Enforcing Regulation when Violations are Heterogeneous: Empirical Evidence from U.S. Stationary Emissions Policy*

Almira Salimgarieva†

Wesley Blundell‡

November 9, 2023

Abstract

Enforcement of environmental regulation in the United States is often delegated from the federal level to local authorities. This devolution of responsibility poses a significant challenge when local regulators differ in their knowledge and priorities regarding the harm posed by environmental violations. Using plant-level data from the Environmental Protection Agency (EPA), we exploit variation in the application of a 2014 revision to the criteria for classifying severe violations under the Clean Air Act (CAA). We find that following the revision, plants located in states most impacted by the policy exhibited a greater decrease in emissions. As a result, the overall emissions-related damages from stationary sources of air pollution decreased by 2.5%, equivalent to \$2.4 billion annually. These results provide quasi-experimental evidence on the effectiveness of limiting regulatory discretion and the importance of marginal deterrence in enforcement.

*We thank Trudy Cameron, Ruoyu Chen, Joseph Cook, Ben Cowan, Dietrich Earnhart, Ana Espinola-Arredondo, Mary Evans, Timo Goeschl, Jill McCluskey, Felix Munoz-Garcia, and Will Wheeler as well as seminar participants at Washington State University, the 2022 Western Economic Association International Summer Conference, and the 2021 Association of Environmental and Resource Economists Summer Conference. Special thanks to Michael Greenstone for providing the TRI chemical-to-pollutant mapping data.

†almira.salimgarieva@wsu.edu, Washington State University

‡wesley.blundell@wsu.edu, Washington State University

1 Introduction

Laws and regulations often have varying types of violations and degrees of possible infringement. For example, drivers can choose how much they exceed the speed limit as well as whether they ignore traffic signs. According to the economic literature, optimal enforcement in these settings is characterized by tailoring the level of monitoring or punishment for a violation according to its marginal harm [Mookherjee and Png, 1994, Polinsky and Shavell, 1998]. However, in practice it is often difficult to achieve this ideal. The marginal damage from a violation may vary across jurisdictions where the regulation is enforced. Furthermore, regulatory resources and costs may vary across jurisdictions.

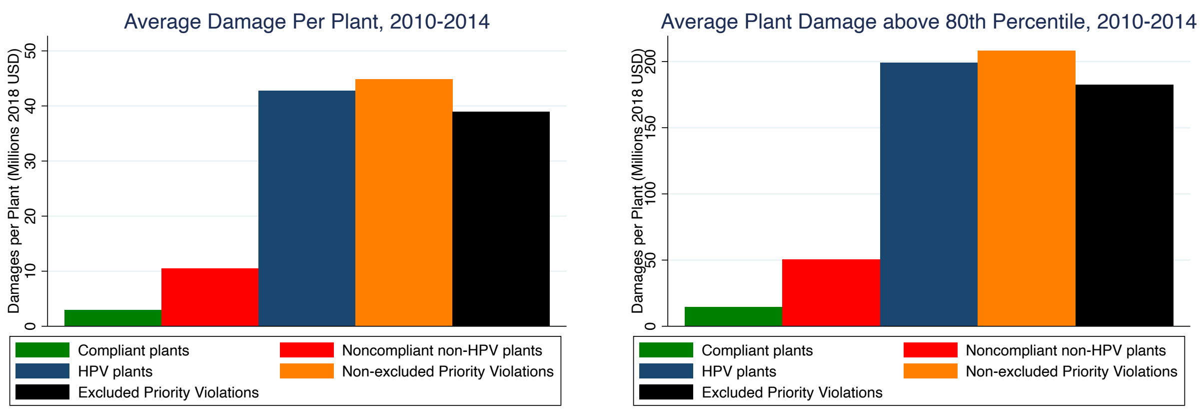
In light of these difficulties, there has been an extensive debate on the degree to which the enforcement of environmental regulation should be centralized and how much discretion to be granted to local authorities [Oates, 1999]. If local authorities have greater knowledge regarding the types of violations most harmful to their jurisdiction, then greater enforcement discretion will improve local welfare. Conversely, differing incentives or interests of local authorities may prevent optimal enforcement across violation types, even when damages are known. Currently, there is limited empirical evidence on whether reducing regulatory discretion in the enforcement and monitoring of heterogeneous violations could reduce environmental harm [Zhang et al., 2018, Kang and Silveira, 2021, Mu et al., 2021]. Recent decreases in environmental enforcement resources [Kelderman et al., 2019, Gray and Shimshack, 2011, Evans, 2016] suggest that the question of how and whether local regulators optimally reallocate enforcement resources when faced with reduced discretion is of significant policy interest.

In this paper, we examine this question by using quasi-experimental variation in the criteria for enforcement of violations with heterogeneity in damages. This experiment follows the EPA's 2014 revision of the criteria for defining high priority violations (HPV) under the Clean Air Act (CAA). Specifically, the revision redefined which plants are classified as priority violators by delisting a number of violations from the listing criteria, with the intent to "focus on CAA violations that experience shows are most likely to be significant for human health and the environment" [EPA, 2014]. Consistent with this stated intent, the most common types of delisted criteria are those focused on the failure of a facility to submit a Clean Air Act operating permit or other certification, violations that appear unlikely to cause significant harm as compared to the non-excluded criteria focused on exceeding emissions limits.

Using data on the marginal damages for various pollutants from Clay et al. [2019], along with EPA compliance and emissions information, we confirm this is also the case empirically.[1] Figure 1 illustrates that prior to the policy change, the average estimated annual damages for priority violators was almost four times

[1] These damage estimates are based on a number of factors including mortality, decreased agricultural output, and morbidity.

the damages from plants with other types of violations. Furthermore, among these priority violators, plants with the delisted violations had lower damages than priority violators whose classification stemmed solely from violations that were not reclassified. These gaps are also present at the 80th percentile of damages for each group illustrated in Figure 1. Even at the top end of the damage distribution, priority violators with non-excluded violations present the highest level of harm. Given the higher levels of regulatory scrutiny for priority violators compared to plants with other types of violations [Blundell et al., 2020], these figures are consistent with the EPA's stated intent of the policy.



(a) All plants

(b) 80th Percentile in Compliance Group

Figure 1: Differences in damages by compliance status

Note: This figure shows the average plant-level damages by regulatory status. The last two bars present mean damages for HPVs excluded under the 2014 policy and non-excluded HPVs, respectively.

Using a highly detailed dataset on environmental enforcement and performance for 81,432 industrial facilities in the United States from 2010 to 2019, we estimate a "fuzzy" difference-in-differences (DID) model [De Chaisemartin and d'Haultfoeuille, 2018] by comparing plants from states with a high incidence of the delisted violations to plants from states with low incidences of the delisted violations. We consider plants from states with a high incidence as having a higher level of treatment, since regulators from these states should experience a greater change in their enforcement priorities and practices following the policy change. We provide suggestive evidence that treatment corresponds with an 78% increase in their average penalties for priority violations. We find no similar increase in penalties for non-HPV violations, consistent with the understanding that treatment should correspond with greater enforcement toward the violations that posed the greatest harm.

Our primary result is that the reduced discretion for classifying priority violators from the policy led to a 2.4% reduction in emissions and a 2.5% reduction in damages for plants on average. These percentage effects are comparable to other recent estimated air pollution impacts of policy changes found in the literature [Auffhammer et al., 2009, Walter and Raff, 2019, Gibson, 2019, Zou, 2021, Bi, 2017]. In addition, our estimated damage effect translates to a \$2.4 billion (2018 \$) reduction in the annual damages from particulate matter (PM2.5), volatile organic compounds (VOCs), nitrogen oxides (NOX), sulfur dioxide (SO2), and ammonia (NH3).[2] We find qualitatively similar results when using alternative pollution data on particulate matter and VOCs for a subset of high pollution facilities, available from the Toxic Release Inventory. Overall, these results indicate that there are substantial benefits from policies that focus enforcement resources on violations with the highest level of damages, by reducing regulatory discretion.

Estimating the causal impact of the policy revision on plant emissions and damages poses a number of challenges. First, the policy could have been implemented due to a differential trend in the environmental performance of plants located in states with a high incidence of the delisted violations or treatment. Second, recent work in the literature indicates there is potential pollution substitution from plants subject to greater environmental enforcement to other plants within the same parent firm with less oversight [Gibson, 2019, Rijal and Khanna, 2020]. Finally, there is potential for general equilibrium effects with the substitution of production from plants in treated states to plants located in states with a lower intensity of treatment [Evans et al., 2018].

We address these challenges through several robustness checks and alternative specifications. Using an event study design, we rule out the existence of differential trends in the environmental damages from plants in states with a high treatment intensity compared to plants from states with a low treatment intensity prior to the policy. Further, we find statistically significant effects of the policy that persist to the end of our sample time frame. In addition, we find that state-level treatment intensity is uncorrelated with the individual state environmental enforcement agency budgets or federally evaluated performance. Next, results from a subsample of plants with single-state parent firms yield qualitatively similar estimates to our primary results. Finally, we compare the effect of the policy on plants in sub-industries with a high level of treatment heterogeneity to that on plants in sub-industries with little heterogeneity and find no evidence that general equilibrium effects bias our results.

Our primary contribution is to the literature on environmental federalism, the devolution of primary enforcement responsibility to individual states or local enforcement agencies [Oates, 1999]. Specifically, our finding that the environmental performance of individual plants improved after the reclassification of priority violators at the federal level contributes to the recent empirical debate on the value of regulatory discretion. Using structural modeling, Kang and Silveira [2021] and Duflo et al. [2018] find that the elimination

[2] Particulate matter, nitrogen oxides, and sulfur dioxide comprise three of the six criteria air pollutants under the CAA, while VOCs are considered a direct precursor to secondary PM2.5 and another criteria pollutant, ozone.

of individual regulator's discretion in penalties or inspections would raise enforcement costs while lowering environmental compliance, for the contexts of Clean Water Act enforcement in the U.S. and air regulation in India. In contrast, using a natural experiment based on plants in China, Zhang et al. [2018] find that centralized supervision and oversight increases the environmental performance of polluting facilities. Similarly, Zhang and Khanna [2021] and Mu et al. [2021] find that local regulators do not properly allocate enforcement resources and monitoring to areas in violation of federal air emissions standards in the United States. Our work differs from these previous studies in two important ways. First, unlike much of the previous literature, our analysis focuses on multiple air pollutants rather than water pollution or a single air pollutant. Second, we quantify the monetary damages and benefits of decreased regulatory discretion by incorporating local marginal damage estimates from Clay et al. [2019] for the various air pollutants considered in our sample.

Second, we contribute to the empirical literature on the enforcement of environmental regulation. Findings from this literature have established relationships between enforcement intensity and environmental performance both in terms of regulatory compliance [Shimshack and Ward, 2008, Evans, 2016] and overall emissions [Hanna and Oliva, 2010, Earnhart, 2004]. In particular, our suggestive finding that after the 2014 policy change regulators increased enforcement of the most severe types of violations contributes to recent discussions on the impact of differences in enforcement priorities between state and federal regulators [Earnhart and Friesen, 2021a,b] as well as discussions regarding the value of marginal deterrence in regulatory enforcement [Blundell et al., 2020, Leisten and Vreugdenhil, 2023]. Analogously, this paper contributes to the literature on criminal depenalization by examining the elimination of some criteria for classifying severe environmental violators. To the best of our knowledge, the current literature on depenalization largely focuses on drug crime, and the empirical results on depenalization are mixed. MacCoun et al. [2001] document modest benefits across studies of depenalization, while a more recent work, Adda et al. [2014], has found negative effects.

The remainder of the paper proceeds as follows. Section 2 describes the regulatory setting and policy background. Section 3 proposes a model of how restricting the criteria for classifying priority violators affects compliance and emissions. Section 4 reports the data sources and defines the monetary emissions damages for plants. Sections 5 presents the empirical strategy and results. Section 6 provides the results of alternative specifications and robustness checks, while section 7 concludes.

2 Institutional Details

2.1 Clean Air Act Enforcement

Originally, the CAA was passed in 1963 with the intent to reduce the level of air pollutants harmful to human health through research and monitoring of emission sources. Continued amendments to the act between 1965 and 1990, including the founding of the federal Environmental Protection Agency (EPA) in 1970, have led to a system of enforced standards that cover a variety of air pollutants. These standards are primarily command and control, mandating the installation of pollution abatement technology to reduce the emission of criteria or hazardous air pollutants, which are known to be harmful to human health and the environment.[3]

Enforcement of these standards is carried out by individual state environmental protection agencies through a system of inspections and penalties to address violations. In certain cases, federal regulation will directly specify a minimum frequency and magnitude for enforcement actions. For example, the regularity of inspections for large industrial facilities depends on whether the facility resides in a county that is in non-attainment with federal National Ambient Air Quality Standards. Smaller facilities, or those located in areas designated as in attainment, may face relatively infrequent rates of inspection, or have no regular monitoring schedule. Enforcement actions can also be undertaken directly by the federal EPA through regional offices. However, there is generally a significant amount of discretion in the type of enforcement approach taken, as federal EPA guidelines explicitly state that "regions and states can take varied approaches to improving state enforcement programs" [EPA, 1991]. Further, it is often the case that the more severe cases of noncompliance subject to joint enforcement efforts between state, regional, and federal regulators initially began with a state-level regulatory investigation.

Although individual state agencies have discretion in CAA enforcement, the federal EPA ensures a minimum level of environmental quality and enforcement through the state review framework (SRF). Specifically, individual states are evaluated every five years with respect to their performance regarding inspections, penalties, identification of violations, record keeping, and timely enforcement of violations. If a state agency is found deficient, it is expected to rectify it by the next review. If the state agency fails to abide by SRF recommendations, potential punishments can range from retraining of the agency staff to loss of CAA primacy, the responsibility of enforcing the CAA within the state. However, states are rarely penalized under the SRF since the majority of deficiencies are corrected by the next review. For example, 85% of recommendations from the first round of the SRF, which began in 2004, were corrected by the second round, and to date no state has lost primacy with CAA enforcement. Often, these reviews are focused on the enforcement of a

[3]Pollutants identified as criteria air pollutants are particulate matter (PM), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen oxides (NO_x), ground-level ozone, and lead.

particular subset of plants, those classified as having high priority violations (HPVs) under the CAA.

2.2 HPV Criteria

Under the initial guidance of the HPV policy in 1998, the HPV classification was "designed to direct scrutiny to those violations that are most important" [EPA, 1999a].[4] Specifically, if a violation met one of the ten general criteria or four matrix criteria for HPV classification, it likely posed greater potential harm than other types of violations and therefore merited greater enforcement priority. It has been documented in the literature that HPVs receive higher levels of regulatory scrutiny as compared to other plants [Blundell et al., 2020].

The criteria for HPV classification can range from "Failure to obtain a Prevention of Significant Deterioration or New Source Review permit" to "violation of allowed emissions limit" [EPA, 1998] A subset of these criteria (e.g., "Violation by a chronic or recalcitrant violator") allow for both regulatory discretion as well as dynamic enforcement, that is, the differential treatment of plants according to their previous violation history. In addition to the potential discretion in classification, penalties for violators themselves are also subject to some degree of regulatory discretion. Specifically, factors within the EPA's penalty guidelines, such as the plants "degree of willfulness or negligence" and "degree of cooperation", are inherently subjective [EPA, 1991]. The variation and discretion in these criteria make it plausible that there is significant heterogeneity in environmental outcomes, although according to the SRF each state should individually meet federal expectations for enforcement.

Following 16 years of HPV enforcement, in 2014 the EPA revised the criteria for the classification of HPVs. In particular, the revision resulted in the complete removal of all matrix criteria and three general classification criteria. Although the number of matrix criteria pre-2014 varies by one between the original policy [EPA, 1998] and later documents [EPA, 1999a], the fifth criterion is encapsulated within the original four criteria and we do not treat it separately, which is consistent with the documentation of our data [EPA, 2015]. In the appendix we list the revised criteria in Table A1 alongside the original criteria in Table A2, with the criteria numbered according to how they are described in the remainder of this subsection. To our knowledge there is no published mapping between the original 1998 HPV criteria and the revised 2014 criteria. However, for non-excluded criteria there is a reasonable mapping between the two sets of criteria.

First, although there are only six revised general criteria, it is our view that new criterion six encapsulates both old general criteria nine and ten. For the case of old general criterion nine, the note by the EPA in the original documentation [EPA, 1998] states that *persistent or recalcitrant* is noncompliant behavior when "it is mutually agreed by the region and the delegated agency that the source should be bumped up into

[4] We use a priority *violation* and a priority *violator* interchangeably to refer to HPV.

HPV status" fits with the discretionary nature of the sixth criterion in the 2014 revision. The section 112 requirements for the tenth general criterion also fit with the discretionary nature of the new sixth criterion as “The section 112(g) provision is designed to ensure that emissions of toxic air pollutants do not increase if a facility is constructed or reconstructed" and is meant to be a “transitional measure" until the modifications to the polluting facility are completed and final emissions limits are established [EPA, 1999b].

Both the old and revised criteria one are primarily concerned with the installation of "Best Available Control Technology" (BACT) and “Lowest Available Emissions Reductions" (LAER). Old general criteria two, three, and eight are concerned with violations of emissions limits for differing regulations. These excess emissions criteria are encapsulated by the revised criteria two, three, and four, to the extent that they “recur (or recurred) regularly or intermittently for at least seven days." Lastly, for the non-excluded criteria, the focus of the original seventh criterion on interference with enforcement or determining a source's compliance or emissions fits within the framework of the new fifth criterion.

This leaves general criteria four, five, and six, along with the original matrix criteria as excluded from the revised HPV classification framework. Since the original fifth and sixth criteria are concerned with “failure to submit a certification" or “failure to submit a permit application" which does not fit with any of the descriptions in the revised criteria nor does their inclusion fit with the EPA's stated intent in revising the HPV criteria to focus “on CAA violations that experience shows are most likely to be significant for human health and the environment" [EPA, 2014]. Similarly, we exclude the original fourth general criterion as violation of a consent decree does not necessarily represent a significant or immediate harm to human health and the environment. In our view, the area with the greatest level of ambiguity in this mapping process is with the matrix criteria. We are confident that the fourth matrix criterion, a standard regarding opacity, likely represent a criterion that does not focus on violations that impose immediate harm to human health and the environment. However, matrix criteria one and three are often focused on emissions limit violations. We do not view them as applicable to the revised criteria two, three, and four since in the full classification document [EPA, 1998], these matrix criteria can be triggered for emissions violations under seven days. Falling under this seven-day threshold is what restricts their inclusion in the list of the revised criteria, in our opinion. However, in the appendix we test the assumption regarding whether matrix criteria one and three should be excluded and find that these results do not qualitatively differ from findings in our primary analysis. Finally, matrix criterion two is excluded because it is clear in documentation that the timeframe for emissions violations under this criterion is well under seven days [EPA, 1998].

We consider the 2014 policy a reduction in the state regulator's discretion for two reasons. First, it eliminated some criteria with inherent discretion ("violation of any substantive term of any local, state, or federal order."), reducing the ways in which a plant could be classified as an HPV and subjected to joint federal and state enforcement activity. Second, it reduced the potential for heterogeneity in how plants classified as HPV were treated. The existence of the watch list early in our sample period and previous work in the literature [Evans and Stafford, 2019] indicate that state regulators often prioritized some HPV plants over others.

In the following section, we provide a theoretical justification for how policies limiting the threshold for classification as HPVs can increase the overall level of environmental performance, while section 4 describes our empirical strategy for determining whether this policy revision led to improved environmental performance for a sample of large polluting facilities.

3 Theoretical Model

This section presents a model of regulatory enforcement when violations are heterogeneous. The goal is to establish a simple framework for understanding optimal regulatory enforcement policy when monitoring and punishment are costly and firms can select the degree of violation. The model helps to understand how specifying the set of violations that merit higher levels of enforcement may reduce environmental damages, thereby yielding testable predictions that we apply to the data in section 5.

3.1 Model

We develop a model in the spirit of Mookherjee and Png [1994] with multiple types of firms $i \geq 2$ and a single regulator j. Firms decide on the level of violation $a_i \in [0, 1]$ to maximize expected profits. We assume that the level of violation is a continuous variable. Examples include, but are not limited to, violating an emission limit for sulfur dioxide at a coal power-generating facility or exceeding a mandated water concentration limit for chlorine from pulp bleaching during paper manufacturing. A regulator chooses a threshold of a violation above which actions are seen as more dangerous and firms undergo intense scrutiny. A regulator imposes policy enforcement through levels of monitoring μ, prosecuting $p(a_i)$, and punishing $f(a_i)$ violators, with monitoring and prosecution costs represented by $c_\mu > 0$ and $c_p > 0$, respectively. Once detected via monitoring, the level of the violation becomes observable to the regulator.

In this setup, each firm obtains a private benefit from a violation, $b(a_i)$. We assume that $b(a_i)$ is increasing in the degree of violation, a: $\frac{\partial b(a)}{\partial a} \geq 0$. In addition, we introduce an upper bound on a benefit function, $b(a_i) \leq B$, to secure deterrence since there are decreasing marginal benefits to the degree of violation. Besides providing benefit for firms, each violation results in a social harm, $h(a_i)$, and it is increasing in the level of a violation: $\frac{\partial h(a)}{\partial a} \geq 0$. Under the condition that enforcement is costless and the firms' types are observable, the regulator could set an individual enforcement policy for each firm such that it causes them to select the first-best level of violation where marginal private benefit equals marginal social harm, $b'(a_i) = h'(a_i)$.

However, for our setting we are concerned with the case where enforcement is costly and there is a strict upper bound on the fines imposed, $0 \leq f(a_i) \leq F$.[5] Therefore, the regulator solves for the optimal expected enforcement policy, $e(a^*) = \mu p(a^*) f(a^*)$, such that it internalizes both the firms' private benefits and the social costs of violations:

$$W = \sum_i (b(a_i) - h(a_i)) - \mu c_\mu - \mu \sum_i p(a_i) c_p(f(a_i))$$

$$s.t.\, 0 \leq \mu \leq 1,\ 0 \leq p(a_i) \leq,\ 0 \leq f(a_i) \leq F,\ f(0) = 0$$

The first element of W captures the difference between each firm's private benefit and the social harm from a level of violation, a_i, aggregated across all firm types in the market. The second term is the regulator's cost of monitoring, while the third term is the variable cost of enforcement. The regulator has to internalize the set of violating actions across types and set a level of monitoring and enforcement subject to the constraint imposed by the maximum penalty, F. The optimal policy determines a range of potential violations, with the most severe violations committed by the highest type and lesser violations by the lowest. Intuitively, if society wishes to lower the maximum observed violation to some level a_τ, then the maximum penalty for all actions greater than a_τ would have to exceed the benefit for all types, i.e., $f(a_{i>\tau}) > b(a_{i>\tau})$.

A firm of type i maximizes its private benefit accounting for the expected penalty from the violation a_i:

$$V = \max_{a_i}\ b(a_i) - E\, e(a_i)$$

Under the conditions of costly enforcement, a finite maximum penalty, and a well-behaved distribution of types, Mookherjee and Png [1994] arrive to the following conclusions regarding the optimal enforcement policy in this setting.[6] First, harmful acts below the optimal threshold should be legalized since enforcement is costly. Second, there is a range of violations where the marginal expected penalty is less than the marginal social harm given that firms differ in the benefits they receive from each level of violation. In other words, actions below the optimal threshold should be underpenalized in order to increase the marginal deterrence of more harmful violations. Third, a regulator should raise the enforcement threshold when enforcement costs increase. Moreover, all acts below the new increased threshold should see a reduction in penalties,

[5] This is a common assumption in the literature and is consistent with the idea that there are political limitations to the degree to which regulators can punish firms.

[6] Specifically, Mookherjee and Png assume the distribution of firm types has three characteristics. First, the number of firms with the lowest type is finite. Second, the number of firms with the highest cost type is positive. Finally, the inverse hazard rate is increasing in type.

while penalties for acts above the new threshold will increase. Intuitively, the regulator is allocating more enforcement resources to more severe violations.

The EPA's 2014 revision that raised the threshold for HPV status could alter the behavior of firms in several ways. First, there is an increase in the marginal benefit of committing a non-HPV-level violation relative to an HPV. This is because the expected penalties for being classified as HPV relative to other noncompliance statuses increases. Second, if the previous threshold for HPV status was set too low, the revision should result in a lower average level of environmental harm.[7] However, this discussion hinges on the federal EPA correctly delisting the violations with the lowest environmental harm. In section 4.2, we present evidence showing that with the 2014 revision the federal EPA appears to have delisted the violation types with the lowest harm from HPV enforcement.

Predictions

These theoretical predictions translate to the following testable predictions regarding the EPA's 2014 policy revision for our empirical analysis.

> **Prediction I:** After the policy change, the level of enforcement for HPVs relative to other non-compliant plants will increase. This is because the composition of plants in each regulatory status will change, with fewer but more harmful plants retaining an HPV status.
>
> **Prediction II:** Following the policy change, the overall emissions damages from plants will decrease.

This is consistent with the understanding that prior to the policy revision, the threshold for HPV status was too low, and the policy sufficiently rectified this.

In the following section, we describe our data for empirically investigating these predictions. We then describe how the policy change provides a quasi-experimental setting to test the validity of the model predictions. Specifically, the federal EPA did in fact eliminate associated HPV criteria with the lowest predicted impact on environmental damages. Finally, we describe our empirical framework for formally testing our theoretical predictions.

4 Data and Empirical Framework

In this section, we first describe the construction of the data used for our analyses. Second, we use these data to investigate whether the EPA's policy revision is consistent with its goal of focusing on the most harmful plant violations. Third, we provide descriptive evidence for why the implementation of the EPA's 2014

[7]In the appendix, we derive a simple two-firm one-regulator example that shows changes in the aggregate pollution level given a revision to the high violation status threshold. Following a change in the threshold, we observe that on average firms choose a level of violation below the cutoff, which leads to a reduction in total pollution.

policy revision is a natural experiment for the evaluation of marginal deterrence and regulatory discretion. Finally, we describe our formal empirical framework for estimating the impact of the policy revision on plant environmental damages.

4.1 Data

We construct a novel dataset on state environmental enforcement characteristics and plant-level outcomes by merging several databases, primarily from the EPA. The state policy treatment intensity data are derived from the EPA's Air Facility System "Actions" database by calculating the proportion of HPVs with delisted criteria among all enforced violations within a state between the beginning of 2010 and ending in 2014, the time period prior to the policy. We infer that states with a higher portion of delisted violations will more significantly alter their enforcement in the post-policy period, as the decrease in discretion shifts their priorities away from the delisted violations, observed in the pre-policy period, toward more harmful violations. We believe this definition is preferable to using the portion of facilities that receive a delisted violation, since this alternative may not adequately capture differences in regulator's priorities as a number facilities received multiple violations in the pre-policy period.[8] To evaluate whether our definition of treatment intensity corresponds with other state-level enforcement characteristics, we incorporate state environmental agency budget data from the Environmental Integrity Project [Kelderman et al., 2019] and SRF results data provided by the EPA.

Our primary facility-level pollution results are derived from the EPA's National Emissions Inventory (NEI). This contains facility-level information on the emission of criteria air pollutants, the subset of harmful pollutants considered under the Clean Air Act's "National Ambient Air Quality Standards," as well as other common or hazardous pollutants. This data is based on information collected from federal, state and, local environmental enforcement agencies. Although this data is only publicly released triennially, we utilize an annual point source set of this data made available after a request made using the EPA's Air Emissions Inventories contact form. In our view, this data provides a comprehensive overview of U.S. industrial point source emissions as it contains information on both large facilities such as electric power plants, as well as small facilities such as dry cleaners.

Plant-level compliance and enforcement data are obtained from the Integrated Compliance Information System (ICIS), which provides detailed information on industry, plant location, inspections, warnings, fines and their amount, and compliance status (HPV, noncompliant but non-HPV, in compliance). We also supplement the CAA compliance information with plant CAA watch list designations from the EPA. Plant pollution damages in monetary terms are calculated using the air integrated assessment model (AP3) from

[8] Although in the appendix we do consider this alternative definition for use in our primary specification.

Clay et al. [2019]. The model provides county-level damages per ton of selected pollutants, including particulate matter 2.5 (PM2.5), volatile organic compounds (VOCs), sulfur dioxide (SO_2), nitrogen oxides (NOX), and ammonia (NH_3). These damages are based on each pollutant's impact on human health, agricultural output, timber yield, and morbidity. We use the "Low Stack" category for damage approximation.[9] Finally, for robustness we incorporate parent company data using a plant's Dun and Bradstreet number or parent company name according to the Facility Registry System. This allows us to determine which plants have a sister facility in the same sub-industry to which they could potentially substitute their own pollution.

In supplementary analysis within the appendix, we incorporate data on toxic chemical emissions from the Toxic Release Inventory (TRI). The TRI provides self-reported emission data from primarily larger facilities in industries like electric power generation and metal mining.[10] Plants report each chemical and the amounts released during their production every year. Since there is no clear transfer from a chemical to a pollutant regulated by the CAA, we adopt the chemical-to-pollutant conversion method provided by Greenstone [2003] to derive a secondary measure of plant particulate matter and VOC emissions. This chemical-to-pollutant conversion process requires two major assumptions for our appendix analysis. First, it assumes that the conversion of TRI chemicals for the iron and steel industry, the industry analyzed in Greenstone [2003], is applicable to a wider range of industries (electricity generation, mining, etc.). If the industrial process used to emit specific TRI chemicals varies substantially across sectors, with respect to their correspondence with criteria air emissions, then this would lead to mismeasurement of plant particulate matter or VOCs. Second, although this conversion methodology doesn't distinguish between types of particulate matter, we assume that all particulate matter measured under this process is PM (2.5) to map to the damages under the AP3 model.

We combine these datasets to build an extensive annual-level panel consisting of operating facilities between 2010 and 2019. As a result, the final panel includes 699,013 annual observations for 81,432 plants across the US. To alleviate concerns about plant entry and exit, this sample group is conditioned on a plant being observed both prior to and following the policy. In addition, we drop extreme outliers in terms of emissions, those with more than 5,000 tons. This omitted group is less than 0.5% of the observations and is likely reflective of the disaggregated non-publicly available nature of the sample. For example, dropped observations include a generically name "WORKS NO 4". Although we provide estimations of our primary specification without these sampling conditions, with qualitatively similar results to our main findings.

Table 1 provides a comparison of key variable mean values before and after the 2014 policy change.

[9] This choice was motivated by the fact that for the vast majority of counties, low stack damages are the middle between the other common damage categories of "Medium Stack" and "Area."

[10] As discussed in Gibson [2019] these data are subject to a number of caveats, the most prominent of which is that they only covers large facilities and are based on engineering estimates.

Table 1: Policy impact on enforcement, compliance, and damages across plants

	Mean pre-policy (se)	Mean post-policy (se)	Group Diff (p-value)
	(1)	(2)	(3)
Emissions (tons)	43.667	39.380	-4.286***
	(224.537)	(199.663)	(0.000)
Penalties ($)	775.100	599.239	-175.861**
	(33,985.965)	(28,503.068)	(0.020)
Inspections (#)	0.387	0.332	-0.056***
	(1.872)	(1.418)	(0.000)
Damages (mln $)	1.453	1.420	-0.033
	(13.183)	(16.209)	(0.351)
Compliance (%)	0.976	0.970	-0.007***
	(0.152)	(0.171)	(0.000)
HPV	0.090	0.075	-0.015***
	(1.208)	(1.375)	(0.000)
Noncomplaint Non-HPV	0.019	0.049	0.029***
	(0.288)	(0.627)	(0.000)
N	363,679	335,334	699,013

Notes: The table shows mean values for key variables before and after the HPV policy change in 2014. Column (1) corresponds to mean plant-annual pre-policy change observations. Column (2) corresponds to average plant-annual observations after the policy change. Column (3) presents the difference in group mean values. The time frame is 2010–2019 using data from ICIS and NEI databases. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 2: Policy impact on enforcement by regulatory status

	HPV Plants			Other Noncompliant Plants		
	Mean pre (se)	Mean post (se)	Group Diff (p-value)	Mean pre (se)	Mean post (se)	Group Diff (p-value)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All Plants						
Penalties	9,696.034	5,772.842	-3,923.192***	1,343.713	1,970.581	626.869
($ per violation)	(88,064.844)	(40,007.301)	(0.007)	(14,350.943)	(37,899.543)	(0.435)
Inspections	1.144	0.729	-0.415***	1.305	1.416	0.111*
(# per violation)	(2.516)	(1.718)	(0.000)	(3.210)	(2.234)	(0.072)
N	6,242	4,184		2,359	5,980	
Panel B: Plants Top Quintile of Treatment						
Penalties	7,954.534	8,029.073	74.538	3,055.888	1,649.686	-1,406.202**
($ per violation)	(47,037.176)	(52,864.250)	(0.971)	(17,229.354)	(9,705.680)	(0.040)
Inspections	1.419	0.951	-0.468***	1.988	2.123	0.135
(# per violation)	(3.865)	(2.110)	(0.000)	(4.168)	(2.944)	(0.475)
N	1,291	1,072		373	1,381	
Panel C: Plants Bottom Quintile of Treatment						
Penalties	11,017.862	6,631.218	-4,386.644	969.642	1,370.057	400.415
($ per violation)	(84,943.703)	(44,919.574)	(0.197)	(12,050.431)	(21,635.285)	(0.681)
Inspections	1.126	0.781	-0.345***	2.469	1.345	-1.124***
(# per violation)	(2.214)	(1.600)	(0.000)	(4.994)	(1.772)	(0.000)
N	963	761		555	1,393	

Notes: The table shows mean values for key variables before and after the HPV policy change in 2014. Columns (1)–(3) correspond to HPV values, while columns (4)–(6) present values for the FRVs. Columns (1) and (4) show mean plant-annual pre-policy change observations. Columns (2) and (5) correspond to average plant-annual observations after the policy change. Columns (3) and (6) present the difference in group mean values. The time frame is 2010–2019 using data from ICIS databases. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Compliance is a binary outcome taking a value of one if no violation is observed during the reported year. There are several important aspects of the data worth noting. First, average plant-level damages and emissions went down in the post-policy change period. Second, the rate of HPVs went down in the post-policy period, while the percent of noncompliant but non-HPV went up.[11] This statistic is consistent with restricting the criteria for HPV classification. Finally, the overall level of average total plant penalties went down. In order to understand how this relates to the policy revision, or other contemporaneous factors, we further break down the enforcement data to observations of plants with either an HPV or non-HPV within a calendar year in Table 2.

Panel A of Table 2 presents descriptive statistics of enforcement for HPV and other noncompliant facilities. Consistent with the trend of decreased enforcement noted in Shimshack [2014] and Blundell et al. [2021], we see that both penalties and inspections per violation decreases in the post period for those plants classified with at least one HPV. However, we see no comparable decrease in enforcement for other noncompliant facilities, in fact there is a marginally significant increase in inspections for this group. Panels B and C of Table 2 separate out these plants by whether they reside in a state that is in the bottom or top quintile of treatment. With this subgroup comparison, we find some suggestive evidence consistent with the stated intent of the EPA policy revision. HPV plants located in high treatment states observed no discernible change in the level of penalties, while other noncompliant plants saw a statistically significant decrease in penalties. This is consistent with the understanding that regulators allocated dwindling enforcement resources away from less harmful noncompliant facilities to HPVs. In contrast, at the bottom treatment quintile observed in Panel C, we observe a decrease in penalties for plants with at least one HPV (p-value 0.197), since these plants are impacted little by the policy, this change is reflective of the general downward enforcement trend. We find no consistent pattern with inspections, which is likely reflective of the understanding that many inspections are driven by regular monitoring requirements under the CAA. Overall, Tables 1 and 2 suggest the 2014 policy had its intended impact. In the following subsections, we formally evaluate the exogeneity of the policy's implementation as well as whether the policy's described priorities are consistent with the data. Additionally, in the appendix we provide descriptive statistics using only state regulator enforcement actions, with qualitatively similar results to those in this section.

4.2 Understanding the EPA Policy Revision Using Machine Learning

Before we empirically estimate the causal impact of the policy on plant-level damages in our primary analysis, we use these data to empirically evaluate whether the federal EPA's 2014 policy revision was consistent with their stated goal of focusing "on CAA violations that experience shows are most likely to be significant for human health and the environment" [EPA, 2014]. Specifically, we employ two algorithms from the machine

[11] Formally, these noncompliant but non-HPV cases are defined from Federally Reportable Violations reported in the data that were not HPV.

learning and statistics literature to select important features in predicting pollution damages. This analysis provides an objective assessment of whether the policy, as stated, would shift enforcement priorities away from violations with lower damages to more significant violations. The first machine learning algorithm we use is feature selection [Cai et al., 2018, Furnival and Wilson, 2000] in order to identify which violation classifications had the most suggestive impact on damages. Secondly, we conduct a factor analysis [Hastie et al., 2009] to check whether multiple pre-2014 HPV criteria capture the same underlying violation. By estimating these models we obtain better understanding regarding the HPV policy revision on pollution damages.

We start with a step-wise bi-directional feature selection model with an underlying ordinary least squares (OLS) algorithm to pick the violation types that have a high correlation with the level of pollution damages. Specifically, the algorithm functions by starting with a fixed total number of criteria to be included, then selects the combination from all criteria such that the sum of squared residuals is lower than any other criteria combination. Below is the estimation model, where $GC_{i,j}$ and $M_{i,j}$ represent whether a plant i fell under general criteria and matrix criteria violations, respectively. There are ten general criteria and four matrix criteria violations represented by a subscript j. The outcome we consider is the inverse hyperbolic sine (IHS) of plant damages to account for the many zero-valued observations in our data [Bellemare and Wichman, 2020].

$$IHS(Damage_i) = \sum_{j=1}^{10} \beta_j \cdot GC_{i,j} + \sum_{j=1}^{4} \gamma_j \cdot M_{i,j} + \varepsilon_i$$

We use this feature selection model to provide suggestive evidence on which violations from the original criteria should be kept if the total were restricted to seven. The choice of seven remaining criteria is consistent with the number of criteria the EPA maintained under the policy revision. We find that when the number of violations is restricted to seven, the results are not surprising, as the majority of the features selected (four) overlaps with the seven criteria kept by the federal EPA in its 2014 revision. Coincidentally, seven is the total number of criteria selected by the algorithm when using the Bayes-Information-Criterion. As compared to models which use fewer HPV criteria, or more, the choice of seven performs best on this standard model selection metric.

The key difference between our feature selection results and the EPA policy is that our algorithm consistently selects the emissions matrix criteria that correspond highly with damages. However, this is likely because these matrix criteria are rarely independently observed from the other general criteria maintained under the policy revision. Therefore, in the appendix, we use a factor analysis approach to formally investigate the importance of these matrix criteria, relative to the other HPV criteria. The underlying premise of the factor analysis is that there are latent variables or noncompliant events, such as a scrubber failure, that correspond with multiple criteria [Harman, 1976]. This analysis provides evidence of the uniqueness of any HPV criterion with respect to these latent variables (compliance events) or whether it is redundant in the determination of noncompliance. The factor analysis results indicate that the subset of matrix criteria identified as corresponding with high damages in our feature selection models have the same latent variable as the general criteria the EPA kept in its policy revision.

A more detailed explanation of the factor analysis and the corresponding results are found in the appendix. Overall, the results indicate that the EPA's policy revision was consistent with its stated goal of focusing on the violations that posed the greatest potential harm. This motivates the formal analysis of the impact of this policy revision discussed in the following subsections.

4.3 Quasi-Experimental Variation

The central empirical challenge of this article is to test whether the 2014 policy reduced plant environmental damages. Credibly estimating the impact of the policy requires that the intensity of state-level treatment is not associated with unobserved determinants of plant-level damages over time. In this subsection, we provide evidence suggesting that this is indeed the case, that is, the variation in state-level treatment intensity is exogenous to state characteristics.

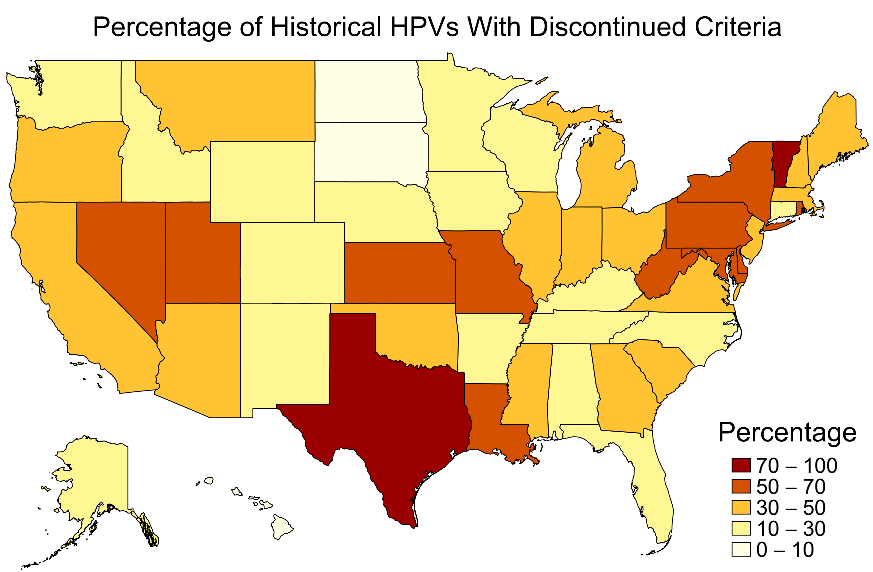


Figure 2: Distribution of discontinued HPV criteria

Note: This figure shows the pre-policy distribution of discontinued HPV criteria across states. The darker shade refers to greater rate of delisted HPV violations.

Starting with the state-level characteristics, Figure 2 shows the pre-policy distribution of delisted viola-

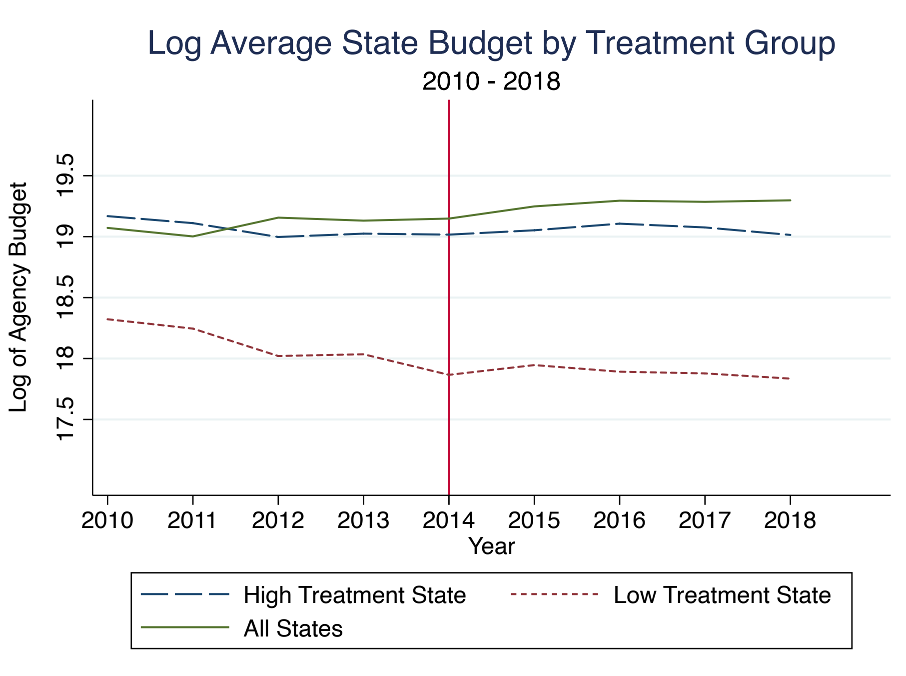


Figure 3: State budgets by treatment group

Note: This figure shows the log of average budgets for those states that underwent significant changes in enforcement of HPVs after the policy change in 2014, with the top quintile of treatment intensity in blue, and states in the bottom quintile, with a low treatment intensity in red. The individual state budget data captures the period from 2010 to 2018.

Table 3: State characteristics and treatment intensity

	(1) Lax enforcement (%)	(2) Lax penalty (%)	(3) Lax inspection (%)	(4) Avg. manufacture	(5) Log Budget
Policy	-0.121	-0.140	-0.189	0.124	0.100
	(0.268)	(0.260)	(0.119)	(0.120)	(0.124)
N	50	50	50	50	48
R^2	0.069	0.049	0.017	0.028	0.052

Notes: This table shows OLS regression estimates of the relationship between state regulator characteristics and the intensity of treatment during the pre-treatment time period 2010–2014. Specifically, all of these estimates are from a regression of the outcome on a constant and our continuous measure of policy intensity. Columns (1)–(3) display estimates for the relationship between treatment intensity enforcement problems, problems with punishment, and issues with inspections in a state, respectively. We define a state regulator as insufficient (*lax*), taking a value of one, if the regulator did not meet a required level of performance in the corresponding area under the State Review Framework prior to the policy and zero otherwise. Column (4) displays the estimated relationship between state-regulated facilities in manufacturing and treatment intensity. Column (5) shows the estimated relationship between state treatment intensity and the log of the state regulator's budget. White–Huber standard errors are shown in parentheses. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.
Sample: A cross-section of all 50 states using average characteristics for the pre-treatment 2010–2014 time period.

tions across states. The darker shade refers to the greater portion of delisted violations. Importantly for our analysis, the figure shows significant variation in treatment intensity across states, demonstrating that the policy was plausibly exogenous to the geographic region. Figure 3 depicts the trend in agency budgets for the time period of 2010–2018. The red line represents the average environmental agency budgets for states with a low treatment intensity, in the bottom quintile. Meanwhile, the blue line represents the corresponding budget data for the top quintile, those with a high treatment intensity. It is important to note that the figure demonstrates no differential trend in the agency budgets between states with different levels of intensity after the policy was implemented, nor does it show any deviation in the trend for the full sample. This suggests that treatment from the policy is not predictive of state enforcement resources. In fact, the post policy difference in budget growth between states in the lowest and highest quintiles of treatment is less than 0.13% of the pre-policy average. Next, Table 3 indicates that the state-level treatment intensity is uncorrelated with outcomes such as federal evaluations of the state agency, the proportion of manufacturing plants within the state, or the log of the state budget. Federal evaluations of the individual state agency are defined by the State Review Framework, which contains information on the quality, accuracy, and timeliness of data, inspections, violations, penalties, and enforcement for each state regulator on a five-year basis. The specific outcomes considered are whether a state regulator enforced violations in a timely manner (*enforcement*), whether the calculation, assessment, and collection of penalties is accurate (*penalty*), and whether the regulator met its inspection and coverage commitments (*inspection*).

4.4 Two-Way Fixed Effects Estimation

Given the quasi-experimental nature of the policy implementation discussed in the previous subsection, we use a dose-response or “fuzzy" difference-in-differences design for estimation. This type of research design relies on the treatment rate or treatment dosage being higher for some states and allows for no state to remain fully untreated [De Chaisemartin and d'Haultfoeuille, 2018, de Chaisemartin et al., 2019]. The heterogeneity in the use of de-listed violations prior to the 2014 policy revision detailed in Figure 2 provides the variation needed to estimate the effect of an increased threshold for priority violations on plant pollution damages under this framework.

In our primary specification, the outcomes of interest are plant damages, $Damage_{i,t}$, and emissions, $Emissions_{i,t}$. Specifically, $Damage_{i,t}$ is the annual pollution damage from plant i in year t expressed in 2018 USD dollars. While $Emissions_{i,t}$ is the emission of the damaging pollutants from plant i in year t. We take the inverse hyperbolic sine of these two outcome variables to account for the many zero-valued observations in our data [Bellemare and Wichman, 2020], while still giving it an interpretation equivalent to taking the log of the outcome. However, we investigate the robustness of this choice in section 6 and the appendix. Our base specification is the following:

$$Y_{i,t} = \alpha_0 + \alpha_1 \cdot Policy_s \cdot Post_t + \gamma_i + \delta_t + \phi_{j,t} + \varepsilon_{i,t} \tag{1}$$

$Policy_s$ represents the intensity of treatment from the policy change in a state s. It ranges from zero for states that never enforced the altered HPV types and one for states where 100% of the pre-policy priority violations enforced were of the reclassified type. $Post_t$ is a dummy variable that takes a value of one every year after the 2014 policy change. This is consistent with the understanding that since the policy was announced in the third quarter of 2014, there should not be any significant changes in plant emissions or damages that year. In addition, we include plant and year fixed effects, γ_i and δ_t, along with industry-year fixed effects, $\phi_{j,t}$.[12] These terms should allow us to capture plant-specific time-invariant heterogeneity, such as plant age, as well more general trends in the emissions from an entire industry, potentially due to factors such as new technology. Finally, the error term $\varepsilon_{i,t}$ captures the unobserved determinants of the outcome. We cluster at the plant level, since given the size and unbalanced nature of our panel, there are likely too few clusters (48) to support clustering at the level of treatment, the state.[13] In the appendix, we demonstrate that our primary results are robust to the use of standard errors clustered at both the state and state-year level.

By using a continuous measure of treatment, the proportion of HPVs prior to the policy change which correspond to the excluded criteria, we require several assumptions for identification. Following the discussions in De Chaisemartin and d'Haultfoeuille [2018] and de Chaisemartin et al. [2019], the first assumption needed for identification of α_1 in equation 1 is common trends. Similar to the standard parallel trends assumption used in differences-in-differences specifications, plants with a lower policy treatment intensity represent the counterfactual trend in damages for higher treatment intensity plants in the absence of the policy. Second, the treatment effect of the policy is stable over time, there are no substantial dynamic effects. These two assumptions, along with conditions met by the continuous and varied implementation of the policy in figure one, will allow $\hat{\alpha_1}$ to be a consistent weighted estimate of the local average treatment effects across the different treatment levels observed in our setting.[14]

We understand that the assumption of a stable treatment effect over our sample period is rather strong, plants may take time to adjust to the policy. Therefore, we relax this assumption by also estimating the "time-corrected" (TC) and "changes-in-changes" (CIC) fuzzy differences-in-difference estimators from De Chaisemartin and d'Haultfoeuille [2018]. Mechanically, these two estimators for $\hat{\alpha_1}$ are calculated by averaging the pre and post policy changes in the outcome for different treatment levels under different weighting schemes. The former "time-corrected" estimator requires a stronger trends assumption, parallel trends at every dosage or level of treatment. Specifically, regardless of the realized level of treatment, plants

[12] We define industry at the five-digit NAICS code level.

[13] This is consistent with the discussion in Cameron and Miller [2015] "there is no clear-cut definition of "few". Depending on the situation "few" may range from less than 20 clusters to less than 50 clusters in the balanced case, and even more clusters in the unbalanced case."

[14] Formally, the assumptions are 1 - 5 in de Chaisemartin et al. [2019].

with a lower level of treatment would have the same trend in outcomes as plants with a higher treatment level, in the event both groups received the higher (or lower) treatment.[15] Instead of a common trends assumption, the latter "changes-in-changes" estimator requires that the potential outcome are monotonically increasing with treatment, the unobservable determinants of the outcome are stationary, and the distribution of potential outcomes is independent of the initial level of treatment.[16] These latter assumptions could be violated if there is state-level heterogeneity in the industrial composition of plants that both made them more likely to qualify for the delisted violations prior to the policy and determined how effective the state-level response to the policy was in reducing damages. For simplicity, the identification assumptions of equation 1 are met if policy intensity was strictly exogenous to the time varying unobservables that determine our emissions related outcomes, which is supported by the analysis in the prior section and by a series of robustness checks in section 6.

5 Results

In this section, we first graphically illustrate the effect of the policy on plant emissions and damages for both the balanced sample and the subset of plants in the lowest and highest quintiles of treatment intensity. Following this, we report the damages and emissions results from estimating equation 1. Next, we use our estimation framework to investigate the association between the intensity of policy treatment and changes in regulators' enforcement priorities. Finally, we use our damage estimates to provide back-of-the-envelope calculations for the total reduction in damages from the EPA's 2014 policy revision.

5.1 Graphical Results and Common Trends

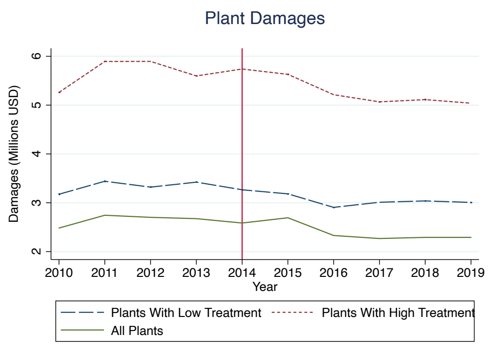
Figure 4a illustrates the impact of the policy on plant damages. This plot displays average plant level damages for the fully balanced sample, the average level of damages for plants in the lowest quintile of treatment intensity (blue line) and plants in the highest quintile of treatment intensity (red line). We use the fully balanced sample in order to avoid issues regarding differences in composition between the publicly available NEI (2011, 2014, 2017) and the other years.[17] The plot highlights that before the policy took effect, the plants in the top quintile of treatment had a similar trend in damages to plants from states in the lowest quintile of treatment intensity. This similarity lends credibility to the common trends assumptions needed for identification in the baseline and time-corrected estimators.

After the policy took effect, there is a steady decline in plant damages for the full sample. This provides

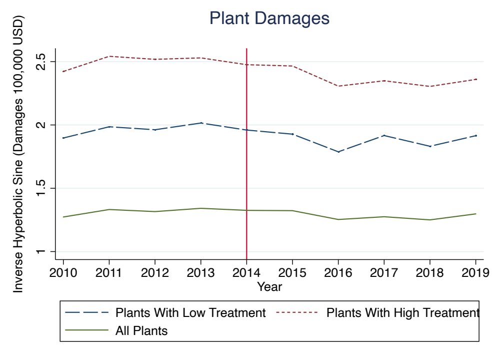
[15] Formally, the assumptions are 1 - 3 and 4' in de Chaisemartin et al. [2019].
[16] Formally, the assumptions are 1 - 3, 6, and 7 in de Chaisemartin et al. [2019].
[17] These differences are likely accounted for by the use of plant and year fixed effects in our regression estimation

preliminary evidence that the HPV criteria revision is associated with a 7.7% decrease in plant damages.[18] Further, based on this graphical analysis, we can see the gap in damages between plants in the highest and lowest quintiles shrank following the policy's implementation. The larger decline in damages for plants in the highest quintile supports the understanding that damages are decreasing in treatment intensity. The comparison provided in Figure 4b using the inverse hyperbolic sine of damages, our primary outcome, yields a qualitatively similar damage decline of 3.4% following the criteria revision.

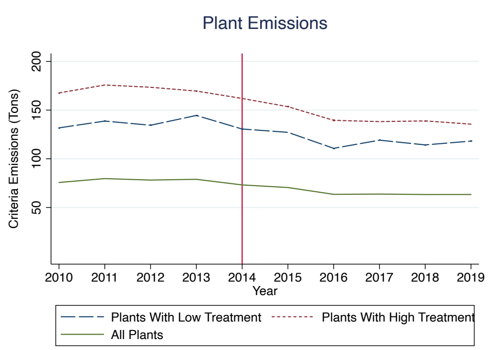
Figure 4: Damages and Emissions from NEI Plants



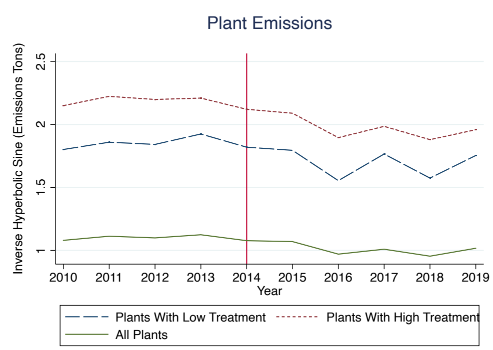
(a) Damages



(b) IHS(Damages)



(c) Emissions



(d) IHS(Emissions)

Source: National Emissions Inventory, 2010 - 2019

To consider the impact of the policy on emissions, Figure 4c plots firms' average emissions before and after the policy. Figure 4d displays the graphical results for the inverse hyperbolic sine of emissions. These plots illustrate that prior to the policy, plants from both the high and low treatment quintiles had a similar slightly downward trend in emissions. Similar to Figures 4a and 4b, following the implementation of the policy, the decrease in the emissions outcome is greater for plants located in high treatment intensity states. Figures 4c and 4d show that the overall decline in the level emissions is 11% and inverse hyperbolic sine of

[18]The percentage comparisons in this subsection are calculated by taking the difference between the 2014 outcome (the year the policy was announced) and average annual outcome in the post policy period, then dividing by the 2014 level.

emissions is 6.7% for the full sample.

Overall, the four figures together lend preliminary support to the hypothesis that the decrease in emission related outcomes is associated with the intensity of treatment. In addition, these figures provide some suggestive evidence in favor of the assumption needed for identification of equation 1, plants that received different levels of treatment have common trends in the outcome. We next investigate these findings more rigorously through the estimation of equation 1, which controls for plant, industry, and time fixed effects.

5.2 Primary Regression Results

Columns (1) to (3) of Table 4 presents results from estimating our primary two-way fixed effects model, as specified in equation 1. The outcome variable is the inverse hyperbolic sine of damages. Column (1) reports the baseline estimate using the full sample of plants. The estimated coefficient for the full impact of the policy is -0.069, or a 6.9% decrease in damages. Since the average treatment intensity is 0.357 for this sample, this result implies the policy decreased pollution damages by 2.5%.[19] This average is statistically indistinguishable from the 3.4% average decline observed in Figure 4b, despite the inclusion of an extensive set of fixed effects. Columns (2) and (3) of Table 4 report results from the alternative "time-corrected" and "changes-in-changes" estimation strategies. As we discussed in section 4.4, identification under these approaches requires different assumptions regarding common trends, unobservables, and the potential outcomes under treatment. Both estimated coefficients are again negative and statistically significant and imply that the policy reduced plant damages between 3.5% and 7.4%.

Columns (4) to (6) of Table 4 presents regression results for the outcome of emissions. The coefficient estimates for the policy range from -6.7% to -41% and are statistically significant. This translates to an average decrease in emissions from the policy between 2.4% and 14.6%. Which is consistent with the 6.7% decrease in average emissions displayed in Figure 4d. Furthermore, the estimates from this emissions analysis are similar to the 1.6% to 42.6% range of estimated emissions impacts from other changes in air emissions policies found in the literature [Auffhammer et al., 2009, Walter and Raff, 2019, Gibson, 2019, Zou, 2021].

Overall, our results are consistent with Prediction II from the theoretical model, which states that the policy should lead to an overall reduction in the environmental harm posed by plants. The results in columns (4) to (6) of Table 4 indicate that emissions fell following the implementation of the policy. The drop in emissions is associated with a decrease in damages, as shown in columns (1) to (3) of Table 4. In the following subsection, we examine a possible mechanism explaining the post-policy decrease in environmental harm by investigating the relationship between the intensity of treatment and changes in state regulators' enforcement priorities.

[19] $\bar{Policy} \cdot \hat{\alpha_1} = 0.357 \cdot -0.069 \approx -0.025$.

Table 4: Primary Results: Damages and Emissions

	Damages			Emissions		
	Base (1)	TC (2)	CIC (3)	Base (4)	TC (5)	CIC (6)
Policy · Post	-0.069***	-0.097***	-0.208***	-0.067***	-0.276***	-0.410***
	(0.010)	(0.017)	(0.069)	(0.009)	(0.016)	(0.089)
N	699,013	699,013	699,013	699,013	699,013	699,013
Adj. R^2	0.932	-	-	0.933	-	-
Unique Plants	81,432	81,432	81,432	81,432	81,432	81,432
Plant FE	Y	Y	Y	Y	Y	Y
Industry-Year FE	Y	Y	Y	Y	Y	Y

Notes: This table reports estimates of our baseline equation 1 on our primary outcomes. Columns (1) - (3) consider the outcome of the inverse hyperbolic sine of damages. Columns (4) to (6) consider the inverse hyperbolic sine of emissions. Base specification is our baseline DID estimate, TC stands for "Time-Corrected" estimate for equation 1 according to de Chaisemartin et al. [2019], and CIC is the "changes-in-changes" estimate for α_1 from equation 1 according to de Chaisemartin et al. [2019]. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.
Sample: Plant-year level data from the National Emissions Inventory beginning in 2010 and ending in 2019. We only include plants observed prior to and following policy implementation.

5.3 Enforcement Priorities

To better understand the underlying mechanisms for how the policy decreased damages and emissions, we now use our two-way fixed effects framework from equation 1 to investigate the association between policy intensity and changes in regulator enforcement priorities. Specifically, we investigate whether the intensity of the policy corresponds with changes in the penalty and inspection practices for HPV and other noncompliant plants. The goal of this secondary analysis is to clarify whether the observed reduction in environmental harm from the application of the policy can be explained by changes in enforcement practices.

Table 5 shows regression results for penalties according to a noncompliant plant's violation status. Column (1) shows the average penalty dollars per violation for plants designated as a priority violator. Among this group, the coefficient estimate indicates that the penalty per violation increased by $21,167.22 in the post-policy period, and this effect is statistically significant at the 10% level. Overall, this effect translates to a 78% increase in penalty ($) per violation relative to the pre-policy period baseline. Column (2) shows a similar result using the alternative "time-corrected" estimator of $\hat{\alpha_1}$, although this effect is statistically insignificant. The estimated effect from column (3) of Table 5 is positive, albeit both economically and statistically insignificant.

Columns (4) to (6) present the analogous results for noncompliant but nonpriority plants, the subset of plants not targeted by the policy. For this secondary group of noncompliant plants, the policy also corresponded with a small increase in their penalty ($) per violation. However, the column (4) to (6) estimates are not statistically significant at conventional levels. The finding of a positive but statistically insignificant effect from the policy for non-HPVs is in line with the understanding that the revision of

HPV criteria may have pushed many former HPVs into this non-HPV group, thereby increasing the relative severity and penalties for this group in states with a higher treatment intensity. Overall, the penalty analysis in Table 5 provides suggestive evidence that penalties were one mechanism for meeting the stated goal of the policy to target enforcement resources toward violations that present the greatest environmental harm. Penalties for plants classified as having the most severe violations increased more in states where the policy would have a greater impact.

Table 6 shows regression results for inspections according to a plant's violation status. Columns (1) to (3) show the impact of the policy on the average inspections per violation for plants designated as a priority violator, with the results indicating the policy corresponded with a moderate decrease in monitoring for these plants. For these specifications, the estimated coefficients range from -0.453 to -0.829, but are otherwise statistically indistinguishable from zero. Similarly, the results for non-HPVs in columns (4) to (6) suggest that the policy corresponded with a moderately sized decrease in the frequency of inspections, and except for column (6), these estimates are also statistically insignificant at conventional levels. Given that this time period corresponded with a decrease in overall enforcement resources [Kelderman et al., 2019], one way to interpret these results is that they are suggestive of regulators maintaining the frequency of inspections for HPVs while possibly reducing inspections at other facilities. Further, it is possible that regulator's substituted from more costly to less costly types of inspections, a strategic behavior that has been documented with hazardous waste enforcement [Blundell et al., 2021], but is not observable with our data.

Overall, these results are consistent with Prediction I in section 3, which states that the policy will increase the value of being a non-priority plant relative to being classified as a priority violator. Following the policy change, priority plants in states with greater treatment intensity experienced a larger increase in their regulatory scrutiny relative to other plants. We acknowledge these results are merely suggestive of a change in the individual state regulators' priorities, as many HPV penalties and inspections result from joint enforcement activities between the state regulator and federal EPA. For large violations, such as HPVs, the federal EPA will often take over the later stages of an enforcement case. However, in the appendix, we find qualitatively similar results for the limited subset of HPV cases that were carried out from start to finish by the individual state regulators.

5.4 Back-of-the-Envelope Estimation

We now use our regression estimates to calculate a lower bound on the reduction in air pollution damages from the 2014 policy. Using our baseline specification estimate (Table 4, column (1)), which shows that

Table 5: Primary Results: Penalties per Violation

	HPV Plants			Other Noncompliant Plants		
	Base	TC	CIC	Base	TC	CIC
	(1)	(2)	(3)	(4)	(5)	(6)
$Policy \cdot Post$	21,167.22*	17,856.89	2,145.202	2,175.377	8,589.232	25,350.59
	(12,386.27)	(54,670.43)	(49,691.56)	(4,171.94)	(14,096.37)	(35,996.23)
N	7,686	7,686	7,686	3,959	3,959	3,959
R^2	0.249	-	-	0.506	-	-
Unique Plants	1,759	1,759	1,759	1,476	1,476	1,476
Plant FE	Y	Y	Y	Y	Y	Y
EPA Region-Year FE	Y	Y	Y	Y	Y	Y
Industry-Year Trend	Y	Y	Y	Y	Y	Y

Notes: This table reports estimates of our baseline equation 1 with fine amount per violation as estimated outcome. Base specification is our baseline DID estimate, TC stands for "Time-Corrected" estimate for equation 1 according to de Chaisemartin et al. [2019], and CIC is the "changes-in-changes" estimate for α_1 from equation 1 according to de Chaisemartin et al. [2019]. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Sample: Plant-year observations from the ICIS beginning in 2010 and ending in 2019. Columns (1) to (3) consider plants with at least one HPV in the current year. Columns (4) to (6) consider plants that are noncompliant but non-HPV in the current year.

Table 6: Primary Results: Inspections per Violation

	HPV Plants			Other Noncompliant Plants		
	Base	TC	CIC	Base	TC	CIC
	(1)	(2)	(3)	(4)	(5)	(6)
$Policy \cdot Post$	-0.453	-0.829	-1.537	-0.521	-3.524	-5.702*
	(0.618)	(1.511)	(1.375)	(0.933)	(2.260)	(2.923)
N	7,686	7,686	7,686	3,959	3,959	3,959
R^2	0.531	-	-	0.818	-	-
Unique Plants	1,759	1,759	1,759	1,476	1,476	1,476
Plant FE	Y	Y	Y	Y	Y	Y
EPA Region-Year FE	Y	Y	Y	Y	Y	Y
Industry-Year Trend	Y	Y	Y	Y	Y	Y

Notes: This table reports estimates of our baseline equation 1 with the number of inspections per violation as estimated outcome. Base specification is our baseline DID estimate, TC stands for "Time-Corrected" estimate for equation 1 according to de Chaisemartin et al. [2019], and CIC is the "changes-in-changes" estimate for α_1 from equation 1 according to de Chaisemartin et al. [2019]. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Sample: Plant-year observations from the ICIS beginning in 2010 and ending in 2019. Columns (1) to (3) consider plants with at least one HPV in the current year. Columns (4) to (6) consider plants that are noncompliant but non-HPV in the current year.

hypothetically moving from zero treatment under the policy to fully treated reduces damages from air pollution by 6.9%. We translate this effect to total damages reduced via multiplying it by the average treatment intensity (0.357), the average pre-treatment damages for each plant according to column (1) of Table 1, and the number of post-policy observations. This calculation indicates the policy resulted in $12 billion (2018 $) in total damages over the post-policy period (2015-2019), or roughly $2.4 billion annually.

This back-of-the-envelope estimate is consistent with the literature on the estimated effects of various U.S. emissions policies. For example, Chan et al. [2018] estimate a $280.5 million (2018 $) annual benefit from the Acid Rain Program in 2002, while Deschenes et al. [2017] find that the NO_x budget trading program has reduced medical expenses by an estimated $848 million (2018 $) annually. Our estimate of the 2014 HPV revision is larger because it impacts a wider range of pollutants ($PM_{2.5}$, SO_2, NO_x, NH_3,VOCs) as compared to the policies considered in these prior studies. Further, our policy estimate is consistent with general estimates of the overall benefits of the CAA. For example, Bento et al. [2015] estimate an $8.58 billion annual property value benefit from the CAA amendments during the 1990–2000 time period, while Chay and Greenstone [2005] find that the designation of non-attainment status during the 1970–1980 time period had an estimated $6.39 billion annual property value benefit.

Overall, our analysis suggests that the 2014 policy led to a substantial reduction in the damages from air pollution through the strategic reallocation of limited enforcement resources to the types of violations that posed the greatest environmental harm.

6 Robustness and Falsification

We believe that our two-way fixed effects approach provides substantial evidence that the 2014 federal policy led to a reduction in damages caused by air pollution. However, for completeness, we demonstrate the robustness of this result in a number of ways.

6.1 Industry Spillover, The EPA Watch List and Multi-Plant Firms

Recent empirical work examining the enforcement of environmental regulation in the United States points to potential spillovers to entities not initially subject to enforcement. In particular, facilities subject to an increase in regulatory scrutiny may substitute their emissions to other non-scrutinized facilities within the same parent company [Gibson, 2019, Rijal and Khanna, 2020], or via competitive general equilibrium effects, to other facilities in the same industry [Evans et al., 2018].

It is apparent that this potential substitution could threaten the validity of the common trends assumption of our empirical strategy. If plants in higher treatment intensity states substitute their emissions to sister facilities within the same parent company that are located in lower treatment areas, our baseline estimates for equation 1 may be biased. In addition, our estimates may be biased if there is substantial asymmetry in the distribution of plants within an industry across high and low treatment states, as increased regulatory costs in areas with a higher intensity of treatment will lead to that industry's production and emissions being redistributed toward plants in areas with a lower treatment intensity. Therefore, to address these concerns, we expand equation 1 as follows:

$$Y_{i,t} = \alpha_0 + \alpha_1 \cdot Policy_s \cdot Post_t + \alpha_2 \cdot Policy_s \cdot Post_t \cdot Treat_i + \alpha_3 \cdot Treat_i \cdot Post_t + \gamma_i + \delta_t + \phi_{j,t} + \varepsilon_{i,t} \quad (2)$$

Here, $Treat_i$ represents some other factor that coincides with the policy. The key coefficient of interest is α_2, the interaction between $Treat_i$ and a plant's policy treatment intensity in the post-2014 time period. By testing for whether α_2 is statistically distinguishable from zero, we can provide evidence of whether some alternative factor, $Treat_i$, explains any differential change in damages for plants under the 2014 policy. A finding that α_2 is statistically indistinguishable from zero would indicate that there is no plant heterogeneity represented by $Treat_i$ that violations the underlying identification assumptions for our primary policy estimates obtained from the equation 1 specification.

For the analyses regarding potential within-parent company substitution, we define $Treat$ in equation 2 as the difference between the treatment intensity for a state that individual plant (i) is located in and the average treatment intensity for other plants within the same parent company. The coefficient on the triple interaction represents the magnitude of substitution from plants in located in areas with higher treatment relative to its sister facilities. Our results in columns (1) and (3) of Table 7 indicate that the estimated differential substitution effect is not statistically significant at conventional levels, while the primary effect of the policy identified by α_1 remains negative and statistically significant. In addition, in columns (2) and (4) of Table 7 we find similar results for the impact of the policy when we re-estimate our primary baseline specification in equation 1 using the sub-sample of firms whose plants are only located in a single state. When focusing on the sub-sample of firms in multi-plant firms that have a higher level of treatment than their sister facilities in columns (3) and (6) of Table 7, we find larger estimated reductions from the policy. Although the difference between these coefficient estimates and those from our baseline specification in columns (1) and (4) in Table 4 are statistically insignificant.

For the analyses regarding general equilibrium effects and substitution of emissions and production from high treatment intensity states to low treatment intensity states, we define $Treat$ in equation 2 as the difference between the treatment intensity for the state that individual plant (i) is located in and the average treatment intensity for other plants within the same industry according to the six-digit NAICS code. The coefficient on the triple interaction represents the magnitude of substitution from plants in high treatment intensity states with high treatment relative to other facilities in the same industry, their most direct competitors. We focus on plants within the manufacturing sector, consistent with Evans et al. [2018], which comprises a significant portion of our sample and 364 sub-industries at the six-digit NAICS code level. The results in columns (1) and (3) of Table 8 indicate the policy still had a negative and statistically significant impact on emission damages. Furthermore, there is no evidence of differential policy impact among plants with significant industry heterogeneity in treatment intensity, the estimate for α_2 is statistically indistinguishable from zero. Interestingly, we find that in the case of damages the estimate for the impact of industry geographic heterogeneity in the post-policy period, α_3, is positive and statistically significant. We interpret this latter result as evidence that the finding in Evans et al. [2018], geographic heterogeneity increases environmental harm due to spillovers, was exacerbated later in our sample period with by the decline in regulatory enforcement resources. Most importantly for our setting though, this spillover effect appears unrelated to our analysis of the EPA's revision to the HPV criteria. Overall, these first two robustness analyses demonstrate that the estimated policy effects from our primary analysis are not likely to be biased by emissions substitution.

Finally, we use equation 2 to determine whether a state's involvement with EPA watch list facilities and the delisting of the policy biases our policy estimates. For the specifications in columns (2) and (4) of Table 8, we define $Treat$ as the percentage of plants that appeared on the EPA watch list out of all plants in that state prior to the policy change. In this case, the coefficient on the triple interaction represents the differential effect of policy treatment for plants located in states heavily impacted by the EPA watch list. We find that our primary results are robust even under these specifications, and the estimated coefficient on α_1 in equation 2 is negative and statistically significant. Further, the coefficient for α_2 is statistically insignificant, indicating that the significance of our primary results are not driven by the EPA watch list or its contemporaneous discontinuation. However, the impact of the watch list on polluting facilities [Evans, 2016] likely explains the statistically significant estimates for α_3.

6.2 Alternative Research Design

As an alternative specification, we consider a dynamic treatment effects approach that exploits within-treatment intensity and year variation in the policy. Specifically, we compare the behavior of plants in the states with higher levels of treatment intensity to their lower treatment intensity counterparts for each individual year prior to and following the policy. There are a few advantages from using this dynamic treatment effects approach. First, it allows us to formally test for the existence of common trends between

Table 7: Robustness Results: Multi-Plant Firms

	Damages			Emissions		
	(1)	(2)	(3)	(4)	(5)	(6)
$Policy \cdot Post \cdot Treat$	-0.052			-0.146		
	(0.248)			(0.221)		
$Policy \cdot Post$	-0.052***	-0.080*	-0.124***	-0.050***	-0.071*	-0.116***
	(0.009)	(0.045)	(0.039)	(0.008)	(0.040)	(0.034)
$Treat \cdot Post$	-0.033			-0.001		
	(0.113)			(0.101)		
N	699,013	73,637	108,009	699,013	73,637	108,009
Adj. R^2	0.943	0.901	0.905	0.945	0.899	0.906
Unique Plants	81,432	8,181	11,834	81,432	8,181	11,834
Plant FE	Y	Y	Y	Y	Y	Y
Industry-Year FE	Y	Y	Y	Y	Y	Y
Parent-Year FE	Y	-	-	Y	-	-

Notes: Columns (1) and (4) of this table reports estimates of equation 2 on our primary outcomes of the inverse hyperbolic sine of damages and emissions. For these specifications, $Treat$ is defined as as the difference between the treatment intensity for a state that individual plant (i) is located in and the average treatment intensity for other plants within the same parent company. Columns (2), (4), (5), and (6) of this table reports estimates of equation 1 on our primary outcomes of the inverse hyperbolic sine of damages and emissions. All specifications include industry-year and plant fixed effects, with columns (1) and (4) considering parent-year fixed effects additionally. Standard errors are clustered at the plant-level. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.
Sample: Plant-year level data from the National Emissions Inventory beginning in 2010 and ending in 2019. Columns (1) and (4) only include plants observed prior to and following policy implementation. Columns (2) and (5) consider the subset of plants where the multi-plant parent is located in a single state. Columns (3) and (6) consider the sample of multi-plant parent firms where the firm receives a higher than parent average treatment intensity.

Table 8: Robustness: Industry Spillovers and the EPA Watch List

	Damages		Emissions	
	(1)	(2)	(3)	(4)
$Policy \cdot Post \cdot Treat$	0.043	0.359	0.023	-2.271
	(0.063)	(1.669)	(0.058)	(1.480)
$Policy \cdot Post$	-0.270***	-0.049***	-0.238***	-0.026*
	(0.033)	(0.017)	(0.025)	(0.015)
$Treat \cdot Post$	0.090**	-2.586***	0.012	-1.473**
	(0.039)	(0.773)	(0.033)	(0.665)
N	173,752	699,013	173,752	699,013
adj. R^2	0.915	0.932	0.916	0.933
Unique Plants	20,260	81,432	20,260	81,432
Plant FE	Y	Y	Y	Y
Industry-Year FE	-	Y	-	Y
Industry-Year Trends	Y	-	Y	-

Notes: This table reports estimates of equation (2) on our primary outcomes of the inverse hyperbolic sine of damages and emissions. In columns (1) and (3), we define $Treat$ as the difference between the state treatment intensity where the plant is located and the average treatment intensity for other plants within the same industry according to the six-digit NAICS code. Columns (2) and (4) report results for the specification when $Treat$ is defined $Treat$ as the percentage of plants that appeared on the EPA watch list out of all plants in that state prior to the policy change. All columns include plants fixed effects, while columns (1) and (3) include industry-year trends additionally, and columns (2) and (4) consider industry-year fixed effects. Standard errors are clustered at the plant-level. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.
Sample: Columns (2) and (4) consider plant-year level data from the National Emissions Inventory beginning in 2010 and ending in 2019. We only include plants observed prior to and following policy implementation. Columns (1) and (3) consider the subset of these plants that are from manufacturing industries.

plants with varying levels of treatment prior to the policy. Second, by examining individual treatment effects for each of the post-treatment years, we can better understand how plants and regulators responded to the policy.

Our econometric specification for the dynamic treatment effects analysis is as follows:

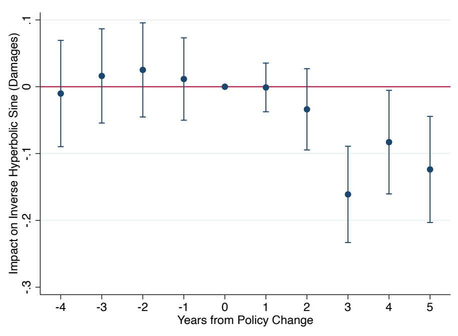
$$Y_{i,t} = \alpha_0 + \sum_{\tau \in [-4,5]} \beta_\tau \cdot \mathbb{1}(t = \tau) \cdot Policy_s + \gamma_i + \delta_t + \phi_{j,t} + \varepsilon_{i,t} \tag{3}$$

The key coefficients β_τ represent changes in the outcome with respect to the level of treatment intensity in year τ. The identified treatment effects window is 2010–2019 with 2014, the year the policy was announced, as the baseline year.

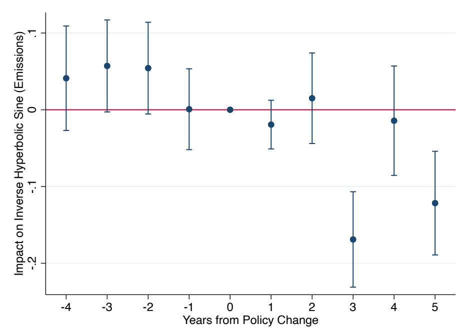
We normalize with respect to this baseline year by setting $\beta_0 = 0$, which is consistent with equation 1 and the understanding that the policy should impact plants in the years following the announcement. We also use the fully balanced subsample of plants for this analysis to avoid differences in sample composition between years biasing the β_τ estimates. Figure 5a plots the estimated coefficients of β_τ for the impact of the policy on damages and Figure 5b provides analogous results for emissions. Although the corresponding regression results for these figures in appendix Table A3 indicate that two of the pre-treatment coefficients for emissions are statistically significant at the 5% level (but not the 1% level), these pre-policy effects are positive suggesting that treatment intensity corresponded with worse emissions prior to 2014. Therefore, these figures largely support the conjecture that prior to the policy, treatment intensity did not correspond with any differential trends for our primary outcomes. However, following policy implementation, we observe negative and statistically significant treatment effects for both damages and emissions beginning with the third full year of treatment in 2017. This indicates that plants and/or state regulators may have taken time to adapt to the policy. Overall, these results indicate our findings are not driven by the choice of research design and are largely supportive of the common trends assumption necessary for a subset of our primary specifications.

6.3 Robustness of the Primary Specification

Given the significance of our primary two-way fixed effects approach, it is important to demonstrate the robustness of our results to a number of specification choices. In the appendix, we present additional results including: (i) testing the importance of functional form by considering the log of the outcome plus one and the level outcome (Table A4), (ii) examining the importance of including or excluding extreme outliers from our analysis and the inclusion of six-digit NAICS by year fixed effects (Table A5), (iii) using a fully balanced panel to examine the importance of the exclusion of plants that opened or closed during the sample



(a) IHS(Damages)



(b) IHS(Emissions)

Figure 5: Dynamic Effects

Source: National Emissions Inventory, 2010 - 2019

time period (Table A5), and (iv) testing the sensitivity of the baseline results to alternative definitions of treatment. Specifically, in Table A6 we first redefine treatment by including matrix criteria one and three as part of the revised criteria and as a second alternative we redefine treatment using the proportion of all plants in a state that received an HPV using the delisted criteria. The robustness of our results under these alternative sample and specification choices confirms the nature of our findings.

Additionally, in the appendix we provide further evidence of the robustness of our results regarding to other research design choices. These appendix analyses include (i) testing the statistical significance of the baseline results using standard errors clustered at the state level and standard errors clustered at the state-year level (Table A7), (ii) using a binary definition to estimate a standard difference-in-differences model comparing plants in the highest and lowest quartiles (deciles) of treatment (Table A9), (iii) examining the importance of the NEI dataset by instead estimating our baseline specification with the alternative TRI dataset (Table A8), with a discussion of this data in section 4.1, (v) Using alternative damage and emissions outcomes including or excluding certain pollutants (Table A10). Specifically, we consider alternative outcomes that separately distinguish NO_x and SO_2 from the other pollutants, since NO_x and SO_2 are disproportionately emitted by the energy sector which is subject to a variety of other regulatory factors. Finally, we investigate the presence of nonlinear treatment effects of the policy using a cubic spline (Figure A1). Overall, these alternative research design choices confirm both the nature of our primary results and the validity of our original research design.

7 Conclusion

Heterogeneity in violations and the damages they cause is one of the major difficulties in designing an enforcement policy for environmental regulation. Typically, regulators must prioritize more costly environmental violations over others when faced with limited enforcement resources. Difficulties with the design of enforcement policy are further compounded in settings such as the United States, where federal policies are often enforced by separate local regulators, which may vary in their preferences or knowledge. When local regulators have greater knowledge about what violations pose the greatest potential harm, increased discretion may improve welfare. In contrast, discretion may lower welfare if local regulators have interests that differ from those of the central government.

This paper seeks to understand whether environmental regulators reallocate enforcement resources to the most harmful violations when faced with reduced regulatory discretion and ultimately decrease the level of environmental harm. We answer this question using a quasi-experiment emerging from a 2014 reduction in the number of criteria for classifying air-polluting facilities as a priority, the highest tier of noncompliance. The stated intent of the policy by the federal EPA was to "focus on CAA violations that experience shows are most likely to be significant for human health and the environment" [EPA, 2014]. States that had previously devoted significant resources to the delisted criteria were treated by the policy to a greater extent, as they had more resources to allocate toward other enforcement priorities. The results suggest that after the policy change, plants in states with a higher degree of treatment exhibited a greater increase in their average penalties per priority violation, suggesting that enforcement resources were reallocated to the most harmful violations. Furthermore, this change in enforcement practices corresponded with a larger reduction in emissions and damages from plants located in those treated states.

In recent years, there has been a persistent downward trend in the resources available for the enforcement of environmental regulation at both the U.S. state and federal level [Kelderman et al., 2019, Blundell et al., 2021]. This paper demonstrates the important role of policies that focus limited enforcement resources on violations that pose the greatest environmental harm. In our setting, environmental damages decreased after the discretion of local regulators to classify plants with the highest level of scrutiny was reduced. Further work is needed to clarify the role of local regulators' preferences versus knowledge as well as the applicability of our results to other settings.

References

Jérôme Adda, Brendon McConnell, and Imran Rasul. Crime and the depenalization of cannabis possession: Evidence from a policing experiment. *Journal of Political Economy*, 122(5):1130–1202, 2014.

Maximilian Auffhammer, Antonio M Bento, and Scott E Lowe. Measuring the effects of the clean air act amendments on ambient pm10 concentrations: The critical importance of a spatially disaggregated analysis. *Journal of Environmental Economics and Management*, 58(1):15–26, 2009.

Marc F Bellemare and Casey J Wichman. Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):50–61, 2020.

Antonio Bento, Matthew Freedman, and Corey Lang. Who benefits from environmental regulation? evidence from the clean air act amendments. *Review of Economics and Statistics*, 97(3):610–622, 2015.

Xiang Bi. "cleansing the air at the expense of waterways?" empirical evidence from the toxic releases of coal-fired power plants in the united states. *Journal of Regulatory Economics*, 51(1):18–40, 2017.

Wesley Blundell, Gautam Gowrisankaran, and Ashley Langer. Escalation of scrutiny: The gains from dynamic enforcement of environmental regulations. *American Economic Review*, 110(8):2558–85, 2020.

Wesley Blundell, Mary F Evans, and Sarah L Stafford. Regulating hazardous wastes under us environmental federalism: The role of state resources. *Journal of Environmental Economics and Management*, 108: 102464, 2021.

Jie Cai, Jiawei Luo, Shulin Wang, and Sheng Yang. Feature selection in machine learning: A new perspective. *Neurocomputing*, 300:70–79, 2018.

A Colin Cameron and Douglas L Miller. A practitioner's guide to cluster-robust inference. *Journal of human resources*, 50(2):317–372, 2015.

H Ron Chan, B Andrew Chupp, Maureen L Cropper, and Nicholas Z Muller. The impact of trading on the costs and benefits of the acid rain program. *Journal of Environmental Economics and Management*, 88: 180–209, 2018.

Kenneth Y Chay and Michael Greenstone. Does air quality matter? evidence from the housing market. *Journal of political Economy*, 113(2):376–424, 2005.

Karen Clay, Akshaya Jha, Nicholas Muller, and Randall Walsh. External costs of transporting petroleum products: Evidence from shipments of crude oil from north dakota by pipelines and rail. *The Energy Journal*, 40(1), 2019.

Cl'ement De Chaisemartin and Xavier d'Haultfoeuille. Fuzzy differences-in-differences. *The Review of Economic Studies*, 85(2):999–1028, 2018.

Cl'ement de Chaisemartin, Xavier D'Haultfoeuille, and Yannick Guyonvarch. Fuzzy differences-in-differences with stata. *The Stata Journal*, 19(2):435–458, 2019.

Olivier Deschenes, Michael Greenstone, and Joseph S Shapiro. Defensive investments and the demand for air quality: Evidence from the nox budget program. *American Economic Review*, 107(10):2958–2989, 2017.

Esther Duflo, Michael Greenstone, Rohini Pande, and Nicholas Ryan. The value of regulatory discretion: Estimates from environmental inspections in india. *Econometrica*, 86(6):2123–2160, 2018.

Dietrich Earnhart. Panel data analysis of regulatory factors shaping environmental performance. *Review of Economics and Statistics*, 86(1):391–401, 2004.

Dietrich Earnhart and Lana Friesen. Enforcement federalism: comparing the effectiveness of federal punishment versus state punishment. *Environmental and Resource Economics*, 78(2):227–255, 2021a.

Dietrich Earnhart and Lana Friesen. Use of competitive endogenous audit mechanisms by federal and state inspectors within environmental protection agencies. *Journal of Environmental Economics and Management*, 109:102476, 2021b.

EPA. Clean Air Act stationary source civil penalty policy, October 1991.

EPA. The Timely and Appropriate (T&A) enforcement response to High Priority Violations (HPVs). Technical report, 1998.

EPA. Timely and Appropriate (T&A) enforcement response to High Priority Violations (HPVs). Technical report, Office of Enforcement and Compliance Assurance, 1999a.

EPA. Section 112(g) of the Clean Air Act. Technical report, Environmental Protection Agency, 1999b.

EPA. Revision on Timely and Appropriate (T&A) Enforcement Response to High Priority Violations. Technical report, Environmental Protection Agency, 2014.

EPA. AFS data element dictionary, February 2015.

Mary F Evans. The clean air act watch list: An enforcement and compliance natural experiment. *Journal of the Association of Environmental and Resource Economists*, 3(3):627–665, 2016.

Mary F Evans and Sarah L Stafford. The clean air act watch list and federal oversight of state enforcement efforts. *Journal of Environmental Economics and Management*, 93:170–184, 2019.

Mary F Evans, Scott M Gilpatric, and Jay P Shimshack. Enforcement spillovers: Lessons from strategic interactions in regulation and product markets. *The Journal of Law and Economics*, 61(4):739–769, 2018.

George M Furnival and Robert W Wilson. Regressions by leaps and bounds. *Technometrics*, 42(1):69–79, 2000.

Matthew Gibson. Regulation-induced pollution substitution. *Review of Economics and Statistics*, 101(5): 827–840, 2019.

Wayne B Gray and Jay P Shimshack. The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence. *Review of Environmental Economics and Policy*, 2011.

Michael Greenstone. Estimating regulation-induced substitution: The effect of the clean air act on water and ground pollution. *American Economic Review Papers and Proceedings*, 93(2):442–448, 2003.

Rema Nadeem Hanna and Paulina Oliva. The impact of inspections on plant-level air emissions. *The BE Journal of Economic Analysis & Policy*, 10(1), 2010.

Harry H Harman. *Modern factor analysis*. University of Chicago press, 1976.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

Karam Kang and Bernardo S Silveira. Understanding disparities in punishment: Regulator preferences and expertise. *Journal of Political Economy*, 129(10):2947–2992, 2021.

Keene Kelderman, Eric Schaeffer, Tom Pelton, Ari Phillips, and Courtney Bernhardt. The thin green line: Cuts in state pollution control agencies threaten public health. *Environmental Integrity Project*, 2019.

Matthew Leisten and Nicholas Vreugdenhil. Dynamic regulation with firm linkages: Evidence from texas. 2023.

Robert J MacCoun, Peter Reuter, et al. *Drug war heresies: Learning from other vices, times, and places*. Cambridge University Press, 2001.

Dilip Mookherjee and Ivan PL Png. Marginal deterrence in enforcement of law. *Journal of Political Economy*, 102(5):1039–1066, 1994.

Yingfei Mu, Edward A Rubin, and Eric Zou. What's missing in environmental (self-) monitoring: Evidence from strategic shutdowns of pollution monitors. Technical report, National Bureau of Economic Research, 2021.

Wallace E Oates. An essay on fiscal federalism. *Journal of economic literature*, 37(3):1120–1149, 1999.

A Mitchell Polinsky and Steven Shavell. On offense history and the theory of deterrence. *International Review of Law and Economics*, 18(3):305–324, 1998.

Binish Rijal and Neha Khanna. High priority violations and intra-firm pollution substitution. *Journal of Environmental Economics and Management*, 103:102359, 2020.

Jay P Shimshack. The economics of environmental monitoring and enforcement. *Annu. Rev. Resour. Econ.*, 6(1):339–360, 2014.

Jay P Shimshack and Michael B Ward. Enforcement and over-compliance. *Journal of Environmental Economics and Management*, 55(1):90–105, 2008.

Jason M Walter and Zach Raff. When the regulator goes home: The effectiveness of environmental oversight. *Available at SSRN 3394867*, 2019.

Bing Zhang, Xiaolan Chen, and Huanxiu Guo. Does central supervision enhance local environmental enforcement? quasi-experimental evidence from china. *Journal of Public Economics*, 164:70–90, 2018.

Ruohao Zhang and Neha Khanna. Strategic local regulators and the efficacy of pollution standards. 2021.

Eric Yongchen Zou. Unwatched pollution: The effect of intermittent monitoring on air quality. *American Economic Review*, 111(7):2101–2126, 2021.

Appendix: Supplementary Analysis, Figures and Tables

Additional Tables

In this section, we first present the results of policy impact on our primary outcomes obtained via a dynamic treatment effects framework (Table A3). Second, we estimate the sensitivity of our primary results for damages by altering the functional form of the dependent variable (Table A4). Third, we re-estimate our primary specification using an alternative set of fixed effects, the balanced subsample of plants and samples both including and excluding extreme outliers (Table A5). Fourth, A6 provides the baseline specification results using alternative continuous definitions of treatment. Fifth, Table A7 provides alternative sets of the standard errors for our baseline specifications. Sixth, we show the impact of the policy on damages and emissions for the TRI sample (Table A8. Seventh, A9 provides the baseline specification results using binary definitions of treatment. Eighth, Table A10 provides the baseline specification results using alternative combinations of pollutants to comprise total damages and emissions. Finally, we test the validity of our results for enforcement practices by examining the impact of the policy on state–only penalties and inspections for priority and non-priority violation plants, Table A12 and Table A13.

Table A1: Post-2014 enforced criteria

Criterion		**Violation**
	1	*Failure to obtain a New Source Review (NSR) permit (for either attainment or non-attainment areas) and/or install Best Available Control Technology (BACT) or Lowest Available Emission Reductions (LAER) (and/or obtain offsets) for any new major stationary source or major modification at a major stationary source.*
	2	*A violation of any federally enforceable emission limitation, emission standard or operating parameter, which is a surrogate for emissions, that was issued pursuant to Title I, Part C or D, of the CAA and the implementing regulations, or the equivalent provision(s) in an EPA-approved implementation plan (state, local, territorial or tribal) where such violation continued (or is expected to continue) for at least seven days.*
General	**3**	*A violation of any emission limitation, emission standard or operating parameter, which is a surrogate for emissions, in an applicable Standards of Performance for New Sources (NSPS) (Part 60) or in an analogous regulation adopted by state, local, tribal or territorial authorities and the EPA has granted delegation to enforce such regulations in lieu of the NSPS where such violation continued (or is expected to continue) for at least seven days.*
	4	*A violation of any emission limitation, standard or surrogate parameter (emission or operating) of an applicable National Emission Standards for Hazardous Air Pollutants (NESHAP) (Parts 61 and 63) or in an analogous regulation adopted by state, local, tribal or territorial authorities and EPA has granted delegation to enforce such regulations in lieu of the NESHAP where such violation continued (or is expected to continue) for at least seven days.*
	5	*A violation that involves federally enforceable work practices, testing requirements, monitoring requirements, recordkeeping or reporting that substantially interferes with enforcement of a requirement or a determination of the source's compliance. The determination of what is substantial shall be part of a case-by-case analysis/discussion between the EPA Region and the enforcement agency.*
	6	*Any other violations specifically identified and communicated to enforcement agencies from time to time by the Director, Air Enforcement Division (AED), U.S. EPA (general applicability) or as mutually agreed upon between the enforcement agency and corresponding EPA region (case-by-case). For example, an enforcement agency believes an emission violation warrants designation as an HPV even thought the violation lasted (or will last) for less than seven days.*

Notes: This table provides brief summary of the enforced post-2014-revision HPV criteria.

Table A2: Pre-2014 enforced criteria

Criterion		**Violation**
General	**1**	*Failure to obtain a PSD permit (and/or to install BACT), an NSR permit (and/or to install LAER or obtain offsets) and/or a permit for a major modification of either.*
	2	*Violation of an air toxics requirement (i.e., NESHAP, MACT) that either results in excess emissions or violates operating parameter restrictions.*
	3	*Violation by a synthetic minor of an emission limit or permit condition that affects the source's PSD, NSR or Title V status (i.e., fails to comply with permit restrictions that limit the source's potential emissions below the appropriate thresholds; refers only to pollutants for which the source is a synthetic minor. It is not necessary for a source's actual emissions to exceed the NSR/PSD/Title V thresholds.)*
	4	*Violation of any substantive term of any local, state or federal order, consent decree or administrative order.*
	5	*Substantial violation of the source's Title V certification obligations, e.g., failure to submit a certification.*
	6	*Substantial violation of the source's obligation to submit a Title V permit application. (i.e., failure to submit a permit application within sixty (60) days of the applicable deadline.)*
	7	*Violations that involve testing, monitoring, record keeping or reporting that substantially interfere with enforcement or determining the source's compliance with applicable emission limits.*
	8	*A violation of an allowable emission limit detected during a reference method stack test.*
	9	*Clean Air Act (CAA) violations by chronic or recalcitrant violators* [20].
	10	*Substantial violation of Clean Air Act Section 112(r) requirements (for permitting authorities that are not implementing agencies under Section 112(r) program, limited to source's failure to submit Section 112(r) risk management plan).*
Matrix	**1A**	*Violation of allowable emissions limitations detected via reference method stack testing*
	1B	*Violation of allowable emissions limitations detected via coatings analysis, fuel samples or other process material sampling.*
	2A	*Violation of parameter emissions limitations detected via continuous/periodic parameter monitoring.*
	3A	*Violation of applicable standards (non-opacity) under continuous emissions monitoring (where the CEM is certified under federal performance specifications.)*
	4A	*Violation of applicable standards (opacity) under continuous opacity monitoring.*
	4B	*Violation of applicable standards (opacity) under method 9 visual emissions readings.*

Notes: This table provides brief summary of the enforced prior to 2014 HPV criteria.

Table A3: Dynamic Effects

	IHS(Damages)	IHS(Log Emissions)
	(1)	(2)
Year 4 prior	-0.010	0.041
	(0.031)	(0.026)
Year 3 prior	0.016	0.057**
	(0.027)	(0.023)
Year 2 prior	0.025	0.054**
	(0.027)	(0.023)
Year 1 prior	0.012	0.001
	(0.024)	(0.020)
Year 1 post	-0.001	-0.019
	(0.014)	(0.012)
Year 2 post	-0.034	0.015
	(0.024)	(0.023)
Year 3 post	-0.161***	-0.169***
	(0.028)	(0.024)
Year 4 post	-0.083***	-0.014
	(0.030)	(0.028)
Year 5 post	-0.124***	-0.122***
	(0.031)	(0.026)
N	313,926	313,926
Adj. R^2	0.936	0.940
Unique plants	31,476	31,476
Plant FE	Y	Y
Industry-Year FE	Y	Y

Notes: This table shows the dynamic treatment effect results from equation 3 for the impact of the implementation of the policy on the inverse hyperbolic sine of pollution damages and emissions. We show estimated coefficients four years prior and five years after the HPV policy revision. Both specifications include plant and industry-level fixed effects. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.
Sample: We consider the fully balanced subsample of plants that appeared at in every year of our estimation sample and the 2008 NEI.

Table A4: Robustness: Functional Form

	Damages			Emissions		
	(1)	(2)	(3)	(4)	(5)	(6)
$Policy \cdot Post$	-184,358	-135,235**	-0.194***	-6.811***	-7.611***	-0.082***
	(131,316.616)	(52,769.535)	(0.031)	(2.298)	(2.027)	(0.012)
N	699,013	697,655	699,013	699,013	697,655	699,013
Adj. R^2	0.750	0.869	0.882	0.895	0.900	0.938
Plant FE	Y	Y	Y	Y	Y	Y
Industry-Year FE	Y	Y	Y	Y	Y	Y

Notes: This table reports estimates of equation 1 on damages and emissions without considering the inverse hyperbolic sine transformation. Specifically, columns (1), (2), (4), and (5) consider the level outcome. Columns (3) and (6) consider the log of the damages or emissions plus one. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. **Sample:** Columns (1), (3), (4), and (6) consider annual plant observations from the NEI. Columns (2) and (6) excludes the extreme outliers, those with more than 100 million in damages.

Table A5: Robustness: Fixed Effects and Outliers

	Damages				Emissions			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Policy · Post	-0.063***	-0.058***	-0.112***	-0.070***	-0.063***	-0.055***	-0.114***	-0.068***
	(0.010)	(0.010)	(0.023)	(0.010)	(0.009)	(0.009)	(0.020)	(0.009)
N	698,037	813,477	336,280	697,655	698,037	813,477	336,280	697,655
Adj. R^2	0.933	0.938	0.933	0.930	0.934	0.939	0.937	0.932

Notes: This table reports estimates of equation 1 on our primary outcomes of the inverse hyperbolic sine of damages and emissions. Columns (1) and (5) use industry-year fixed effects at the six digit NAICS code level. Columns (2) and (6) consider the full sample without omitting the extreme outliers in emissions. Columns (3) and (6) consider the fully balanced sample of plants that appear in every NEI from 2010–2019. Columns (4) and (8) consider the primary estimation sample but omitting the small set of extreme outliers with damages greater than $100 million USD. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. **Sample:** Plant-year level data from the National Emissions Inventory beginning in 2010 and ending in 2019. We only include plants observed prior to and following policy implementation.

Table A6: Robustness: Alternative Definitions of Treatment

	Damages		Emissions	
	(1)	(2)	(3)	(4)
Policy · Post	-0.040***	-1.443***	-0.026**	-1.635***
	(0.014)	(0.127)	(0.012)	(0.112)
N	699,013	699,013	699,013	699,013
adj. R^2	0.932	0.932	0.933	0.933
Plant FE	Y	Y	Y	Y
Industry-Year FE	Y	Y	Y	Y

Notes: This table reports estimates of equation 1 on our primary outcomes of the inverse hyperbolic sine of damages and emissions. Columns (1) and (3) revise the definition of treatment to remove matrix criteria one and three from the excluded criteria of the policy revision. Columns (2) and (4) redefine the definition of treatment to be the pre-treatment proportion of plants within a state with at least one delisted violation. All specification include plant and industry-level fixed effects. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. **Sample:** Plant-year level data from the National Emissions Inventory beginning in 2010 and ending in 2019. We only include plants observed prior to and following policy implementation.

Table A7: Robustness: Alternative Standard Errors

	IHS(Damages) (1)	IHS(Emissions) (2)
$Policy \cdot Post$	-0.069	-0.067
Facility Clustered SE	(0.010)	(0.009)
State Clustered SE	(0.034)	(0.038)
State-Year Clustered SE	(0.017)	(0.019)
N	699,013	699,013
Adj. R^2	0.932	0.932
Unique Plants	81,432	81,432
Plant FE	Y	Y
Industry-Year FE	Y	Y

Notes: This table reports estimates of equation 1 on our primary outcomes of the inverse hyperbolic sine of damages and emissions. The standard errors under differing clusters are presented including facility-clustered, state-clustered, and state-year clustered errors. Each specification includes plant and industry-level fixed effects. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. **Sample:** Plant-year level data from the National Emissions Inventory beginning in 2010 and ending in 2019. We only include plants observed prior to and following policy implementation.

Table A8: Baseline Results: TRI Data

	IHS(Damages) (1)	IHS(Emissions) (2)
$Policy \cdot Post$	-0.269***	-0.176***
	(0.050)	(0.030)
N	112,282	112,282
Adj. R^2	0.912	0.931
Unique plants	13,394	13,394
Plant FE	Y	Y
Industry-Year FE	Y	Y

Notes: This table reports estimates of equation 1 on our primary outcomes of the inverse hyperbolic sine of damages and emissions using TRI data. Column (1) and (2) consider damages and emissions, respectively, as an outcome. Both specifications include plant and industry-level fixed effects. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. **Sample:** We consider full sample of plants that appeared at in the TRI between 2010 - 2019.

Table A9: Robustness: Binary Treatment Definition

	Damages		Emissions	
	(1)	(2)	(3)	(4)
$Policy \cdot Post$	-0.032***	-0.028***	-0.041***	-0.020***
	(0.008)	(0.005)	(0.007)	(0.005)
N	148,710	330,291	148,710	330,291
Adj. R^2	0.933	0.931	0.933	0.931
Plant FE	Y	Y	Y	Y
Industry-Year FE	Y	Y	Y	Y

Notes: This table reports estimates of equation 1 on our primary outcomes of the inverse hyperbolic sine of damages and emissions using binary treatment definition. Columns (1) and (3) revise the definition of treatment to take a value of one if a plant is in the top decile of treatment and zero if the plant is in the bottom decile of treatment. Columns (2) and (4) revise the definition of treatment to take a value of one if a plant is in the top quartile of treatment and zero if the plant is in the bottom quartile of treatment. All specification include plant and industry-level fixed effects. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. **Sample:** Plant-year level data from the National Emissions Inventory beginning in 2010 and ending in 2019. We only include plants observed prior to and following policy implementation.

Table A10: Robustness: Alternative Outcomes

	Damages		Emissions	
	(1)	(2)	(3)	(4)
$Policy \cdot Post$	-0.054***	-0.051***	-0.044***	-0.054***
	(0.008)	(0.009)	(0.007)	(0.008)
N	699,013	699,013	699,013	699,013
Adj. R^2	0.923	0.928	0.921	0.932
Plant FE	Y	Y	Y	Y
Industry-Year FE	Y	Y	Y	Y

Notes: This table reports estimates of equation 1 on alternatively defined outcomes. Columns (1) - (2) consider the outcome of damages, while columns (3) and (4) consider the outcome of emissions. Columns (1) and (3) define outcome without NO_X and SO_2. Columns (2) and (4) consider damages and emissions from NO_X and SO_2 only as an outcome. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level. **Sample:** Plant-year level data from the National Emissions Inventory beginning in 2010 and ending in 2019. We only include plants observed prior to and following policy implementation.

Table A11: Policy impact on state enforcement by regulatory status

	HPV Plants			Other Noncompliant Plants		
	Mean pre (se)	Mean post (se)	Group Diff (p-value)	Mean pre (se)	Mean post (se)	Group Diff (p-value)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All Plants						
State penalties	2,881.042	2,286.419	-594.623	741.429	857.297	115.869
($ per violation)	(23,576.164)	(12,207.019)	(0.133)	(6,805.154)	(10,672.958)	(0.625)
State inspections	0.671	0.534	-0.137***	1.109	1.286	0.177***
(# per violation)	(2.055)	(1.540)	(0.000)	(3.122)	(2.127)	(0.003)
N	6,242	4,184		2,359	5,980	
Panel B: Plants Top Quintile of Treatment						
State penalties	3,915.008	3,137.375	-777.632	1,870.002	1,172.271	-697.731
($ per violation)	(17,183.564)	(15,996.271)	(0.259)	(11,635.036)	(7,593.055)	(0.165)
State inspections	1.177	0.843	-0.335***	1.517	1.953	0.436**
(# per violation)	(3.438)	(1.981)	(0.005)	(4.133)	(2.748)	(0.016)
N	1,291	1,072		373	1,381	
Panel C: Plants Bottom Quintile of Treatment						
State penalties	4,794.229	3,417.397	-1,376.831	503.390	868.206	364.816
($ per violation)	(37,085.871)	(16,968.375)	(0.343)	(5,478.806)	(19,416.947)	(0.663)
State inspections	0.695	0.623	-0.072	2.401	1.235	-1.166***
(# per violation)	(1.856)	(1.133)	(0.347)	(4.980)	(1.754)	(0.000)
N	963	761		555	1,393	

Notes: The table shows mean values for key state enforcement variables before and after the HPV policy change in 2014. Columns (1)–(3) correspond to HPV values, while columns (4)–(6) present values for the FRVs. Columns (1) and (4) show mean plant-annual pre-policy change observations. Columns (2) and (5) correspond to average plant-annual observations after the policy change. Columns (3) and (6) present the difference in group mean values. The time frame is 2010–2019 using data from ICIS databases. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A12: Robustness: State Penalties per Violation

	HPV Plants			Other Noncompliant Plants		
	Base	TC	CIC	Base	TC	CIC
	(1)	(2)	(3)	(4)	(5)	(6)
Policy · Post	10,083.44	10,167.77	14,097.73	1,148.186	11,246.11	10,061.13
	(8,820.43)	(13,865.58)	(15,253.35)	(3147.902)	(11,146.4)	(9,079.76)
N	7,686	7,686	7,686	3,959	3,959	3,959
R^2	0.342	-	-	0.516	-	-
Unique Plants	1,759	1,759	1,759	1,476	1,476	1,476
Plant FE	Y	Y	Y	Y	Y	Y
EPA Region-Year FE	Y	Y	Y	Y	Y	Y
Industry-Year Trend	Y	Y	Y	Y	Y	Y

Notes: This table reports estimates of our baseline equation 1 with state fine amount per violation as estimated outcome. Base specification is our baseline DID estimate, TC stands for "Time-Corrected" estimate for equation 1 according to de Chaisemartin et al. [2019], and CIC is the "changes-in-changes" estimate for α_1 from equation 1 according to de Chaisemartin et al. [2019]. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Sample: Plant-year observations from the ICIS beginning in 2010 and ending in 2019. Columns (1) to (3) consider plants with at least one HPV in the current year. Columns (4) to (6) consider plants that are noncompliant but non-HPV in the current year.

Table A13: Robustness: State Inspections per Violation

	HPV Plants			Other Noncompliant Plants		
	Base	TC	CIC	Base	TC	CIC
	(1)	(2)	(3)	(4)	(5)	(6)
Policy $\cdot$ *Post*	-0.459	-0.785	-0.395	-0.461	-2.926	-4.765
	(0.591)	(1.332)	(1.199)	(0.927)	(2.199)	(2.908)
N	7,686	7,686	7,686	3,959	3,959	3,959
R^2	0.345	-	-	0.640	-	-
Unique Plants	1,759	1,759	1,759	1,476	1,476	1,476
Plant FE	Y	Y	Y	Y	Y	Y
EPA Region-Year FE	Y	Y	Y	Y	Y	Y
Industry-Year Trend	Y	Y	Y	Y	Y	Y

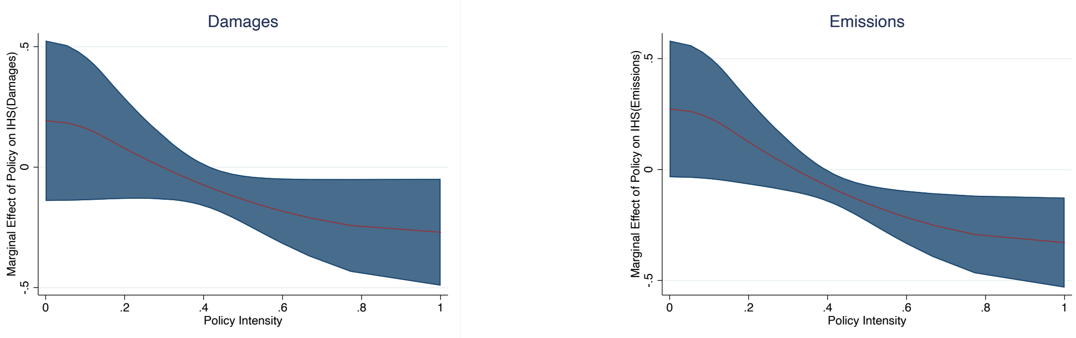
Notes: This table reports estimates of our baseline equation 1 with the number of state inspections per violation as estimated outcome. Base specification is our baseline DID estimate, TC stands for "Time-Corrected" estimate for equation 1 according to de Chaisemartin et al. [2019], and CIC is the "changes-in-changes" estimate for α_1 from equation 1 according to de Chaisemartin et al. [2019]. We use plant-level clustered standard errors. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Sample: Plant-year observations from the ICIS beginning in 2010 and ending in 2019. Columns (1) to (3) consider plants with at least one HPV in the current year. Columns (4) to (6) consider plants that are noncompliant but non-HPV in the current year.

Table A14: Policy impact on enforcement, compliance, and damages across plants

	Mean pre-policy (se)	Mean post-policy (se)	Group Diff (p-value)
	(1)	(2)	(3)
Panel A: Plants in Top Quintile of Treatment			
Emissions (tons)	71.436	63.912	-7.524***
	(279.785)	(253.559)	(0.000)
Penalties ($)	1,598.679	1,551.978	-46.701
	(34,015.727)	(50,450.875)	(0.868)
Inspections (#)	0.913	0.789	-0.124***
	(3.569)	(2.254)	(0.000)
Damages (mln $)	2.394	2.321	-0.073
	(15.665)	(15.611)	(0.481)
Compliance (%)	0.965	0.944	-0.021***
	(0.183)	(0.229)	(0.000)
HPV	0.200	0.155	-0.045***
	(1.876)	(1.363)	(0.000)
Noncomplaint Non-HPV	0.046	0.107	0.061***
	(0.524)	(0.802)	(0.000)
Observations	47,744	44,141	91,885
Panel B: Plants in Bottom Quintile of Treatment			
Emissions (tons)	66.593	58.469	-8.124***
	(283.602)	(252.157)	(0.000)
Penalties ($)	638.685	544.015	-94.670
	(20,855.664)	(24,175.762)	(0.473)
Inspections (#)	0.571	0.501	-0.070***
	(1.952)	(1.617)	(0.000)
Damages (mln $)	1.542	1.435	-0.108**
	(8.641)	(8.638)	(0.033)
Compliance (%)	0.975	0.962	-0.013***
	(0.157)	(0.191)	(0.000)
HPV	0.064	0.057	-0.006
	(0.692)	(0.673)	(0.107)
Noncomplaint Non-HPV	0.018	0.057	0.038***
	(0.219)	(0.440)	(0.000)
Observations	60,124	56,587	116,711

Notes: The table shows mean values for key variables before and after the HPV policy change in 2014. Column (1) correspond to values before the policy change, while column (2) presents the mean values post-2014 HPV revision. Column (3) presents the difference in group mean values. The time frame is 2010–2019 using data from ICIS databases. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.



(a) IHS(Damages) (b) IHS(Emissions)

Figure A1: Spline Estimation

Notes: This is cubic spline formulation of equation 1 that replaces $Policy \cdot Post$ with three dosage points for treatment intensity (0, 0.15, and 1). The data used is our plant-year data from the National Emissions Inventory beginning in 2010 and ending in 2019.

Feature Selection and Factor Analysis

The federal EPA in its 2014 HPV policy revision eliminated all matrix criteria violations and reduced the number of general criteria violations from ten to seven. Given that the stated intent of the policy was to prioritize violation classifications with greater potential harm to human health and environment, we should observe that the remaining HPV criteria correspond with violations that had the highest observed damages, pre-policy. In section 4.2, we employed bi-directional step-wise feature selection to execute this task. This analysis largely confirmed that the federal EPA eliminated violations that are associated with lower damages. However, the results from this feature selection model below indicate there were three total criteria selected by our algorithm as being associated with a high level of plant damages that nonetheless were delisted in the policy revision, two of these three were matrix criteria. In this section, we investigate whether the inclusion of these matrix criteria by our algorithm was an error, due to these violations often occurring in conjunction with more severe types of violations that were maintained under the revised policy.

Feature Selection Results

$$GC1, GC2, \underline{GC6}, GC8, GC9, \underline{M1}, \underline{M3}$$

Recall that matrix criteria often appear in our data in combination with other violations only, hence it is hard to estimate the marginal damages per matrix criterion violation ($M1, M2, M3, M4$). Therefore, we use factor analysis (FA) to identify any latent variables that include matrix criteria violations.

Table A15 provides the FA results. The factor analysis identifies nine latent variables , meaning that there are nine major types of violations that are captured by originally enforced 14 general and matrix criteria. The interest to us is that the two matrix criteria $M1$ and $M3$ have low uniqueness values of 0.177 and 0.431. This translates to these criteria being redundant as compared to the use of other criteria for the classification of noncompliance events that pose significant environmental harm. Similarly, we find a low uniqueness score of 0.192 for the other criteria chosen by our feature selection model, GC6, which was eliminated in the policy revision.

A Simple Model

Here we outline predictions of the theoretical model described in section 3 under the setting of two firms of different types and one regulator. For simplicity we assume that one of the firms is of high-benefit (low-cost) type and another is of low-benefit (high-cost) type. In the empirical setting that we are studying, one can think of a high-benefit-type firm being more likely to become a priority violator while low-benefit-type firm

Table A15: Factor Analysis Results

	F1	F2	F3	F4	F5	F6	F7	F8	F9	Uniqueness
$GC1$						0.943				0.218
$GC2$			0.946							0.177
$GC3$							0.919			0.224
$GC4$			0.325							0.543
$GC5$		-1.013								0.118
$GC6$					0.960					0.192
$GC7$	0.320		-0.334							0.097
$GC8$	-1.018									0.101
$GC9$					0.399					0.577
$GC10$								0.478	-0.310	0.590
$M1$				0.959						0.177
$M2$				0.343			0.326	-0.357		0.518
$M3$								0.759		0.431
$M4$									0.972	0.121

Notes: This table reports the rotated factor loadings of each violation in relation to the principal components or factors. These loadings represent how each variable is weighted for each factor, while uniqueness identifies the amount of variance that is not explained by the factors. Blanks represent absolute loadings less than 0.3.
Sample: We consider HPV classification violations and their damages before the policy revision in 2014.

being a non-HPV. Benefit functions from a level of violation, a_i, for high-benefit and low-benefit types are $b(a_{high}) = 4a_{high} - (1 - a_{high})^2$ and $b(a_{low}) = a_{low} - (1 - a_{low})^2$, respectively. For simplicity a rate of monitoring is equal to one, $\mu = 1$, since a regulator makes regularly monitoring all firms in the market a priority. A regulator cannot prosecute all violations above the threshold, a_τ, due to limited enforcement resources, hence prosecution efforts are proportional to the number of violators, i.e. $p(a_i) = \frac{1}{n} = \frac{1}{2}$. In addition, we assume that the penalty is equal to one ($f = 1$) for any level of violation above the optimal threshold. As a result, the expected enforcement is equal to $\frac{1}{2}$, $E[e(a_i)] = 1 * \frac{1}{2} * 1 = \frac{1}{2}$ for any violation above the threshold level if both firms are in noncompliance. Otherwise, $E[e(a_i)] = 1 * \frac{1}{1} * 1 = 1$ if only one firm is in violation.

We analyze two separate cases. First, the case where the threshold, a_τ, is zero and any violation level above this threshold is penalized by regulator. Second, the case where the threshold gets raised to 0.75, and only a firm committing a violation above 0.75 gets penalized.

Case $a_\tau = 0$.

Consider the case when both firms are violators. A high-benefit type firm chooses the level of violation, a_{high}, by solving its profit maximizing problem:

$$\max_{a_{high}} \ b(a_{high}) - E[e(a_{high})] \Rightarrow 4a_{high} - (1 - a_{high})^2 - \frac{1}{2}$$

The first order condition w.r.t. a_{high} is:

$$4 + 2(1 - a_{high}) = 0 \Rightarrow a_{high} = 3 \ \Rightarrow a_{high} = 1$$

Since $a_i \in [0, 1]$, a firm chooses $a_{high} = 1$ and obtains an expected profit of 3.5.[21] On the other hand, a firm can choose not to violate altogether, $a_{high} = 0$. In this case a firm does not receive a penalty, hence its profit is equal to -1.[22] Note that for a high-benefit-type firm choosing zero level of violation yields lower profit than choosing some level of violation when threshold is low, $a_\tau = 0$. This is true when a firm is one of the two violators, $E[e(a_{high})] = \frac{1}{2}$ as well as when a firm is the only violator, $E[e(a_{high})] = 1$. A high-benefit firm receives a profit of 3 when a $E[e(a_{high})] = 1$, which is greater than the profit from choosing a level of violation equal to zero.[23]

A low-benefit-type firm chooses the level of violation, a_{low}, by solving the following profit maximizing function:

$$\max_{a_{low}} \ b(a_{low}) - E[e(a_{low})] \Rightarrow a_{low} - (1 - a_{low})^2 - \frac{1}{2}$$

The FOC w.r.t. a_{low} is:

$$1 + 2(1 - a_{low}) = 0 \Rightarrow a_{low} = \frac{3}{2} \ \Rightarrow a_{low} = 1$$

However, again $a_i \in [0, 1]$, a firm chooses $a_{low} = 1$ and obtains a profit equal to $\frac{1}{2}$ when there are two violators in the market.[24] On the other hand, the low type firm can choose not to violate and comply with the regulations. In this case, it obtains the profit of -1.[25] Note that a low-benefit firm still chooses a level of violation of 1 even when it is the only violating firm in the market, $E[e(a_{low})] = 1$, since the profit from $a_{low} = 1$ is 0 and greater than -1.[26] We conclude that even for the low-benefit-type firm, committing a violation yields higher profits than staying in compliance when the threshold is too low.

To sum up, under the case when the violation threshold level is too low and enforcement resources are limited, both firms commit the highest level of violation possible. It results in 2 units of pollution in the market with all enforcement resources being exhausted.

Case $a_\tau = 0.75$.

[21] $\pi(a_{high} = 1) = 4(1) - (1 - 1)^2 - \frac{1}{2} = \frac{7}{2}$
[22] $\pi(a_{high} = 0) = 4(0) - (1 - 0)^2 = -1$
[23] $\pi(a_{high} = 1) = 4(1) - (1 - 1)^2 - 1 = 3$
[24] $\pi(a_{low} = 1) = 1 - (1 - 1)^2 - \frac{1}{2} = \frac{1}{2}$
[25] $\pi(a_{low} = 0) = 0 - (1 - 0)^2 = -1$
[26] $\pi(a_{low} = 1) = 1 - (1 - 1)^2 - 1 = 0$

In this case, a plant is fined only if its level of violation is above 0.75. Each type of a firm solves their profit maximizing problem as defined above and compares the profits from committing a violation at the level greater than 0.75 to the profit from committing a level of violation below or equal to the threshold.

A high-benefit-type firm chooses between the profit when $a_{high} = 1$ and the profit at the highest non-prosecuted level of a violation, $a_{high} = 0.75$. Profit from $a_{high} = 1$ is 3.5 considering both firms are violators, while the profit from $a_{high} = 0.75$ is 2.9375.[27] [28] When a high-benefit firm is the only violator and it is aware that it is the only one to be fined, the profit from $a_{high} = 1$ becomes 3.[29]

Despite the fact that violations below 0.75 are no longer penalized, the high-benefit-type firm obtains more profit (even in the worst case scenario of receiving a fine of 1) from committing the highest level of violation, $a_{high} = 1$.

The low-benefit-type firm faces the same choice: $a_{low} = 1$ or $a_{low} = 0.75$. The profit from level of violation equal to one, $a_{low} = 1$, for a low type firm is $\frac{1}{2}$ when there are two firms violating.[30] Profit from $a_{low} = 0.75$ is 0.6875 and is clearly greater than the profit from committing a violation at level of 1.[31] In the case when the low-benefit-type firm is the only violator, its profit with $E[e(a_{low})] = 1$ becomes 0.[32] Hence, for the low-type firm choosing the level of violation below or equal the new threshold is always more profitable.

As a result, the total level of pollution is 1 with only a high-type firm choosing a level of violation above the new threshold. It is due to the benefit that a firm receives from $a_{high} = 1$ despite the fact that it has to pay a fine. On the other hand, under the new threshold it becomes unprofitable for the low-type firm to commit a violation above the defined threshold.

Based on this simple example, we conclude that increasing the level of priority violations threshold leads to a lower level of violations, and therefore, pollution.

[27] $\pi(a_{high} = 1) = 4(1) - (1-1)^2 - \frac{1}{2} = \frac{7}{2}$
[28] $\pi(a_{high} = 0.75) = 4(0.75) - (1-0.75)^2 = 2.9375$
[29] $\pi(a_{high} = 1) = 4(1) - (1-1)^2 - 1 = 1$
[30] $\pi(a_{low} = 1) = 1 - (1-1)^2 - \frac{1}{2} = \frac{1}{2}$
[31] $\pi(a_{low} = 0.75) = 0.75 - (1-0.75)^2 = 0.6875$
[32] $\pi(a_{low} = 1) = 1 - (1-1)^2 - 1 = 0$