongside investments in efficiency improvements. This suggests that emissions reduction could be more effectively targeted by mandating specific energy input use rather than imposing caps on emissions followed by continuous monitoring. At the same time, it highlights the need for coordinated policies on decarbonization. Even if the shift toward electricity does not necessarily steer away from coal at present, it could pave the way for such a transition when it becomes technically and economically viable to green the power grid, thereby facilitating a smoother transition to cleaner fuels. Moreover, our results point to aggregate costs in terms of lower business dynamism, potentially impairing competitiveness in the global economy.

Related Literature. Our paper contributes to the literature that quantifies the impacts of environmental regulation. We focus on an emissions capping regulation aimed at industrial clusters. Prior research has focused on regulations that often take the form of either command and control or cap-and-trade policies (Bartram, Hou, and Kim, 2022; Fowlie, 2010; Harrison, Hyman, Martin, and Nataraj, 2019; Ivanov, Kruttli, and Watugala, 2023). They are often localized, targeting specific geographic regions, industries, and/or pollutants. A key insight from this body of work is that firms in developed economies substitute their emissions within and across firms to other regions with less stringent regulations, have spillover effects on unregulated firms, or pass it along their supply chains (Ben-David, Jang, Kleimeier, and Viehs, 2021; Dai, Duan, Liang, and Ng, 2021a; Dai, Liang, and Ng, 2021b; Duchin, Gao, and Xu, 2022; Kim and Xu, 2021; Schiller, 2018). In contrast, we focus on regulations targeting firms in industrial clusters, common in both advanced and developing economies, responsible for 15-20 percent of global carbon emissions.

At the same time, several papers have examined the firm-level impacts of environmental regulations (Berman and Bui, 2001; Greenstone, List, and Syverson, 2012; Harrison, Hyman, Martin, and Nataraj, 2019; He, Wang, and Zhang, 2020; Kala and Gechter, 2023). There is mixed evidence on the impact of these regulations on key outcomes such as productivity (Duflo, Greenstone, Pande, and Ryan, 2013; Kala and Gechter, 2023; Kalmenovitz and Chen, 2021) and financial performance (Fan, Zivin, Kou, Liu, and Wang, 2019; Lenox and Eesley, 2009; Naaraayanan, Sachdeva, and Sharma, 2021; Servaes and Tamayo, 2013). Our evidence suggests that firms reorganize their production processes by adjusting inputs and outputs to maintain profitability and increase productivity. However, this conceals significant heterogeneity, whereby less-regulated and smaller firms, on average, make abatement investments and display higher productivity. In contrast, more regulated and larger firms primarily respond by adjusting their energy inputs and product portfolios away from coal. Without comprehensive data on production responses, understanding margins of adjustment for emissions reduction and accurately tracing who bears the disproportionate costs of regulation would be empirically challenging.

Importantly, while our focus is on India due to the availability of granular data and quasi-natural experimental variation, our results have implications for other contexts considering emissions reductions in industrial clusters. For example, the World Economic Forum recently launched a global initiative aiming to reduce heavy industry asset emissions in regional industrial zones

(World Economic Forum, 2023).[3] Notably, these industrial clusters account for a large fraction of global CO_2 emissions, making them an attractive target for emission reductions worldwide. Our study suggests that an alternative national strategy of regulating industrial clusters by capping their emissions may be effective in aligning national efforts towards achieving net-zero targets, but these regulations likely impede economic competitiveness.

1 Institutional Background

This paper focuses on a regulation in India targeting pollution from industrial clusters—dense concentration of firms associated with positive productivity and innovation spillovers. The Central Pollution Control Board (CPCB), the principal national regulator, implemented a regulation in 2009 to curb emissions and their health impacts. They developed a Comprehensive Environmental Pollution Index (CEPI) that identified 88 prominent industrial clusters in consultation with the Ministry of Environment, Forest and Climate Change (MoEF&CC). To arrive at the index values, they conducted a comprehensive environmental analysis and data-gathering effort in these clusters through recognized environmental laboratories. Through the analysis, the CPCB developed an index that takes a value between 0 and 100, to characterize the environmental quality at a given location. The CEPI combined proxies for (i) the amount and toxicity of pollutants, (ii) the potential impact of that pollution on humans and ecosystems, and (iii) an assessment of the quality of actions already taken by cluster firms to capture or adequately dispose of emissions. Figure 1 describes the construction of the CEPI. See Central Pollution Control Board of India (2009) for a complete discussion of the components and construction of the CEPI.

[Figure 1 here.]

The CPCB used these index values to enforce emission reductions for firms located in these clusters based on whether the cluster CEPI values exceeded pre-defined thresholds. In our empirical setting, we exploit the resulting discontinuities in enforcement intensity, both within and across industrial clusters. Specifically, the CPCB classified those clusters with a CEPI at or above 60 and below 70 as "Severely Polluted Areas" (henceforth, $CEPI^{[60,70)}$). These became subject to

[3] Nine leading industrial clusters in China, Indonesia, Japan, Spain, and the United States have joined the World Economic Forum initiative, "Transitioning Industrial Clusters towards Net Zero," to help industries reduce emissions.

central monitoring. Specifically, regulators installed online continuous emission/effluent monitoring systems in these clusters and instituted in-person quarterly audits. Additionally, the CPCB classified industrial clusters with a CEPI of at least 70 as "Critically Polluted Areas" (henceforth, $CEPI^{[70,100]}$). Firms within clusters with CEPI values of at least 70 were subject to the same monitoring treatment as $CEPI^{[60,70)}$ clusters.

As part of the regulation, the CPCB also mandated firms to submit remedial action plans for approval detailing the actions and timelines for emissions reduction.[4] If a firm failed to comply with the directives of the action plan, then it would lose its Environmental Clearance and Consent to Operate permits that allow firms to function within the formal economy.[5] Moreover, Consent to Establish permits could not be issued to new operations if they do not fully comply with the cluster regulations and action plans. Out of the 88 industrial clusters subject to the regulation in 2009, 43 industrial clusters in 17 States had a CEPI value of 70 or above. A further 32 industrial clusters had a CEPI value between 60 and below 70. Online Appendix B provides additional details and examples of CPCB monitoring efforts.

A prominent concern among policymakers and academics relates to the lax enforcement of environmental regulations driven by concerns about limited institutional and governance capacity (Duflo, Greenstone, Pande, and Ryan, 2013, 2018; Greenstone, Pande, Sudarshan, and Ryan, 2022). We present evidence that, in our setting, these have a limited role. First, we rely on CPCB audits that updated the CEPI values twice in the following five years to assess the effectiveness of the action plans.[6] We hand-collected the results of this follow-up monitoring that recalculated the CEPI for the $CEPI^{[70,100]}$ clusters (according to the 2009 ranking) in 2011 and 2013. Our analyses demonstrate that the CEPI regulation effectively reduced emissions at industrial clusters. The top panel of Figure 2 reproduces the original 2009 CEPI distribution. The vertical line represents the cutoff at 70. The bottom panel reports the distributions of the recalculated CEPI values for the 2009 $CEPI^{[70,100]}$ clusters in 2011 and 2013. The distributions shifted to the

[4]The Supreme Courts have been the highest-reputation enforcers of environmental regulation in India since the 1980's, see Greenstone and Hanna (2014); Harrison et al. (2019) for a discussion on these.

[5]All firms, except for those in a small number of nonpolluting sectors, are required to apply for and receive approval from their respective State-level Pollution Control Boards (SPCBs). New activities require a permit called Consent To Establish, while new activities and renewals require one called Consent to Operate (Bhat, 2010; Fenske, Haseeb, and Kala, 2023; Kapur and Khosla, 2019).

[6]Treated clusters were monitored continuously and audited quarterly, but the more extensive full re-analysis of the CEPI value took place at intervals of two years.

left in both follow-up assessments, with significant improvement continuing between 2011 and 2013. This evidence accords with the CPCB's narrative of the regulation as part of an ongoing process of long-term emissions reduction investment. We also observe that while the average cluster improved, a sizable number of clusters continued to have CEPI values above the cutoff in 2013, indicating the difficulties of mandating pollution below a threshold. Second, in Section 4, we provide additional evidence documenting a significant reduction in cluster-level air emissions using satellite readings, and in Section 5, we show that firms alter their production decisions to reduce their product-level CO_2 emissions.

[Figure 2 here.]

As an additional source of variation in pre-existing enforcement intensity, we rely on MoEF&CC designated industries as "highly polluting" (HPI). Firms in these industries were subject to stricter monitoring standards in 2003.[7] For example, monitoring stations were more likely to be placed near HPI plants and HPI firms in $CEPI^{[70,100]}$ clusters and likely subject to firm-specific abatement investment mandates under the Supreme Court action plans (Harrison, Martin, and Nataraj, 2017). We use this additional variation and exploit differences across clusters to shed light on which firms win and lose from the regulation.

2 Data and Summary Statistics

2.1 Data

Policy Data. We hand-collected data on the 2009 regulation from policy documents published by the CPCB. These include cluster-level CEPI values and their corresponding component values broken down by medium (air, water, and land). Additionally, we collected data from two follow-up rounds of monitoring by the CPCB in 2011 and 2013 in clusters with a 2009 CEPI value at or above 70. This data also includes information on the institutional details, monitoring, and enforcement mechanisms of the regulation. See Online Appendix B for more details.

[7]The specific industries include aluminum, copper, iron and steel, and zinc smelting. Also included are the production of chlor-alkali, cement, dyes, fertilizer, pesticides, petrochemicals, pharmaceuticals, sugar, and pulp and paper, as well as tanneries, distilleries, oil refineries, and thermal power plants. IAA1 presents differences in firm and product characteristics, split by HPI and non-HPI industries. In the year before the CEPI regulation, we do not see a statistically significant discontinuity across several characteristics, except for the proportion of coal used as an input.

Industrial Clusters. We hand-collected the location of the near-universe of industrial clusters in 2009 from the CPCB. We compile a dataset of more than 2,000 clusters by name and location. As industrial clusters are a dense agglomeration of SMEs, we map them to the most granular regional unit available across different datasets. Specifically, we standardize the cluster names, extract their geolocation from Google API, and match them to pincodes and cities.[8] We aggregate CEPI to the city level by assigning each city the maximum index value among all the industrial clusters within it. We can match 61 of the 88 industrial clusters for which the CPCB released data on the CEPI. Overall, the firms in our matched sample represent a significant share of economic activity, producing over 70% of output between 2005 and 2015.

Emissions Datasets. To measure cluster-level changes in air emissions, we rely on the Emission Database for Global Atmospheric Research (EDGAR). EDGAR is a comprehensive global database that documents human-caused emissions of greenhouse gases and other pollutants. We use the highest resolution data available: emissions measured in $0.1° \times 0.1°$grids at a monthly frequency. This data presents several advantages for assessing emissions impact. First, EDGAR's figures are derived independently, using consistent international statistics and an established IPCC methodology.[9] Second, the data provide emissions data for various distinct emissions separately. We use emissions data for nitrous oxide (NO_x), particulate matter less than 2.5 μmin diameter ($PM_{2.5}$), and particulate matter less than 10 μm in diameter (PM_{10}). Finally, the data are adjusted to separate emissions from industrial activities from other sources such as fires. This separation allows us to focus exclusively on emissions from industrial sources, including industry. These features of the EDGAR dataset make it well-suited to study changes in emissions, which has been a substantial limitation in prior research on environmental regulations in emerging markets (Greenstone and Jack, 2015).

We link emissions from EDGAR to exact cluster locations and involves two main steps. Ini-

[8]For our primary analyses, the city is the most granular regional unit. However, we match the location of industrial clusters to the more granular PIN code level in the business registration dataset. In India, a PIN code, which stands for Postal Index Number code, is a numerical code used by the postal system to facilitate the sorting and delivery of mail. PIN codes are employed to specify precise locations for mail delivery. On average, a single Post Office serves an area of approximately 21 square kilometers and a population of around 10,000.

[9]See https://edgar.jrc.ec.europa.eu/ and (Crippa, Guizzardi, Muntean, Schaaf, and Oreggioni, 2019; Crippa, Solazzo, Huang, Guizzardi, Koffi, Muntean, Schieberle, Friedrich, and Janssens-Maenhout, 2020; European Commission, Joint Research Centre (EC-JRC)/Netherlands Environmental Assessment Agency (PBL), 2020) for more information.

tially, we define the area around each plant as a circle with a 5-kilometer radius, creating what we refer to as a 'footprint.' This concise footprint ensures that we attribute changes in air emissions to firms operating within each distinct industrial cluster. Then, we allocate the monthly measurements from the grid to this footprint. If a footprint spans multiple grid cells, we calculate the pollution using a weighted average based on the respective land area of those cells. See Online Appendix C for more details.

We supplement these analyses using data on fine particulate matter ($PM_{2.5}$) created by Van Donkelaar, Martin, Spurr, and Burnett (2015). These data are constructed by combining Aerosol Optical Depth (AOD) data from several satellite sources and then calibrating the readings to pollution monitor data using a Geographically Weighted Regression (GWR). The data are available monthly at the spatial resolution of 1km $\times$ 1km.

Firm Financials. We utilize firm and product-level data from Prowess, a database maintained by the Centre for Monitoring the Indian Economy (CMIE). Several prior studies on Indian firms have used this dataset, including Bertrand, Mehta, and Mullainathan (2002), Gopalan, Nanda, and Seru (2007), Lilienfeld-Toal, Mookherjee, and Visaria (2012), Gopalan, Mukherjee, and Singh (2016), and Naaraayanan and Wolfenzon (2023). We extract data from the latest vintage of Prowess, which is free from survivorship bias, as highlighted by Siegel and Choudhury (2012).

The CMIE gathers data from balance sheets and income statements for approximately 37,000 publicly listed and private firms. The covered firms account for more than 70% of the industrial output, 75% of corporate taxes, and over 95% of excise taxes collected by the Government of India and are representative of large and medium-sized firms in India (Bau and Matray, 2023).[10] Moreover, in addition to headline firm financial statements, Prowess data includes firm abatement expenditures, a critical proxy to measure one of the key levers the CEPI regulation used to combat cluster emissions. Therefore, the data is particularly well-suited for examining how firms adjust over time in response to environmental regulations.

[10]It is worth considering how alternative datasets compare to Prowess. Most prominently, prior research on India has used the Annual Survey of Industries (ASI) to examine impact of reforms on the manufacturing sector. Most notably, Prowess is a firm-level panel dataset while ASI is a establishment-level dataset which surveys a repeated cross-section of 30,000 establishments per year (Martin, 2011; Sivadasan, 2009) Moreover, ASI limited in terms of panel coverage, making it particularly ill-suited to study *within-firm* responses to environmental regulations.

Product-level Inputs and Outputs. To shed light on *within* firm decisions, we utilize detailed product-level data made available due to the disclosure requirements set out in the Companies Act of 1956. On the output side, the dataset captures total product sales and total quantity sold at the firm-product level, allowing us to compute unit prices and quantities. In addition, it provides information on capacities, production, and sales from company annual reports (see, Goldberg, Khandelwal, Pavcnik, and Topalova (2010) and Bau and Matray (2023) for more details).[11] There are 1,700 distinct products in our final dataset, where the definition of a product is Prowess's internal product classification, which in turn relies on the National Industrial Classification (NIC). We construct a panel of product-level output and prices, with unit-level prices for each product defined as the total unit sales over total unit quantity.

On the input side, Prowess captures product-wise energy consumption reported in company annual reports.[12] The data are at the firm-product-year-energy source level and are expressed in energy input units per reported production unit. We transform energy input into CO_2 output by making assumptions about each energy source's energy content and CO_2 output. We calculate tonnes of CO_2 emitted per reported production unit for each firm-product-year-energy source and collapse to the firm-product-year level across energy sources (see, Online Appendix D). To our knowledge, no other data source offers comparable granularity and scope to understand the intersection of firm production and energy consumption.

Plant Announcements. We use data on new and abandoned plant announcements from the CapEx database maintained by CMIE. This dataset contains information on all new and abandoned plants announced in India since 1990. Specifically, it provides information on the project announcement date, location, ownership, cost, and industry classification. CMIE obtains the data

[11] A limitation of the dataset is that firms choose which type of units to report so that they are not always standardized across firms or even within firms over time. Thus, we standardize units within and across firms and drop observations where there is insufficient information to reconcile changes in unit types within a firm-product over time.

[12] Sub-section (1), clause (e) of section 217 within the Companies Act of 1956 stipulates that all companies are obligated to report their *total* energy consumption in a specified format. Nevertheless, there is no legal requirement for companies to disclose their product-specific energy consumption per production unit. Consequently, one limitation of the analysis on changes in product-level energy consumption is that firms can decide whether to disclose this information. Not all firms choose to do so. In Appendix D we explore the representativeness of this data. Importantly, we observe in our data that when a firm initiates the reporting of product-level energy inputs, it typically does so consistently throughout the entire period. Moreover, while we find that there is a reduction in the probability of filing energy inputs at all in the post-regulation period but no discontinuity in this probability at the treatment thresholds.

from multiple sources, including annual reports, news articles, and government press releases. The database is updated daily and contains information on the entire project life cycle whenever information is available. Typically, projects costing more than INR 100 million (approximately USD 2 Million) are included in the database (Alok and Ayyagari, 2020; Naaraayanan and Wolfenzon, 2023).

Other Data Sources. We use the near-universe of firm registrations from the Ministry of Corporate Affairs (MCA), allowing us to track business formation across all formal firms in the economy. Further, we use data from the 2001 Population Census to examine whether observables differ significantly in treated and control clusters around the CEPI score treatment thresholds and to test the assumption that the CEPI thresholds are economically meaningful to firms because of the 2009 regulation and not because they correspond to other policy or economically-relevant thresholds. Finally, we convert the data into real values using the capital deflator series from the Ministry of Statistics and Programme Implementation (MOSPI).

2.2 Final Sample and Summary Statistics

Our primary focus is understanding the impact of regulation targeting pollution from industrial clusters, which feature dense agglomeration of manufacturing firms. Our estimation sample, therefore, comprises manufacturing firms located in the 66 clusters for which we have a CEPI score. We focus on a three-year window around the 2009 regulation. Moreover, as we aim to shed light on *within* firm response to the regulation, we focus on mutliproduct firms, allowing us to better understand different margins of adjustment. These firms represent over 95% of the output during the sample period.[13]

Table 1 presents the descriptive statistics for the sample of manufacturing firms in 2008. Panel A reports summary statistics at the firm-year level from 2007 to 2012. The data represents 5,688 unique manufacturing firms, of which 961 list securities on a stock exchange. The average (median) firm has 3.5 (0.6) million INR in total assets and 3.4 (0.8) million INR in total sales. The sample firms are, on average, moderately indebted, with average (median) leverage ratios

[13] This focus on multiproduct firms is consistent with prior studies (De Loecker, 2011; De Loecker, Goldberg, Khandelwal, and Pavcnik, 2016; Eckel and Neary, 2010).

(bank borrowing scaled by total assets) of 0.28 (0.25). Approximately 27% (25%) of total assets for the average (median) firm are fixed assets (property, plant, and equipment). Focusing on the listed firms in our sample, the average (median) market capitalization is 19 (2.3) million INR. The average (median) company exports, deriving 17% (2.5%) of its total sales from exports of goods. The average (median) firm produces 3 (2) distinct products a year.[14] The average (median) firm is also moderately profitable, reporting 11% (10%) of the value of year-before sales in new earnings before interest, taxes, depreciation, and amortization.

[Table 1 here.]

Panel B of Table 1 describes the product-year level panel dataset in 2008. Our dataset has 18,605 unique products, and a single firm can produce multiple products. The overall picture is of significant heterogeneity in operations. Product profit margins—defined as *(unit price - unit cost)/unit price*—tell us that the average (median) firm-product commands a 0.10 (0.16) INR per unit produced. This granular evidence is consistent with the firm-level profitability distribution. Finally, the distribution of product sales, cost, and price are all highly skewed, as is the distribution of unit-level CO_2 emissions, calculated for those firms that report product-level energy inputs. Overall, the panel of manufacturing firms involves a broad cross-section of firms and is consistent with industrial clusters comprised of a few large firms and many medium-sized ones.

3 Empirical Methodology

3.1 Difference-in-Discontinuities (DiRD)

Our analysis exploits the cross-sectional variation in environmental regulatory costs emanating from implementing the regulation. As described in Section 1, firms in clusters just to the left of the CEPI value of 60 faced no sanctions, while those in clusters with CEPI just to the right of 60 were subject to heightened emissions monitoring. Finally, those firms located in clusters just to the right of the second CEPI treatment threshold at 70 were mandated to take more targeted steps by Supreme Court administered action plans.

[14]Note that we drop firms that produce one product throughout the sample but not firms that switch between being single- and multi-product producers so that we preserve variation from the extensive margin.

Therefore, at each cutoff, there is a discontinuous jump in treatment intensity. In the empirical specification, for identification, we exploit the resultant discontinuities and the time variation around the regulation implementation with a difference-in-discontinuities (DiRD) design. The DiRD design allows us to difference out the effect of any potential pre-existing discontinuity at the treatment cutoffs and by focusing on the variation at the threshold, it further allows us to circumvent concerns often associated with the difference-in-differences approach, where control firms might not serve as an appropriate counterfactual for treated firms.

Specifically, we estimate:

$$\begin{aligned} Y_{kijcst} = \beta_1 Post_t \times CEPI_c^{[70,100]} + \beta_2 Post_t \times CEPI_c^{[60,70)} + \\ + \beta_3 CEPI_c + \beta_4 Post_t + \gamma_i + \kappa_{jst} + \epsilon_{kijcst} \end{aligned} \tag{1}$$

where k, i, j, c, s, and t represent a product, firm, industry, city, state, and year, respectively. Our running variable, $CEPI_c$, is the pollution index value, defined at the city level as described in Section 2. We assign firms to clusters based on their headquarters city as of the 2009 regulation. Specifically, $CEPI^{[70,100]}$ takes the value of one if the firm's headquarters city has a maximum industrial cluster CEPI value at or above 70 and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the firm's headquarters city has a maximum industrial cluster CEPI value greater than or equal to 60 and below 70, and zero otherwise. The omitted category includes firms whose headquarters city has a maximum industrial cluster CEPI value below 60. Including both $CEPI_c^{[70,100]}$ and $CEPI_c^{[60,70)}$ group indicators captures the greater intensity of treatment for firms located in cities where the maximum CEPI score is greater than or equal to 70.[15] In our dataset, there are 33 cities with a maximum CEPI value greater than or equal to 70 in 2009 and an additional 20 cities with a maximum 2009 CEPI value greater than or equal to 60 and below 70.

The variable $Post_t$ is an indicator variable taking the value of 1 for all years after and including 2009, the year in which the CEPI regulation was implemented. Finally, the granularity of

[15] Our treatment assignment procedure is the most granular possible. It may result in the misclassification of some firms by labeling them as 'treated' when they are actually 'control,' and vice versa. However, such misclassification likely biases our estimates *against* finding any effect by narrowing the estimated difference between outcomes in the two treatment groups. Moreover, we provide evidence below that the identification assumptions are satisfied, which likely mitigates other sources of bias.

the data allows us to address concerns about location-specific and industry-specific effects that may differentially affect firms' production and emission decisions using the empirical specification. Specifically, we include firm fixed effects (γ_i) to control for unobserved time-invariant firm characteristics.[16] We include state-by-industry-year fixed effects (κ_{jst}) to control for time-varying industry shocks within the same state. The stringency of these fixed effects allow us to rule out several location- and industry-specific concerns such as technical innovation, regulation which vary considerably over state and industries. We cluster standard errors at the city level, which is the level at which we define treatment (Abadie, Athey, Imbens, and Wooldridge, 2017; Bertrand, Duflo, and Mullainathan, 2004; Roberts and Whited, 2013).

The coefficient β_1 quantifies the effect of being located in a cluster with a CEPI value of at least 70 on the outcome Y_{ijcst} relative to the effect of being located in a cluster to the left of the cutoff with a CEPI value below 60, in addition to any effect of being assigned to treatment ($Post \times CEPI^{[60,70)}$). The coefficient β_2 quantifies the regulation's effect on firms located in clusters with CEPI values between 60 and below 70 relative to firms located in clusters with values below 60. Thus, the total treatment effect for a firm with a CEPI value at or above 70 is $\hat{\beta}_1 + \hat{\beta}_2$. Note that a significant advantage of decomposing the treatment effect into two groups – firms with CEPI values of at least 70 and those with values between 60 and below 70 – is that it allows us to shed additional insight on whether the treatment effects predominantly relate to the extensive margin of treatment (crossing the 60 CEPI threshold) or the intensity of treatment, which incurs additional regulation and consequences (crossing the 70 CEPI threshold). Moreover, leveraging quasi-random variation around these two thresholds expands the scope of the estimated local average treatment effects and enhances the generalizability of our findings.

3.2 Identification Assumptions

The primary identification assumptions of DiRD are parallel trends and that potential outcomes are smooth around the cutoffs. While parallel trends assumption is fundamentally untestable, we present several pieces of evidence in support of it. Moreover, the latter assumption requires that firms do not perfectly manipulate the policy thresholds used to assign treatment and control

[16]Note that the main effect of $CEPI_c$, which is invariant within a firm, drops out with the inclusion of the firm fixed effects while the main effect of $Post_t$ drops out with the inclusion of state-by-industry-year fixed effects.

groups and that the thresholds are not economically important for reasons other than treatment assignment (Grembi, Nannicini, and Troiano, 2016).

Across several tests, we provide evidence supporting these assumptions. Specifically, we graphically show similar trends for treatment and control group firms across key outcomes in the pre-regulation period. Further, we present evidence of no manipulation of the CEPI value by industrial clusters, no discontinuity in firm-, product-, and cluster-level characteristics at the CEPI thresholds before the regulation.

We begin by testing the assumption that industrial clusters could not influence their position around the thresholds at 60 and 70 by manipulating their CEPI values, our running variable. This is a crucial identification assumption because it underpins our ability to assume that firms whose industrial clusters have pollution rankings on opposite sides of the cutoffs are otherwise comparable. We test for ranking manipulation around the cutoffs at the industrial cluster level in 2009.

As a summary test, we combine the CEPI thresholds of 60 and 70 into one variable through a normalized measure of CEPI value, which we create by subtracting the closest threshold from each cluster's CEPI value.[17] Specifically, we fit the distribution of the ranking variable on either side of the pooled cutoffs and then test if those distributions differ statistically (McCrary, 2008). Figure 3 reports the results for the pooled sample.

[Figure 3 here.]

We do not find evidence of bunching around the cutoffs, and the *p*-value from a two-sided test is 0.33. Thus, we fail to reject the null hypothesis of no manipulation of the CEPI ranking. A remaining concern is that omitted variables affect the composition of industrial clusters. We assuage these concerns in Section 7.3, where we document no evidence of changes in mergers and acquisitions activity leading firms to exit the industrial clusters in response to the stringent environmental regulations.

Second, in Figure 4, we present the geographical variation in the industrial clusters selected by CPCB for the environmental assessment relative to the location of all industrial clusters as of

[17] We have limited observations in the cluster to the left of the CEPI threshold at 60. As a result, we do not have enough power to examine the two cutoffs separately. Hence, to be consistent, we present results with normalized thresholds throughout these tests.

2009 (gray dots). We find that the industrial clusters targeted by the CPCB are representative of clusters in general, with above- and below-cutoff clusters coming from geographically proximate regions within each state.

[Figure 4 here.]

Next, we test the identifying assumption that there are no discontinuous jumps in our key outcomes and firm- and product-level characteristics at these thresholds in the pre-regulation year.[18] The intuition is that if we observe discontinuities around the thresholds even before the CPCB evaluation, then it is likely that some other policy differentially affected firms at these cutoffs, making it hard to isolate the effect of the environmental regulation from these other policies.

[Figure 5 here.]

Figure 5 presents the scatter plots of means of several firm-level covariates, defined as of 2008, by different bins (each of size 1) around the pooled threshold. We normalize the thresholds to zero by subtracting off their respective threshold values from the CEPI value of the industrial cluster. We find no evidence of discontinuities in baseline covariates. The characteristics we examine include total assets, total sales, bank borrowings, net fixed assets as a fraction of total assets, exporting intensity, and market capitalization (for listed firms). Panel A of Table 2 demonstrates that none of these characteristics are statistically different across the cutoffs even in a regression setting.

[Table 2 here.]

Similarly, Figure 6 tests for discontinuities at the pooled thresholds for product-level covariates defined as of 2008. Again, we see no significant discontinuous jumps in the product characteristics. Panel B of Table 2 reports the associated average differences of products of firms in industrial clusters with CEPI values below versus above the threshold and the coefficient and associated p-value of the regression discontinuity specification at the threshold between $CEPI^{[60,70)}$ and $CEPI^{[70,100]}$ cluster status.

[18] This is analogous to the regression discontinuity assumption that potential outcomes are smooth around the cutoffs.

[Figure 6 here.]

Lastly, in Online Appendix Table IAA2, we examine if there are discontinuous jumps in cluster-level covariates defined as of 2008 and taken from the population Census and Harari (2020). As before, we see no evidence of significant ex-ante discontinuities in various proxies for economic activity—both demand and supply for goods—and determinants of pollution extent and impact.

The preponderance of evidence thus suggests that the treatment thresholds do not proxy for pre-existing differences in policies that affected firms in the same way as the environmental regulation. Furthermore, they alleviate concerns that firms in the control group are not a valid counterfactual for treated firms, thereby allowing us to cleanly identify the effect of environmental regulations.

4 Changes in Cluster-level Air Emissions

At the cluster level, our primary outcome of interest is emissions from industrial activities within the clusters. As discussed in Section 2, we extract emissions data, measured in milligrams per month within a spatial resolution of 0.1°x 0.1°, from the EDGAR dataset. We build a monthly panel of emissions from industrial activity split by the type of pollutant at the cluster address-level and estimate the following event study difference-in-differences (DiD) specification:

$$\begin{aligned} Emissions_{pcst} = \sum_{k \in \{-4,-2\}} \beta_k D_k \times CEPI_c^{[60,100]} + \sum_{k \in \{0,5\}} \beta_k D_k \times CEPI_c^{[60,100]} + \\ CEPI_c + \gamma_{cp} + \gamma_{pst} + \epsilon_{pcst} \end{aligned} \tag{2}$$

where p is pollutant, c is cluster, s is state, and t is year-month. Figure 7 plots the estimated coefficients (β_k) normalized to the fiscal year of 2008 and their corresponding 95% confidence intervals, comparing the evolution of emissions in treated clusters relative to others. The vertical gray dotted line indicates the regulation year of 2009. Standard errors are clustered the address level.[19]

[19]We link emissions to specific cluster locations using their exact address including PIN codes, as extracted from

[Place Figure 7 here.]

As evident from Panel (a) of Figure 7, there are no differential pre-trends in the average emissions emanating from industrial activities between treated and control clusters, suggesting that targeting by regulators did not, on average, coincide with the differential improvement in air quality.[20] These parallel pre-trends support the DiRD identification assumption that outcomes of the treated and control groups would have evolved similarly in the absence of treatment, locally around the thresholds. Moreover, in the post-regulation period, there is an immediate and persistent decrease in the average emissions levels within treated clusters. In Panel (b) of Figure 7, we find a find a significant decrease in $PM_{2.5}$ emissions. Note that the relatively large decrease suggests that post-regulation emission levels in treated clusters, which had higher levels to begin with in the pre-regulation period, become similar to emission levels in clusters with a CEPI value below 60.

[Place Table 3 here.]

Table 3 presents these results in a regression framework. Column 1 combines all pollutants into a single regression and accounts for their differential impact by including high-dimensional fixed effects. These fixed effects interact cluster and state $\times$ year-month fixed effects with the specific pollutant type. Consistent with Figure 7, we find a statistically significant decrease in relative emission levels for treated clusters compared to industrial clusters with a CEPI value below 60. In terms of economic magnitude, relative to the pre-regulation mean in the control group, this change represents a 25.1% decrease in the $CEPI^{[70,100]}$ clusters and a 26.6% decrease in the $CEPI^{[60,70)}$ clusters. The average effect across the two clusters is statistically significant, with a p-value of 0.018. Separating the results by different pollutants, we find a similar decrease among other hazardous air pollutants.

Lastly, in Online Appendix Table IAA3, we supplement these analyses using satellite data on measurements of fine particulate matter ($PM_{2.5}$) at a granular level (Van Donkelaar, Martin,

the CPCB. In India, a PIN code, which stands for Postal Index Number code, is a numerical code used by the postal system to facilitate the sorting and delivery of mail. PIN codes are employed to specify precise locations for mail delivery. On average, a single Post Office serves an area of approximately 21 square kilometers and a population of around 10,000.

[20]For this analysis, we pooled the emissions data across several pollutants: NO_x (nitrogen oxides), $PM_{2.5}$, and PM_{10}.

Spurr, and Burnett, 2015). These data are constructed by combining Aerosol Optical Depth (AOD) data from several satellite sources and then calibrating to pollution monitor data using a Geographically Weighted Regression (GWR). The data are available monthly at the 1km $\times$ 1km resolution. As before, we build a monthly panel and find a statistically significant decrease in relative emission levels for treated industrial clusters in a five-kilometer and 500-meter radius around the center of each cluster. In terms of economic magnitude, relative to the pre-regulation mean in the control group, this change represents a 3.2% decrease.[21]

Taken together, our results suggest that firms in industrial clusters evaluated by the CPCB in 2009 lowered their emissions, with a larger increase in emissions reduction observed among firms in industrial clusters with higher treatment intensity. This finding is consistent with the CPCB's assertion that clusters with a CEPI value at or above 70 responded to the Supreme Court Action Plans, which mandated emissions abatement investments.

5 Product Outcomes

Treated clusters significantly lowered their emissions in response to the stringent regulation imposed by the CPCB. We now analyze our granular product-level data to understand the drivers behind the aggregate reduction in emissions and document the operational response of manufacturers. We focus on the input side in Subsection 5.1. We analyze product-level energy consumption and estimated carbon emissions. In Subsection 5.2, we analyze the output side, including the extensive- and intensive margins of the production response, product portfolio mix, and pricing and profitability impacts of the regulation.

5.1 Energy Inputs

In this subsection, we focus on the sample of manufacturing firms that reported energy inputs at the product level.[22] We confirm that their estimated CO_2 emissions at the product level also

[21] The differences in economic magnitude relative to the results presented in Table 3 are likely an artifact of the low correlation between $PM_{2.5}$ readings across the two datasets (ρ = -0.06). These differences could arise from variations in calibration methods, in addition to differences in spatial resolution, and the level of industrial activity captured. We believe that, given the low correlation, both data sources provide orthogonal measurements that help establish a robust reduction in emission levels in response to the CPCB regulation.

[22] As we discuss in more detail in Section 2.1, the advantage of this data is that it offers a unique window into the energy margins of adjustment that firms take in response to emissions regulation. The downside is that firms choose

decrease. We find that firms lower emissions by changing their energy inputs, shifting output towards lower-emission products, and purchasing rather than producing electricity. We exploit cross-sectional differences in CPCB monitoring of firms in high-polluting industries versus others within clusters to show that results are stronger for firms in high-polluting industries within treated industrial clusters relative to others.

First, Panel A of Table 4 reports the impact of the 2009 CEPI regulation on firm-level energy inputs. In Model (1), the outcome is the INR value of product-level energy inputs. We see that the average treated firm dramatically lowers the amount spent on energy, controlling for the quantity of the product they produced in the same year. The treated firm in a city with a 2009 CEPI score between 60 and below 70 by approximately 63% on an annual basis, and the average firm in a city with a 2009 CEPI score of at least 70 reduces energy inputs by 84%, significant at the 1% confidence level.[23] Relative to the 2008 average energy expenditure of 8.907 million INR per product, the magnitude is a reduction of about 7.5 million INR on the average product of firms located in clusters with CEPI of at least 70, or a reduction of about 146,859 in 2008 U.S. dollars.

[Place Table 4 here.]

Coal is the primary energy input in our data and a key source of industrial cluster air emissions. In Model (2) of Panel A of Table 4, we test for the use of coal by firms in treated relative to control clusters. We find a considerable decrease of 29% in the use of coal in firms in $CEPI^{[60,70)}$ clusters and a relatively 30% larger average effect in $CEPI^{[70,100]}$. For firms in clusters with a 2009 CEPI of at least 70, this represents a drop from 17% of inputs to 10% of inputs for the average product, relative to the 2008 control use of coal as an input per product.

Treated firms, on average, reduced energy use per product relative to control firms. One mechanism of this reduction was reducing dependence on coal as an energy input. Model (3) of Panel A of Table 4 indicates that another is shifting emissions outside the cluster. We see a striking shift from producing electricity to purchasing it. Specifically, the average treated company increases the proportion of energy that they purchase from the electrical grid by 22 (for firms in $CEPI^{[60,70)}$) to 34% (for firms in $CEPI^{[70,100]}$).

whether to report product energy inputs. See Appendix D for further discussion and analyses.

[23]Calculated as $\exp -1.824 - 1 = -0.839$.

In the data, the majority of firms use electricity. In 2008, the average control firm purchased 46% of the electricity used to make the typical product and produced the remaining 64% by burning coal, diesel, oil and gas, or biomass, in that order of frequency. The treatment magnitude for the highest-treatment group ($CEPI^{[70,100]}$ cluster firms) is a shift from purchasing 46% of electricity used to produce the average product to purchasing 62%.

We observe reduced emissions in all clusters, but it is unclear whether all firms are reducing emissions. A notable feature of the CEPI regulation is that it targets clusters rather than individual firms, opening up the possibility that firms in the same cluster receive different treatments. To get at this margin, we exploit an *intra-cluster* source of treatment variation from the institutional detail that treatment affected firms in "highly-polluting industries" more intensely than firms in other industries located in the same cluster. To do so, we estimate the following triple difference-in-discontinuities specification:

$$\begin{aligned} Y_{kijcst} = \beta_1 Post_t \times CEPI_c^{[70,100]} + \beta_2 Post_t \times CEPI_c^{[60,70)} + \\ \beta_3 Post_t \times CEPI_c^{[70,100]} \times \text{High-Polluting}_j + \\ \beta_4 Post_t \times CEPI_c^{[60,70)} \times \text{High-Polluting}_j + \\ \beta_5 CEPI_c + \beta_6 Post_t + \beta_7 \text{High-Polluting}_j + \gamma_i + \kappa_{jst} + \epsilon_{kijcst} \end{aligned} \tag{3}$$

In other words, we interact treatment indicators $CEPI_c^{[70,100]}$ and $CEPI_c^{[60,70)}$ in Equation 3 with High-Polluting$_j$. High-Polluting$_j$ is one if the firm's main industry in 2008 was one of the seventeen industries considered highly-polluting industries by the CPCB, and zero otherwise.[24]

Panel B of Table 4 reports the results of estimating Equation 3 on the input data. Though the tests are underpowered, the relative magnitudes of the coefficients suggest that treated firms in high-pollution industries drive the shift towards lower energy input and from purchasing to producing electricity that we document in Panel A. Moreover, Online Appendix Table IAA1 shows that there is no differential discontinuity in either firm or product characteristics across HPI vs non-HPI industries.

Next, Table 5 tests if product-level CO_2 emissions decrease with the energy input changes. We transform energy input into CO_2 output by making assumptions about each energy source's

[24]See Section 1 for further details about this classification and how it interacts with treatment intensity.

energy content and CO_2 output.[25] We then calculate tonnes of CO_2 emitted per reported production unit for each firm-product-year-energy source and collapse to the firm-product-year level across energy sources. The outcome for Model (1) is the natural logarithm of the total CO_2 emissions of each product-year. The outcome for Model (2) is the natural logarithm of the per-unit CO_2 emissions, calculated as the ratio of total annual CO_2 emissions and total production units for the product year. In both models, we control for concurrent changes in production quantity. Emissions fall sharply relative to treated firms, and more for the treated in clusters with CEPI at least 70 than for the average treated firm. These findings are consistent with the cluster-level emissions evidence in Section 4.

[Place Table 5 here.]

Specifically, Model (1) in Panel A of Table 5 reports that product CO_2 emissions decrease by approximately 87% for the average firm in the highest-treatment group from clusters with a CEPI of at least 70. The decrease is approximately 66% for treated in clusters with 2009 CEPI between 60 and below 70. Interpreted relative to the 2008 control mean of 162,230 tonnes of CO_2 per product-year, this is a reduction of between 107 and 141 thousand tonnes of CO_2. Model (2) of Table 5 tells a similar story, but this time considering tons of CO_2 per reported production unit (e.g., tonnes of CO_2 per tonnes of widgets produced). Here we see a 69 to 79 percentage point reduction in emissions per unit for firms in $CEPI_c^{[60,70)}$ and $CEPI_c^{[70,100]}$ clusters, respectively. Finally, Model (3) of Table 5 reports that the average firm shifted its product portfolio away from its highest-coal input product in 2008. While the average control firm weighted its highest coal input product at 0.78 in 2008, the average treated firm in a cluster with CEPI of at least 70 reduced that weight by 36 percentage points to 0.43. The reduction is significant at the 95% confidence level, indicated by the reported p-value against the null hypothesis that $\beta_1 + \beta_2 = 0$.

Panel B of Table 5 runs the triple DiRD specification by interacting the treatment group indicators with an indicator for being in a highly polluting industry. As in Panel B of Table 4, the tests lack power. Nonetheless, the estimates' direction and relative magnitude support the view that firms in highly polluting industries within treated clusters are the population adjusting their product energy inputs and CO_2 emissions. In Model (3), where the outcome is the weight of the

[25] See Section 2.1 and Online Appendix D.

firm's highest coal input product in 2008, we find that HPI firms in the highest-treatment clusters respond significantly more than firms in the same cluster in non-HPI industries. The difference is significant at the 95% confidence level.

5.2 Outputs

In this subsection, we test the impact of the 2009 CEPI regulation on firm output at the product level. Ex ante, it is not clear there should be any treatment effect, given that the manufacturers in our sample are primarily upstream in their supply chains and have long-term production relationships with the firms they supply. The sample is all manufacturing firms in the clusters with 2009 CEPI values within a bandwidth of 10 index units around the CEPI cutoffs at 60 and 70.

Panel A of Table 7 reports that there is no differential response, on average, between the production on the intensive (Model (1)) margin of firms in treated and control clusters, as proxied by the natural logarithm of the product-level production quantity. Model (2) of Panel A shows no evidence of any differential extensive margin response, proxied for by the natural logarithm of the number of product lines at the firm level.

However, Panel B of Table 7 reveals that the average effect presented in Panel A masks heterogeneity within the treatment groups. In particular, treated firms in highly polluting industries increase their production (Model (1)) and slightly decrease the number of products they produce (Model (2)). Precisely, firms in highly polluting industries in the clusters subject to Supreme Court action plans increase production in the average product by 45% following the CEPI regulation. Though not precisely estimated, Model (2) suggests that the average treated firm in a highly polluting industry is also more likely to reduce its product offerings relative to control firms and treated firms in the same cluster operating in non-HPI industries.

[Place Table 7 here.]

In Models (3) and (4) of Table 7, we test for the effect of the CEPI regulation on product variety in the following years. In aggregate, as reported in Panel A, treated firms are significantly less likely to introduce new products. The reduction is about a 20% decrease in the probability that a firm in the highest-treatment clusters with CEPI above 70 will introduce a new product in the five years following the CEPI regulation in 2009. The probability of dropping a product is positive but insignificant in aggregate.

In Panel B, we see that the reduction in the probability of introducing a new product is driven by all treatment firms, with no differential effect for those in highly polluting industries. However, Firms in highly polluting industries are approximately 13 percentage points more likely to remove a product in the years following the CEPI regulation, significant at the 95% confidence level. Overall, the evidence points to decreased product variety in treated clusters, especially among firms in highly polluting industries, who appear to double down on their existing products.

To see the divergence between the production and HPI and non-HPI firms more clearly, we investigate the dynamics of their portfolio decisions. First, Figure 8 plots the effect over time on the weight of the treated firm's highest-margin product in 2008, relative to control firms, both overall (Panel (a)) and separately for firms in HPI and non-HPI sectors. Specifically, Panel (a) of this figure plots $\widehat{\beta}_k$ from the regression:

$$\begin{aligned} W_{ijcst} = \beta_{-2} D_{-2} \times CEPI_c^{[60,100]} + \sum_{k \in \{0,3\}} \beta_k D_k \times CEPI_c^{[60,100]} \\ + CEPI_c + \gamma_i + \gamma_{jst} + \epsilon_{ijcst} \end{aligned} \tag{4}$$

Note that the treatment indicator is pooled across the two treatment thresholds (the graphical counterpart to the tests of $\beta_1 + \beta_2 = 0$ in the regression tables). We normalize coefficients to one period before the regulation.

[Figure 8 here.]

The figure demonstrates that the shift towards the highest margin product begins at the regulation and increases over time.[26] Panel (b) splits the coefficient into the effect on firms in high-pollution industries (gray diamonds) and firms in other industries (black dots). Now, we observe a stark divergence between the shift toward high-margin products in non-HPI firms

[26]the weight on the highest margin product is marginally significantly less than zero three years before the regulation in 2006, but there is no clear trend in the coefficients, and the pooled pre-period does not differ from zero, with quite precise error bounds.

and the shift away from the highest-margin product among HPI firms in the same cluster that the aggregate coefficients had masked. The difference widens over time, suggesting that the regulation has long-lasting impacts on the production decisions of the highest-treatment group (the HPI group) while less-treated firms (the non-HPI group) make production decisions that increase profitability over time.

In contrast, Figure 9 plots the portfolio weight on the firm's highest emission product in 2008. Panel (a) displays that firms shift away from their dirtiest product, on average, relative to control firms. However, Panel (b) of Figure 9 demonstrates a start divergence between HPI firms and other industries. Namely, only firms in highly polluting industries shift production away from their highest-emission product, while firms in other industries increase their weight on their dirtiest product. This divergence in the behavior of HPI and non-HPI firms masked in the aggregate model further underlies the importance of observing detailed production decisions to accurately estimate the effects of emissions regulation and determine winners and losers.

6 Abatement Expenditures

The 2009 CEPI regulation employed two main policy levers to address industrial cluster emissions. The first was enhancing emissions monitoring. The second was to mandate emissions abatement investments, specifically in the most critically polluted clusters, those with CEPI at or above the 70 CEPI treatment threshold. In the Supreme Court action plans describing the abatement to be undertaken, some investments are mandated at the cluster level and some are to be undertaken by specific plants. Prowess' data series on firm abatement expenditures is the best proxy to measure firm exposure to these investment mandates because it captures the firm's expenditures even when it is contributing to a cluster-wide investment.

We find that firms increase their abatement expenditures, on average, both on the extensive and the intensive margin. Specifically, Model (1) in Panel A of Table 6 reports the extensive margin response of abatement investments to the 2009 CEPI regulation. The outcome $1_{Abatement}$ is an indicator that takes the value of one if the firm's abatement expenditure is non-zero and not missing and zero if the firm's abatement expenditure is zero. We see that only firms in the clusters with 2009 CEPI of at least 70 significantly increase their likelihood of making any abatement investment relative to the abatement investment rate of firms in control clusters. The interpretation relative to the 2008 level of abatement among firms in control clusters is an increase in the probability of making any abatement investment from 4.8% to 7.7%, or a 160% relative increase from baseline following the reform.

[Table 6 here.]

In contrast to the extensive margin results in Model (1), we see in Model (2) that abatement expenditures increase significantly among all firms in treated clusters relative to firms in control clusters. The outcome is abatement investments as a proportion of total firm assets. The effect is relatively stronger in clusters with a 2009 CEPI of at least 70, which is as expected since these clusters are those facing mandated abatement investments. Specifically, firms in clusters with CEPI of at least 70 increase their abatement expenses as a ration of their total assets by 8% relative to control firms. This is a very large effect; at baseline, control firms spent 1.4% of the value of their assets on abatement expenses. Thus, in relative terms, the regulation caused an almost 600% increase in abatement expenses.

We now turn to Panel B of Table 6 where we interact the intensive and extensive margin models with an indicator for the firm's primary industry being classified as highly polluting by the CPCB.

Now again see in Model (1) that the increase in abatement activity on the extensive margin is driven entirely by firms in clusters with CEPI of at least 70 subject to mandatory abatement. However, while the interaction with HPI status in these clusters is positive, there is no statistically significant differential impact between firms in HPI and non-HPI categories.

Turning to Model (2) of Panel B in Table 6, we find that all treated firms are increasing their abatement investments but that firms in HPI sectors are increasing by less. In particular, while the average non-HPI firm in a cluster with CEPI at least 70 increases their abatement investment relative to their total assets by approximately 10%, the average HPI firm in the same cluster increases it by 7%. In summary, the empirical evidence suggests that (1) abatement investments increase on the intensive and extensive margins; (2) abatement was concentrated in clusters subject to Supreme Court action plans, though there is a relative increase in existing abatement investments in clusters only subjected to enhanced monitoring as well; and (3) abatement investments were not allocated based on HPI status.[27]

7 Cluster Dynamism and Aggregate Impacts

Most existing empirical evidence finds that emissions regulation harms growth, albeit with differing estimates on costs. However, a burgeoning viewpoint posits that costs are short-lived and should not argue against using rigorous emissions standards as a policy tool since, in the long term, firms will optimize their energy inputs and adopt green technology, simultaneously reducing emissions and increasing productivity (Linn, 2008; Newell, Jaffe, and Stavins, 1999; Wu, Yu, Jiaxing, and Zhou, 2023). We now speak to this debate by looking at the impacts of the CEPI regulation on firm productivity, pricing, and aggregate cluster firm dynamics.

7.1 Productivity and Profitability

We start by studying the effect of the CEPI regulation on firms' total factor productivity (TFP), or the efficiency with which companies turn inputs into outputs. We construct our primary measure for TFP following Levinsohn and Petrin (2003). When we estimate the production function, we measure output using the total sales value, including income earned by the firm from selling industrial goods and their raw materials, byproducts, stores, and waste.[28]

Model (1) in Panel A of Table 8 reports that productivity increases following the regulation for treated firms, significant at the 95% confidence level. The effect is not affected by treatment intensity. Indeed, in Panel B, where we interact with an indicator for HPI industry membership, we find that firms in the less polluting industries drive the effect while there is no significant change in the TFP of treated HPI firms. Thus, while the headline result would seem to support the hypothesis that emissions regulation increases productivity, subsample analyses show that firms that do not shift production to lower emissions have a higher average relative efficiency

[27] We cannot, unfortunately, observe the bargaining process to agree on the Supreme Court action plans and how they allocated regulatory costs.

[28] We proxy for capital using gross fixed assets, including tangible assets, such as land, building, plant, and machinery, and intangible assets, such as goodwill, software, etc. Labor is compensation to employees that includes all cash and payments in kind made by a company to its employees. Material inputs are the sum of intermediate inputs, proxied by the combined value of raw materials, power, and fuel consumption; raw materials, which we define as the sum of expenses on raw materials, stores, spares, and tools firms use in production. Finally, energy inputs are proxied as power and fuel, including the firms' expenses on power, fuel, and water. We follow best practices by controlling for firm size and deflating all numbers using industry deflators to reflect real values.

in converting inputs to revenue after the regulation, not those that respond to the regulation by cutting emissions and energy input. We report the 2008 average productivity for the control group to aid the interpretation of the coefficients. For every INR of input—measured as capital, labor, and material inputs, as described above—the average control multi-product manufacturer generated approximately 3 INR of revenue. Thus, after the 2009 regulation, HPI firms in treated clusters earned about 0.7 INR more in revenue per INR of input than control firms in 2008.

[Table 8 here.]

Consistent with this, Model (2) of Panels A and B in Table 8 reveal that the average firm preserves headline profitability, as measured by firm earnings before interest, taxes, depreciation, and amortization. Though the relative difference is not statistically significant, the sign of the coefficients points to firms in non-HPI industries as those able to preserve profits following the regulation.

Model (3) in Panel B reports that raw materials costs, which include energy inputs, are lower on average for non-HPI treated firms (significant at the 90% confidence level) while they are relatively higher, on average, for HPI firms (significant at the 99% confidence level).

Finally, Model (4) Panel A asserts that the average treated firm shifts significantly towards the product that had commanded the highest profit margin among the firm's portfolio in 2008. Specifically, treated in clusters with 2009 CEPI between 60 and below 70 increased their weight on their highest margin product by 13% relative to the weight control firms put on their highest margin products. Firms in clusters with 2009 CEPI of at least 70 increase that weight by 28%. Relative to an already relatively high weight on the most profitable product for the average control firm in 2008 of 0.73, firms in the highest-treatment group weighted their highest-margin product after the regulation by over 0.9.

Model (4) of Panel B demonstrates that a sizeable relative portfolio shift among non-HPI firms towards their highest margin product drives the aggregate shift in Panel A. Recall that HPI firms were instead the only group to shift away from their high-coal use products. Thus, Table 8 demonstrates that among treated firms, those in HPI and non-HPI made diverging production decisions in response to the CEPI regulation.

The results in Table 8 are consistent with declining profitability of HPI firms following the regulation relative to non-HPI firms. To further test this, we turn to product-level profit margins in Table 9, which zooms into the product-level pricing decisions of the average treated firm (Panel A) and the relative decisions of treated firms in HPI and non-HPI sectors (Panel B).

The average net result of the product portfolio changes is higher product margins (Model 1). Consistent with the results from Table 8, this effect is driven by non-HPI firms, particularly those in the highest-treatment industrial clusters. While insignificant, Models (2) and (3) of Table 9 demonstrate similar patterns, with only treated HPI firms in industrial clusters with a CEPI of at least 70 facing increased marginal costs and the necessity to increase prices.

[Table 9 here.]

Finally, in Appendix Table IAE13, we follow De Loecker et al. (2016) to construct *quantity* productivity (TFPQ), measuring the efficiency of turning inputs into outputs. Examining quantity and revenue productivity together provides insight into what aspect of value-creation is affected by the regulation while minimizing the disadvantages of each measure.[29] We see that, unlike *revenue* productivity, the efficiency of converting inputs to revenue, there is no significant relative difference in quantity productivity for treated and control firms. If anything, the sign of the (insignificant) effect is negative. While we are only able to calculate quantity TFP for a sub-set of treated firms, the overall conclusion is similar to the picture suggested by Tables 8 and 9 that it is primarily the productivity of non-HPI firms that are affected by the emissions regulation, with no evidence of productivity benefits.[30]

The balance of the evidence suggests that the CEPI emissions regulation did not increase average firm productivity. Instead, firms in highly polluting industries that complied with the regulation and changed their production decisions to lower emissions became less productive and profitable. Further supporting the assertion that HPI firms are the ones that respond by changing energy inputs, only this group experienced a significant increase in raw material expenditures.

[29] Atkin, Khandelwal, and Osman (2019) find that TFPQ performs poorly at measuring quantity productivity, shows excessive dispersion across firms, and correlates negatively with quality productivity. They attribute this to the difficulty of adjusting for product specifications and quality to make apples-to-apples comparisons. They find that TFPR does better than TFPQ at capturing broad firm capabilities. However, TFPR suffers from being unable to separate effects from changes in productivity, markups, the firm product mix, and product quality. See Appendix E for more details about the assumptions and methods we use and for evidence that the assumptions hold in our setting.

[30] In Appendix Table IAA5 we also see that treated are not making more investments or increasing their R&D. Thus, we do not find any evidence of a technology adoption channel underlying the theoretical channel from emissions capping to enhanced productivity.

Conversely, non-HPI firms within the same cluster that received relatively less treatment increased product margins and their efficiency in translating inputs into revenue, preserving their overall profitability even after the regulation. These results point to a potential impact on the competitiveness and dynamism of clusters that is a function of their composition, a hypothesis we turn to in the next section.

7.2 Firm Entry

In this subsection, we consider aggregate effects at the cluster level. A key motivation for organizing manufacturing activity using industrial clusters is to boost firm productivity through economies of scale and positive spillovers. We focus on firm entry as a summary measure of the agglomeration benefits since it reflects the net benefits and costs of agglomeration in a given cluster that are unobservable and observable in our dataset.

Table 10 reports the results from our DiRD design on firm entry run at the cluster level. The sample for Panel A is all firms from the formal firm registry of the Ministry of Corporate Affairs, not just the relatively larger firms present in the Prowess dataset. We interpret these results as the effect entry of the average, small firm into a cluster. As in the cluster-level emissions tests, the control group comprises clusters for which the CPCB constructed a CEPI score in 2009 but whose scores are below the lowest treatment threshold at a CEPI of 60.

[Table 10 here.]

The regulation decreases firm entry, driven by the effect in $CEPI^{[70,100]}$ clusters. Model (1) presents a linear probability model on the incidence of new firm creation. We see an approximately 1.8% drop in new firm registrations in clusters with a CEPI of at least 70. The effect for clusters with a CEPI between 60 and below 70 is half that effect, though it is not statistically significant after controlling for the differential effect on the highest-treatment clusters.

Model (2) of Table 10 demonstrates the effect on the natural logarithm of the number of firms in $CEPI^{[70,100]}$ versus control areas. We again see a significant reduction. The magnitude is large; company entry more than halves after the regulation in $CEPI^{[70,100]}$ clusters relative to control clusters within the bandwidth. Column (3) runs the same test except that the outcome is the inverse hyperbolic sine function (i.e., $asinh(x) = ln(x + sqrt(x * x + 1))$), which, unlike the natural logarithm, is well-defined at zero, while Column (4) presents a model with the levels of the number of firms in each cluster as the outcome. Results are consistent with those of Columns (1) and (2). The implication is that the regulation's costs were sufficient to deter entry.

In Panel B of Table 10, we repeat the exercise using only firms in the Prowess database. These are the firms in our regression samples. The interpretation of this panel is the effect on the average large firm. We also see in this subsample that (1) firm entry significantly decreases; (2) Clusters with a CEPI of at least 70 again drive the effect; and (3) the magnitude is about a 50 percent decrease in firm entry relative to the ex-ante pattern in control firms.

In summary, we document a significant decrease in firm entry across the firm size distribution. This effect is strongest in industrial clusters with the greatest enforcement intensity. These findings suggest a dampening competitive pressure within the cluster, with lowered potential agglomeration benefits. These results are most consistent with the part of the literature that contends that emissions regulation can lower emissions but at a cost to economic dynamism.[31]

7.3 Other Margins of Adjustment

We consider alternative margins of adjustment for lower emissions among firms in treated clusters. Firms may relocate or expand their operations to regions with less stringent environmental regulations to reduce their production costs, particularly those related to environmental compliance (Copeland and Taylor, 1994, 2004). Prior work has documented that localized policies shift emissions within and across geographies (Bartram, Hou, and Kim, 2022; Ben-David, Jang, Kleimeier, and Viehs, 2021).

We examine two margins of adjustment by firms beyond altering their production decisions. First, we focus on changes in merger and acquisition activity around the regulation. Specifically, we examine whether firms in the targeted industrial cluster are more likely to be acquired or merged with firms outside the cluster. In Online Appendix Table IAA6, we show that, on average, firms in treated clusters are unlikely to be a target or be acquired in merger and acquisition around the regulation, with the probabilities being similar across treated and control clusters.

[31] A full, general equilibrium analysis is beyond the scope of this paper, limited as we are by less information on the direct costs of the regulation and the counterweight from health benefits from lower emissions. We do find that costs and benefits of the regulation (1) occur over different horizons, with costs accumulating in the more medium- and long-term in the data, and (2) there are winners and losers, so accounting for heterogeneity is key in any welfare analysis.

Second, instead of entirely relocating their operations through mergers and acquisitions, firms may relocate their production activities by building new plants and expanding capacity elsewhere. We explore this possibility in Online Appendix Table IAA7. Specifically, our findings indicate that, on average, firms are unlikely to announce new plant constructions or abandon the expansion of existing plants.

8 Discussion

Our results open up the 'brown box' of how firms change their inputs and outputs in response to environmental mandates. We find that these regulations lowered emissions by prompting a shift away from high-emission energy sources alongside investments in abatement. However, in the aggregate, these regulations lowered business dynamism, potentially impairing competitiveness. Our results point to the mechanisms that policymakers could target to balance these costs against lower emissions.

First, it is firms subject to higher monitoring intensity who change their energy inputs. Alternative emissions targeting could mandate specific energy input use rather than imposing caps on emissions followed by continuous monitoring. In contrast, it was firms subject to relatively more diffuse monitoring that make abatement investments. This suggests that command-and-control mechanisms may not be the most effective means to cost-effectively incentivize investment and technology adoption.

Second, another dimension that constrained regulators should consider when allocating their monitoring efforts is where in the supply chain they are focusing regulation. Our sample is concentrated upstream in the supply chain and this is perhaps the primary reason we find firms internalizing rather than passing on costs to their customers and suppliers.

More broadly, our results speak to the importance of disclosure of emissions at different stages of production. Our study informs policymakers about how to effectively target and monitor firms using these data. Policymakers are already considering such disclosure requirements around the world. For example, Gary Gensler, the Chair of the United States Securities and Exchange Commission, has called for more detailed emissions disclosures including indirect and supply chain emissions (Gensler, 2023). At the same time in the European Union, the Corporate Sustainability Reporting directive requires all public firms to report their greenhouse gas emissions, including from Scope 3 (European Commission, 2021). Our results suggest that these disclosures will enhance the effectiveness of environmental regulations and play an important role in national efforts to combat climate change.

Finally, our results highlight the need for coordinated policies on decarbonization. Arguably, the shift toward purchasing rather than producing electricity that we document does not necessarily steer away from coal at present. However, incentivizing firms to use electricity could facilitate a smoother transition to cleaner fuels as it becomes technically and economically viable to green the power grid.

9 Conclusion

Policymakers are increasingly setting emission targets for industrial clusters while considering the impact on firm productivity and global competitiveness. However, evidence on the regulation's effectiveness and cost is mixed. Particularly, research on climate regulation's effect on within-firm adaptation is limited. Estimating emission cap impacts is challenging due to opaque within-firm responses, including operational, product, and energy adjustments.

This paper explores within-firm production responses, inputs, and outputs to emissions reduction, using a novel environmental regulation in India in a difference-in-discontinuities (DiRD) specification. We show that this regulation significantly decreased emissions, as seen in regulatory data and satellite readings. Our detailed analysis of product-level firm data reveals that on the input side, firms optimize energy use and shift from producing to buying electricity, investing significantly in emissions abatement. On the output side, there's a shift from high-coal, high-emission products to higher-margin ones, with no evidence of operations moving outside the cluster, unlike in developed economies.

Our analysis within industrial clusters shows that high-pollution industries primarily drive operational changes and bear the brunt of emissions reduction, leading to decreased productivity. In contrast, other sectors see productivity and profitability gains. This indicates a trade-off between emissions reduction and productivity. Comparing firms within the same cluster allows control over many unobservables and alternative narratives. Overall, while there are both winners and losers, the aggregate impact suggests a trade-off at the industrial cluster level, marked by reduced economic dynamism, fewer new firm entries, and less product variety.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge, 2017, When should you adjust standard errors for clustering?, Technical report, National Bureau of Economic Research.

Abadie, Alberto, and Matias D Cattaneo, 2018, Econometric methods for program evaluation, *Annual Review of Economics* 10.

Alok, Shashwat, and Meghana Ayyagari, 2020, Politics, state ownership, and corporate investments, *The Review of Financial Studies* 33, 3031–3087.

Atkin, David, Amit K Khandelwal, and Adam Osman, 2019, Measuring productivity: Lessons from tailored surveys and productivity benchmarking, in *AEA Papers and Proceedings*, volume 109, 444–449, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Bartram, Söhnke M, Kewei Hou, and Sehoon Kim, 2022, Real effects of climate policy: Financial constraints and spillovers, *Journal of Financial Economics* 143, 668–696.

Bau, Natalie, and Adrien Matray, 2023, Misallocation and capital market integration: Evidence from india, *Econometrica* 91, 67–106.

Ben-David, Itzhak, Yeejin Jang, Stefanie Kleimeier, and Michael Viehs, 2021, Exporting pollution: where do multinational firms emit co2?, *Economic Policy* 36, 377–437.

Berman, Eli, and Linda TM Bui, 2001, Environmental regulation and productivity: evidence from oil refineries, *Review of Economics and Statistics* 83, 498–510.

Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan, 2007, Obtaining a driver's license in india: an experimental approach to studying corruption, *Quarterly Journal of Economics* 122, 1639–1676.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, 2004, How much should we trust differences-in-differences estimates?, *Quarterly Journal of Economics* 119, 249–275.

Bertrand, Marianne, Paras Mehta, and Sendhil Mullainathan, 2002, Ferreting out Tunneling: An Application to Indian Business Groups, *Quarterly Journal of Economics* 117, 121–148.

Bhat, Sairam, 2010, *Natural resources conservation law* (SAGE Publications Ltd).

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell, 2020, Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs, *Econometrics Journal* 23, 192–210.

Central Electricity Authority, 2008, Baseline carbon dioxide emission database, version 4.0, https://cea.nic.in/cdm-co2-baseline-database/, Government of India.

Central Pollution Control Board of India, CPCB, 2009, Comprehensive environmental assessment of industrial clusters, Technical report, Technical report.

Copeland, Brian R, and M Scott Taylor, 1994, North-south trade and the environment, *Quarterly Journal of Economics* 109, 755–787.

Copeland, Brian R, and M Scott Taylor, 2004, Trade, growth, and the environment, *Journal of Economic Literature* 42, 7–71.

Crippa, Monica, Diego Guizzardi, Marilena Muntean, Edwin Schaaf, and Gabriel Oreggioni, 2019, Edgar v5.0 global air pollutant emissions, Date accessed: 16 April 2020.

Crippa, Monica, Efisio Solazzo, Ganlin Huang, Diego Guizzardi, Ernest Koffi, Marilena Muntean, Christian Schieberle, Rainer Friedrich, and Greet Janssens-Maenhout, 2020, High resolution temporal profiles in the emissions database for global atmospheric research, *Nature Scientific Data* 7, 1–17.

Dai, Rui, Rui Duan, Hao Liang, and Lilian Ng, 2021a, Outsourcing climate change, *Working Paper* .

Dai, Rui, Hao Liang, and Lilian Ng, 2021b, Socially responsible corporate customers, *Journal of Financial Economics* 142, 598–626.

De Loecker, Jan, 2011, Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity, *Econometrica* 79, 1407–1451.

De Loecker, Jan, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik, 2016, Prices, markups, and trade reform, *Econometrica* 84, 445–510.

Duchin, Ran, Janet Gao, and Qiping Xu, 2022, Sustainability or greenwashing: Evidence from the asset market for industrial pollution, *Working Paper* .

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, 2013, Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India, *Quarterly Journal of Economics* 128, 1499–1545.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, 2018, The value of regulatory discretion: Estimates from environmental inspections in India, *Econometrica* 86, 2123–2160.

Eckel, Carsten, and J Peter Neary, 2010, Multi-product firms and flexible manufacturing in the global economy, *Review of Economic Studies* 77, 188–217.

Energy and Resources Institute, 2008, *The Energy and Resources Institute Energy Data Directory & Yearbook 2008/09* (Energy and Resources Institute), teri.

European Commission, 2021, Corporate sustainability reporting directive (csrd), Official Journal of the European Union, Accessed on: [insert date of access].

European Commission, Joint Research Centre (EC-JRC)/Netherlands Environmental Assessment Agency (PBL), 2020, Emissions database for global atmospheric research (edgar), release edgar v5.0 (1970 - 2015) of april 2020, Date accessed: 27 February 2021.

Fan, Haichao, Joshua S Graff Zivin, Zonglai Kou, Xueyue Liu, and Huanhuan Wang, 2019, Going green in china: Firms' responses to stricter environmental regulations, Technical report, National Bureau of Economic Research.

Fenske, James, Muhammad Haseeb, and Namrata Kala, 2023, How rules and compliance impact organizational outcomes: Impact from delegation in environmental regulation, Technical report, Working paper.

Fowlie, Meredith, 2010, Emissions trading, electricity restructuring, and investment in pollution abatement, *American Economic Review* 100, 837–869.

Gensler, Gary, 2023, Remarks at the financial stability oversight council, Securities and Exchange Commission, Retrieved from Securities and Exchange Commission website: https://www.sec.gov/news/speech/gensler-remarks-fsoc-climate-072823.

Goldberg, Pinelopi K, Amit K Khandelwal, Nina Pavcnik, and Petia Topalova, 2010, Multiproduct firms and product turnover in the developing world: Evidence from India, *Review of Economics and Statistics* 92, 1042–1049.

Gopalan, Radhakrishnan, Abhiroop Mukherjee, and Manpreet Singh, 2016, Do debt contract enforcement costs affect financing and asset structure?, *Review of Financial Studies* 29, 2774–2813.

Gopalan, Radhakrishnan, Vikram Nanda, and Amit Seru, 2007, Affiliated firms and financial support: Evidence from Indian business groups, *Journal of Financial Economics* 86, 759–795.

Greenstone, Michael, 2002, The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 clean air act amendments and the census of manufactures, *Journal of Political Economy* 110, 1175–1219.

Greenstone, Michael, and Rema Hanna, 2014, Environmental regulations, air and water pollution, and infant mortality in India, *American Economic Review* 104, 3038–72.

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

Greenstone, Michael, John A List, and Chad Syverson, 2012, The effects of environmental regulation on the competitiveness of us manufacturing, Technical report, National Bureau of Economic Research.

Greenstone, Michael, Rohini Pande, Anant Sudarshan, and Nicholas Ryan, 2022, The benefits and costs of emissions trading: Experimental evidence from a new market for industrial particulate emissions, *Working Paper* .

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano, 2016, Do fiscal rules matter?, *American Economic Journal: Applied Economics* 8, 1–30.

Harari, Mariaflavia, 2020, Cities in bad shape: Urban geometry in india, *American Economic Review* 110, 2377–2421.

Harrison, Ann, Benjamin Hyman, Leslie Martin, and Shanthi Nataraj, 2019, When do firms go green? Comparing command and control regulations with price incentives in India, Technical report, National Bureau of Economic Research.

Harrison, Ann, Leslie A Martin, and Shanthi Nataraj, 2017, Green industrial policy in emerging markets, *Annual Review of Resource Economics* 9, 253–274.

He, Guojun, Shaoda Wang, and Bing Zhang, 2020, Watering down environmental regulation in china, *Quarterly Journal of Economics* 135, 2135–2185.

IPCC, 2023, Climate change 2023: Synthesis report. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change, Technical report, IPCC, Geneva, Switzerland.

Ivanov, Ivan, Mathias S Kruttli, and Sumudu W Watugala, 2023, Banking on carbon: Corporate lending and cap-and-trade policy, *Available at SSRN 3650447* .

Jaffe, Adam B, and Karen Palmer, 1997, Environmental regulation and innovation: a panel data study, *Review of economics and statistics* 79, 610–619.

Juhász, Réka, Nathan Lane, Emily Oehlsen, and Verónica C Pérez, 2022, The who, what, when, and how of industrial policy: A text-based approach, *Working Paper* .

Kala, Namrata, and Michael Gechter, 2023, Firm presence, environmental quality, and agglomeration: Evidence from a large-scale randomized relocation, Technical report, Working paper.

Kalmenovitz, Joseph, and Jason Chen, 2021, The environmental consequences of pay inequality, *NYU Stern School of Business Forthcoming* .

Kapur, Devesh, and Madhav Khosla, 2019, *Regulation in India: Design, capacity, performance*, volume 24 (Bloomsbury Publishing).

Kim, Taehyun, and Qiping Xu, 2021, Financial constraints and corporate environmental policies, *Review of Financial Studies* forthcoming.

Kugler, Maurice, and Eric Verhoogen, 2012, Prices, plant size, and product quality, *Review of Economic Studies* 79, 307–339.

Lanjouw, Jean Olson, and Ashoka Mody, 1996, Innovation and the international diffusion of environmentally responsive technology, *Research policy* 25, 549–571.

Lenox, Michael J, and Charles E Eesley, 2009, Private environmental activism and the selection and response of firm targets, *Journal of Economics & Management Strategy* 18, 45–73.

Levinsohn, James, and Amil Petrin, 2003, Estimating production functions using inputs to control for unobservables, *Review of Economic Studies* 70, 317–341.

Lilienfeld-Toal, Ulf von, Dilip Mookherjee, and Sujata Visaria, 2012, The distributive impact of reforms in credit enforcement: Evidence from Indian debt recovery tribunals, *Econometrica* 80, 497–558.

Linn, Joshua, 2008, Energy prices and the adoption of energy-saving technology, *Economic Journal* 118, 1986–2012.

Lyubich, Eva, Joseph S Shapiro, and Reed Walker, 2018, Regulating mismeasured pollution: Implications of firm heterogeneity for environmental policy, in *AEA Papers and Proceedings*, volume 108, 136–142, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Martin, Leslie A, 2011, Energy efficiency gains from trade: Greenhouse gas emissions and India's manufacturing sector, *Mimeograph, Berkeley ARE* .

Mayer, Thierry, Marc J Melitz, and Gianmarco IP Ottaviano, 2014, Market size, competition, and the product mix of exporters, *American Economic Review* 104, 495–536.

McCrary, Justin, 2008, Manipulation of the running variable in the regression discontinuity design: A density test, *Journal of Econometrics* 142, 698–714.

Melitz, Marc J, and Gianmarco IP Ottaviano, 2008, Market size, trade, and productivity, *Review of Economic Studies* 75, 295–316.

Naaraayanan, S Lakshmi, Kunal Sachdeva, and Varun Sharma, 2021, The real effects of environmental activist investing, *European Corporate Governance Institute–Finance Working Paper* .

Naaraayanan, S Lakshmi, and Daniel Wolfenzon, 2023, Business group spillovers, *Review of Financial Studies* Forthcoming.

Newell, Richard G, Adam B Jaffe, and Robert N Stavins, 1999, The induced innovation hypothesis and energy-saving technological change., *Quarterly Journal of Economics* 114.

Pollution Control Board of Andhra Pradesh, CPCB, 2010, Final action plan for improvement of environmental parameters in critically polluted areas of "patancheru-bollaram cluster," andhra pradesh, Technical report, Technical report.

Porter, Michael E, and Claas van der Linde, 1995, Toward a new conception of the environment-competitiveness relationship, *Journal of Economic Perspectives* 9, 97–118.

Roberts, Michael R, and Toni M Whited, 2013, Endogeneity in empirical corporate finance1, in *Handbook of the Economics of Finance*, volume 2, 493–572 (Elsevier).

Schiller, Christoph, 2018, Global supply-chain networks and corporate social responsibility .

Servaes, Henri, and Ane Tamayo, 2013, The impact of corporate social responsibility on firm value: The role of customer awareness, *Management Science* 59, 1045–1061.

Siegel, Jordan, and Prithwiraj Choudhury, 2012, A reexamination of tunneling and business groups: New data and new methods, *Review of Financial Studies* 25, 1763–1798.

Sivadasan, Jagadeesh, 2009, Barriers to competition and productivity: Evidence from india, *The BE Journal of Economic Analysis & Policy* 9.

Van Donkelaar, Aaron, Melanie S Hammer, Liam Bindle, Michael Brauer, Jeffery R Brook, Michael J Garay, N Christina Hsu, Olga V Kalashnikova, Ralph A Kahn, Colin Lee, et al., 2021, Monthly global estimates of fine particulate matter and their uncertainty, *Environmental Science & Technology* 55, 15287–15300.

Van Donkelaar, Aaron, Randall V Martin, Robert JD Spurr, and Richard T Burnett, 2015, High-resolution satellite-derived pm2. 5 from optimal estimation and geographically weighted regression over north america, *Environmental Science & Technology* 49, 10482–10491.

World Economic Forum, 2023, Transitioning industrial clusters towards net zero, *World Economic Forum* .

Wu, Xiting, Xiaoyun Yu, You Jiaxing, and Clara Zhou, 2023, Leveling up your green mojo: The benefits of beneficent investment, *Working Paper* .

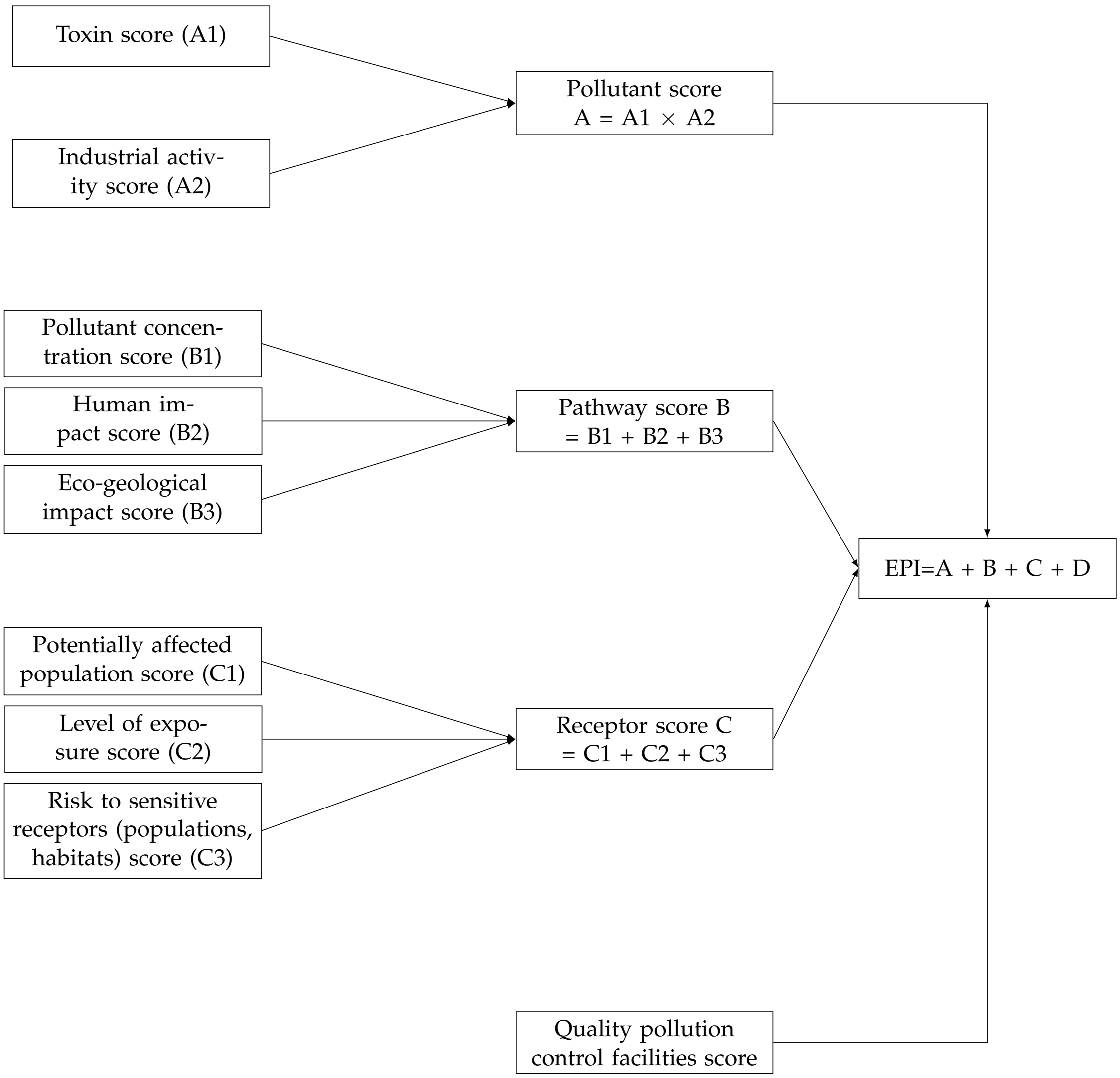
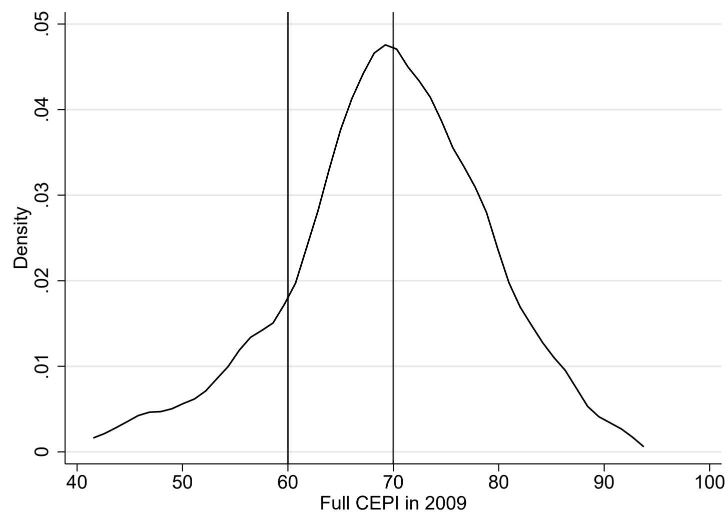
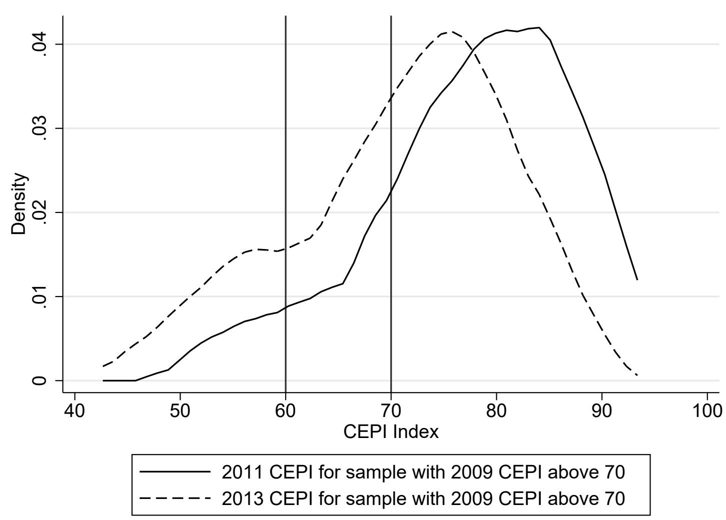


FIGURE 1: THE COMPREHENSIVE ENVIRONMENTAL POLLUTION INDEX (CEPI)

Notes: This figure displays how different components lead to the construction of the CEPI which was used to separate targeted industrial clusters into: Critically Polluted Areas (CPA) and Severely Polluted Areas (SPA).



(a) CEPI in 2009



(b) Comparison between 2011 and 2013

FIGURE 2: EVOLUTION OF THE RANKING VARIABLE

Notes: Panel A presents the distribution of rankings for all 88 industrial clusters for which the CPCB computed a CEPI value in 2009. The vertical line corresponds to the treatment thresholds, CEPI = 60 and CEPI = 70. Panel B reports the distribution of the CEPI in follow-up government studies in 2011 (solid line) and 2013 (dashed line) for the subsample of industrial clusters that were classified as having a CEPI value $>$ 70 in 2009.

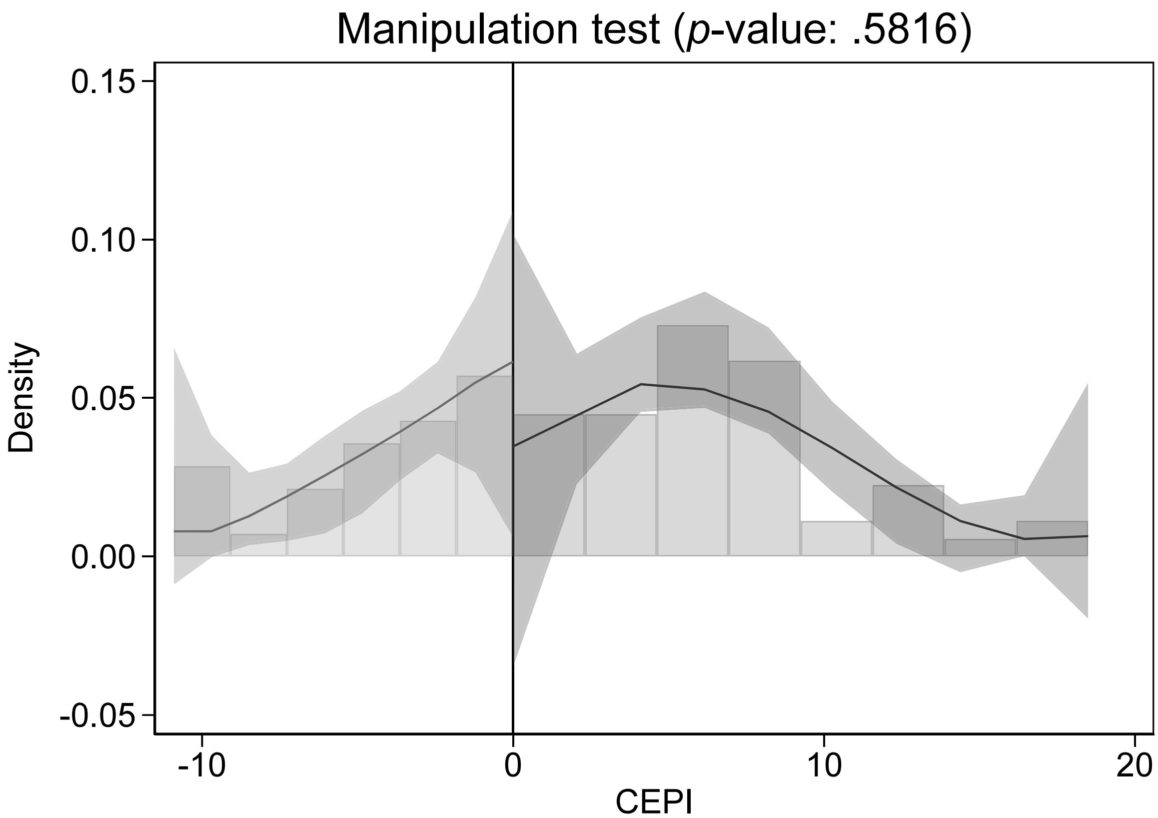


FIGURE 3: TESTING MANIPULATION OF THE RANKING VARIABLE

Notes: This figure tests for manipulation of the industrial cluster pollution ranking variable, around the pooled thresholds at the CEPI values of 60 and 70. The *p*-value is from a two-sided test with the null hypothesis that the distributions of the rankings do not differ across the cutoff (Abadie and Cattaneo, 2018; McCrary, 2008).

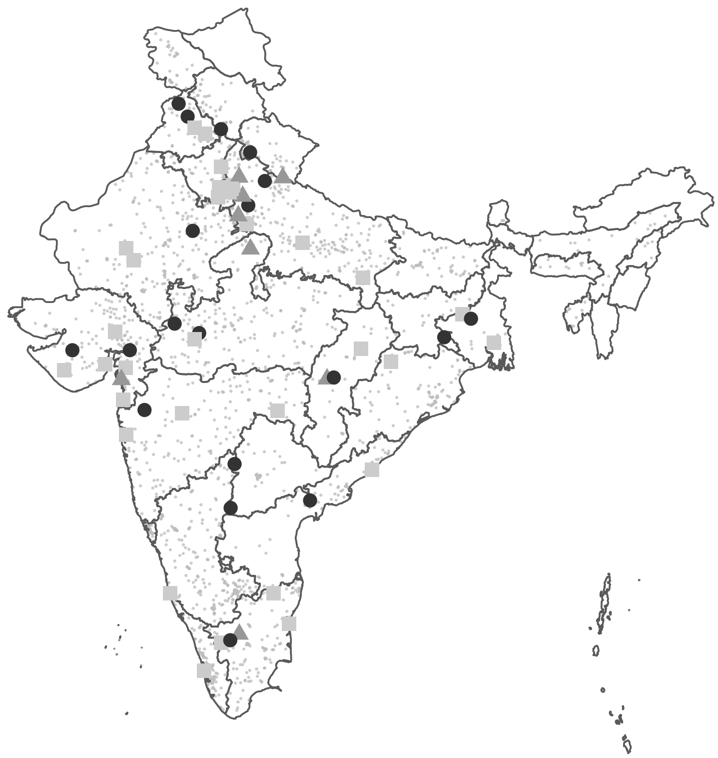


FIGURE 4: GEOGRAPHIC VARIATION OF INDUSTRIAL CLUSTERS

Notes: This figure presents the geographic variation of all industrial clusters as of the year 2009. Small gray dots illustrate the location of all industrial clusters. Larger black circles correspond to clusters with CEPI values at or above the 70 threshold, triangles correspond to clusters with index values between the 60 and 70 thresholds, and squares correspond to clusters with index values below the 60 threshold.

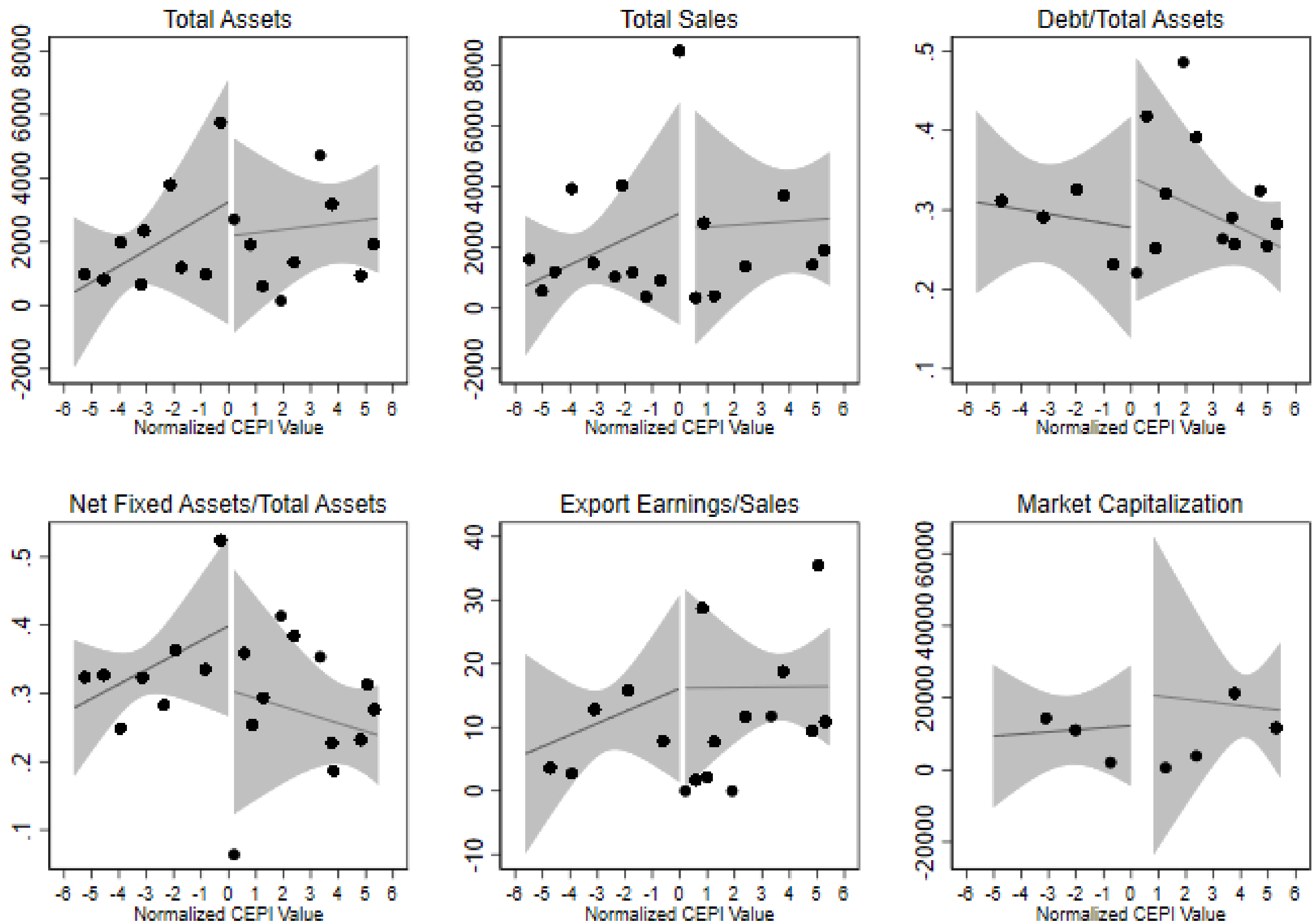


FIGURE 5: FIRM CHARACTERISTICS PRIOR TO THE INTRODUCTION OF THE CEPI

Notes: This figure presents regression discontinuity estimates for baseline firm characteristics. Specifically, the average firm characteristic from 2008, one year before the introduction of CEPI regulation, is graphed in CEPI bins around the cutoff. Data are pooled across the two CEPI thresholds at 60 and 70 (indicated by zero in the figures). Figures include 95% confidence intervals.

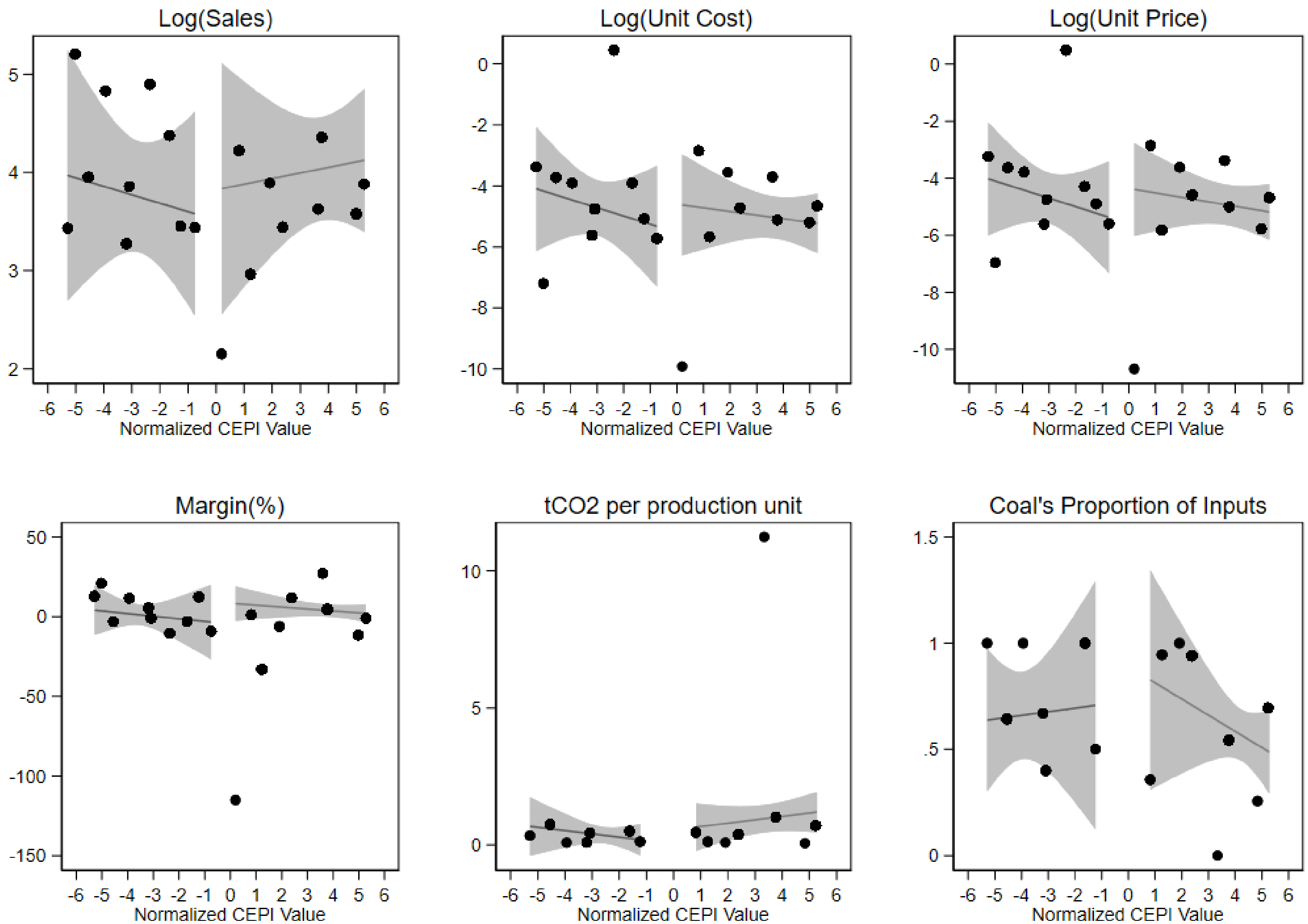


FIGURE 6: BASELINE PRODUCT CHARACTERISTICS PRIOR TO THE INTRODUCTION OF CEPI

Notes: Specifically, the average firm-product characteristics from 2008, one year before the CEPI regulation, is graphed in CEPI bins around the cutoff. Data are pooled across the two CEPI thresholds at 60 and 70 (indicated by zero in the figures). Figures include 95 percent confidence intervals. All ratios are winsorized at 1% tails.

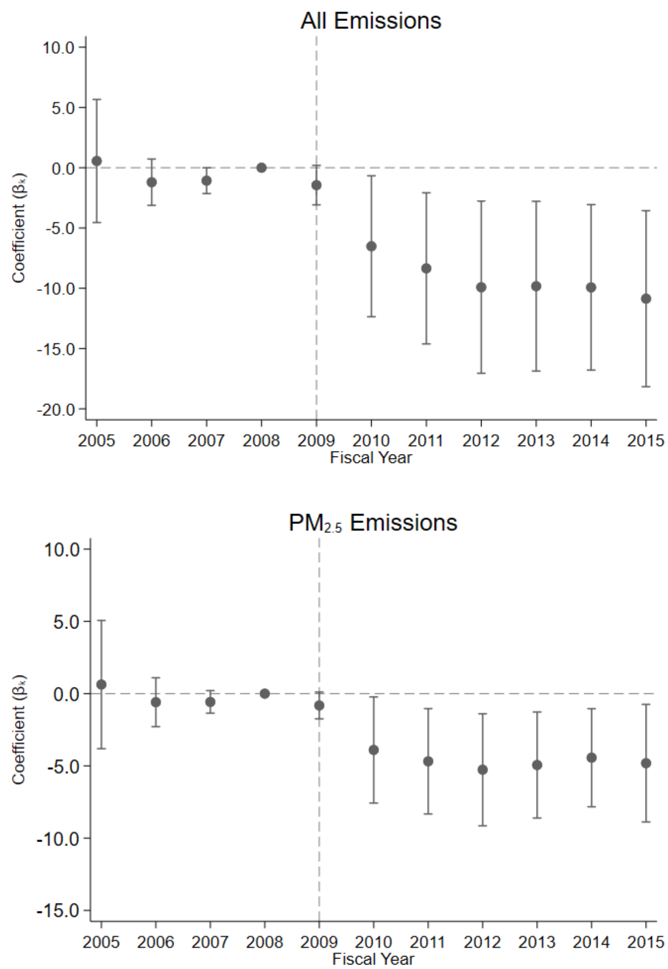


FIGURE 7: CHANGES TO CLUSTER-LEVEL INDUSTRIAL AIR EMISSIONS

Notes: The figure plots the dynamic coefficients (β_k) and the corresponding 95% confidence intervals from the following difference-in-differences specification:

$$Emissions_{pcst} = \sum_{k\in\{-4,-2\}} \beta_k D_k \times CEPI_c^{[60,100]} + \sum_{k\in\{0,6\}} \beta_k D_k \times CEPI_c^{[60,100]} + CEPI_c + \kappa_{cp} + \gamma_{pt} + \epsilon_{cst}$$

where c is cluster, s is state, and t is year-month. The top figure plots the coefficient and their 95% confidence intervals for all emissions while the bottom figure focuses on fine particulate matter < 2.5 microns in diameter. All coefficients are plotted relative to emission levels at k=-1, the year before the introduction of the CEPI regulation in 2009, indicated by the dotted vertical line, which is normalized to zero. $CEPI^{[60,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 60, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. Data source: Emissions Database for Global Atmospheric Research (Crippa et al., 2019, 2020; European Commission, Joint Research Centre (EC-JRC)/Netherlands Environmental Assessment Agency (PBL), 2020).

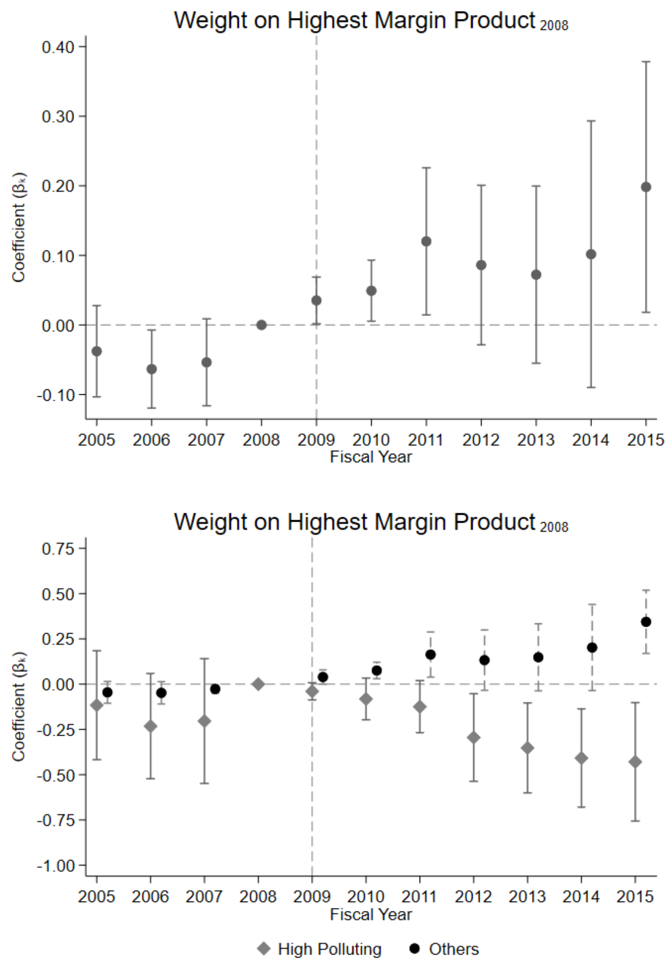


FIGURE 8: CHANGES TO PRODUCT PORTFOLIO BY PROFITABILITY

Notes: The top figure plots coefficients (β_k) and their 95% confidence intervals estimated from the following specification:

$$W_{ijcst}^{\text{Highest Margin Product, 2008}} = \sum_{k\in\{-4,-2\}} \beta_k D_k \times CEPI_c^{[60,100]} + \sum_{k\in\{0,6\}} \beta_k D_k \times CEPI_c^{[60,100]} + CEPI_c + \gamma_i + \gamma_{sjt} + \epsilon_{ijcst} \quad (5)$$

where i is firm, j is product, c is cluster, s is state, and t is year. The top figure plots the coefficient and their 95% confidence intervals for all firms while the bottom figure separates between firms in High-Polluting and other industries. All coefficients are plotted relative to k=-1, the year before the introduction of the CEPI regulation in 2009 indicated by the dotted vertical line, which is normalized to zero. $CEPI^{[60,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 60, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. Data source: CMIE Prowess.

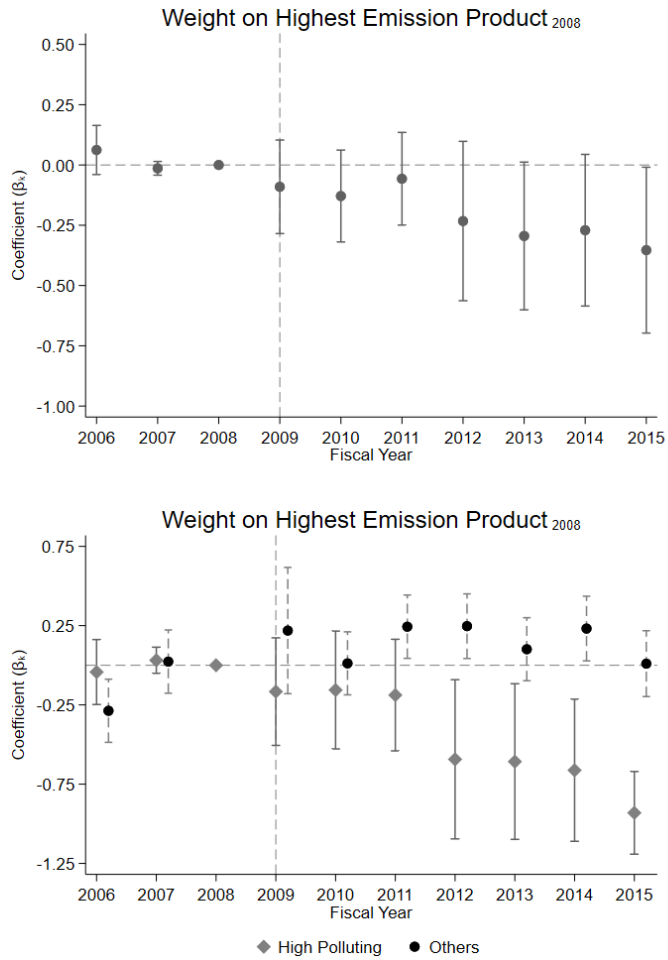


FIGURE 9: CHANGES TO PRODUCT-LEVEL EMISSIONS

Notes: The top figure plots coefficients (β_k) and their 95% confidence intervals estimated from the following specification:

$$W_{ijcst}^{\text{Highest Emission Product, 2008}} = \sum_{k \in \{-3,-2\}} \beta_k D_k \times CEPI_c^{[60,100]} + \sum_{k \in \{0,6\}} \beta_k D_k \times CEPI_c^{[60,100]} + CEPI_c + \gamma_i + \gamma_{sjt} + \epsilon_{ijcst} \quad (6)$$

Where i is firm, j is product, c is cluster, s is state, and t is year. The top figure plots the coefficient and their 95% confidence intervals for all firms while the bottom figure separates between firms in High-Polluting and other industries. All coefficients are plotted relative to k=-1, the year before the introduction of the CEPI regulation in 2009 indicated by the dotted vertical line, which is normalized to zero. $CEPI^{[60,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 60, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. Data source: CMIE Prowess.

TABLE 1: SUMMARY STATISTICS

The table presents descriptive statistics for the firms and products in our baseline sample. Panel A summarizes the firm-year panel dataset. *Assets* and *Sales* is in thousands of INR. *Leverage* is the sum of short- and long-term debt obligations scaled by contemporaneously reported Assets. *Tangibility* is property plant and equipment less deprecation scaled by contemporaneously reported Total Assets. *Exporting intensity* are firm earnings from exports of goods plus services scaled by contemporaneous total sales. *Market capitalization* is reported for listed firms and is the total number of listed shares multiplied by the share price at the end of the fiscal year. *Ln(Productivity)* is the natural log of firm productivity, which is calculated following Levinsohn and Petrin (2003) and controls for firm size. *Profitability* are Earnings Before Interest, Taxes, Depreciation and Amortization as a ratio of the prior year sales. Panel B summarizes the firm-product-year dataset. *Ln(Product Sales)* and *Ln(Unit Cost)* is the per-product production sales and cost, respectively. *Ln(Unit Price)* is the natural logarithm of the per-unit price (where unit is unique within but not across firm). *Margin* is (unit price - unit cost)/unit price. *Ln(Per Unit CO_2 Emissions)* are the author-calculated CO_2 emissions per reported unit of production. All variables are defined in Appendix Table IAA8.

	Panel A: Firm characteristics					
	Obs	Mean	Std. dev.	Min.	Median	Max.
	(1)	(2)	(3)	(4)	(5)	(6)
Assets (000 INR)	11,452	3524	8864	6.70	621	52,664
Sales (000 INR)	11,452	3282	7274	3.90	755	40,262
Leverage	10,307	0.27	0.20	0.00	0.25	1.13
Tangibility	11,057	0.27	0.17	0.01	0.25	0.81
Exporting intensity	11,452	16.30	26.09	0.00	1.64	97.84
Market capitalization	1,949	22247	57844	35.04	2362	323121
Log(Revenue Productivity)	11,452	3.07	1.86	1.02	2.54	8.63
Number Product Lines	11,452	2.84	2.02	1.00	2.00	22.00
Profitability	11,452	0.11	0.08	-0.09	0.10	0.30

	Panel B: Firm-product characteristics					
	Obs	Mean	Std. dev.	Min.	Median	Max.
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Product Sales)	30,137	4.45	2.80	-2.30	4.76	13.20
Ln(Unit Cost)	15,589	-4.93	4.04	-21.44	-3.85	15.97
Ln(Unit Price)	16,328	-4.87	4.05	-21.32	-3.73	16.98
Margin (%)	15,589	0.01	0.70	-5.67	0.14	0.64
Ln(Per Unit CO_2 Emissions)	1,163	-2.32	2.93	-10.64	-1.85	5.13

TABLE 2: COVARIATE BALANCE

The table presents tests for differences in firm and product characteristics before the 2009 reform for the regression sample. Panel A reports balance at the firm level while Panel B reports balance at the firm-product level. In both panels, columns 1 and 2 present the unconditional means for cities below the treatment threshold, and cities above the treatment threshold, respectively. Column 3 presents the difference in means between cities below the treatment threshold and cities above the treatment threshold. Additionally, in Panel A, column 4 shows the regression discontinuity estimate of the effect of being above the treatment threshold on the baseline variable. The model is estimated within a bandwidth of 10 units of the CEPI around the treatment thresholds at 60 and 70. Finally, column 5 is the *p*-value for this estimate, using bias-corrected, heteroskedasticity robust standard errors of Calonico, Cattaneo, and Farrell (2020). All variables are defined in Appendix Table IAA8.

	Panel A: Firm characteristics					
	All	Below	Above	Difference	RD Estimate	*p*-value
	(1)	(2)	(3)	(4)	(5)	(6)
Assets (000 INR)	2,335	1,955	2,402	-448	-489	0.84
Sales (000 INR)	2,268	1,730	2,363	-633	-90	0.97
Leverage	0.28	0.3	0.28	0.023	0.034	0.71
Tangibility	0.28	0.34	0.27	0.069	-0.047	0.66
Exporting Intensity	15	12	16	-4.2	0.45	0.95
Market Capitalization	21,757	20,777	21,867	-1,091	16,282	0.71
Log(Revenue Productivity)	3.3	3.2	3.3	-0.051	-0.34	0.71
Number of Products	2.9	2.9	2.9	-0.0022	-0.25	0.54
Profitability	0.11	0.12	0.11	0.0027	0.065	0.14

	Panel B: Firm-product characteristics					
	All	Below	Above	Difference	RD Estimate	*p*-value
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Product Sales)	4.3	4.0	4.4	-0.41	-0.02	0.97
Ln(Unit Cost)	-4.9	-4.8	-5.0	-0.13	1.1	0.32
Ln(Unit Price)	-4.9	-4.7	-4.9	0.19	0.71	0.45
Margin (%)	-0.66	-0.48	-0.71	0.23	0.55	0.39
Ln(Per Unit CO_2 Emissions)	-1.5	-1.4	-1.5	0.14	-1.8	0.62
Coal's Proportion of Inputs	0.65	0.57	0.67	-0.096	0.3	0.58

TABLE 3: CHANGES IN CLUSTER INDUSTRIAL EMISSIONS BY POLLUTANT

This table reports the impact of CEPI reform on industrial emissions using data from EDGAR. The unit of analysis is at the cluster-yearmonth level. The dependent variable is the measurement of emissions from the database within a 5 kilometer radius circle around the centroid of the industrial cluster. In column 1, we focus on all pollutants whereas we break them down: $PM_{2.5}$ (column 2), PM_{10} (column 3) and NO_x (column 4). *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. All specifications include cluster-address fixed effects and State $\times$ yearmonth fixed effects. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year of 2008. The standard errors are clustered at the cluster-address level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: The Emissions Database for Global Atmospheric Research (EDGAR).

Dependent variable:	Emissions (milligrams per month)			
Pollutant(s):	All	$PM_{2.5}$	PM_{10}	NO_x
	(1)	(2)	(3)	(4)
Post $\times CEPI^{[70,100]}$ (β_1)	-7.109**	-3.489*	-7.669	-10.169*
	(3.225)	(1.813)	(4.748)	(5.937)
Post $\times CEPI^{[60,70)}$ (β_2)	-7.232**	-3.686*	-7.113	-10.898*
	(3.597)	(2.054)	(5.653)	(6.536)
p-value [$\beta_1 + \beta_2 = 0$]	0.033	0.058	0.145	0.092
2008 Dependent Variable Mean (Control)	23.09	16.86	38.95	13.45
Fixed effects:				
Cluster $\times$ Pollutant	Yes	Yes	Yes	Yes
State $\times$ year-month $\times$ Pollutant	Yes	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes	Yes
Adjusted-R^2	0.932	0.949	0.946	0.836
Observations	54,648	18,216	18,216	18,216

TABLE 4: IMPACT ON FIRM INPUTS

This table reports the changes in firm inputs around the 2009 CEPI emissions regulation. The unit of analysis is firm-product-year. The dependent variable in column 1 is the natural logarithm of the input energy value while in column 2 it is the proportion of electricity purchased (as opposed to produced) for each product. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. All specifications include firm and State $\times$ two-digit industry $\times$ year fixed effects. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

	Panel A: All Industries		
Dependent variable	Ln(Value Energy Input)	$\mathbb{1}_{\text{Coal Use}}$	Proportion Purchased Electricity
	(1)	(2)	(3)
Post $\times$CEPI$^{[70,100]}$ (β_1)	-0.818**	-0.301***	0.100**
	(0.294)	(0.092)	(0.036)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-1.006***	-0.289*	0.196***
	(0.219)	(0.150)	(0.059)
Ln(Production Quantity)	-0.208	0.033	-0.034
	(0.300)	(0.027)	(0.036)
p-value [$\beta_1 + \beta_2 = 0$]	0.000	0.019	0.001
2008 Dependent Variable Mean (Control)	8.906M	0.17	0.46
Fixed effects:			
Firm	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.795	0.496	0.786
Observations	901	565	901

Dependent variable	Panel B: Industries Split by High-Polluting vs. Others		
	Ln(Value Energy Input)	$\mathbb{1}_{\text{Coal Use}}$	Proportion Purchased Electricity
	(1)	(2)	(3)
Post $\times$CEPI$^{[70,100]}$ (β_1)	-0.464	-0.476***	0.023
	(0.464)	(0.143)	(0.065)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-0.539	-0.341	0.173
	(0.338)	(0.208)	(0.119)
Post $\times$CEPI$^{[70,100]}$ $\times$ High-Polluting (β_3)	-0.600	0.371	0.157*
	(0.673)	(0.238)	(0.079)
Post $\times$CEPI$^{[60,70)}$ $\times$ High-Polluting (β_4)	-1.100	-0.327	0.002
	(0.712)	(0.321)	(0.147)
Ln(Production Quantity)	-0.200	0.027	-0.036
	(0.292)	(0.025)	(0.036)
p-value [$\beta_1 + \beta_2 = 0$]	0.201	0.024	0.201
p-value [$\beta_3 + \beta_4 = 0$]	0.141	0.928	0.381
2008 Dependent Variable Mean (Control)	8.906M	0.17	0.46
Fixed effects:			
Firm	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.796	0.506	0.787
Observations	901	565	901

TABLE 5: IMPACT ON EMISSIONS

This table reports the changes in firm inputs around the 2009 CEPI emissions regulation. The unit of analysis is firm-product-year. Panel A reports the average effect across all industries whereas Panel B reports treatment effects split by High-Polluting vs. Other industries. Across both panels, the dependent variables are: Product-level emissions (column 1), Product-level emissions scaled by production quantity (column 2), and firm-level weight of the product that uses the highest proportion of coal in its energy input mix in the firm's overall product portfolio, measured in percentage points. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. All specifications include firm and State × two-digit industry × year fixed effects. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

	Panel A: All Industries		
Dependent variable:	Ln(Product CO_2 Emissions)	Ln(Per Unit CO_2 Emissions	Highest Coal Product Weight_{2008}
	(1)	(2)	(3)
Post $\times \text{CEPI}^{[70,100]}$ (β_1)	-0.944**	-0.687**	-0.139
	(0.346)	(0.270)	(0.114)
Post $\times \text{CEPI}^{[60,70)}$ (β_2)	-1.083***	-0.885***	-0.309**
	(0.283)	(0.306)	(0.123)
Ln(Production Quantity)	0.801**		
	(0.334)		
p-value [$\beta_1 + \beta_2 = 0$]	0.001	0.003	0.054
2008 Dependent Variable Mean (Control)	162,229.6	2.788	0.780
Fixed effects:			
Firm	Yes	Yes	Yes
State × industry × year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.893	0.774	0.855
Observations	901	901	705

	Panel B: Industries Split by High Polluting vs. Others		
Dependent variable:	Ln(Product CO_2 Emissions)	Ln(Per Unit CO_2 Emissions	Highest Coal Product Weight_{2008}
	(1)	(2)	(3)
Post $\times$CEPI$^{[70,100]}$ (β_1)	-0.515	-0.273	-0.007
	(0.526)	(0.646)	(0.102)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-0.591	-0.369	-0.060
	(0.391)	(0.469)	(0.090)
Post $\times$CEPI$^{[70,100]}$ $\times$ High-Polluting (β_3)	-0.750	-0.725	-0.175*
	(0.638)	(0.877)	(0.083)
Post $\times$CEPI$^{[60,70)}$ $\times$ High-Polluting (β_4)	-1.112	-1.196	-0.531*
	(0.779)	(0.945)	(0.262)
Ln(Production Quantity)	0.811**		
	(0.325)		
p-value [$\beta_1 + \beta_2 = 0$]	0.219	0.555	0.734
p-value [$\beta_3 + \beta_4 = 0$]	0.118	0.226	0.033
2008 Dependent Variable Mean (Control)	162,229.6	2.788	0.780
Fixed effects:			
Firm	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.893	0.775	0.862
Observations	901	901	705

TABLE 6: IMPACT ON ABATEMENT EXPENDITURES

This table reports the changes in firm-level abatement expenditures around the 2009 CEPI emissions regulation. The unit of analysis is firm-year. Panel A reports the average effect across all industries whereas Panel B reports treatment effects split by High-Polluting vs. Other industries. Across both panels, the dependent variable in column 1 is an indicator variable if the firm report environment and pollution control related expenses in that year while in column 2 is the intensive margin, defined as the ratio of expenses and total assets winsorized at 1% tails. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. For ease of interpretation, we multiply the coefficients by 100 in column 2. All specifications include firm and State $\times$ two-digit industry $\times$ year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

	Panel A: All Industries	
Dependent variable:	$\mathbb{1}_{\text{Abatment}}$ (1)	Abatement/Assets (2)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.077**	0.038**
	(0.029)	(0.016)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.048	0.039*
	(0.031)	(0.020)
p-value $[\beta_1 + \beta_2 = 0]$	0.039	0.027
2008 Dependent Variable Mean (Control)	0.048	0.014
Fixed effects:		
Firm	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes
Bandwidth	Yes	Yes
R^2	0.725	0.753
Observations	10,752	10,752

	Panel B: Industries Split by High Polluting vs. Others	
Dependent variable:	$\mathbb{1}_{\text{Abatment}}$ (1)	Abatement/Assets (2)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.071**	0.044**
	(0.031)	(0.017)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.046	0.052***
	(0.029)	(0.019)
Post $\times$CEPI$^{[70,100]}$ $\times$ High-Polluting (β_3)	0.026	-0.025***
	(0.019)	(0.007)
Post $\times$CEPI$^{[60,70)}$ $\times$ High-Polluting (β_4)	0.011	-0.046**
	(0.029)	(0.020)
p-value [$\beta_1 + \beta_2 = 0$]	0.055	0.005
p-value [$\beta_3 + \beta_4 = 0$]	0.352	0.004
2008 Dependent Variable Mean (Control)	0.048	0.014
Fixed effects:		
Firm	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes
Bandwidth	Yes	Yes
R^2	0.725	0.754
Observations	10,752	10,752

TABLE 7: IMPACT ON PRODUCT PORTFOLIO

This table reports the changes to firm-level product portfolios around the 2009 CEPI emissions regulation. The unit of analysis is firm-year. Panel A reports the average effect across all industries whereas Panel B reports treatment effects split by High-Polluting vs. Other industries. Across both panels, the dependent variable in column 1 (column 2) is the natural logarithm of the total quantity produced for each product (total number of products produced by a firm) in a year. The dependent variable in column 3 is an indicator for whether the firm added a product in a year while in column 4 it is an indicator for whether the firm dropped a product in that year. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. All specifications include firm and State $\times$ two-digit industry $\times$ year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

	Panel A: All Industries			
Dependent variable:	Ln(Product-level Production	Ln(No. of Products)	$\mathbb{1}_{\text{Add Product}}$	$\mathbb{1}_{\text{Remove Product}}$
	(1)	(2)	(3)	(4)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.030	0.007	-0.057*	0.023
	(0.130)	(0.072)	(0.034)	(0.030)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-0.110	0.013	-0.117***	0.003
	(0.182)	(0.078)	(0.041)	(0.036)
p-value [$\beta_1 + \beta_2 = 0$]	0.760	0.888	0.011	0.689
2008 Dependent Variable Mean (Control)	35,914.871	2.714	0.214	0.190
Fixed effects:				
Firm	Yes	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes	Yes
R^2	0.582	0.746	0.263	0.242
Observations	15,521	10,752	10,752	10,752

	Panel B: Industries Split by High-Polluting vs. Others			
Dependent variable:	Ln(Product-level Production	Ln(No. of Products)	$\mathbb{1}_{\text{Add Product}}$	$\mathbb{1}_{\text{Remove Product}}$
	(1)	(2)	(3)	(4)
Post $\times \text{CEPI}^{[70,100]}$ (β_1)	-0.008	0.015	-0.051	0.013
	(0.137)	(0.073)	(0.034)	(0.032)
Post $\times \text{CEPI}^{[60,70)}$ (β_2)	-0.331	0.003	-0.141***	-0.028
	(0.235)	(0.076)	(0.041)	(0.042)
Post $\times \text{CEPI}^{[70,100]}$ $\times$ High-Polluting (β_3)	0.090	-0.036*	-0.030	0.036*
	(0.105)	(0.019)	(0.025)	(0.019)
Post $\times \text{CEPI}^{[60,70)}$ $\times$ High-Polluting (β_4)	0.621***	0.025	0.073	0.107**
	(0.222)	(0.083)	(0.052)	(0.050)
p-value $[\beta_1 + \beta_2 = 0]$	0.266	0.899	0.006	0.835
p-value $[\beta_3 + \beta_4 = 0]$	0.009	0.904	0.471	0.012
2008 Dependent Variable Mean (Control)	35,914.871	2.714	0.214	0.190
Fixed effects:				
Firm	Yes	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes	Yes
R^2	0.583	0.746	0.263	0.243
Observations	15,521	10,752	10,752	10,752

TABLE 8: IMPACT ON REVENUE PRODUCTIVITY AND PROFITABILITY

This table reports the changes in firm profitability and revenue productivity around the 2009 CEPI emissions regulation. The unit of analysis is firm-year. Panel A reports the average effect across all industries whereas Panel B reports treatment effects split by High-Polluting vs. Other industries. Across both panels, the dependent variable in column 1 is the natural logarithm of total factor productivity estimated following Levinsohn and Petrin (2003) that controls for firm size. The dependent variable in column 2 is firm profitability defined as the ratio of firm earnings before interest, taxes, depreciation, and amortization (EBITDA) and sales winsorized at 1% tails, in column 3 it is the ratio of raw material expenses to firm-level net sales, and in column 4 it is the weight of the highest margin product in the firm's overall product portfolio, measured in percentage points. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. All specifications include firm and State × two-digit industry × year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

	Panel A: All Industries			
Dependent variable:	Ln(TFP)	EBITDA/ Sales	Raw Material Expense	Highest Margin Product Weight$_{2008}$
	(1)	(2)	(3)	(4)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.127***	0.008	-0.034	0.124***
	(0.039)	(0.014)	(0.027)	(0.046)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.100	0.004	-0.033	0.120**
	(0.075)	(0.015)	(0.030)	(0.050)
p-value [$\beta_1 + \beta_2 = 0$]	0.025	0.688	0.213	0.213
2008 Dependent Variable Mean (Control)	2.767	0.100	0.54	0.734
Fixed effects:				
Firm	Yes	Yes	Yes	Yes
State × industry × year	Yes	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes	Yes
R^2	0.851	0.638	0.641	0.880
Observations	10,752	10,752	10,752	7,995

	Panel B: Industries Split by High Polluting vs. Others			
Dependent variable:	Ln(Revenue Productivity)	EBITDA/ Sales	Raw Material Expense	Highest Margin Product Weight$_{2008}$
	(1)	(2)	(3)	(4)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.146***	0.009	-0.039	0.129***
	(0.043)	(0.015)	(0.027)	(0.047)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.131*	0.008	-0.061**	0.166***
	(0.074)	(0.015)	(0.030)	(0.053)
Post $\times$CEPI$^{[70,100]}$ $\times$ High-Polluting (β_3)	-0.076	-0.004	0.017	-0.015
	(0.054)	(0.007)	(0.013)	(0.017)
Post $\times$CEPI$^{[60,70)}$ $\times$ High-Polluting (β_4)	-0.114	-0.016	0.095***	-0.122**
	(0.161)	(0.011)	(0.032)	(0.058)
p-value [$\beta_1 + \beta_2 = 0$]	0.008	0.550	0.066	0.066
p-value [$\beta_3 + \beta_4 = 0$]	0.279	0.168	0.001	0.035
2008 Dependent Variable Mean (Control)	2.767	0.100	0.54	0.734
Fixed effects:				
Firm	Yes	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes	Yes
R^2	0.851	0.639	0.641	0.880
Observations	10,752	10,752	10,752	7,995

TABLE 9: IMPACT ON PRODUCT PROFITABILITY

This table reports the changes in product-level profitability around the 2009 CEPI emissions regulation. The unit of analysis is firm-product-year. Panel A reports the average effect across all industries whereas Panel B reports treatment effects split by High-Polluting vs. Other industries. Across both panels, the dependent variable in column 1 is the product-level profit margin computed as the difference between price and cost per unit as a fraction of price per unit, winsorized at 1% tails. The dependent variable in column 2 is the natural logarithm of product unit price computed as the ratio of total product sales to the total product quantity sold while in column 3 it is the natural logarithm of product unit cost computed as the ratio of product cost of goods sold to the total product quantity sold. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. The table also reports the *p*-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. All specifications include firm and State × two-digit industry × year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

	Panel A: All Industries		
Dependent variable:	Product Margins	Ln(Unit Price)	Ln(Unit Cost)
	(1)	(2)	(3)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.147***	-0.129	-0.221
	(0.054)	(0.220)	(0.197)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.037	-0.059	-0.016
	(0.081)	(0.225)	(0.194)
p-value [$\beta_1 + \beta_2 = 0$]	0.123	0.670	0.532
2008 Dependent Variable Mean (Control)	0.002	0.722	0.892
Fixed effects:			
Firm	Yes	Yes	Yes
State × industry × year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.722	0.592	0.599
Observations	15,225	15,984	15,225

	Panel B: Industries Split by High Polluting vs. Others		
Dependent variable:	Product Margins	Ln(Unit Price)	Ln(Unit Cost)
	(1)	(2)	(3)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.157***	-0.160	-0.255
	(0.052)	(0.220)	(0.200)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.018	-0.055	0.024
	(0.096)	(0.218)	(0.193)
Post $\times$CEPI$^{[70,100]}$ $\times$ High-Polluting (β_3)	-0.042	0.112	0.137
	(0.032)	(0.082)	(0.123)
Post $\times$CEPI$^{[60,70)}$ $\times$ High-Polluting (β_4)	0.043	0.003	-0.084
	(0.078)	(0.185)	(0.207)
p-value [$\beta_1 + \beta_2 = 0$]	0.167	0.615	0.540
p-value [$\beta_3 + \beta_4 = 0$]	0.988	0.591	0.842
2008 Dependent Variable Mean (Control)	0.002	0.722	0.892
Fixed effects:			
Firm	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.722	0.592	0.599
Observations	15,225	15,984	15,225

TABLE 10: CHANGES IN CLUSTER-LEVEL FIRM ENTRY

This table reports changes in cluster-level firm entry around the 2009 CEPI emissions regulation. The unit of analysis is cluster-industry-year. Panel A reports changes in firm entry using the universe of business registration from the Ministry of Corporate Affairs while Panel B reports changes in firm entry from CMIE Prowess. Across both panels, the dependent variable in column 1 is an indicator for whether atleast one manufacturing firm incorporates in the cluster in a given industry in the year. The dependent variable in column 2 is one plus the natural logarithm of the number of newly registered manufacturing firms in that year while in column 3 it is the inverse hyperbolic sine of the number of newly registered manufacturing firms in that year. Column 4 uses the raw number of newly registered manufacturing firms in each cluster in a specific industry in that year. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. In both panels, columns 1 through 3 are estimated using Ordinary Least Squares (OLS) while column 4 is estimated using Pseudo-Poisson Maximum Likelihood (PPML). All specifications include cluster, two-digit Industry × year, and State × year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: Ministry of Corporate Affairs and CMIE Prowess.

	Panel A: All firms from business registry			
Dependent variable:	$\mathbb{1}_{\text{New Firm}}$	Ln(No. of firms)	*asinh*(No. of firms)	No. of firms
	(1)	(2)	(3)	(4)
Post $\times$CEPI$^{[70,100]}$ (β_1)	-0.018*	-0.016*	-0.020*	-0.185*
	(0.010)	(0.009)	(0.012)	(0.104)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-0.009	-0.011	-0.014	-0.105
	(0.011)	(0.010)	(0.013)	(0.132)
p-value [$\beta_1 + \beta_2 = 0$]	0.158	0.135	0.138	0.163
2008 Dependent Variable Mean (Control)	0.080	0.202	0.202	0.202
Fixed effects:				
Firm	Yes	Yes	Yes	Yes
State × industry × year	Yes	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes	Yes
R^2	0.449	0.570	0.570	
Observations	33,534	33,534	33,534	19,958

	Panel B: Large firms in CMIE Prowess			
Dependent variable:	$\mathbb{1}_{\text{New Firm}}$	Ln(No. of firms)	*asinh*(No. of firms)	No. of firms
	(1)	(2)	(3)	(4)
Post $\times$CEPI$^{[70,100]}$ (β_1)	-0.041*	-0.035*	-0.045*	-0.795**
	(0.021)	(0.018)	(0.023)	(0.370)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-0.003	0.001	0.001	-0.289
	(0.017)	(0.016)	(0.021)	(0.440)
p-value [$\beta_1 + \beta_2 = 0$]	0.229	0.231	0.230	0.149
2008 Dependent Variable Mean (Control)	0.011	0.011	0.011	0.011
Fixed effects:				
Firm	Yes	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes	Yes
R^2	0.411	0.439	0.440	
Observations	4,416	4,416	4,416	678

INTERNET APPENDIX
FOR ONLINE PUBLICATION

Appendix A Additional figures and tables

TABLE IAA1: SUMMARY BY HIGHLY POLLUTING INDUSTRY STATUS

The table presents summary statistics separately for firms in Highly Polluting Industries, as defined by the CPCB. Panel A summarizes the firm-year panel dataset. *Assets* and *Sales* is in thousands of INR. *Leverage* is the sum of short- and long-term debt obligations scaled by contemporaneously reported Total Assets. *Tangibility* is property plant and equipment less deprecation scaled by contemporaneously reported Total Assets. *Exporting intensity* are firm earnings from exports of goods plus services scaled by contemporaneous total sales. *Market capitalization* is reported for listed firms and is the total number of listed shares multiplied by the share price at the end of the fiscal year. *Log(Productivity)* is the natural log of firm productivity, which is calculated following Levinsohn and Petrin (2003) and controls for firm size. *Profitability* are Earnings Before Interest, Taxes, Depreciation and Amortization as a ratio of the prior year sales. Panel B summarizes the firm-product-year dataset. *Log(Product Sales)* and *Log(Product COGS)* is total product sales and production cost, respectively. *Unit Price* is the per-unit price (where unit is unique within but not across firm). *Margin* is (unit price - unit cost)/unit price. *Log(Per Unit CO_2 Emissions)* are the author-calculated CO_2 emissions per reported unit of production.

	Panel A: Firm characteristics					
	All	Not HPI	HPI	Difference	Above × HPI Estimate	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
Assets (000 INR)	3,334	2,919	4,550	-1,631	494	0.8
Sales (000 INR)	3,108	2,694	4,321	-1,626	333	0.81
Leverage	0.27	0.27	0.28	-0.014	-0.028	0.45
Tangibility	0.27	0.27	0.28	-0.014	0.016	0.64
Exporting Intensity	16	18	11	7.2	-0.56	0.89
Market Capitalization	22,092	18,302	32,952	-14,650	50,744	0.042
Log(Revenue Productivity)	3.1	3.1	3	0.18	0.35	0.29
Number of Products	2.8	2.7	3.3	-0.63	-0.1	0.86
Profitability	0.1	0.11	0.095	0.012	0.0053	0.73

	Panel B: Firm-product characteristics					
	All	Not HPI	HPI	Difference	Above × HPI Estimate	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Product Sales)	4.2	4.1	4.4	-0.29	0.6	0.13
Ln(Unit Cost)	-5.0	-5.4	-3.9	-1.6	-0.51	0.55
Ln(Unit Price)	-4.9	-5.4	-3.8	-1.6	-0.42	0.61
Margin	-0.53	-0.55	-00.48	-0.075	-1.1	0.35
Ln(Per Unit CO_2 Emissions)	-1.7	-1.8	-1.6	-0.27	-1.9	0.4
Coal's Proportion of Inputs	0.58	0.65	0.49	0.17	0.59	0.042

TABLE IAA2: CLUSTER-LEVEL COVARIATE BALANCE

The table presents mean values for baseline city characteristics, as recorded in Population Census. Column 1 presents the unconditional mean while column 2 (column 3) presents mean for clusters below (above) the treatment threshold. Column 4 presents the difference in means between cities below the treatment threshold and cities above the treatment threshold. Additionally, column 5 shows the regression discontinuity estimate, following the main estimating equation, of the effect of being above the treatment threshold on the baseline variable and column 6 is the *p*-value for this estimate, using heteroskedasticity robust standard errors. Data sources: Population Census and Harari (2020).

	All	Below	Above	Difference	Estimate	*p*-value
	(1)	(2)	(3)	(4)	(5)	(6)
City roads, km, 1981	337.206	268.936	391.822	-122.886	-297.226	0.486
Log(population), 2001	13.330	13.015	13.572	-0.556	0.403	0.693
Population density (000 per Sq. km), 2001	8.632	9.387	7.993	1.394	-1.081	0.802
Average rent (per Sq. m.), 2008	953.696	907.779	990.430	-82.651	356.072	0.422
Proximity index, 2008	0.071	0.003	0.116	-0.113	-0.004	0.970
Nearest waterway (km), 2008	13.901	17.660	10.948	6.713	-16.694	0.182
Potential yields (tons/ha), 2008	1.440	1.485	1.406	0.079	0.111	0.123
Diameter from center (km), 2008	4.863	3.866	5.706	-1.840	2.223	0.514
Area footprint (Sq. km.), 2008	187.831	114.277	250.069	-135.792	184.250	0.485

TABLE IAA3: IMPACTS ON FINE PARTICULATE MATTER ($PM_{2.5}$)

This table reports the impact of CEPI reform on fine particulate matter using data from Van Donkelaar, Martin, Spurr, and Burnett (2015). The unit of analysis is at the cluster-yearmonth level. The dependent variable is the measurement of fine particulate matter ($PM_{2.5}$) in μm^3In column 1, we focus on measurements within a 5 kilometer radius circle around the centroid of the industrial cluster while in column 2, we focus on measurements within a 500 meter radius circle. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. All specifications include cluster-address fixed effects and State $\times$ year-month fixed effects. The table also reports the *p*-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year of 2008. The standard errors are clustered at the cluster-address level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: Van Donkelaar, Martin, Spurr, and Burnett (2015).

Dependent variable:	Fine $PM_{2.5}$ (μ g/m^3)	
Radii of circle:	5 kilometers	500 meters
	(1)	(2)
Post $\times$CEPI$^{[70,100]}$ (β_1)	-2.311***	-1.893**
	(0.775)	(0.743)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-1.018	-0.560
	(0.756)	(0.673)
p-value [$\beta_1 + \beta_2 = 0$]	0.025	0.069
2008 Dependent Variable Mean (Control)	84.0	84.0
Fixed effects:		
Cluster	Yes	Yes
State $\times$ year-month	Yes	Yes
Bandwidth	Yes	Yes
R^2	0.963	0.959
Observations	17,952	18,216

TABLE IAA4: CHANGES IN CLUSTER ENERGY EMISSIONS BY POLLUTANT

This table reports the impact of CEPI reform on the power generation sector using data from EDGAR. The unit of analysis is at the cluster-year-month level. The dependent variable is the measurement of emissions from the database within a 5 kilometer radius circle around the centroid of the industrial cluster. In column 1, we focus on all pollutants whereas we break them down: $PM_{2.5}$ (column 2), PM_{10} (column 3) and NO_x (column 4). *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. All specifications include cluster-address fixed effects and State $\times$ year month fixed effects. The table also reports the *p*-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year of 2008. The standard errors are clustered at the cluster-address level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: The Emissions Database for Global Atmospheric Research (EDGAR).

Dependent variable:	Pollution Measurement			
Pollutant(s):	All	$PM_{2.5}$	PM_{10}	NO_x
	(1)	(2)	(3)	(4)
Post $\times CEPI^{[70,100]}$ (β_1)	-0.169	-0.181	-0.184	-0.143
	(0.755)	(0.304)	(0.549)	(1.520)
Post $\times CEPI^{[60,70)}$ (β_2)	-0.229	-0.112	-0.170	-0.405
	(0.715)	(0.274)	(0.542)	(1.415)
p-value [$\beta_1 + \beta_2 = 0$]	0.770	0.582	0.719	0.843
2008 Dependent Variable Mean (Control)	8.18	1.78	3.34	19.43
Fixed effects:				
Cluster $\times$ Pollutant	Yes	Yes	Yes	Yes
State $\times$ year-month $\times$ Pollutant	Yes	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes	Yes
Adjusted-R^2	0.756	0.795	0.823	0.734
Observations	29,808	9,936	9,936	9,936

TABLE IAA5: CHANGES IN FIRM-LEVEL FACTORS OF PRODUCTION

This table reports the changes in factors of production around the 2009 CEPI emissions regulation. The unit of analysis is firm-year. Panel A reports the average effect across all industries whereas Panel B reports treatment effects split by High-Polluting vs. Other industries. Across both panels, the dependent variable in column 1 is the total wage bill as a fraction of net sales while in column 2 it is the ratio of raw material expenses to firm-level net sales. In column 3, the dependent variable is investment, defined as the ratio of gross fixed assets to total assets. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. All specifications include firm and State $\times$ two-digit industry $\times$ year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

	Panel A: All Industries		
Dependent variable:	Wage Bill	Raw Material Exp.	Investment
	(1)	(2)	(3)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.001	-0.034	0.020
	(0.007)	(0.027)	(0.024)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-0.005	-0.033	0.019
	(0.009)	(0.030)	(0.031)
p-value [$\beta_1 + \beta_2 = 0$]	0.777	0.213	0.429
2008 Dependent Variable Mean (Control)	0.05	0.54	0.89
Fixed effects:			
Firm	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.638	0.641	0.826
Observations	10,752	10,752	9,643

	Panel B: Industries Split by High Polluting vs. Others		
Dependent variable:	Wage Bill	Raw Material Exp.	Investment
	(1)	(2)	(3)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.002	-0.039	0.018
	(0.007)	(0.027)	(0.024)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-0.004	-0.061**	0.028
	(0.011)	(0.030)	(0.032)
Post $\times$CEPI$^{[70,100]}$ $\times$ High-Polluting (β_3)	-0.003	0.017	0.009
	(0.008)	(0.013)	(0.017)
Post $\times$CEPI$^{[60,70)}$ $\times$ High-Polluting (β_4)	-0.003	0.095***	-0.027
	(0.010)	(0.032)	(0.032)
2008 Dependent Variable Mean (Control)	0.05	0.54	0.89
Fixed effects:			
Firm	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.638	0.641	0.826
Observations	10,752	10,752	9,643

TABLE IAA6: SHIFTING PRODUCTION: NO IMPACT ON MERGERS AND ACQUISITIONS

This table reports changes in mergers and acquiisitions around the 2009 CEPI emissions regulation. The unit of analysis is firm-year. The dependent variable in column 1 is an indicator for whether the firm in a given cluster was a target in a merger and acquisition in that year while in column 2 it is an indicator for whether the firm in a given cluster was acquired in a merger and acquisition in that year. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. All specifications include firm and State × two-digit Industry × year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

Dependent variable:	$\mathbb{1}_{\text{Target}}$	$\mathbb{1}_{\text{Acquired}}$
	(1)	(2)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.009	0.005
	(0.009)	(0.007)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.018	-0.000
	(0.012)	(0.008)
p-value [$\beta_1 + \beta_2 = 0$]	0.165	0.757
2008 Dependent Variable Mean (Control)	0.000	0.000
Fixed effects:		
Firm	Yes	Yes
State × industry × year	Yes	Yes
Bandwidth	Yes	Yes
R^2	0.193	0.148
Observations	10,752	10,752

TABLE IAA7: SHIFTING PRODUCTION:NO IMPACT ON PLANT ANNOUNCEMENTS

This table reports changes in the probabilities of plant announcements around the 2009 CEPI emissions regulation. The unit of analysis is firm-year. The dependent variable in column 1 is an indicator for whether the firm announced a new plant in the year while in column 2 it is an indicator for whether the firm announced that it is abandoning a new plant that it had announced in the past years. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. The table also reports the p-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. All specifications include firm and State $\times$ two-digit industry $\times$ year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data sources: CMIE Prowess and CapEx.

Dependent variable:	$\mathbb{1}_{\text{New Plant}}$	$\mathbb{1}_{\text{Abandon Plant}}$
	(1)	(2)
Post $\times$CEPI$^{[70,100]}$ (β_1)	-0.010	-0.004
	(0.011)	(0.010)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.008	0.003
	(0.013)	(0.011)
p-value $[\beta_1 + \beta_2 = 0]$	0.925	0.990
2008 Dependent Variable Mean (Control)	0.0	0.0
Fixed effects:		
Firm	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes
Bandwidth	Yes	Yes
R^2	0.350	0.284
Observations	10,752	10,752

TABLE IAA8: VARIABLE DEFINITIONS

Notes: EDGAR refers to the 'Emissions Database for Global Atmospheric Research' data, available at https://edgar.jrc.ec.europa.eu/dataset_htap_v3. PROWESS refers to the 'Performance and Ownership with Excellence' available at https://prowessdx.cmie.com/. CBCB refers to the 'Central Pollution Control Board', with supporting data available at https://cpcb.nic.in/.

Variable	Description	Data Source
Panel A: Pollution		
All Pollution	Summation of pollution measures for a given area	EDGAR
$PM_{2.5}$	Particulate Matter with a diameter of 2.5 micrometers or less, measured in Mg/month	EDGAR
PM_{10}	Particulate Matter with a diameter of 10 micrometers or less, measured in Mg/month	EDGAR
NO_x	Nitrous oxides, measured in Mg/month	EDGAR
$PM_{2.5}$	Fine Particulate Matter with a diameter of 2.5 micrometers or less, measured in Mg/month	Van Donkelaar et al. (2015)
Panel B: Firm Characteristics		
Assets (million INR)	Total assets of firm operations in INR	PROWESS
Sales (million INR)	Total revenue from goods and services sold in missions of INR	PROWESS
Leverage	The sum of short- and long-term debt obligations scaled by contemporaneously reported Total Assets	PROWESS
Tangibility	The property plant and equipment less deprecation scaled by contemporaneously reported Total Assets	PROWESS
Exporting intensity	Firm earnings from exports of goods plus services scaled by contemporaneous total sales	PROWESS
Market capitalization (million INR)	The total number of listed shares multiplied by the share price at the end of the fiscal year	PROWESS
Ln(Productivity)	The natural log of firm productivity, which is calculated following Levinsohn and Petrin (2003) and controls for firm size	PROWESS
Profitability	Earnings Before Interest, Taxes, Depreciation, and Amortization as a ratio of the prior year sales	PROWESS
Number of Products	Number of unique products for a given firm in a given year	PROWESS
$\mathbb{1}_{\text{File Energy Inputs}}$	Indicator for whether the firm reports inputs	PROWESS
$\mathbb{1}_{\text{New Plant}}$	Indicator for whether the firm announced a new plant in the year	CapEx
$\mathbb{1}_{\text{Abandon Plant}}$	Indicator for whether the firm announced that it abandoned the plant in the year	CapEx
Panel C: Product Characteristics		
Ln(Product Sales)	The natural logarithm of the per-product sales	PROWESS
Ln(Product COGS)	The natural logarithm of the per-product cost of goods sold (COGS)	PROWESS
Unit Price	The natural logarithm of the per-unit price, where unit is unique within but not across firm	PROWESS
Margin	Is measured as (unit price - unit cost)/unit price	PROWESS
Ln(Per Unit CO_2 Emissions)	Author-calculated CO_2 emissions per reported unit of production	PROWESS
TFPQ	Natural logarithm of quantity-based total factor productivity estimated following De Loecker, Goldberg, Khandelwal, and Pavcnik (2016)	PROWESS
TFPR	Natural logarithm of total factor productivity estimated following Levinsohn and Petrin (2003) that controls for firm size	PROWESS
Panel D: Cluster Characteristics		
$CEPI^{[70,100]}$	CEPI equal or greater than 70, and less or equal to 100	PROWESS and CPCB
$CEPI^{[60,70)}$	CEPI equal or greater than 60, and less than 70	PROWESS and CPCB

Appendix B Additional background on the regulation

By 2009, India was an acknowledged industrial powerhouse. However, significant environmental degradation accompanied impressive growth. This pollution concentrated in industrial clusters, which shared infrastructure, administrative structures, and proximity to major population centers made desirable locations for manufacturing and industrial production. District and state authorities have regulated industrial cluster emissions since environmental regulation began in the 1980s. However, enforcement has been uneven, emissions measurements and regulatory thresholds were not standardized, and firms were often allowed to self-monitor, rather than be subjected to independent auditors. Moreover, the government lacked even basic information on industrial environmental impact for most locations.

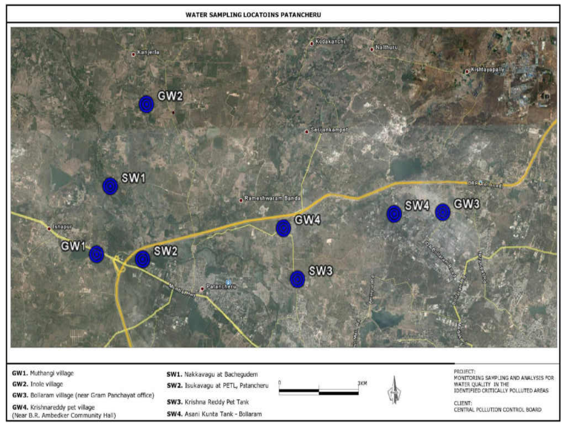
Against this background, the Central Pollution Control Board (CPCB) of the Ministry of Environment, Forest and Climate Change conducted a comprehensive environmental assessment of industrial clusters. The aims were to enhance, standardize and centralize pollution monitoring. The first step was to design a measure of pollution: The Comprehensive Environmental Pollution Index (CEPI hereafter). Figure 1 describes its construction. The CEPI combines proxies for (1) the amount and toxicity of pollutants, (2) the potential impact of that pollution on humans and ecosystems, and (3) an assessment of the quality of actions already taken by cluster firms to capture or adequately dispose of emissions. We include a complete discussion of each component and its construction as of the 2009 regulation in Central Pollution Control Board of India (2009).

Of the over two thousand industrial clusters in India, the CPCB reported CEPI scores for the 88 worst-polluting clusters. The CPCB classified those clusters with a CEPI above 60 as Severely Polluted Areas (SPA). These became subject to central monitoring at the national level rather than the relatively weak local control. Moreover, the CPCB classified industrial clusters with a CEPI of at least 70 as Critically Polluted Areas (CPA), which were additionally mandated to submit a remedial action plan for approval detailing the actions and timelines at the cluster and firm levels.

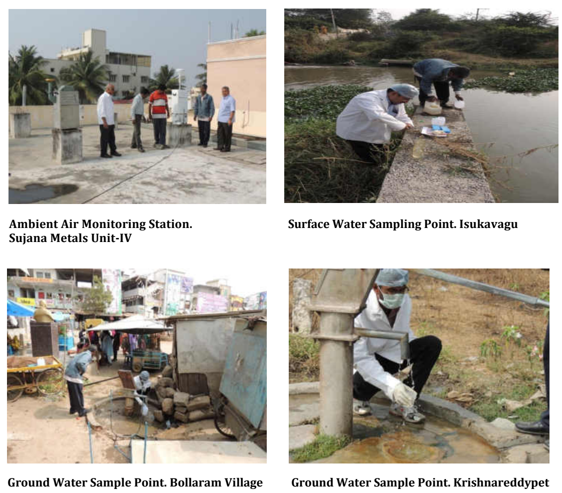
If a firm within a Critically Polluted Area failed to comply with the directives of the action plan, then they would lose their Environmental Clearance and Consent to Operate permits that allow firms to function within the formal economy. Moreover, Consent to Establish permits could not be issued to new operations if they do not fully comply with the cluster regulations and action plans.

Consider the example of the action plan for the CPA Patancheru-Bollaram Cluster in Andhra Pradesh (Pollution Control Board of Andhra Pradesh (2010)), which contains the operations of 106 establishments and whose CEPI, at 70.07, was just over the cutoff between being classified as a Severely or Critically Polluted Area. The Action Plan specifies a lengthy list of specific actions and deadlines agreed to by the firms of the cluster. For example, the cluster agreed to build a common effluent treatment plant, a treatment storage and disposal facility, and alternative drainage systems so no firm would outlet emissions into significant water bodies. In addition, firms operating in specific high-polluting industries would no longer be allowed to expand, and new firms in these industries could not be established in the cluster. The plan also listed self-policing mechanisms the cluster agreed to in order to prevent illegal dumping. In addition, the cluster agreed to pay compensation to local farmers affected by pollution and to supply drinking water to affected villages. The action plan then details a long list of agreed investments and recorded progress for each of the 106 individual establishments in the cluster.

The CPCB also installed continuous remote pollution sensors for air, water, and land pollution, video cameras (including night cameras) on the premises of factories at the point of their process emissions, and instigated tri-annual CPCB audits (January-February, May-June, and September-October) and quarterly audits from district and state level monitoring committees. These reports were released to the public annually via the CPCB website. Each report specifies the longitude and latitude of the air and water sampling locations, the laboratories used to carry out sampling and analyze the samples, and the date of the sampling. Then for each of air, groundwater and surface water samples the report specifies a particular pollutant (e.g., lead) or parameter (e.g., color (Hazen units)), the measurements of that pollutant and the test method used. Finally, the report includes photographs of the measurements. To give readers an idea, Figure IAB1 provides an example of the monitoring documentation for the Patancheru-Bollaram Cluster in Andhra Pradesh. Panel (a) reports the locations of water sampling locations superimposed over a map of the cluster. Panel (b) reproduces a few of the sampling documentation photographs of air, surface and groundwater sampling in the Patancheru-Bollaram Cluster.



(a) Water sampling locations



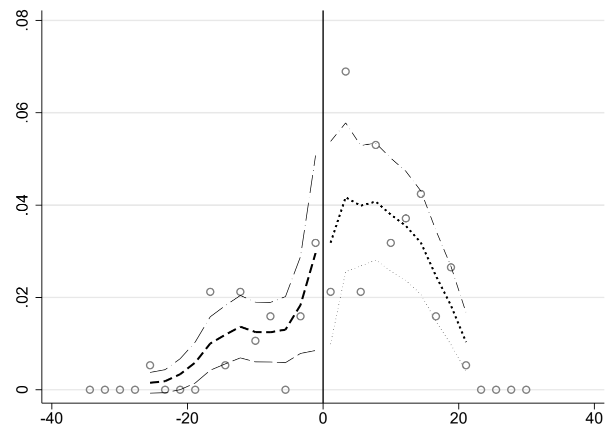
(b) Sampling documentation photos

FIGURE IAB1: PANTANCHERU-BOLLARAM CLUSTER POLLUTION MONITORING

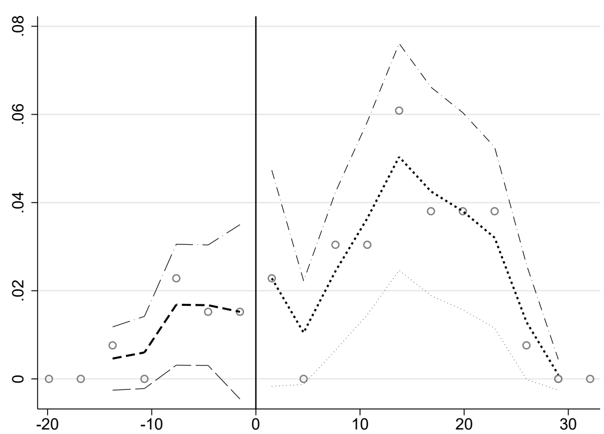
Notes: In Panel (a), blue dots signify water collection sites. Source: CPCB annual reports, "Sampling and Analysis of Ambient Air Quality and Water Quality in Industrial/Cluster Areas."

Finally, we complement the analysis presented in Figure 3 testing for manipulation of the original, 2009 CEPI with complementary McCrary (2008) density tests for the re-calculated CEPI values in 2011 and 2013. Note that in 2011 and 2013 only the CEPI values for those firms that had been ranked CPA in 2009 were published, so we are testing for manipulation in these updated scores among the 2009 CPA sub-population.

In summary, in India, as elsewhere, holding firms accountable for their environmental impact was difficult with unreliable data and weak enforcement. Accordingly, the CPCB centralized control and broadened its scrutiny of emissions. It automated monitoring to the extent possible, increased auditor independence, and instigated overlapping monitoring regimes. Moreover, the CPCB increased engagement Well-defined cutoffs defined escalating severity of regulatory scrutiny.



(a) 2011 CEPI



(b) 2013 CEPI

FIGURE IAB2: TEST FOR MANIPULATION OF THE CEPI IN 2011 AND 2013

Notes: Panel (a) reports the fitted distribution of the 2011 CEPI update around the cutoff at 70 (normalized to zero) for clusters with a 2009 CEPI of at least 70 (Critically Polluted Areas). Panel (b) reports the fitted distribution of the 2013 CEPI update around the (updated) cutoff atf 60 (normalized to zero) for the same sample. Source: CPCB.

Appendix C Measuring cluster-level pollution

The appendix provides a comprehensive overview of the methodology employed to compile the pollution data panel for each industrial site. We first discuss the Emission Database for Global Atmospheric Research (EDGAR) data, followed dataset by Van Donkelaar, Hammer, Bindle, Brauer, Brook, Garay, Hsu, Kalashnikova, Kahn, Lee, et al. (2021) (henceforth, Van Donkellaar). Finally, we detail the steps undertaken to construct the data panel.

C.1 Emission Database or Global Atmospheric Research (EDGAR)

Our primary pollution data comes from the Emission Database or Global Atmospheric Research (EDGAR), with particular emphasis on the Hemispheric Transport of Air Pollution ($HTAP_{v3}$) mosaic. This mosaic is designed to enhance the temporal range, sectoral breakdown, and geographical coverage of existing official data.[32] We use pollution data for nitrous oxide (NO_x), particles less than 2.5 µm in diameter ($PM_{2.5}$), and particles less than 10 µm in diameter (PM_{10}), are available, and each is processed distinctly. The monthly data with the highest resolution (0.1°x 0.1°) is downloaded. Upon reading this raster file, we keep only the industrial pollution layer, given its relevance to our study.

C.2 Fine Particulate Matter ($PM_{2.5}$) from Van Donkelaar et al. (2021)

We also use data from Van Donkelaar et al. (2021), as it offers monthly high-resolution (0.01° x 0.01° grid) estimates of ground-level fine particulate matter ($PM_{2.5}$). These pollution estimates are calculated by merging Aerosol Optical Depth (AOD) data from NASA's MODIS, MISR, and SeaWiFS instruments with outputs from the GEOS-Chem chemical transport model. The dataset is refined through calibration with global ground-based observations via Geographically Weighted Regression (GWR).

C.3 Measurement Procedure

The process of measuring pollution data at the industrial cluster level is broken down into four steps.

1. Extract the cluster location from the PDF titled "Assessment of the Need from Common Effluent Treatment Plants."
2. Geocode each identified location and construct corresponding circles around industrial areas/estate locations.
3. Using the location and pollution data from the previous step, we compute the weighted overlap between the designated circular region and the pollution raster layer.

(1) Extraction of Industrial Clusters: We use the document "Assessment of the Need from Common Effluent Treatment Plants," which is published by the CPCB under the Ministry of Environmental & Forest, Govt. of India. Starting from page 22 in Annexure II, it presents a list of industrial areas and estates of new locations categorized by state. Using this document, we extract these addresses into Excel using PDF converters. Given the document's inconsistencies, research assistants meticulously review the output by hand to guarantee accurate extraction.

(2) Geocoding and Shape Construction: Next, we pinpoint the latitude and longitude of industrial clusters. We send each address to the Google Maps API, and retrieve their geocodes. This helps exclude duplicate locations, proposals that weren't realized, and entries with incomplete information. After this step, we are left with 2914 locations. Using this refined list, a geometry, specifically a circle with a 500m radius, is constructed around each location.

(3) Weight Pollution Data: The final step involves calculating the pollution at each industrial location. For this purpose, we utilize raster files from EDGAR and Van Donkelaar, as discussed above, to assess pollution levels surrounding these sites.

[32] See https://edgar.jrc.ec.europa.eu/dataset_htap_v3 for more information.

A critical aspect to consider is that the vicinity of an industrial location can span multiple grid cells. To account for this, we calculate a weighted average of the pollution values. This involves determining the proportion of the industrial area's footprint overlapping each grid cell, which then serves as the weight for that cell. By summing these weighted values across the industrial area, we obtain a comprehensive dataset detailing pollution levels by industrial area, month, and pollutant type.

Annexure II

List of Industrial Areas /Estates

State/UT: Haryana

	Ambala Cant
1	HSIIDC Ambala
2	IGC Food Park, Phase-I, Saha
3	IGC Phase-II, Saha
	Bhiwani
4	HUDA Sec-21
5	HUDA Sec-21

FIGURE IAC3: SAMPLE FROM THE ASSESSMENT OF THE NEED FROM COMMON EFFLUENT TREATMENT PLANTS

Notes: The figure presents an excerpt from the Assessment of the Need from Common Effluent Treatment Plants document. It presents the first 5 observations from page 22.

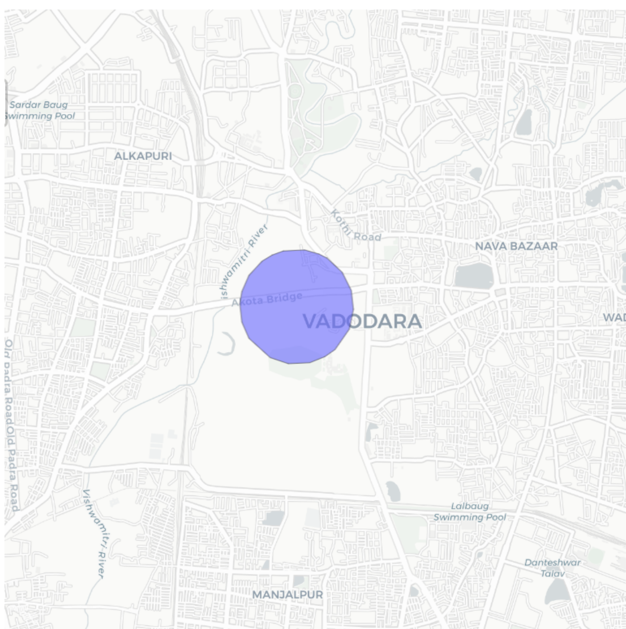


FIGURE IAC4: VADODARA, GUJARAT

Notes: The figure presents a shape drawn for a given industrial area/estate using a 500m radius.

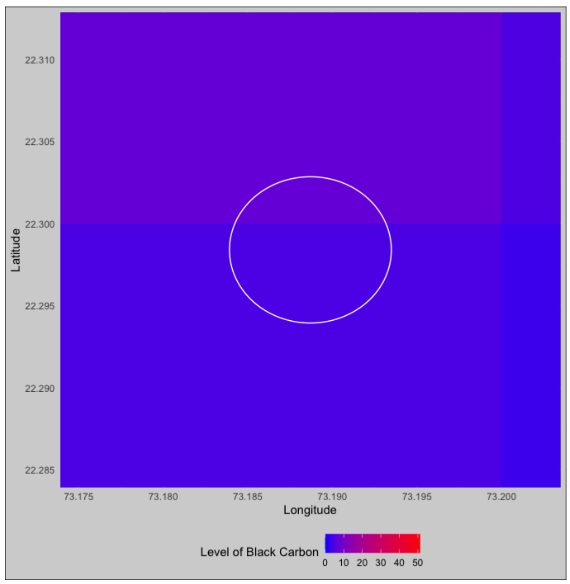


FIGURE IAC5: EXAMPLE OF INTERSECTION

The figure illustrates pollution calculations from December 2012, depicted as a circle divided into two segments: the top segment represents 27%, and the bottom segment represents 73%, indicating the proportional weight calculations

Appendix D Product-Level CO_2 Emissions

This appendix describes how we clean the product-level energy inputs from the Prowess database and transform them into product-level CO_2 emissions. First, a word about the source of this data. The Prowess product-wise energy consumption data are from company disclosures in their annual reports. Clause (e) of sub-section (1) of section 217 of the Companies Act of 1956 mandates that every company disclose *total* energy consumption in a prescribed format. However, there is no legal obligation to disclose the product-level energy consumption per unit of production. Thus, a limitation of this data is that firms choose whether or not to disclose it, and not every firm chooses to do so. Of the 1,001 manufacturing firms in our regression sample, 230 report product-level energy inputs in at least one year between 2007 and 2013. Once a firm starts reporting product-level energy inputs, however, it tends to continue to do so through the entire period. Note that this changes the interpretation of our results to be most directly applicable to these types of firms but is unlikely to violate the identification assumption that there is no discontinuity in the probability of reporting product-level energy assumption at CEPI treatment thresholds. Figure IAD6 supports the reasonableness of this identification assumption.

TABLE IAD9: PROBABILITY OF FILING ENERGY INPUTS

This table reports the effect of the 2009 CEPI emissions regulation on the probability of reporting product-level energy inputs in firm annual reports. The unit of analysis is firm-year. Model (1) is on all firms in the Prowess database. Models (2) and (3) are on the regression dataset comprising manufacturing firms in clusters with CEPI within a bandwidth of 10 pollution index units around the cutoffs at 70 and 60. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. All specifications include firm and State × two-digit Industry × year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

Dependent variable:	$\mathbb{1}_{\text{File Energy Inputs}}$		
Sample:	All	Regression	
	(1)	(2)	(3)
Post	-0.007***		
	(0.001)		
Post $\times$CEPI$^{[60,100]}$		0.004	
		(0.004)	
Post $\times$CEPI$^{[70,100]}$			0.005
			(0.004)
Post $\times$CEPI$^{[60,70)}$			-0.000
			(0.010)
Fixed effects:			
Firm	Yes	Yes	Yes
State × industry × year	No	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.408	0.448	0.448
Observations	119,943	19,688	19,688

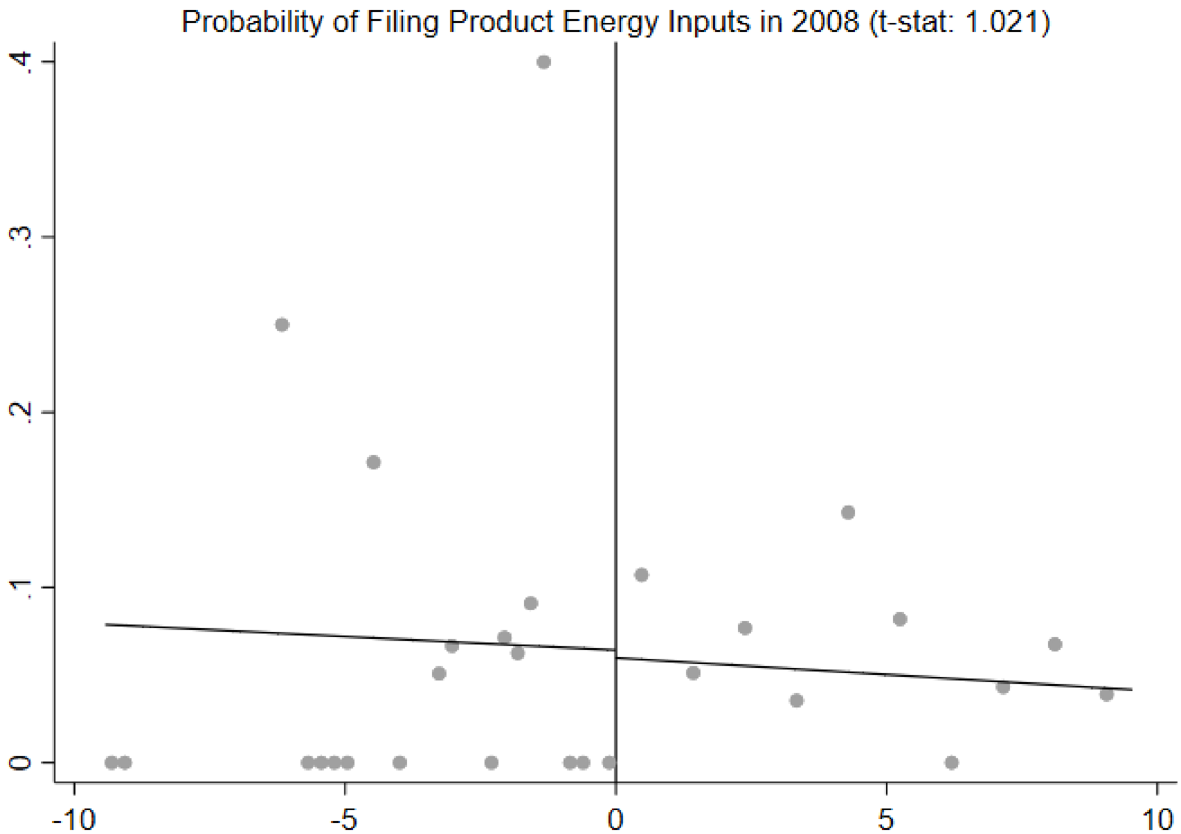


FIGURE IAD6: DISCONTINUITY IN THE PROBABILITY OF FILING PRODUCT ENERGY INPUTS AT BASELINE

Notes: The figure presents the average probability that firm reports product-level energy inputs in its annual statement in 2008 around the CEPI treatment thresholds. We pool across a 10 CEPI-index value window around the two thresholds at CEPI 60 and 70, normalized in the figure to zero.

Now that we are clear about the limitations of this data let us note its advantages. To our knowledge, it this dataset offers unique access product-level energy inputs for such a large cross-section of firms. We exploit this unique data to bring new insight into how emission-capping regulations impact production decisions along the input dimension.

The data are at the firm-product-year-energy source level and are expressed in energy input units per reported production unit. For example, A.B.G. Cement Ltd. reported using 70.28 kWh of purchased electricity, 0.14 tonnes of coal, and 3.3 KWh of firm-produced electricity from a diesel generator per tonne of cement produced in 2014. Since regulators do not mandate a particular reporting standard, there exists a lot of variation in reporting units in the raw data. We therefore first separate the energy and production units and then standardize them. For example, we transform all production units reported in "lakh liters" into "liters" by using the fact that one lakh liter is 100,000 liters. Ultimately, we express all energy inputs in kcal per production unit. This conversion allows us to test for shifts in energy use across energy sources as a proportion of the total energy input in kcal.

Next, we transform energy input into CO_2 output. This exercise requires assumptions about each energy source's energy content and CO_2 output. We use the conversion factors and unit assumptions from the Central Electricity Authority (CEA) of India for 2008, the year before the regulation.CITE This choice fixes the energy technology just before the regulation. We assume that in the three-year window around the regulation there are not drastic changes to technology that would change the CO_2 emissions of each fuel type significantly.

Specifically, we use the CEA's assumptions on gross calorific value and CO_2 emission factors per fuel source that this regulator mandated that electricity plants use to quantify their CO_2 emissions in 2008 (Central Electricity Authority, 2008). We supplement this source from the 2008 *Commercial Energy Balance Tables and Conversion Factors* from the Energy and Resources Institute (Energy and Resources Institute, 2008). The latter gives us fuel-specific conversions between, e.g., mass units and volume units for the type of fuel used in Indian manufacturing firms. Table IAD10 reproduces the calculation inputs. Note that energy input from hydro, solar or wind sources are assumed to have zero CO_2 emissions.

TABLE IAD10: IMPACT ON FIRM COSTS

This table reports the assumptions we used when transforming product-level energy inputs into product CO_2 emissions. Note (*) that we assume that CO_2 emissions from burning bio-waste are based on the idea of a closed carbon cycle —the carbon dioxide emitted when bio-waste is burned is offset by the carbon dioxide absorbed during the growth of the plants that produced the waste so that the amount of CO_2 released is approximately equal to the amount of CO_2 absorbed. Data source: Central Electricity Authority CO_2 Baseline Database 2008 and the Commercial Energy Balance Tables and Conversion Factors from the Energy and Resources Institute.

Fuel	Gross Calorific Value (kcal/kg)	Density (t/kls)	Fuel CO_2 Emission Factor	
			Electricity from Fuel (tCO_2/mWh)	Fuel (gCO_2/MJ)
Coal	3,755	0.95	1.04	90.6
Diesel	10,350	0.83	0.78	69.1
Oil	9,850	0.95	0.66	71.9
Gas	11,300	0.86	0.55	49.4
Lignite	3,000	0.83	1.28	100.5
Naptha	10,750	0.70	0.61	66.0
Bio*	3,625	N/A	0.00	0.0
Hydro	N/A	N/A	0.00	0.0
Solar	N/A	N/A	0.00	0.0
Wind	N/A	N/A	0.00	0.0
Nuclear	N/A	N/A	0.00	0.0

We calculate tonnes of CO_2 emitted per reported production unit for each firm-product-year-energy source. Now that all energy sources have the same units on the energy input side (kcal/production unit) and the CO_2 emissions side (tonnes of CO_2/production unit), we can collapse to the firm-product-year level across energy sources.

Using the unique firm and product codes, we merge with our regression dataset of manufacturing firms, which contains the quantity produced of each product by each firm per reporting year. We next calculate the total CO_2 emissions per firm-product-year. So, in the end, we have a dataset at the firm-year-product level for the fiscal years 2007 to 2013 that tells us the total energy input per product reporting unit, the total CO_2 emissions and the CO_2 emissions per product reporting unit, and the proportion of the total energy from each fuel source. These data are summarized below for our regression sample in Table IAD11.

TABLE IAD11: DESCRIPTIVE STATISTICS OF PRODUCT ENERGY INPUTS AND CO_2 EMISSIONS

The table presents descriptive statistics of the product energy inputs and CO_2 emissions for our baseline sample. *Ln(Total product energy input)* is defined as the natural logarithm of total product-level energy inputs for the firm-year. *Ln(Total product CO_2)* is defined as the natural logarithm of total product-level CO_2 emissions for the firm-year. *CO_2 per production unit* is defined as the ratio of total product in tonnes and the units the product is quoted in on the firm's annual statement. *Proportion purchased electricity* is defined as the ratio of the total product-level energy from purchased electricity and total product energy input. All series are winsorized at the 1% and 99% levels.

	Obs	Mean	Std. dev.	Min	Median	Max.
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Total product energy input)	1,151	13.81	2.96	5.29	14.39	21.48
Ln(Total product CO_2 emissions)	1,151	4.37	4.45	-6.54	4.50	14.79
Tons of CO_2 per production unit	1,151	2.72	17.28	0.00	0.16	168.34
Proportion purchased electricity	1,151	0.56	0.47	0.00	0.99	1.00

Appendix E Quantity Productivity estimation

How should we measure productivity if firms are capital constrained, regulations limit labor adjustment, firms are subject to power blackouts and other infrastructure constraints, and multiple other constraints such that the capital and labor firms are using are at best constrained optimal decisions? Clearly, it requires more heroic assumptions to apply methods for measuring productivity like Levinsohn-Petrin that assume that firms can readily adjust intermediate inputs when faced with a productivity shock to settings were all these constraints limit the ability of firms to adjust. Moreover, this methodology calculates revenue-based total factor productivity (TFPR) that reflect changes in productivity but also reflect changes in markups, the product mix, and product quality. This is problematic since we expect all three to respond to the emissions regulation. On the other hand, if consumers value quality, TFPR may be preferable to TFPQ-based measures since higher prices and revenues may capture the ability to produce high quality (Atkin et al., 2019). Indeed, we confirm that in our data that input prices are an increasing function of product quality.

In this Appendix, we follow De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) in constructing quantity total factor productivity (TFPQ). We control for input price variation across firms using differences in output quality, which we model as an increasing function of output price, product market share, and product dummies.[33] This approach is attractive because it flexibly controls for quality differences to be consistent with a large class of demand models and any degree of passthrough between input and output prices. Further, it allows us to recover firm-product-year estimates of markups and marginal costs. Estimates are corrected for product quality, as proxied by input price variation, and for sample selection using the methodology of De Loecker, Goldberg, Khandelwal, and Pavcnik (2016).

E.1 Estimation assumptions

Following De Loecker, Goldberg, Khandelwal, and Pavcnik (2016), our estimation of firm quantity productivity, and firm-product-level marginal cost and markups rely on several key simplifying assumptions, as described below.

1. All producers of the same product use the same production technology, though productivity in producing the product can differ;
2. Firms are equally productive at producing all its products;
3. Firms can only change output in the short term by adjusting material inputs, but not capital and labor, which are sticky;
4. We model firms as minimizing short-turn costs, taking concurrent (time-t) quantity and input prices as given;
5. The production function coefficients are assumed to be constant over the sample period;
6. The number of products manufactured by firms increases with the firm's productivity.

Assumption 2 is not likely to hold, but is standard in the literature because it allows estimates of markups for multi-product firms. Assumption 3 allows us to ignore cross-elasticities, which we cannot estimate because we only observe labor and capital at the firm-year level. Note that this does not impose that firms cannot substitute between capital and labor in such a way that output remains constant. If assumption 3, that the only variable input is materials, and assumption 4, that firms minimize costs, hold, then markups are computed as the deviation between the elasticity of output with respect to inputs and that input's share of total revenue. Assumption 4 also implies that input prices, our main proxy for product quality, do not depend on input quantities. Note that this is unrealistic in the sense that it rules out static sources of market power in input markets, i.e., monopsony power. As a result, this approach understates the level of markups and is therefore most useful in explaining *changes* in markups. The intuition is that

[33]Intuitively, output prices are highly correlated with input prices since producers of more expensive products also use more expensive inputs, on average (for example, Kugler and Verhoogen (2012)) Following De Loecker, Goldberg, Khandelwal, and Pavcnik (2016), we also assume that input quality is correlated across the factors of production. Intuitively, manufacturing high-quality products requires combining high-quality materials with labor and capital. This assumption allows us to model input prices as a function of a single index of product quality at the firm-product level.

if market power is static or if contemporaneous changes in market power are not correlated with the 2009 emissions regulation shock then the changes in markups will be estimated without bias. Assumption 5 is necessary because we do not have enough data to estimate production functions for different time periods.

E.2 Addressing empirical bias

There are two main sources of bias in estimating TFPQ: (1) the unobserved allocation of inputs across products for firms that produce more than one product and (2) the unobserved quality of products. To address the first, we estimate the production function on single-product firms only. Of course, firms choose if they will produce one or multiple products, introducing selection bias into our estimates. Assuming that the number of products manufactured by firms is an increasing function of firm productivity (assumption 6 above) allows us to control for selection into being a multi-product firm by estimating the probability that a firm continues to produce one product as a function of the firm's productivity forecast and the state variables (number of products, material inputs, and exogenous factors like firm location). The assumption that multi-product firms use the same production technology as single-product firms producing the same product (assumption 1 above) allows us to extrapolate our single-product estimates to our subsample of multi-product firms.

The second bias, that we do not observe the quality of products, is a fundamental problem of productivity estimation. In particular, TFPQ estimations are downward biased when the econometrician does not observe product quality differences across firms.[34] To overcome this, we proxy for output quality by input quality. We do not observe input quality directly either because we do not observe how firms that produce multiple products allocate inputs across those products. To partially address this, we estimate the production function using the subsample of single-product firms.[35] This approach is attractive because it controls for quality differences flexibly so as to be consistent with a large class of demand models and with any degree of passthrough between input and output prices. The approach also allows us to recover firm-product-year level estimates of markups and marginal costs.

The specific steps we take are to:

1. Estimate the production function parameters and recover the product-specific output elasticity with respect to materials from a subsample of single-product firms. We model the production function using a translog functional form;

2. Correct for selection bias from the non-random decision of how many products to produce by estimating the productivity threshold beyond which firms move from producing one to multiple products and then controlling for the probability that the firm will continue to be below the threshold in a given year as a function of firm productivity and the state variables;

3. Proxy for the (unobserved in our data) product-level materials share of total revenue for each product of multi-product firms using the estimated production function coefficients for single product firms and an input price control function that expresses the product-specific allocation of material inputs to each product as a function of the firm-product-year output price, market share, product and location fixed effects, and the firm's export status;

[34] TFPR includes prices, which means that it captures cross-sectional quality differences between firms within narrowly-defined product categories. However, the TFPR measure also includes markups in prices have both demand and supply determinants, biasing estimates of productivity changes and cross-sectional comparisons. The direction and magnitude of this bias are highly dependent on the specific empirical setting.

[35] We confirm that input prices are an increasing function of product quality and therefore we can control for input price variation across firms using differences in output quality across firms, which we model as an increasing function of output price, product market share, and product dummies. Intuitively, output prices have been found to be highly correlated with input prices since producers of more expensive products also use more expensive inputs, on average (for example, Kugler and Verhoogen (2012)) We also assume that input quality is correlated across the factors of production. Intuitively, manufacturing high-quality products requires combining high-quality materials with high-quality labor and capital. This assumption allows us to model input prices as a function of a single index of product quality at the firm-product level.

4. Compute firm-product-year level markups and marginal costs, where the markup is the ratio of the output elasticity of materials to the materials share of total revenue and marginal costs are the ratio of the products price to its markup.

Table IAE12 reports basic summary statistics of the two-digit 1887 NIC industry codes for the Indian manufacturing sector for the period 2007 to 2012. There are 1,840 unique products, and 6,711 unique firms in our dataset, for whom 2,854 are single-product firms. It is on this sub-sample that we estimate the production function coefficients (assumed constant over the period).

TABLE IAE12: SUMMARY STATISTICS BY SECTOR

This table reports summary statistics for the average year in the sample. Column (1) reports the share of the output by sector in the average year. Column (2) reports the number of products by sector in the average year. Columns (3) and (4) report the number of firms and the number of single-product firms manufacturing products in the average year. Data source: Prowess.

Manufacturing sector	Share of total output (1)	Unique products (2)	Unique firms (3)	Unique single-product firms (4)
10 Coal, peat, & lignite	0.6%	24	145	14
21 Food products	10.6%	180	973	457
22 Beverages & tobacco products	0.2%	12	43	16
23 Textiles & apparel	5.0%	144	634	261
27 Wood & wood products	0.3%	22	132	44
28 Paper & printing publishing	0.5%	28	78	11
29 Leather, fur & synthetic leather	1.9%	27	193	151
30 Chemicals (except petroleum & coal)	8.3%	250	813	395
31 Rubber, plastic, nuclear fuel, petroleum & coal	16.9%	285	757	306
32 Non-metallic mineral products	17.9%	65	516	211
33 Basic metal & alloys industries	9.6%	110	682	312
34 Metal products (not machinery & equipment)	4.4%	87	233	84
35 Machinery & equipment (not transport)	9.9%	373	723	276
37 Transport equipment & parts	10.6%	126	396	182
38 Other manufacturing industries	3.2%	97	393	134
	100%	1,830	6,711	2,854

We perform several sanity checks on the data to see if it conforms with our economic intuition and evidence in the literature. Figure IAE7 reports the correlation between demeaned markups and marginal costs and the natural logarithm of product quantity produced.

The left panel of FigureIAE7 demonstrates that quantities and markups are positively related in our sample, indicating that firms producing more output also enjoy higher markups due to their lower marginal costs. The right-hand panel of Figure IAE7 plots marginal costs against production quantities. Our elasticity estimates show that many firms are characterized by increasing returns to scale, an empirical pattern also noted in De Loecker et al. (2016). Consistent with this, we see that there is an inverse relationship between a product's marginal cost and the quantity produced.

Next, we check the reasonableness of our extrapolation of the production function estimates of single-product firms to multi-product firms. Figure IAE8 reports how our estimated firm-product-year markups (left panel) and marginal costs (right panel) vary across products within multi-product firms. Specifically, we de-mean markups and marginal costs using product-year and firm-year fixed effects in order to make these variables comparable across products within firms. We then plot the de-meaned markups and marginal costs against the sales share of the product within each firm.

In the left-hand panel of Figure IAE8, marginal costs rise as a firm moves away from the product with the

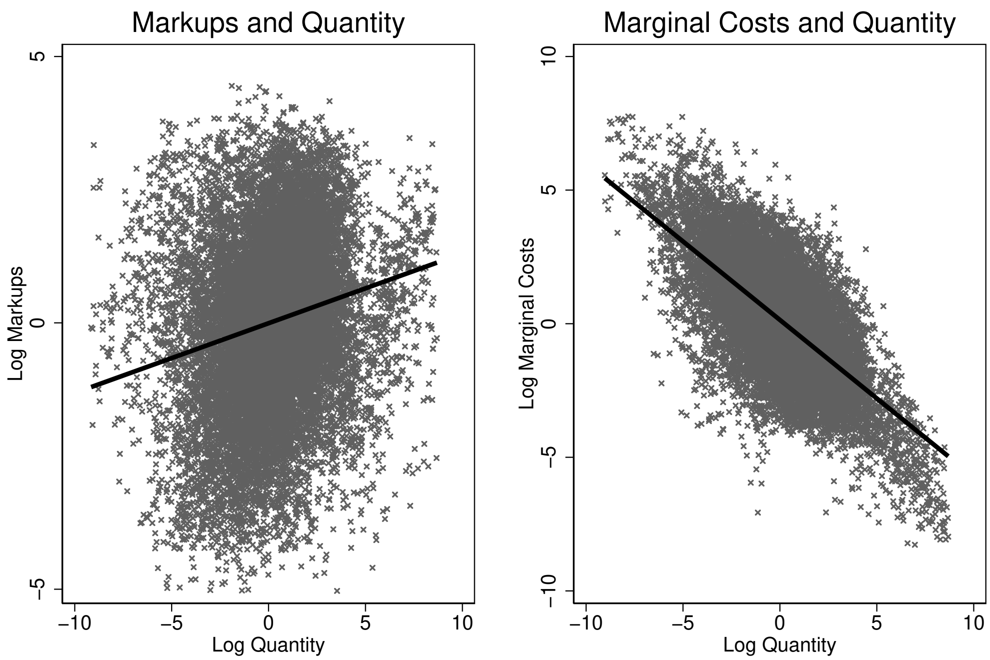


FIGURE IAE7: MARGINAL COSTS, MARKUPS AND QUANTITIES

The left panel presents the correlation between the natural log of product markup and output quantity. The right panel is between the natural log of product marginal cost and output quantity. Data are at the firm-product-year level for the period 2007 to 2012. Data are winsorized at the 3rd and 97th percentiles. Markups, marginal costs, and quantities are demeaned by product-year fixed effects to make them comparable across firms.

lowest within-firm marginal cost (its "core" product). For the other products, marginal costs rise with a product's distance from the core competency. The right panel reports that firms set their highest markups on their core product, and markups decline as they move away from that main product. Although we do not impose any assumptions on the market structure and demand system in our estimation, these correlations are consistent with the theoretical predictions from the multi-product firm literature (Eckel and Neary, 2010; Mayer, Melitz, and Ottaviano, 2014; Melitz and Ottaviano, 2008) and the empirical findings of De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) in the Indian manufacturing sector.

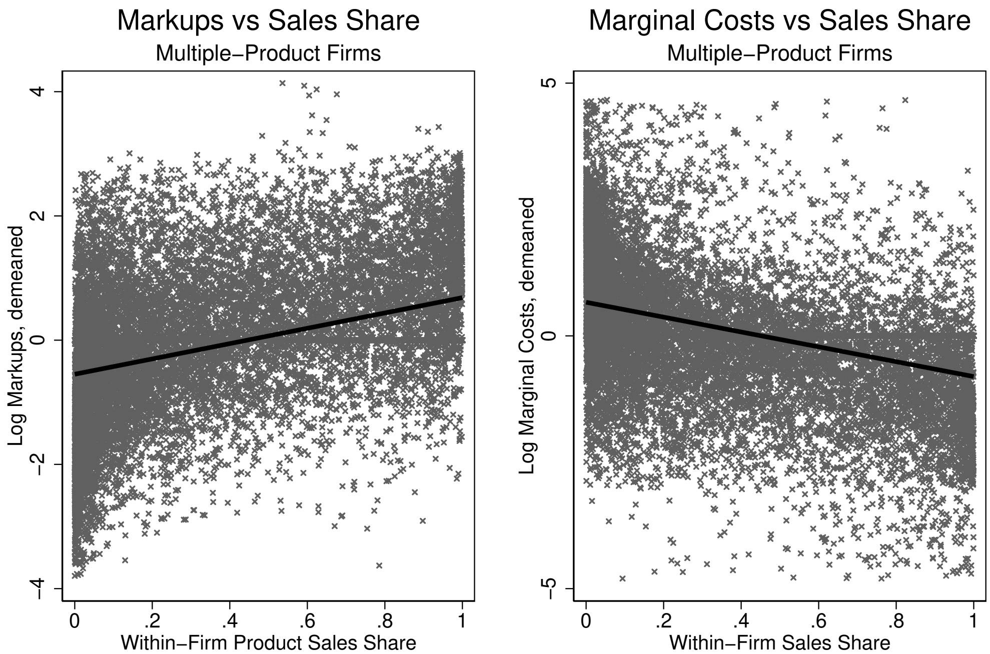


FIGURE IAE8: MARKUPS, COSTS AND PRODUCT SALES SHARE

Notes: The left panel presents the correlation between the natural log of product markup and sales share. The right panel is between the natural log of product marginal cost and sales share. Data are at the firm-product-year level for the period 2007 to 2012. Data are winsorized at the 3rd and 97th percentiles. Markups, marginal costs, and quantities are demeaned by product-year and firm-year fixed effects to make them comparable across firms.

Having estimated the TFPQ measure and convinced ourselves our estimates are reasonable, we consider the impact of the 2009 CEPI reform on *quantity* productivity (TFPQ). In Table 8 we found a significant increase in TFP, driven by firms not operating in highly-polluting industries. In other words, the average treated firm became more productive at turning input into revenue. In Table 8, we see that there is no significant effect on the efficiency with which treated firms turn input into outputs (Model 1). There is also no differential effect of the reform on the TFPQ of firms in highly-polluting firms and firms in other industries (Model 2).

TABLE IAE13: CHANGES IN QUANTITY-BASED PRODUCTIVITY

This table reports the changes in firm profitability and revenue productivity around the 2009 CEPI emissions regulation. The unit of analysis is firm-year. The dependent variable in column 1 is the natural logarithm of quantity-based total factor productivity estimated following De Loecker, Goldberg, Khandelwal, and Pavcnik (2016). Column 2 focuses on the subsample of firms operating in High-Polluting Industries (HPI) while column 3 focuses on the subsample that excludes HPI. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. The table also reports the *p*-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. All specifications include firm and State × two-digit industry × year fixed effects. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix Table IAA8. Data source: CMIE Prowess.

Dependent variable:	Log(Quantity-based Productivity)	
	(1)	(2)
Post $\times$CEPI$^{[70,100]}$ (β_1)	-0.174	-0.118
	(0.153)	(0.164)
Post $\times$CEPI$^{[60,70)}$ (β_2)	-0.287	-0.190
	(0.176)	(0.302)
Post $\times$CEPI$^{[70,100]}$ $\times$ High-Polluting (β_3)		-0.184
		(0.127)
Post $\times$CEPI$^{[60,70)}$ $\times$ High-Polluting (β_4)		-0.189
		(0.376)
p-value [$\beta_1 + \beta_2 = 0$]	0.145	0.471
2008 Dependent Variable Mean (Control)	8.6	8.6
Fixed effects:		
Firm	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes
Bandwidth	Yes	Yes
R^2	0.824	0.825
Observations	1,898	1,898

Together with the results in Table 8, this evidence is suggestive that the main effect of the reform is on profitability, not on operational efficiency. This must be caveated by recognizing the limitations of this analysis, mainly those highlighted in Atkin, Khandelwal, and Osman (2019). Briefly, these authors find that TFPQ is a relatively poor measure of quantity productivity because it shows excessive dispersion across firms and correlates negatively with quality productivity, which they measure off of detailed product quality data in the rug manufacturing sector. The authors attribute this to the difficulty of adjusting for product specifications and quality to make apples-to-apples comparisons. Finally, they find that TFPR does better than TFPQ at capturing broad firm capabilities, even though TFPR suffers from being unable to separate effects from changes in productivity, markups, the firm product mix, and product quality. For this reason we use TFPR as our main measure of productivity.

However, the model also allows us to estimate product-level marginal cost and markup, which can help differentiate these stories. If the treated primarily become better at producing revenue out of a given unit of input, we should expect this to be reflected in pricing. In Appendix Table IAE14 we see results consistent with this. In Panel A Model (1) we see that the sub-sample for which we can calculated TFPQ raise their prices significantly (at the 90% confidence level). From Models (2) and (3) we see this is primarily the result of passing on increased marginal costs. If anything, markups decrease, though the difference relative to control firm markups is not statistically significant. In all cases,

TABLE IAE14: IMPACT ON FIRM PRICING

This table reports the changes in firm pricing around the 2009 CEPI emissions regulation. The unit of analysis is firm-product-year. The dependent variable in column 1 is the natural logarithm of the product price, in column 2 it is the natural logarithm of product marginal cost, and in column 3 it is the natural logarithm of product markup. The marginal cost and markup are computed following De Loecker et al. (2016) and account for unobserved input prices (quality differences), unobserved allocation of inputs across products within multi-product firms, and the endogeneity of the choice to produce multiple products. *Post* is an indicator variable taking the value of 1 for all years including 2009 the year in which reform was implemented and after. $CEPI^{[70,100]}$ takes the value of one if the industrial cluster has a CEPI value at or above 70, and zero otherwise. $CEPI^{[60,70)}$ takes the value of one if the industrial cluster has a CEPI value greater than or equal to 60 and below 70, and zero otherwise. The sample is restricted to the 88 industrial clusters targeted by the CPCB in 2009 with the omitted category including clusters with a CEPI value below 60. All specifications include firm and State × two-digit industry × year fixed effects. The table also reports the *p*-value from the joint test of the coefficients and the mean of the dependent variable in levels in the pre-reform year 2008. The standard errors are clustered at the city level and are robust to heteroscedasticity. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. Data are winsorized at the 1 and 99 percentiles. Data source: CMIE Prowess.

	Panel A: All Industries		
Dependent variable	Ln(Price)	Ln(Marginal Cost)	Ln(Markup)
	(1)	(2)	(3)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.399*	0.783***	-0.375*
	(0.198)	(0.218)	(0.203)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.369**	0.717***	-0.342
	(0.177)	(0.258)	(0.252)
p-value [$\beta_1 + \beta_2 = 0$]	0.043	0.002	0.111
2008 Dependent Variable Mean (Control)			
Fixed effects:			
Firm	Yes	Yes	Yes
State × industry × year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.662	0.595	0.322
Observations	8,198	8,198	8,198

Dependent variable	Panel B: Industries Split by High-Polluting vs. Others		
	Ln(Price)	Ln(Marginal Cost)	Ln(Markup)
	(1)	(2)	(3)
Post $\times$CEPI$^{[70,100]}$ (β_1)	0.379*	0.786***	-0.393*
	(0.201)	(0.228)	(0.214)
Post $\times$CEPI$^{[60,70)}$ (β_2)	0.545**	0.855**	-0.306
	(0.212)	(0.324)	(0.273)
Post $\times$CEPI$^{[70,100]}$ $\times$ High-Polluting (β_3)	0.079	0.006	0.061
	(0.068)	(0.072)	(0.069)
Post $\times$CEPI$^{[60,70)}$ $\times$ High-Polluting (β_4)	-0.492	-0.390	-0.098
	(0.334)	(0.342)	(0.172)
Fixed effects:			
Firm	Yes	Yes	Yes
State $\times$ industry $\times$ year	Yes	Yes	Yes
Bandwidth	Yes	Yes	Yes
R^2	0.663	0.595	0.322
Observations	8,198	8,198	8,198

the average effect is increasing in the intensity of treatment, with an additional affect when a cluster's 2009 CEPI score is at or above 70. Panel B of Appendix Table IAE14 tells us that the effect does not differ if the treated firm is in a highly-polluting industry. This is in contrast to the TFP results in Table 8 where the effect is driven by treated firms operating in industries other than highly-polluting ones. Overall, the evidence presented in this appendix supports the hypothesis that firms are responding to the 2009 CEPI regulation by shifting to higher-margin products where they can and passing on costs where they cannot. And there is some weak evidence that the productivity evidence we document is not coming from enhanced efficiency at converting inputs into outputs (TFPQ) but instead from increased revenues generated per input as a result of the production changes. This evidence, however, is only suggestive as we cannot control for quality differences among the products that are emphasized after the reform and there is good reason to think that the highest-margin products are also the highest-quality ones, meaning this bias is be meaningful in our setting and TFPR is a better proxy for firm total factor productivity (Atkin, Khandelwal, and Osman, 2019).