Global Climate Calamity vs. The Market: Nationwide Carbon Trading in China*

Douglas Almond
Columbia University and NBER

Shuang Zhang
Imperial College London and CEPR

This version: March 20, 2025

(Preliminary and work-in-progress: please do not cite)

Abstract

The world's largest carbon emitter, China, launched the world's largest emissions trading scheme (ETS) in the power sector in 2021. The efficacy of China's national ETS is unclear because it offers different incentives to firms compared to Cap-and-Trade. A primary obstacle to impact evaluations is the lack of reliable data at the plant level. We introduce novel methods and data on plant-level output and carbon emissions based on satellite imagery and machine learning. These data provide an independent and objective benchmark to data reported by firms. We find that, consistent with theory, initially high efficiency plants increase their electricity generation and carbon emissions, and there is no evidence that plants reduce their emission intensity. We also document increases in ambient air pollution originated from high efficiency plants, suggesting environmental costs of the program. Our findings offer timely insights to other emerging countries that are likely to adopt a similar program.

*Almond: Columbia University, School of International and Public Affairs & Department of Economics & NBER; email: da2152@columbia.edu. Zhang: Imperial College London, Department of Economics and Public Policy & CEPR; e-mail: s.zhang@imperial.ac.uk. We thank Xiaohan Dong, Ruixue Li, Sophia Qin and Yuting Wang for excellent research assistance. For helpful comments, we thank Albert Park, ...

1 Introduction

In the 2024 AEA Distinguished Lecture, Michael Greenstone observed that emerging economies will drive the future course of climate change (Greenstone, 2024). Projecting that fossil fuels will remain the globally dominant source of energy, Greenstone observed:

> *Among "emerging" countries, energy consumption is projected to grow by 28 percent from 2025–2050 compared to a 12 percent decline among "developed" countries due to greater efficiencies. In total, the "emerging" countries account for 130 percent of the expected growth.*

Among these pivotal emerging economies, China stands out for four reasons. First, it has by far the largest greenhouse gas emissions. Figure 1 shows its CO2 emissions are more than double the US and more than three times those of the EU or India. Second, GHG emissions continue to rise in China, while they are falling in the US, the EU, and other developed countries. Third, China is appealing from a policy perspective in having recently adopted a nationwide, market-based approach to reduce carbon emissions. Such an approach holds the promise of reducing emissions at the lowest possible cost, something economists and policymakers in Beijing have been keen to achieve.[1] At present, we know nothing about the key impacts of the world's largest carbon market, including whether it stems carbon emissions. Fourth, China is a pioneer in carbon trading in Asia and its policies proliferate to other emerging economies (Asia Development Bank, 2016).

China launched its ambitious nationwide ETS in July 2021. Emitters trade emission allowances in the national carbon market. The scale of carbon emissions regulated is unprecedented. The program started in the power sector, which emit about 4.6 billion tonnes of CO_2 emissions, about 15% of global carbon emissions. When expanded to other sectors, the program's coverage is expected to double the total amount of CO_2 emissions covered by all other emission pricing programs. Despite this scale and ambition, the efficacy of the world's most ambitious carbon market remains unknown for mundane (if familiar) reasons: significant challenges in monitoring and data quality in China.

Indeed, the lack of requisite high-quality data hinders the adoption of market-based instruments in emerging economies generally (Greenstone et al., 2023). While China took a leading step in

[1] In contrast, the United States came close but failed to adopt a nationwide carbon market in 2009 under the Waxman-Markey bill. This despite having pioneered market-based approaches to to address other environmental "bads", such as sulfur dioxide, in 1995.

establishing a national carbon market, the challenges of imperfect monitoring and insufficient transparency persist (Karplus, Zhang, and Almond, 2018; Greenstone, He, Jia, and Liu, 2022; Barwick, Li, Lin, and Zou, 2024). Indeed, China's Ministry of Ecology and Environment has documented a number of cases of falsified carbon data submitted by power plants (Reuters, March 2022)[2]. Obviously, the lack of reliable data can undermine the functioning, effectiveness, and evaluation of the China ETS.

A primary obstacle to impact evaluations of China's national ETS is the lack of reliable data on firm output and carbon emissions at a granular level. We leverage instead data that are more difficult to manipulate. Satellite data on carbon emissions have the potential to serve as an objective benchmark to data reported by firms. However, the conventional column CO_2 measures (e.g. from NASA's OCO-2) are useful to measure carbon emissions over a relatively large geographic area, but are not suitable to measure carbon emissions of individual plants because of their sparse and inconsistent coverage over time.[3]

In this paper, we introduce novel methods and data sources to address this key obstacle. In collaboration with Climate Trace, we introduce and analyze plant-level output and carbon emission data from satellite images and a machine learning algorithm. Satellite imagery provides reliable indicators of a power plant's operation, and the data are both high-resolution and high-frequency. In particular, we analyze objective measures on plant-level electricity generation and a first to document how these respond to China's ETS. We also compare how consistent carbon emissions respond to the ETS in data developed from satellite imagery and data reported by plants.

The way China structured its emission trading scheme introduces additional uncertainty into its expected effects and efficacy in reducing emissions. Conventional cap and trade programs – including the EU's – are mass-based, where compliance depends on keeping the absolute level of emissions below a preset cap. In contrast, China's national ETS is rate-based, where compliance depends on emission intensity (the ratio of emissions to current output). Firms can jointly optimize their output level to influence the amount of emission allowances. In theory, a rate-based system could be less cost-effective because it provides little incentive for reduction of intended output (Goulder et al., 2022, 2023). More importantly, without an absolute cap, it is unclear whether China's national ETS

[2]https://www.reuters.com/world/china/china-slams-firms-falsifying-carbon-data-2022-03-14/

[3]These satellite instruments use spectroscopic methods based on the absorption of reflected sunlight to estimate the column-averaged dry-air mole fraction of CO_2.

reduces carbon emissions, obviously the critical, first-order question for climate forcing and thereby global climate change. On the other hand, a rate-based system has potentially better compatibility with economic growth (Goulder et al., 2022), making it appealing to other emerging economises as well. Asia Development Bank (2016) suggests that, high economic growth in developing countries in Asia and the Pacific will remain a high priority for the longer term, and consequently, there is an emphasis on intensity-based targets rather than absolute emission caps. Indonesia launched a rate-based ETS in 2023, and India has announced to start a similar program in 2025.

The current research vacuum is difficult to overstate. At present, researchers know essentially nothing about a rate-based ETS's actual effects on firm output, carbon emissions and air quality. We provide the first empirical research on the efficacy of the first major rate-based ETS in the world's largest carbon emitter. We first evaluate the causal effects of China's National ETS on firm output and carbon emissions. Given coal's dominance in China's power sector, a natural question is whether China's National ETS delivers environmental co-benefits. We further examine the causal effects of the ETS on ambient air pollution, by leveraging the fact that "[m]ost local pollutants are coproduced with greenhouse gases, so many of the policies to reduce GHGs also reduce local pollutants" (Fullerton and Wolfram, 2016).

We document four main findings. First, initially high efficiency plants increase their electricity generation by 7.8% following the National ETS, while initially low efficiency plants have little change in generation, consistent with theoretical predictions. Second, high efficiency plants increase their CO_2 emissions by about 7.7% after the ETS, similarly in both data developed from satellite imagery and data reported by plants. Third, there is no evidence that plants change their emission intensities. Finally, ambient $PM_{2.5}$ on the trajectories of high-efficiency plants increases by about 3.2% after ETS, compared to the trajectory measures of low-efficiency plants. We also find similar increases of PM_{10} and SO_2.

This paper makes three broader contributions to the literature and policy discussions about addressing the global challenge of climate change through market-based climate policy.

First, in the context of economic and political constraints on environmental policy, it advances the literature on the efficacy of market-based instruments. Rate-based systems are particularly appealing to rapidly growing economies that fear impinging their development trajectories. To the best of our knowledge, **Rate-based emissions trading system have only been studied con-**

ceptually/theoretically (Fischer, 2001; Fischer and Newell, 2008; Goulder et al., 2022, 2023). Can rate-based ETS succeed in reducing carbon emissions, and is this a path for other emerging economies to curtail emissions? Likewise in the "second best" policy spirit, the United States has been unable to adopt a mass-based emissions trading system, and has turned instead to subsidizing green industries and purchases, often with domestic content requirements, under the Inflation Reduction Act. Are rate-based trading systems an appealing alternative to the US approach to climate change? Which basic approach is more effective and appealing?

Second, this project introduces novel methods and data to the literature of environment and development. A common theme is that poor monitoring undercuts the effectiveness of environmental policies in developing countries (Duflo et al., 2013, 2018). Can a market work to reduce emissions, where command-and-control regulations are either performing imperfectly or are non-existent? Existing research on emission markets in developed countries does not provide an answer.[4] The methodology we introduce and employ here can provide new and reliable data on carbon emissions both for research beyond China, particularly in countries where monitoring is challenging and GHG emissions are expected to grow most rapidly.

Third, this project is related to a recent literature on the pilot ETS in seven Chinese regions since 2014 (Almond and Zhang, 2021; Cui et al., 2021, 2023; Zhang et al., 2020; Chen and Lin, 2021). One key finding emerging from that literature is that carbon prices needn't be close to their social costs for substantive changes to result (Almond and Zhang, 2021; Cui et al., 2021; Zhang et al., 2020; Chen and Lin, 2021). This possibility depends on the density of carbon-intensive operations that were economically marginal at baseline. Today, are there many such firms and production activities that will likewise be affected by relatively low carbon prices? Additionally, because the pilot programs do not mirror the national ETS, it is difficult to extrapolate from them to predict key impacts of the national ETS.[5]

The rest of the paper is organized as follows. Section 2 describes the basic features of China's

[4] Greenstone et al. (2023) is an exception that experiments with a new particulate matter emissions market in India and built a continuous emissions monitoring system (CEMS) to measure plant-level particulate matter. To monitor a carbon market with *high-quality* carbon data, however, is more complex and to the best of our knowledge has not been done in developing countries.

[5] The pilot programs experimented with a combination of mass-based and rate-based systems. Even when a so-called "mass-based" system were used, some regions had *ad hoc* measures to adjust a regulated firm's emission cap at the end of a compliance period, which is different from a cap and trade program where the cap is preset. Therefore, previous studies on the pilot program cannot distinguish different incentive effects between these two systems.

National ETS and discusses the theoretical literature. Section 3 introduces data and methods for power generation and carbon emissions. Section 3 describes data and methods for ambient air pollution. Section 5 discusses our empirical strategy. Section 6 presents the empirical results. Section 7 discusses the implications of our findings and concludes.

2 China's National ETS

China's National ETS regulates 2,257 power plants that emitted 26,000 tonnes or more of CO_2 in any year of 2013-2018, which cover about 40% of China's total carbon emissions. China's program is a rate-based system where compliance depends on emission intensity, which is the ratio of emissions to output. In every compliance period, each generation unit of a regulated plant receives a free allocation of emission allowances based on the government's assigned benchmark (maximum allowed emission intensity) and its end-of-period power generation.

Assigned benchmarks are fuel specific and are in general set below the average initial emission intensities among regulated facilities. For conventional coal-fired units, the assigned benchmark is 0.877 for units at or above 400 MW of rated power, and 0.979 for units below 400 MW of rated power (EDF, 2021).

China's program shares an important feature of cap-and-trade programs in the introduction of a national carbon market for allowance trading. However, the key difference between China's rate-based system and cap-and-trade programs is how emission allowances are allocated. In cap-and-trade programs, the amount of emission allowances a facility receives is usually exogenous to its end-of-period output. In contrast, under a rate-based system, the number of allowances is endogenous to a facility's chosen level of output. Firms can jointly optimize their output level to influence the number of emission allowances.

Theoretical literature A rate-based system implicitly subsidizes intended output (Fischer, 2001; Fischer and Newell, 2008). Theoretical analysis of China's rate-based ETS (Goulder et al., 2022, 2023) suggests that the implicit output subsidy provides differential incentives to firms.

- For more efficient firms with initial emission intensity below the benchmark, they receive allowances in excess of that's required for compliance. They have incentives to *increase* pro-

duction to earn additional allowances to be sold on the market.

- For less efficient firms with initial emission intensity above the benchmark, they can reduce the cost of buying allowances by either reducing production or reducing emission intensity. However, there is an implicit tax on the reduction of output, as reducing output implies a reduction in allowances received. Consequently, firms rely more on reduction of emission intensity rather than reduction of output.

It is unclear how total electricity generation respond to China's rate-based ETS and whether this program reduces carbon emissions, the latter of which is a first-order question for global climate change.

Challenges in monitoring Under China's National ETS, the process of monitoring, reporting and verifying (MRV) is delegated to local governments, which presents another challenge to the effectiveness of China's National ETS. Regulated facilities are required to submit annual reports of their carbon emissions. Provincial governments then assign third-party verification companies to verify the reported carbon emissions. Potential conflict of interests may arise because local officials' promotion depends on their economic performances and large emitters could be important contributors to local economy.

Carbon prices Figure 2 presents carbon prices in the national program since July 2021. Prices started relatively low at \$6 per ton of CO2, remained stable around \$8 until July 2023, after which they gradually increased to over \$12. Although prices are still much lower than many other existing ETS, Chinese firms were found to have responded in their CO_2 emissions at low prices of \$6 per ton during the pilot program (Cui et al., 2021; Zhang et al., 2020; Chen and Lin, 2021).

The electricity sector Among fossil-based power plants in China, coal-fired units account for 91% of electricity production and 92% carbon emissions, while circulating fluidized bed units and gas-fired units are limited and contribute far less to emissions (Goulder et al., 2022). A coal-fired power plant typically has multiple generation units. A unit's efficiency depends principally on its generation technology. The three main types of technologies in ascending order of efficiency are:

sub-critical, supercritical and ultra-supercritical. Sub-critical units contribute to the largest share of electricity production and carbon emissions.

3 Data and Methods for Power Generation and Carbon Emissions

Understanding of the first-order effects of the first major rate-based ETS has foundered for lack of reliable data. We introduce and analyze innovative and objective metrics that may be more difficult to manipulate than data that are more readily available to researchers. Specifically, visible water vapor plumes in coal-fired power plants serve as signals for electricity generation, directly related to emissions. We collaborate with Climate Trace to adopt a two-step approach based on high-resolution and high-frequency satellite images using the methodology in Couture et al. (2024): a machine learning algorithm predicts electricity generation from water vapor plumes, which is then combined with plant-specific emission intensity to estimate emissions.

3.1 Step 1: Plant-level electricity generation

Direct measurement of CO2 emissions via satellite imagery is challenging as flue gas is generally less visible at plants that are more efficient or have air pollution controls. Instead, the methodology in Couture et al. (2024) uses visible water vapor plumes, which are robust indicators of electricity generation. These plumes primarily arise from two sources: natural draft cooling towers (NDT) and wet flue gas desulfurization (FGD). NDT systems release heat during generation, forming large, visible plumes, while FGD systems produce plumes saturated with water after desulfurization. Machine learning models are trained for NDT and FGD separately for these two structures to address differences in plume characteristics.

Imagery and Labeling High-resolution and high-frequency satellite images from NASA's Landsat-8, ESA's Sentinel-2A/B, and Planet Labs' Planet Scope are used to train machine learning models. The Planet Scope satellites provide daily revisits with a ground passing time between 7:30 and 11:30am, providing images with spatial resolution of 3 meters. Sentinel-2A/B have a 10-day revisit time at about 10:30am, providing images from 10 to 60 meter spatial resolution. Landsat 8 has a 16-day revisit at about 10am, providing images from 30 to 100 meter spatial resolution.

To train the machine learning models, satellite imagery is matched with plant-level, hourly electricity generation data, available for plants in the US, Europe, and Australia. Data are from US EPA Clean Air Markets Program Data, European Network of Transmission System Operators for Electricity, and Australia National Electricity Market. Each satellite image is matched with hourly generation data at the hour of the timestamp and labeled with two variables: 1) the capacity factor, defined as the plant's generation at the given hour divided by its maximum capacity; and 2) an indicator of on/off, where on is defined as capacity factor greater than 5%, and off otherwise.

The visibility of water vapor plumes is influenced by cloud cover and weather conditions. Two main criteria for image selection are applied in Couture et al. (2024). First, cloud coverage affects plume visibility for both NDT and FGD sources. Image features are used to measure cloud coverage, and only images with less than 20% cloud coverage are included. Second, FGD plumes are less visible under high temperatures or low relative humidity. To address this, additional empirically derived thresholds for temperature and relative humidity are applied to the training sample for FGD models. This process retains about 80% of NDT images and 64% of FGD images for training.

Training, Validation and Testing Two types of models are trained in Couture et al. (2024): gradient-boosted decision trees and convolutional neural networks (CNNs). Gradient-boosted decision tree models used statistical features extracted from satellite image patches centered on the emission structures, including metrics like the mean and standard deviation of pixel values. CNN models, on the other hand, uses a multi-layer framework to combine patches (NDT cooling towers and FGD stacks) and aggregate features from different patches to predict outcomes of interests.

Both models are trained to predict two plant-level outcomes: 1) whether the plant is operating or not (on/off); and 2) a capacity factor, which measures what fraction of the plant's potential is being used to generate power.

The model is validated using both cross-validation and external validation. First, for cross-validation, plants in the training sample are rotated into three groups: half for training, a quarter for validation and a quarter for testing. Plants in the validation and testing groups are not used for training. The performance of the model is tested in the testing group. Second, for external validation out of the training sample, out-of-sample plants in Turkey and India are also used.

Couture et al. (2024) reports the performance of the model. In the cross-validation, for the

classification of plant on/off, the mean average precision (mAP) is between 0.93-0.96 for NDT models and between 0.88-0.90 for FGD models. The model to predict the capacity factor outperforms a naive baseline that predicts a constant value of the sample's average capacity factor. In the external validation, the models also outperform their baselines, particularly by a large margin on the classification of on/off.

Aggregating to annual generation Hourly capacity factor predictions are aggregated into rolling 30-day averages. For each plant and day, the average hourly capacity factor is computed over the preceding 30 days. Multiplying this value by the plant's maximum capacity yields daily plant-level electricity generation. Annual electricity generation is calculated by aggregating the daily values over the year.

Application and Validation The methodology in Couture et al. (2024) is applied by Climate Trace to Chinese power plants 2019-2023 to estimate plant-level annual electricity generation. We analyze annual measures for comparison with ground truth data on generation in the validation exercise below.

To ascertain whether data developed from satellite images can serve as a useful benchmark for data on the ground outside of the training sample, we compare our estimates with ground truth data in a sub-sample of Chinese power plants. In general, plant-level generation data of Chinese power plants are not made publicly available. However, publicly listed power companies are required to report their plant-level generation data in their annual reports. About one-third of power plants in our data are owned by public companies. We compile annual reports of public power companies and digitize the plant-level generation data.

We first compare annual generation data from the two sources *prior to* the China ETS in the baseline year of 2020. Figure 3 shows that the cross-sectional correlation is about 0.89. The levels of satellite data are relatively lower than the ground truth data, consistent with findings in other countries out of the training sample, including Turkey and India in Couture et al. (2024). A possible source for the downward bias in the satellite data is that the satellites' local overpass times are in the morning local time. If Chinese power plants generate more electricity at times when the satellites do not capture them, and such patterns are less prevalent in the training sample, our estimations

could be biased downwards.

To evaluate the impacts of the China ETS, it is most important that *trends* in the data based on satellite imagery before and after the ETS are consistent with trends in the ground truth data. Figure 4 shows very similar trends in these two data sources, suggesting that *changes* in satellite-derived measures (instead of levels) provide a useful benchmark to changes in data on the ground.

In addition to validating the model among Chinese power plants, we conduct a similar exercise to coal-fired power plants in India. To compare with data developed from satellite imagery, we collect plant-level annual generation data from India's National Power Portal. Figure A.1 shows parallel trends in these two data sources.

3.2 Step 2: Plant-level carbon emissions

In the second step, plant-level CO_2 emissions are estimated by combining two separately estimated measures: 1) predicted electricity generation from Step 1; and 2) the estimated emission factor. A power plant's emission factor is estimated using plant characteristics, including fuel type, age, capacity, generation technology and associated generation efficiency. We refer it to the *engineering estimate* of carbon intensity.

We also collect data of reported carbon emissions from power plants. Regulated plants are required to submit annual carbon reports since 2019, two years before the launch of the national ETS. In the report, firms report plant-level, annual amount of carbon emissions. These reports are uploaded to a publicly-accessible platform. We compile and digitize these data.

We use reported CO_2 emissions for two purposes. First, we assess how consistent the trends of carbon emissions before and after the ETS emerge across data derived from satellite imagery and data reported by firms. Second, we combine reported CO_2 emissions data and electricity generation data based on satellite imagery to calculate carbon intensity. We refer it to the *reported estimate* of carbon intensity, and we compare its response to the ETS with that of the *engineering estimate* of carbon intensity.

3.3 Summary statistics

Our main analysis sample is a balanced panel of 512 coal-fired plants from 2019 to 2023. To separate the effects of the National ETS from previous pilot ETS programs, we exclude power plants located

in regions where a pilot ETS was implemented prior to the national ETS. Power plants in our analysis sample accounts for about 58% of China's total coal-fired generation in 2019.

A coal-fired power plant typically has multiple generation units. Because emission permits are allocated at the generation unit level, we examine whether generation units in our sample are representative of all coal-fired units in China. In Table A.1, we compare the characteristics of generation units of all power plants vs units of plants in our analysis sample. Generation units in our analysis sample are slightly larger and more likely to have supercritical or ultra-supercritical technology (more efficient technology), but the differences are small. Using the engineering estimate of carbon intensity to measure the fraction of units above assigned benchmarks, we find that the fraction is very similar at about 76% in both samples.

Table 1 reports summary statistics at the plant level. Power plants in our sample have an average total capacity at 1,218 MW. The average annual utilization rate is 40%, suggesting that these plants are far from operating at their full capacity. On average, these plants generate 4,625 GWh of electricity and emit 4,115 kilotonnes of CO_2 annually.

In data collected from firm-level carbon reports, a smaller sample of 280 power plants consistently report data over five years. Table 2 reports the summary statistics. These plants have a total capacity similar to that in the satellite imagery data, about 1,171 MW. On average, they report that they emit about 4,694 kilotonnes of CO_2 annually.

4 Data and Methods for Ambient Air Pollution

4.1 Ambient pollution data

We obtain measures of ambient air pollution from two sources. First, we compile pollution data in 2018-2023 from the real-time monitoring and reporting system of China's Ministry of Ecology and Environment. Hourly concentration readings of $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3 and CO are available from about 2,000 pollution monitors spread across China. Second, to complement ground monitor pollution measures, we also use high-resolution ($0.01^\circ \times 0.01^\circ$) satellite-derived estimates of monthly average PM2.5 in 2018-2022 (Shen et al., 2024).

4.2 Atmospheric dispersal model

To consider transport of air pollution in the environment, we incorporate an atmospheric dispersal model following the recent literature (Grainger and Ruangmas, 2018; Hernandez-Cortes and Meng, 2023). We generate particle trajectories from emitting power plants using the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model, developed by the US National Oceanographic and Atmospheric Administration (NOAA). The HYSPLIT model computes the trajectories of particles or puffs in the air based on observed or forecast meteorological data and given initial conditions. NOAA offers meteorological data sources, and we use the Global Data Assimilation System output data, specifically the GDAS1 dataset.

We feed plant location, average stack height of 150 meters, date, and hour into the HYSPLIT to generate forward trajectories from power plants.[6] For each plant, we generate four trajectories per day at 8am, 12pm, 4pm and 8pm. We let each trajectory run for 72 hours.[7] Each trajectory consists of hourly locations of an air parcel. After obtaining the trajectories, we match pollution monitors to trajectories. Specifically, we match a monitor to a plant if the monitor is "on trajectory" from the plant at a specific hour if the trajectory intersects a 1 km circle centered at the monitor. We also conduct robustness analysis with 5 km as the radius.

We then construct a balanced panel of data from matched pollution monitors that are on the trajectory for each plant during 2017-2024. We exclude monitors that are only on trajectory for less than 1% of the time during the study period. Using pollution readings from these monitors, we generate monthly pollution measures for each plant by taking a weighted average of the hourly pollution readings in a month of all monitors matched to that plant. The weights are the number of times that a monitor is on the plant trajectory by applying the exponential decay following Hernandez-Cortes and Meng (2023).

In addition to using ground monitor pollution readings, we also use satellite-derived monthly PM2.5 data as an alternative measure of ambient pollution. We conduct a similar exercise to match grids in the satellite-derived PM2.5 data to trajectories. A grid is matched to a plant if a plant trajectory passes through the grid. We exclude grids that are only on trajectory for less than 1% of the time during the study period. Similarly, we measure monthly PM2.5 from a plant by the

[6] We also use 200 meters in stack height in robustness checks.
[7] We also let trajectories to run for 120 hours in robustness checks.

weighted average of PM2.5 readings from grids on trajectory by applying the exponential decay.

5 Empirical Strategy

Although almost all coal-fired power plants are regulated, we exploit the variation in generation efficiency of plants prior to the ETS, which determines the differential incentives in how they respond to the policy. As discussed in Section 2, the theoretical literature suggests that more efficient plants whose initial emission intensity falls below the benchmark could be encouraged to increase electricity generation, while less efficient plants whose initial emission intensity above the benchmark are incentivized to rely more on the reduction of emission intensity to comply. We empirical test these predictions by comparing how these two types of plants respond to the ETS.

The efficiency of a generation unit is principally determined by its generation technology. Using our engineering estimate of emission intensity, we find 91% of sub-critical units have emission intensity above the benchmark. In contrast, only 37% of the supercritical and ultra-supercritical units have emission intensity above the benchmark. Convenient for our analysis using plant-level data, the vast majority of power plants use a single type of technology among its generation units: 54% of plants use subcritical technology only, and 35% use super-critical and ultrasuper-critical technology only. Therefore, in our main analysis, we focus on these two types of power plants. We define high efficiency plants as those using super-critical or ultrasuper-critical technology and low efficiency plants as those using subcritical technology.

We use a difference-in-difference framework to compare power plants by their initial generation efficiency before and after ETS implementation:

$$Y_{it} = \alpha H_i * ETS_t + \gamma_i + \theta_t + X_{it}\beta + \epsilon_{it} \tag{1}$$

where i denotes the power plant and t denotes time in year. The dependent variable Y_{it} denotes operation, electricity generation, carbon emissions and emission intensity of plant i in year t. The key independent variable is the interaction term $H_i * ETS_t$. H_i is a binary variable indicating the level of efficiency, which equals 1 for high efficiency plants, and zero for low efficiency plants. The policy indicator ETS_t equals 1 starting 2021, and zero otherwise. We include plant fixed effects γ_i to absorb plant specific factors and calendar year dummies θ_t to capture annual changes common

to all plants. We also include additional controls X_{it}, including plant-by-year generation capacity and the relevant province-by-year GDP.

The key identification assumption is the absence of pre-existing trends in outcomes by initial technology prior to the ETS. We directly examine whether there are pre-existing trends in event-study analysis, and our findings below support this assumption. Furthermore, we also control for technology group specific time trends, or plant specific time trends, to account for technology specific heterogeneity over time.

The remaining concern could be about other changes in 2021, for example, COVID-19 or other environmental policies. We have read through other policies around the time and do not find that other policies provide differential incentives to power plants by generation efficiency. We also do not think COVID-19 has differential impacts. Nevertheless, we further rule out common shocks to firms by including firm-by-year fixed effects, as power firms usually own multiple power plants. In this robustness check, we essentially compare plants using different generation technologies within firm.

Finally, we disentangle the national program from the regional pilot programs, which have been operational since 2014. The seven pilot regions include two provinces and five cities, which contribute a relatively small percentage of China's total carbon emissions. These pilot regions experimented with hybrid versions of rate-based and mass-based systems before switching to a rate-based system under the national program in 2021. We exclude the seven pilot regions in our analysis of the national program.

6 Results

6.1 Electricity generation and carbon emissions

We start by describing plant-level outcomes before and after the national ETS in an event-study analysis (no regression adjustment). Figure 5 shows annual utilization rates for high efficiency plants and low efficiency plants separately. Prior to the national ETS in 2021, the trends between these two types are relatively parallel. After the launch of the ETS in 2021, high efficiency plants increase their utilization rates, while low efficiency plants remain unchanged in 2021 and show small decreases from 2022. Figure 6 displays trends in annual electricity generation. Similarly, parallel

trends between these two types of plants are observed prior to the ETS. From 2021 to 2023, high efficient plants increase electricity generation, while low efficient plants show little change. These findings are consistent with the theoretical literature.

Table 3 reports the difference-in-difference estimates for utilization rates and electricity generation. Column (1) controls for plant fixed effects and year fixed effects, column (2) adding plant-by-year controls, and column (3) further includes plant-specific time trends. These regression-adjusted estimates yield the same results as the raw data displayed in Figure 5 and Figure 6. In Panel A, column (3) estimates suggest that high efficiency plants increase their annual utilization rate by 1.9 percentage points after the ETS, which is about 4.7 percent of the average utilization rate prior to the ETS. Column (3) of Panel B suggests that high efficiency plants increase their annual electricity generation by about 313 mwh after the ETS, about 7.8 percent of the pre-ETS mean.

For CO_2 emissions, we conduct event-study analysis separately for data based on satellite imagery and data reported by plants. Figure 7 reports CO_2 emissions based on satellite imagery. Similar to electricity generation, CO_2 emissions have parallel trends between the two types of plants prior to the ETS. Since the launch of the ETS in 2021, high efficient plants increase CO_2 emissions, while emissions of low efficiency plants remain relatively unchanged. Figure 8 shows similar patterns using CO_2 emissions data reported by plants before and after the ETS.

The estimated effects of the ETS on CO_2 emissions are very similar between these two data sources. Using data based on satellite imagery, column (3) of Table 4 reports similar estimates that CO_2 emissions of high efficiency plants increase by 278 kilotons after the ETS, about 7.7% of the mean before the ETS. Using data reported by plants, Table 5 suggests that high efficiency plants increase emissions by about 7.6 percent of the pre-ETS mean.

To address the concern of potential confounding factors around 2021, we further rule out common shocks to firms by including firm-by-year fixed effects in Table A.2. Column (1) controls for plant FE, year FE, controls, and firm-by-year FE, and column (2) further includes technology group-specific time trends. We find that our main estimates are robust to this across-plant, within-firm specification.

Finally, we estimate the effects of the ETS on emission intensity. As described in Section 3 , we use two measures of carbon intensity: 1) engineering estimate based on plant characteristics, and 2) estimated estimated measured by reported carbon emissions divided by electricity derived from

satellite imagery. Table 6 and Table 7 report results separately for these two measures. In both tables, we find little change in carbon intensity as response to the ETS.

6.2 Ambient Air Pollution

For our pollution measures using the HYSPLIT model, a "proof-of-concept" question is: How is ambient pollution measured at a pollution monitor affected by being on a trajectory of an emitting plant? To answer this question, we obtain the precise hour when a monitor is "on trajectory" from HYSPLIT, which allows us to map the predicted hourly capacity factor at a plant to the hourly $PM_{2.5}$ readings when a trajectory of the plant arrives at the monitor. Figure A.2 shows that ambient $PM_{2.5}$ increases as the emitting plant increases its generation activity.

Using ambient pollution measures from pollution monitors that incorporate the HYSPLIT model as described in Section 4, we perform a difference-in-difference analysis using our most restrictive specification (including plant-specific time trends). We first examine three pollutants that are likely to be affected by emissions from coal-fired power plants: $PM_{2.5}$, PM_{10}, and SO_2. In columns (1)-(3) of Table 8, we find that the ambient $PM_{2.5}$ on the trajectories of high-efficiency plants increases by about 3.2% after ETS, compared to the trajectory measures of low-efficiency plants. Similarly, PM_{10} increases by about 2.8%, and SO_2 increases by about 4.4%. For the placebo test, we examine three additional pollutants that are less likely to be affected by coal-fired power plants: NO_2, O_3 and CO. In columns (4)-(6), indeed, we find little change in these pollutants after the ETS.

For robustness tests, we use a variety of alternative ways to measure ambient air pollution using the HYSPLIT model. In Appendix Table A.3, we report three sets of robustness checks by using different stack heights, having trajectories run for 120 hours (instead of 72 hours in our main results), and using monitors within 5 km from trajectories. We find results very similar to our main findings.

We also conduct a similar analysis without using the HYSPLIT model. We measure monthly PM2.5 from a plant by the weighted average of pollution readings from monitors within a 50-km radius of a plant, using inverse distance weighting (distance from a plant to a monitor). We used 50 km to be comparable with the pollution measures using the HYSPLIT model, where the medium distance from the plant to the monitors on the trajectories is about 79 km. In Table A.4, we do not find statistically significant increases in all pollution measures following the ETS. For $PM_{2.5}$, PM_{10}, and SO_2, estimates are smaller in magnitude compared to those using the HYSPLIT model

in Table 8, and the standard errors are larger, suggesting more noises in the pollution measure without using the HYSPLIT model.

We also use satellite-derived monthly $PM_{2.5}$ data as an alternative measure of ambient pollution incorporating the HYSPLIT model. In Table A.5, column 1 shows the main result that satellite derived $PM_{2.5}$ increase by 2.6%. The robustness check in column 2 shows similar result.

7 Summary

Using innovative data on carbon emissions developed from satellite imagery, we find that initially more efficient plants increase CO_2 emissions by about 7.7% following China's national ETS, while there is little change in emissions between initially less efficient plants. In a back-of-the-envelope calculation, we estimate that, in the first three years of the National ETS, an increase of 142 million tons of CO_2 emissions per year from coal-fired power plants is induced by the policy. During this period, the aggregate CO_2 emissions in China's coal-fired power sector continue to rise. We calculate that the National ETS contributes to about 39% of this increase. The rest of increase in emissions could be due to increase in energy demand in the post-COVID economy and from new coals plants that come online after 2020 and are therefore not in our analysis sample. Furthermore, we also document increases in ambient air pollution originated from high efficiency plants as responses to the ETS, suggesting environmental costs of the program.

Our findings provide timely insights for China's policy makers in planning the next stages of the program, including tightening assigned benchmarks or potentially switching to a cap-and-trade program. These findings also provide insights for other emerging countries that have adopted or will adopt a similar program in consideration of their policy stringency.

References

Almond, D. and S. Zhang (2021). Carbon-trading pilot programs in china and local air quality. *AEA Papers and Proceedings 111*, 391–95.

Asia Development Bank (2016). Emissions trading schemes and their linking: Challenges and opportunities in asia and the pacific. Technical report, Asia Development Bank.

Barwick, P., S. Li, L. Lin, and E. Zou (2024). From fog to smog: The value of pollution information from fog to smog: The value of pollution information. *American Economic Review 114*(5), 1338–81.

Chen, X. and B. Lin (2021). Towards carbon neutrality by implementing carbon emissions trading scheme: Policy evaluation in china. *Energy Policy 157*, 112510.

Couture, H. D., M. Alvara, J. Freeman, A. Davitt, H. Koenig, A. Rouzbeh Kargar, J. O'Connor, I. Söldner-Rembold, A. Ferreira, J. Jeyaratnam, J. Lewis, C. McCormick, T. Nakano, C. Dalisay, C. Lewis, G. Volpato, M. Gray, and G. McCormick (2024). Estimating carbon dioxide emissions from power plant water vapor plumes using satellite imagery and machine learning. *Remote Sensing 16*(7).

Cui, J., C. Wang, Z. Wang, J. Zhang, and Y. Zheng (2023). Carbon leakage within firm ownership networks. *Available at SSRN 4514971*.

Cui, J., C. Wang, J. Zhang, and Y. Zheng (2021). The effectiveness of china's regional carbon market pilots in reducing firm emissions. *Proceedings of the National Academy of Sciences 118*(52).

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2013). Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from india. *Quarterly Journal of Economics 128*(4), 1499–1545.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2018). The value of regulatory discretion: Estimates from environmental inspections in india. *Econometrica 86*(6), 2123–2160.

EDF (2021). 2019-2020 implementation plan for national carbon emission trading total allowances setting and allocation (power generation industry). Technical report, EDF.

Fischer, C. (2001). Rebating environmental policy revenues: output-based allocations and tradable performance strandards. Technical report, Resource for the Future.

Fischer, C. and R. G. Newell (2008). Environmental and technology policies for climate mitigation. *Journal of Environmental Economics and Management 55*(2), 142–162.

Fullerton, D. and C. Wolfram (2016). Report 2016:2 environmental and energy economics. Report 2016:2, National Bureau of Economic Research.

Goulder, L. H., X. Long, J. Lu, and R. D. Morgenstern (2022). China's unconventional nationwide co2 emissions trading system: Cost-effectiveness and distributional impacts. *Journal of Environmental Economics and Management 111*, 102561.

Goulder, L. H., X. Long, C. Qu, and D. Zhang (2023). China's nationwide co2 emissions trading system: A general equilibrium assessment.

Grainger, C. and T. Ruangmas (2018). Who wins from emissions trading? evidence from california. *Environmental and resource economics 71*, 703–727.

Greenstone, M. (2024, May). The economics of the global energy challenge. *AEA Papers and Proceedings 114*, 1–30.

Greenstone, M., G. He, R. Jia, and T. Liu (2022). Can technology solve the principal-agent problem? evidence from china's war on air pollution. *AER: Insights 4*(1), 54–70.

Greenstone, M., R. Pande, A. Sudarshan, and N. Ryan (2023). Can pollution markets work in developing countries? experimental evidence from india.

Hernandez-Cortes, D. and K. C. Meng (2023). Do environmental markets cause environmental injustice? evidence from california's carbon market. *Journal of Public Economics 217*, 104786.

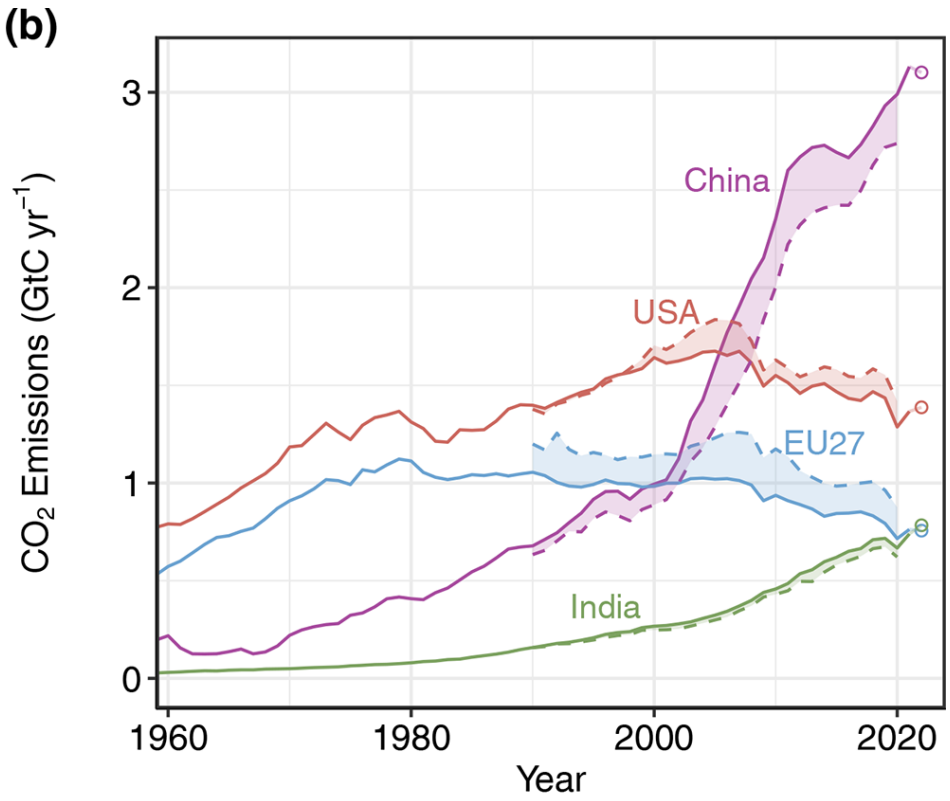
Karplus, V., S. Zhang, and D. Almond (2018). Quantifying coal power plant responses to tighter so2 emissions standards in china. *Proceedings of the National Academy of Sciences 115*(27), 7004–7009.

Shen, S., C. Li, A. van Donkelaar, N. Jacobs, C. Wang, and R. V. Martin (2024, 05). Enhancing global estimation of fine particulate matter concentrations by including geophysical a priori information in deep learning. *ACS ES&T Air 1*(5), 332–345.

Zhang, W., J. Li, G. Li, and S. Guo (2020). Emission reduction effect and carbon market efficiency of carbon emissions trading policy in china. *Energy 196*, 117117.

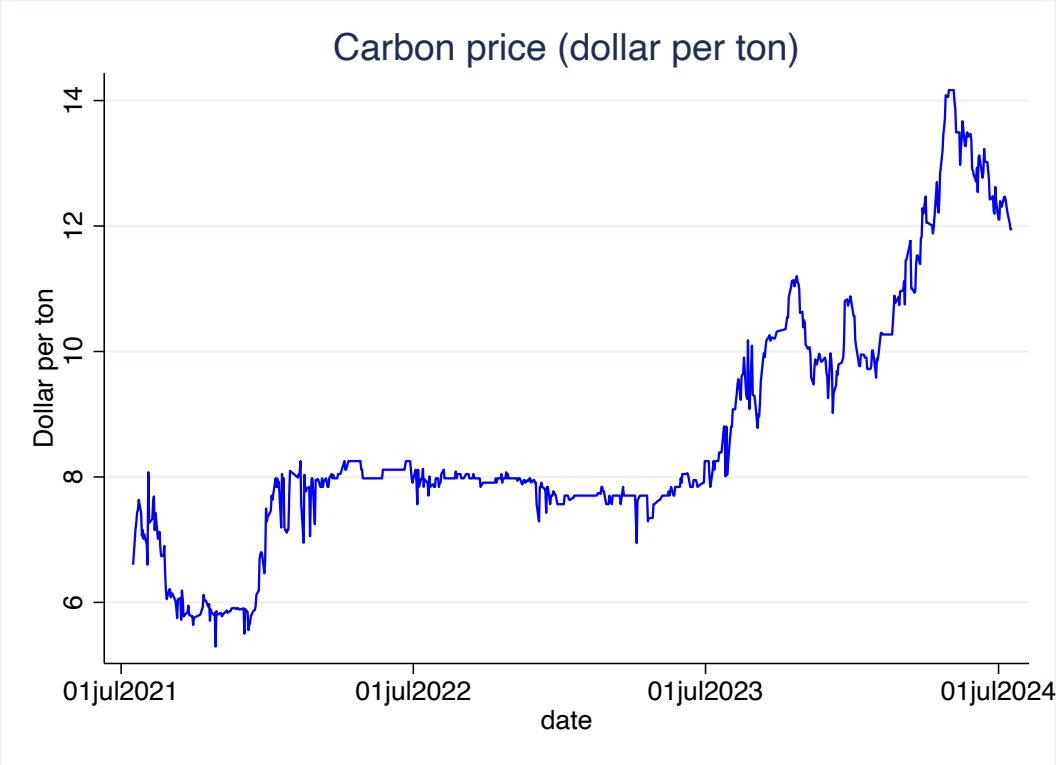
Figures

Figure 1



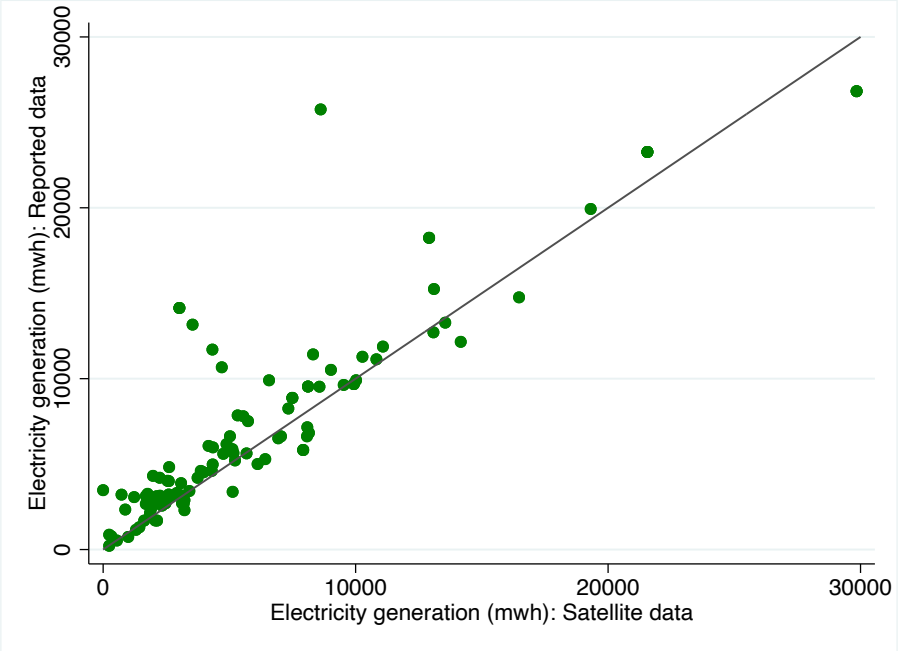
Notes: This figure shows

Figure 2: Carbon price



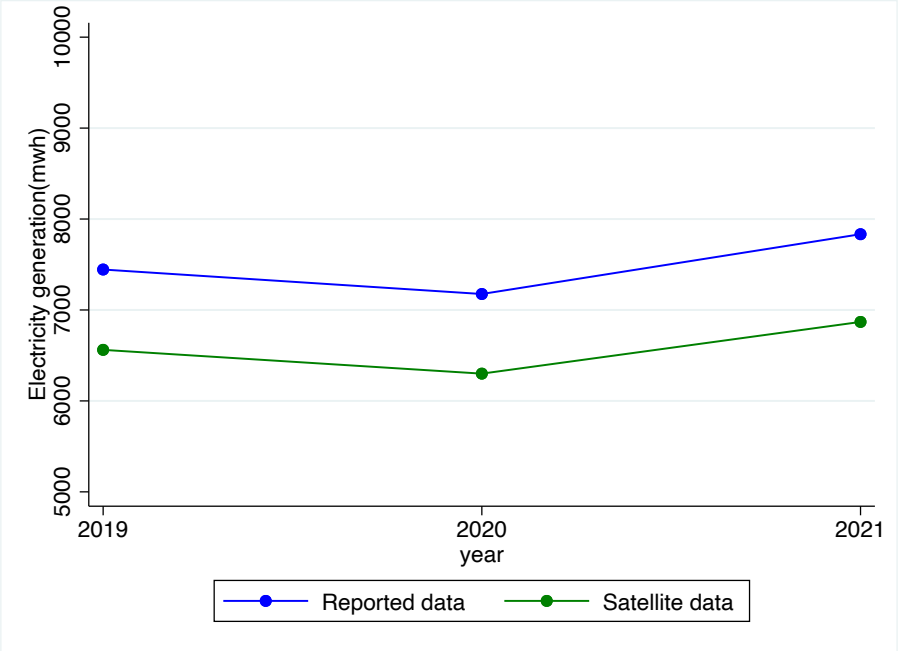
Notes: This figure shows

Figure 3: Electricity generation in 2020



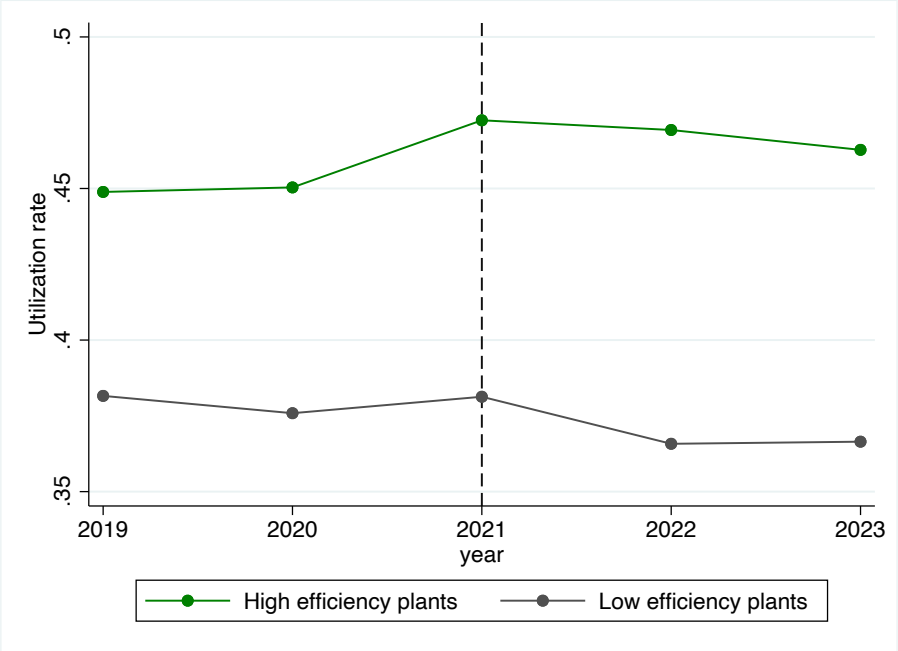
Notes: This figure shows

Figure 4: Electricity generation in 2019-2021



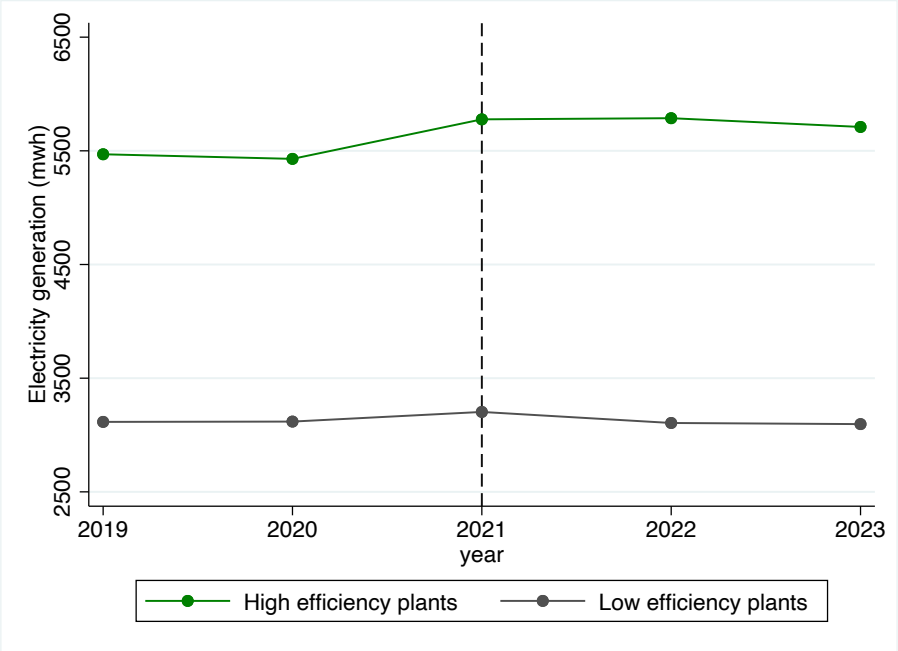
Notes: This figure shows

Figure 5: Utilization rate



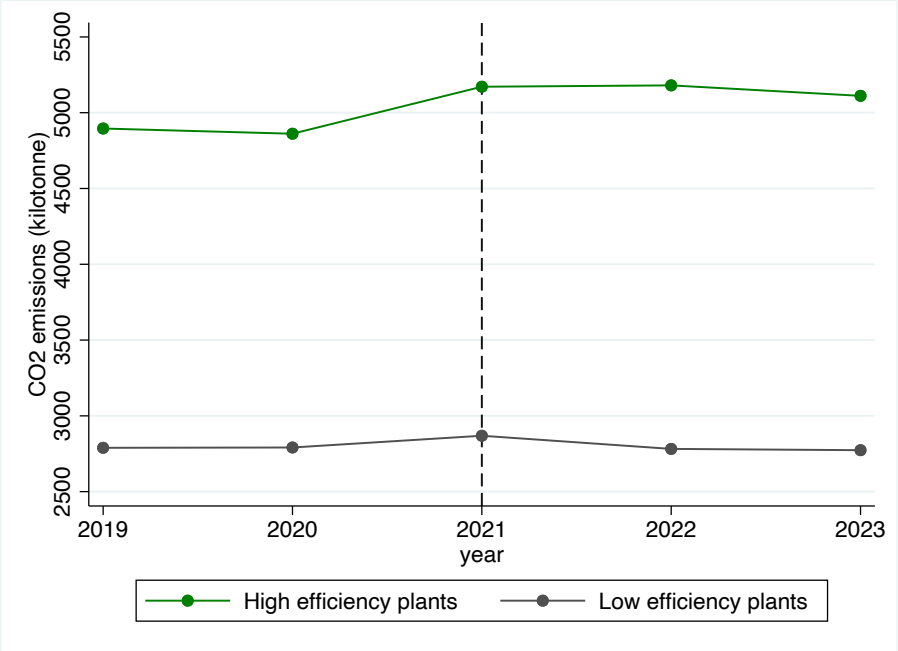
Notes: This figure shows

Figure 6: Electricity generation



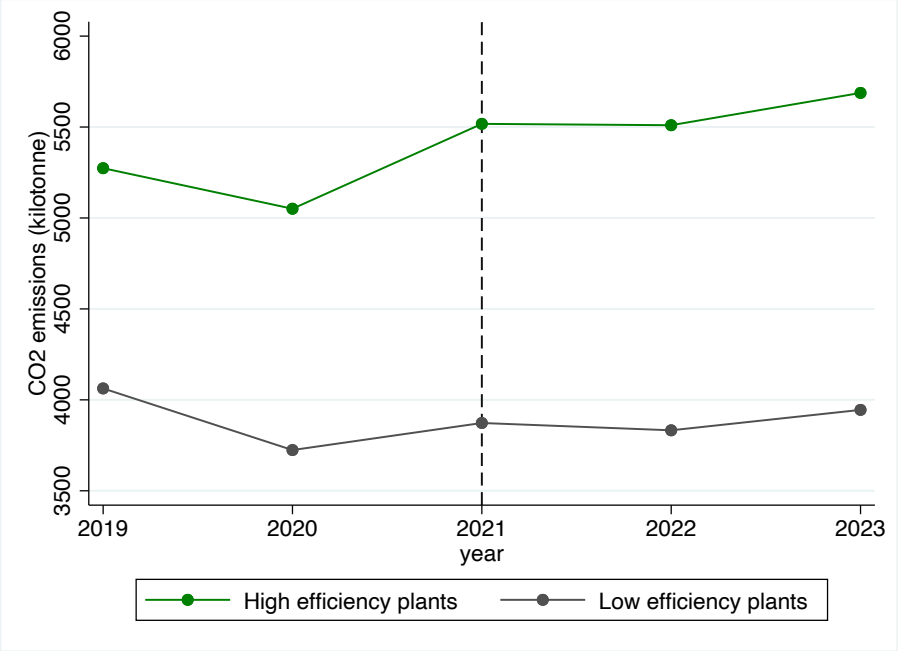
Notes: This figure shows

Figure 7: CO2 emissions: based on satellite imagery



Notes: This figure shows

Figure 8: CO2 emissions: reported by plants



Notes: This figure shows

Tables

Table 1: Summary statistics (data from satellite imagery)

	(1) All plants	(2) More efficient plants	(3) Less efficient plants
Total capacity (MW)	1,199	1,351	877
	(886)	(812)	(712)
Annual utilization rate	0.40	0.46	0.37
	(0.14)	(0.12)	(0.13)
Annual electricity generation (GWh)	4,625	5,635	3,128
	(4,118)	(3,968)	(3,186)
Annual CO2 emissions (Kilotonne)	4,115	5,044	2,801
	(3,758)	(3,547)	(2,860)
Number of power plants	512	163	251

Notes:

Table 2: Summary statistics (data from firm reports)

	(1) All plants	(2) More efficient plants	(3) Less efficient plants
Total capacity (MW)	1,171	1,396	912
	(641)	(671)	(489)
Annual CO2 emissions (Kilotonne)	4,695	5,407	3,873
	(2,480)	(2,710)	(1,875)
Number of power plants	512	150	130

Notes:

Table 3: Utilization rate and electricity generation

	(1)	(2)	(3)
	Panel A: Utilization rate		
H*ETS	0.026***	0.019***	0.019**
	(0.007)	(0.007)	(0.007)
Observations	2,070	2,014	2,014
R^2	0.03	0.03	0.64
Pre-ETS mean	0.407	0.407	0.407
	Panel B: Electricity generation		
H*ETS	209.9***	312.8***	313.4***
	(69.7)	(78.5)	(88.2)
Observations	2,070	2,014	2,014
R^2	0.38	0.38	0.74
Pre-ETS mean	4035.4	4035.4	4035.4
Plant FE	Y	Y	Y
Year FE	Y	Y	Y
Controls	Y	Y	Y
Technology specific time trends	N	Y	N
Plant specific time trends	N	N	Y

Notes:

Table 4: Carbon emissions: based on satellite imagery

	CO2 emissions		
	(1)	(2)	(3)
H*ETS	185.7***	277.8***	278.4***
	(62.6)	(70.2)	(78.9)
Observations	2,014	2,014	2,014
R^2	0.38	0.38	0.74
Pre-ETS mean	3612.5	3612.5	3612.5
Plant FE	Y	Y	Y
Year FE	Y	Y	Y
Controls	Y	Y	Y
Technology specific time trends	N	Y	N
Plant specific time trends	N	N	Y

Notes:

Table 5: Carbon emissions: reported by plants

	CO2 emissions		
	(1)	(2)	(3)
H*ETS	349.7***	350.3**	341.2*
	(121.2)	(159.0)	(176.7)
Observations	1,400	1,400	1,400
R^2	0.14	0.14	0.49
Pre-ETS mean	4568.5	4568.5	4568.5
Plant FE	Y	Y	Y
Year FE	Y	Y	Y
Controls	Y	Y	Y
Technology specific time trends	N	Y	N
Plant specific time trends	N	N	Y

Notes:

Table 6: Carnon intensity: engineering estimate

	CO2 emission intensity		
	(1)	(2)	(3)
H*ETS	-0.0007	-0.0002	-0.0003
	(0.0007)	(0.0003)	(0.0004)
Observations	1,991	1,991	1,991
R^2	0.00	0.00	0.76
Plant FE	Y	Y	Y
Year FE	Y	Y	Y
Controls	N	Y	Y
Plant specific time trends	N	N	Y

Notes:

Table 7: Carnon intensity: reported estimate

	CO2 emission intensity		
	(1)	(2)	(3)
H*ETS	-0.092	-0.093	0.199
	(0.253)	(0.254)	(0.203)
Observations	859	859	859
R^2	0.01	0.01	0.52
Plant FE	Y	Y	Y
Year FE	Y	Y	Y
Controls	N	Y	Y
Plant specific time trends	N	N	Y

Notes:

Table 8: Ambient pollution at pollution monitors (using HYSPLIT)

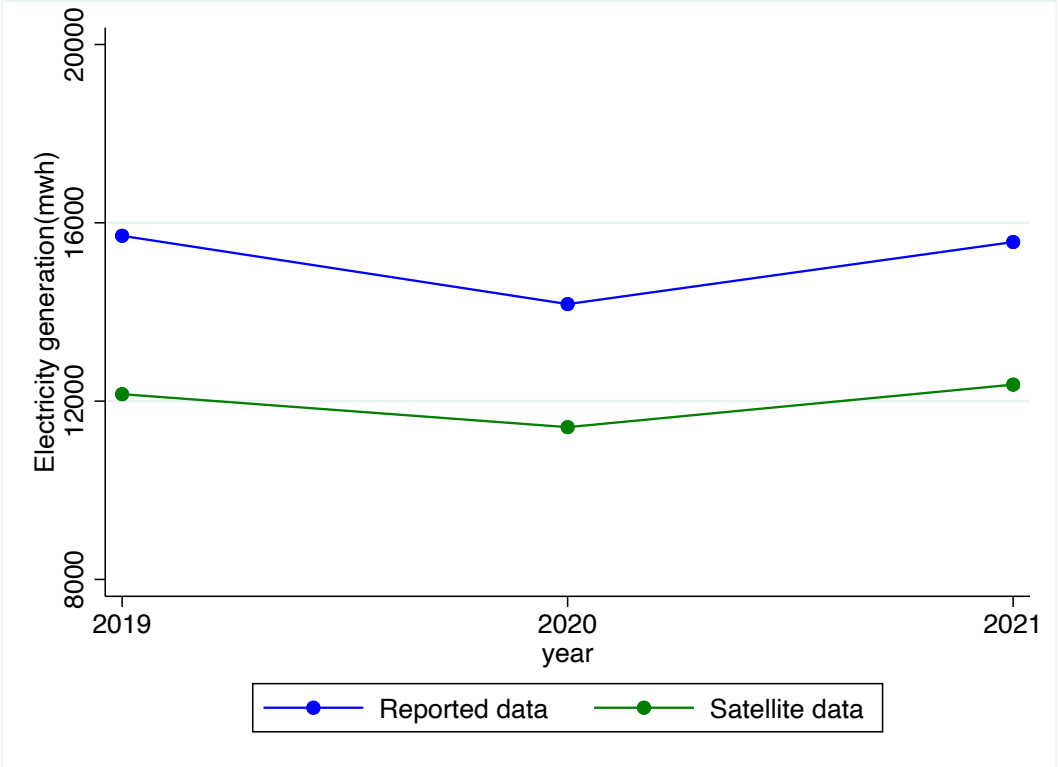
	(1) ln(PM2.5)	(2) ln(PM10)	(3) ln(SO_2)	(4) ln(NO_2)	(5) ln(O_3)	(6) ln(CO)
H*ETS	0.0320**	0.0284*	0.0440**	0.000339	0.00920	0.0100
	(0.0153)	(0.0155)	(0.0205)	(0.0151)	(0.0126)	(0.0154)
Observations	42,290	42,290	42,290	42,290	42,290	42,290
R^2	0.84	0.87	0.86	0.88	0.81	0.86
Plant FE	Y	Y	Y	Y	Y	Y
Year-by-Month FE	Y	Y	Y	Y	Y	Y
Weather controls	Y	Y	Y	Y	Y	Y
Plant specific time trends	Y	Y	Y	Y	Y	Y

Standard errors are clustered at the plant level and reported in parenthesis.

* significant at 10% ** significant at 5%. *** significant at 1%.

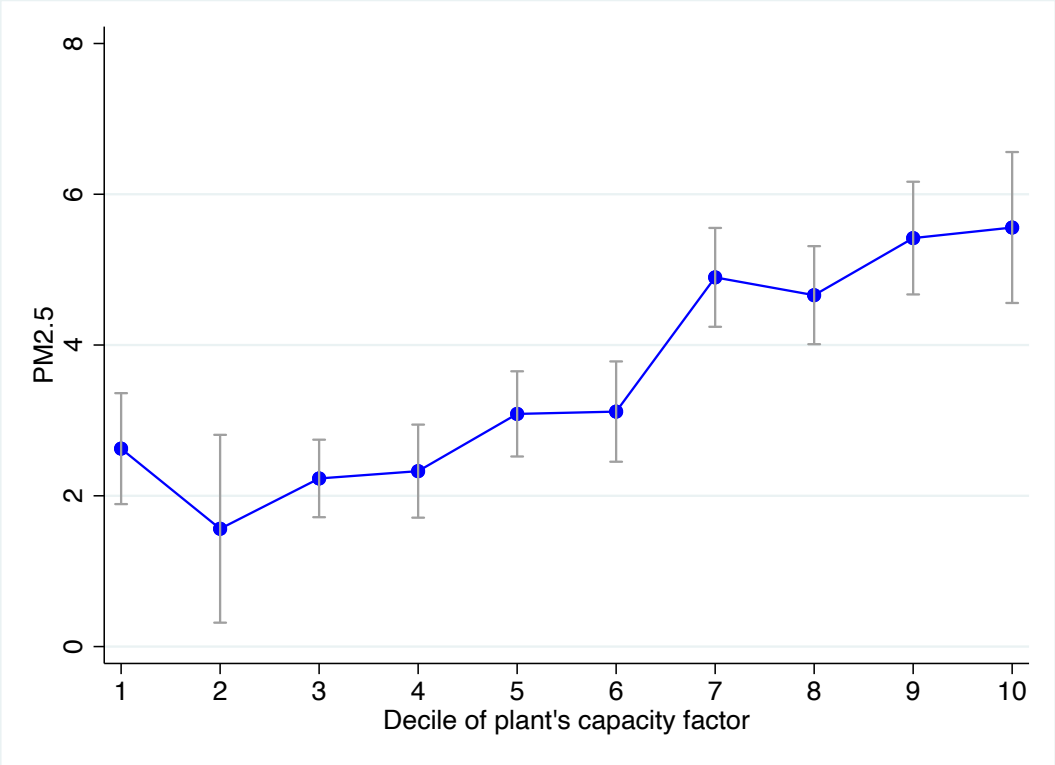
Additional Figures

Figure A.1: Electricity generation in India: 2019-2021



Notes: This figure shows

Figure A.2: Plant's capacity factor and ambient PM2.5: using HYSPLIT



Notes: This figure shows

Additional Tables

Table A.1: Summary statistics of generation units

	(1) Generation units of all power plants	(2) Generation units in the Satellite sample
400 MW or above rated power (%)	23	32
Super/Ultra-supercritical units (%)	27	34
Super/Ultra-supercritical among 400 MW units (%)	78	75
Emission intensity above the benchmark (%)	76	77
Number of generation units	2802	1358

Notes:

Table A.2: Robustness check: include parent firm-by-year FE

	(1)	(2)
	Panel A: Electricity generation	
H*ETS	218.318**	271.021**
	(92.066)	(107.275)
Observations	2,014	2,014
R^2	0.55	0.55
	Panel B: CO2 emissions: based on satellite imagery	
H*ETS	195.7**	240.4**
	(82.6)	(95.8)
Observations	2,014	2,014
R^2	0.55	0.55
	Panel C: CO2 emissions: reported by plants	
H*ETS	278.1**	257.0*
	(114.8)	(143.3)
Observations	1,400	1,400
R^2	0.53	0.53
Plant FE	Y	Y
Year FE	Y	Y
Controls	Y	Y
Parent firm-by-Year FE	Y	Y
Technology specific time trends	N	Y

Notes:

Table A.3: Ambient pollution at pollution monitors: Robustness checks (with HYSPLIT)

	(1) ln(PM2.5)	(2) ln(PM10)	(3) ln(SO_2)	(4) ln(NO_2)	(5) ln(O_3)	(6) ln(CO)
	Panel A: stack height is 200 meters					
H*ETS	0.0299*	0.0266*	0.0411**	0.00131	0.00582	0.00628
	(0.0157)	(0.0159)	(0.0208)	(0.0155)	(0.0130)	(0.0160)
Observations	42,289	42,289	42,289	42,289	42,289	42,289
R^2	0.84	0.87	0.86	0.88	0.81	0.86
	Panel B: trajectories run for 120 hours					
H*ETS	0.0320**	0.0282*	0.0443**	-0.0000738	0.00877	0.0101
	(0.0154)	(0.0156)	(0.0206)	(0.0152)	(0.0127)	(0.0155)
Observations	42,242	42,242	42,242	42,242	42,242	42,242
R^2	0.84	0.87	0.86	0.88	0.81	0.86
	Panel C: monitors within 5 km from trajectories are matched					
H*ETS	0.0325**	0.0284*	0.0441**	0.00345	0.00787	0.0127
	(0.0153)	(0.0157)	(0.0209)	(0.0151)	(0.0127)	(0.0155)
Observations	42,285	42,285	42,285	42,285	42,285	42,285
R^2	0.84	0.87	0.86	0.88	0.81	0.86
Plant FE	Y	Y	Y	Y	Y	Y
Year-by-Month FE	Y	Y	Y	Y	Y	Y
Weather controls	Y	Y	Y	Y	Y	Y
Plant specific time trends	Y	Y	Y	Y	Y	Y

Standard errors are clustered at the plant level and reported in parenthesis.

* significant at 10% ** significant at 5%. *** significant at 1%.

Table A.4: Ambient pollution at pollution monitors within 50-km from plants (without HYSPLIT)

	(1) ln(PM2.5)	(2) ln(PM10)	(3) ln(SO_2)	(4) ln(NO_2)	(5) ln(O_3)	(6) ln(CO)
H*ETS	0.0175	0.0147	0.0321	-0.0128	-0.00930	-0.00227
	(0.0245)	(0.0246)	(0.0335)	(0.0239)	(0.0220)	(0.0262)
Observations	32,523	32,524	32,524	32,524	32,524	32,520
R^2	0.83	0.85	0.82	0.86	0.82	0.82
Plant FE	Y	Y	Y	Y	Y	Y
Year-by-Month FE	Y	Y	Y	Y	Y	Y
Weather controls	Y	Y	Y	Y	Y	Y
Plant specific time trends	Y	Y	Y	Y	Y	Y

Standard errors are clustered at the plant level and reported in parenthesis.

* significant at 10% ** significant at 5%. *** significant at 1%.

Table A.5: Satellite measures of PM2.5 (with HYSPLIT)

	ln(PM2.5)	
	(1) Main	(2) Robustness
H*ETS	0.026***	0.024**
	(0.010)	(0.010)
Observations	26,648	26,830
R^2	0.68	0.68
Plant FE	Y	Y
Year FE	Y	Y
Weather controls	Y	Y
Plant specific time trends	Y	Y