American Economic Journal: Economic Policy 2025, 17(1): 1–36
https://doi.org/10.1257/pol.20220745

Clean Identification? The Effects of the Clean Air Act on Air Pollution, Exposure Disparities, and House Prices†

By Lutz Sager and Gregor Singer*

We assess the US Clean Air Act standards for fine particulate matter ($PM_{2.5}$). Using high-resolution data, we find that the 2005 regulation reduced $PM_{2.5}$ levels by 0.4 $\mu g/m^3$ over five years, with larger effects in more polluted areas. Standard difference-in-differences overstates these effects by a factor of three because time trends differ by baseline pollution, a bias we overcome with three alternative approaches. We show that the regulation contributed to narrowing Urban-Rural and Black-White $PM_{2.5}$ exposure disparities, but less than difference-in-differences suggest. Pollution damages capitalized into house prices, however, appear larger than previously thought when leveraging regulatory variation. (*JEL* D63, K32, Q52, Q53, Q58, R31)

The 1970 Clean Air Act (CAA) and subsequent amendments are the cornerstone of air quality regulation in the United States. The CAA operates through National Ambient Air Quality Standards (NAAQS) set by the US Environmental Protection Agency (EPA), with measures typically targeted at regions found to be in nonattainment of a given NAAQS.[1] The latest air pollutant to be regulated through NAAQS is $PM_{2.5}$, fine particulate matter of diameter smaller than 2.5 micrometers, with regulation coming into effect in 2005. $PM_{2.5}$ is one of the air pollutants most clearly associated with a wide range of adverse health outcomes (Landrigan et al. 2018), productivity losses (Graff Zivin and Neidell 2012) and other nonhealth outcomes (Aguilar-Gomez et al. 2022), and the key driver of the EPA's Air Quality Index. Given the large costs associated with pollution exposure, a central question is how effective policies are at lowering pollution levels.

* Sager: ESSEC Business School (email: sager@essec.edu). Singer: London School of Economics (email: g.a.singer@lse.ac.uk). Lucas Davis was coeditor for this article. We thank Raphael Calel, Roger Fouquet, Josh Graff Zivin, Matthew Neidell, Arik Levinson, and Reed Walker, as well as seminar participants at the LSE Grantham Policy Design and Evaluation workshop, EAERE 2022 Annual Conference, AERE 2022 Annual Conference, SWUFE, CUHK-Shenzhen, University of Hamburg, Queen Mary University London, University of Oslo, and the 2023 LSE/Imperial/King's Workshop in Environmental Economics for helpful comments and discussion. Singer acknowledges support from the Grantham Research Institute on Climate Change and the Environment at the London School of Economics.

† Go to https://doi.org/10.1257/pol.20220745 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] NAAQS are generally implemented at the state level through State Implementation Plans (SIP). States identify nonattainment areas that fail to meet NAAQS for criteria air pollutants, based on methodology set by the EPA. Nonattainment status triggers heightened scrutiny both within state level SIPs and under federal regulation. Since 1970 the spectrum of regulatory instruments has broadened substantially to include national emissions standards for cars and light trucks, various technology mandates and performance standards, offset requirements, fuel standards, as well as market-based instruments.

We estimate the effect of the $PM_{2.5}$ NAAQS nonattainment designations in 2005 on $PM_{2.5}$ concentrations, and assess implications for racial and spatial pollution exposure disparities and house prices in the United States. We use high resolution data from three leading reanalysis projects (Meng et al. 2019a, Di et al. 2019, van Donkelaar et al. 2021a) that estimate $PM_{2.5}$ concentrations by combining ground monitors, satellite data, and chemical transport models for the entire contiguous United States. We combine those with US census data and the EPA-registered $PM_{2.5}$ values (RV) that the agency constructs based on ground level pollution monitor readings and uses to assign nonattainment status.[2] We contribute to a recent literature that uses these $PM_{2.5}$ rules as a setting to study pollution damages or environmental justice (Bishop, Ketcham, and Kuminoff 2023, Jha, Matthews, and Muller 2019, Sanders, Barreca, and Neidell 2020, Currie, Voorheis, and Walker 2023), as well as the broader literature on NAAQS nonattainment effects with three insights.

Our first insight is that standard difference-in-differences (DiD) estimation, despite being popular, significantly overstates nonattainment effects. This is because EPA-registered $PM_{2.5}$ values, and therefore also nonattainment designations, correlate with secular time trends in air quality. Areas that start out with higher levels of pollution also experienced larger pollution reductions over time even in the absence of nonattainment status. Formal placebo tests using only attainment areas confirm that DiD estimations pick up an effect, casting doubt on the parallel trends assumption required for DiD. This pattern holds when we exclude attainment counties that border nonattainment areas, or when we exclude areas that were previously treated as nonattainment with the earlier PM_{10} standard.[3] We find such correlated time trends in all three reanalysis-derived pollution data sources as well as in the EPA's monitor data, for both absolute and relative changes in $PM_{2.5}$ concentrations, and whether or not we control for flexible state-specific time trends.

We propose three alternative strategies that address the systematic relationship between baseline pollution and pollution changes over time. All three produce similar estimates which are substantially smaller than the standard DiD estimates. The first approach augments DiD by controlling for trends correlated with baseline $PM_{2.5}$ directly. We thus call it DiD with baseline (DiDwb). The second approach exploits the fact that we observe census block level pollution which we aggregate to census tracts. Nonattainment is usually assigned at the coarser level of counties and commuting zones. This enables us to employ a matched difference-in-differences (MDiD) strategy comparing tracts from nonattainment and attainment areas that have similar baseline pollution levels. The third approach relies on the discontinuous assignment rule for nonattainment areas, exploiting our collected EPA-registered $PM_{2.5}$ values in a regression discontinuity (RD) design. While placebo tests fail for standard DiD, the placebo tests pass when using these other strategies. Our preferred

[2] To facilitate replication and wider use in future studies we rely exclusively on publicly available data at the most granular level (census blocks) to estimate the effectiveness of the policy in reducing pollution exposure, and note that this is equivalent to using restricted-use microdata and assigning pollution to individuals at the block level.

[3] We show that the pre-trend disappears when assigning areas that have previously also been treated with the earlier PM_{10} standard into the control group as in Currie, Voorheis, and Walker (2023), which, however, requires an implicit assumption of no treatment effects for these units. We test this assumption and show, on the contrary, that areas that have previously been designated into PM_{10} nonattainment experienced larger marginal $PM_{2.5}$ reductions from additional $PM_{2.5}$ nonattainment designation.

specification, MDiD, shows a 0.4 $\mu g/m^3$ reduction in $PM_{2.5}$ between 2001–03 and 2006–08 due to nonattainment status, a third of the standard DiD estimate. This is equivalent to a 3 percent reduction from 2001–03 averages. Bootstrap simulations show that our alternatives are significantly different from DiD but statistically indistinguishable from each other.

Our second insight is that this implies a lower contribution of NAAQS nonattainment areas to narrowing structural pollution exposure disparities. We first confirm the Black-White pollution gap documented in Jbaily et al. (2022) and Currie, Voorheis, and Walker (2023), and that these gaps narrowed, in part due to NAAQS nonattainment areas (Currie, Voorheis, and Walker 2023). We find, however, that the NAAQS' contribution is less than half the size when we use our preferred specification (MDiD) compared to standard DiD. This implies that the Clean Air Act may have contributed less to environmental justice than previously thought, at least with respect to $PM_{2.5}$ pollution. We next document Urban-Rural disparities that are even larger than the racial gap in pollution exposure. Again, we show that the Urban-Rural gap has narrowed, but that the contribution of the 2005 NAAQS is significantly smaller than standard DiD may suggest.[4]

Our third insight is that pollution damages might be even larger than previously thought. We quantify the damages from $PM_{2.5}$ exposure as capitalized in census tract level house prices from the Federal Housing Finance Agency (FHFA), using nonattainment designations as instrumental variable. We find that $PM_{2.5}$ reductions following nonattainment designation were associated with a 6 percent house price increase on average. The implied elasticity with respect to $PM_{2.5}$ of around -1.4 is around twice that found for PM_{10} (Bento, Freedman, and Lang 2015) and up to four times the elasticity for Total Suspended Particles (TSP or PM_{100}) (Chay and Greenstone 2005). Importantly, the simple DiD-IV suggests pollution damages that are substantially smaller than those of our other three alternative approaches, more in line with previous estimates for PM_{10}, which may, however, contain bias. This implies that while simple DiD *overestimates* the effect of nonattainment on $PM_{2.5}$, it *underestimates* the effect of $PM_{2.5}$ on house prices when nonattainment status is used as an instrument for $PM_{2.5}$. The magnitude of our adjustment is important: the house price elasticity changes by a larger increment (from -0.8 to -1.4) after adjusting for these time trends than it changes after accounting for potential endogeneity with instruments relative to a simple OLS regression (from -0.5 to -0.8).

Overall, our results show the importance of accounting for parallel trends violations that overstate air quality improvements from nonattainment designations in standard DiD frameworks. We find similar differences for all three pollution data sources and when looking at the longer term effects until 2011–13. We find evidence of effect heterogeneity, with larger improvements in the most polluted parts of nonattainment areas in line with findings for previous NAAQS (Auffhammer, Bento, and Lowe 2009; Bento, Freedman, and Lang 2015; Gibson 2019). We also show that areas that have previously been treated with PM_{10} nonattainment designation experienced larger marginal effects from $PM_{2.5}$ nonattainment. Finally, we show

[4] We also show that patterns are similar in versions where we allow for heterogeneous NAAQS nonattainment effects by baseline share of Black or urban population across census tracts.

some evidence that the bias we identify is likely to extend to other NAAQS settings, and discuss exceptions in the previous literature that address possible confounding trends (e.g., Greenstone 2004; Chay and Greenstone 2005).

We contribute to the literature on environmental policy analysis generally and the Clean Air Act in particular. Existing literature on nonattainment designations under previous NAAQS include estimated reductions in Ozone (Henderson 1996), sulfur dioxide (SO_2) (Greenstone 2004), TSP (Chay and Greenstone 2005), and PM_{10} concentrations (Auffhammer, Bento, and Lowe 2009).[5] We show that the NAAQS for $PM_{2.5}$ implemented in 2005 were effective, albeit less so than DiD estimation may suggest, an insight that likely extends to other NAAQS. We illustrate the role of the regulation in narrowing pollution exposure disparities, a finding that is relevant for the literature on structural pollution gaps and environmental justice (Currie, Voorheis, and Walker (2023); Jbaily et al. (2022); Banzhaf, Ma, and Timmins 2019, Colmer et al. 2020, Drupp et al. 2021). Finally, we contribute to a growing literature that relies on nonattainment designations as an instrument to quantify pollution damages (Chay and Greenstone 2005; Grainger 2012; Bento, Freedman, and Lang 2015). While we explore the effects on house prices, it appears likely that our adjustments to the first stage to account for correlated time trends are also relevant for other second stage outcomes, such as health (Isen, Rossin-Slater, and Walker 2017, Sanders and Stoecker 2015, Sanders, Barreca, and Neidell 2020, Colmer and Voorheis 2021, Bishop, Ketcham, and Kuminoff 2023).

The rest of the paper begins with a description of the regulatory context and the data we use in Section I. We set up the empirical strategy in Section II along with descriptive statistics that highlight the nuances in identification requirements and their plausibility. Section III shows results from estimating the effects of the CAA 2005 NAAQS rules on $PM_{2.5}$ concentrations. Section IV turns to our two applications, analysing the contribution of nonattainment designations in narrowing structural pollution exposure disparities, and using nonattainment as as instrument for $PM_{2.5}$ to estimate the pollution impact on house prices. Section V discusses the relevancy of our insights for other NAAQS and Section VI concludes.

I. Data and Regulatory Context

A. *The 2005 National Ambient Air Quality Standards for $PM_{2.5}$*

Under the CAA, the EPA primarily regulates air quality through successive NAAQS aimed at specific pollutants. In April 2005, the 1997 NAAQS for $PM_{2.5}$, particulate matter smaller than 2.5 μm in diameter, came into effect.[6] The EPA introduced

[5] Nonattainment designation has also been linked to changes in industrial activity (Henderson 1996; Greenstone 2002), within-product improvements in emission intensity (Shapiro and Walker 2018), and employment (Kahn and Mansur 2013; Walker 2013). Deschênes, Greenstone, and Shapiro (2017) study a non-NAAQS but related CAA policy focusing on nitric oxide (NO_x). Economists have been assessing the benefits and costs of the CAA from its inception, initially using prospective regulatory analyses, but increasingly using retrospective analyses with quasi-experimental methods, as documented in recent surveys by Aldy et al. (2022) and Currie and Walker (2019).

[6] Several litigation procedures from 1999 delayed the implementation of the new regulations escalating up to the Supreme Court (EPA v. American Trucking Assoc., 531 US 457 2001), see US EPA (2005a, 2016). Previous NAAQS regulated coarser particulate matter PM_{10} and TSP, equivalent to PM_{100}.

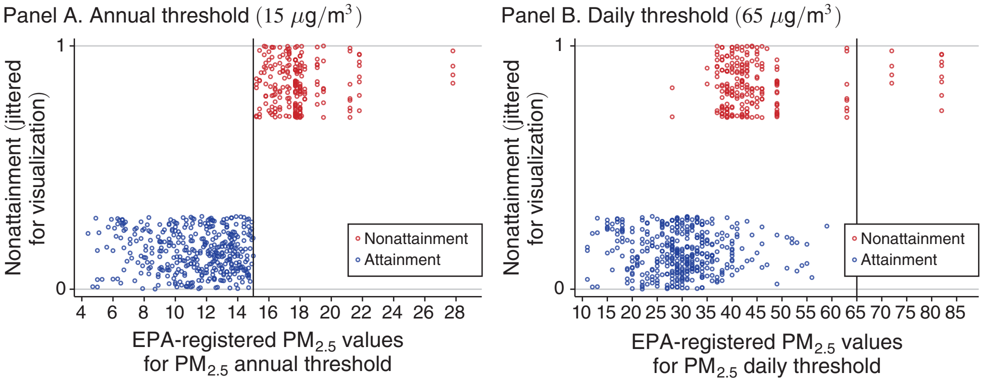


Figure 1. Nonattainment Status and EPA-Registered $PM_{2.5}$ Values

Notes: Both panels plot the EPA-registered $PM_{2.5}$ values of counties and nonattainment status of the NAAQS rules coming into effect in 2005. Panel A shows the EPA-registered $PM_{2.5}$ values for the three-year average of the annual threshold of 15 $\mu g/m^3$ from 2001 to 2003. Panel B shows the EPA-registered $PM_{2.5}$ values for the three-year average of the ninety-eighth percentile daily threshold of 65 $\mu g/m^3$ from 2001 to 2003. The county markers are jittered for visualization. The plots show that the annual threshold in panel A is binding, in the sense that there is no county that meets this requirement, but does not meet the daily requirement in panel B. On the other hand, many counties meet the daily threshold in panel B but are still assigned into nonattainment because they don't meet the annual threshold in panel A. County level RV reflect the RV of the nonattainment area (i.e., are assigned the highest RV within a nonattainment area).

regulation for $PM_{2.5}$ through two new standards: A threshold of 15 $\mu g/m^3$ for the three-year average of annual mean ambient $PM_{2.5}$ concentrations, and a threshold of 65 $\mu g/m^3$ for the three-year average of the 98th percentile of daily (24h) $PM_{2.5}$ concentrations. Areas that failed to meet at least one of these thresholds were designated as nonattainment areas. As Figure 1 shows, whenever an area satisfied the annual requirement, it also satisfied the daily requirement, so we focus on the binding annual requirement for the rest of the analysis.[7] The EPA has several powers to induce air quality improvements in nonattainment areas, for example by reviewing or enforcing air quality improvement plans[8], or by withholding federal funding and denying permits for infrastructure projects or polluting plants. Reclassification from nonattainment to attainment status is usually initiated by requests from states (Sullivan and Krupnick 2018). There was no reclassification to attainment until 2011, and no reclassification into nonattainment based on the 1997 standards (only with new standards, see Footnote 7).

The $PM_{2.5}$ measurements for assigning nonattainment status are based on an incomplete network of ground monitors that the EPA deployed from 1999 to January 2001 (US EPA 2005a). While these monitors were usually placed in more populous counties, they only covered around 20 percent of counties, possibly missing coun-

[7] A 2006 revision of the daily requirement from 65 $\mu g/m^3$ to 35 $\mu g/m^3$ came into effect in December 2009, and designated a few additional counties as nonattainment. Our main analysis focuses on changes until December 2008 before these additional nonattainment designations. A 2012 revision of the annual requirement from 15 $\mu g/m^3$ to 12 $\mu g/m^3$ came into effect in April 2015.

[8] State implementation plans typically include measures such as permits, technological standards such as emission capture, fuel efficiency improvements or retrofits, and surveillance and enforcement rules.

ties that would otherwise be regulated (Sullivan and Krupnick 2018, Fowlie, Rubin, and Walker 2019).[9] The EPA took the three year averages of monitor readings from 2001 to 2003 to calculate the EPA-registered $PM_{2.5}$ values for each area to compare against the regulatory threshold. Since the 1997 NAAQS designations only took effect in April 2005, states were allowed to provide the EPA with updated 2002–2004 measurements, which led to a few counties being reclassified from nonattainment to attainment before 2005. We collect the latest RVs that incorporate these updates.[10]

Most nonattainment areas coincide with county groupings that make up Metropolitan Statistical Areas (MSA) or commuting zones (CZ), but are refined by the EPA on a case-by-case basis using nine decision factors to define air regions.[11] Therefore, the boundaries of nonattainment areas usually extend beyond single counties, motivated by the fact that air pollution can spill over into neighbouring counties. This means that if a county contains a monitor with a RV in excess of one of the $PM_{2.5}$ NAAQS thresholds, the entire air region (usually an MSA) is in nonattainment, including other counties in the area that may have low pollution readings or no monitor at all. In this case, the entire group of counties within a nonattainment area is assigned the RV of the county with the highest RV.[12] In total, the EPA assigned 208 counties into nonattainment in 2005, all violating the annual threshold (and a subset also violating the daily threshold). Figure 2, panel A maps the 208 nonattainment counties based on the EPA air regions.

We use data from the EPA Green Book and Federal Register to identify the nonattainment counties (US EPA 2005a, b, 2021). We obtain the RVs for the annual and daily standards for all counties that are used by the EPA to determine attainment status (US EPA 2018b, a), and importantly, also including those counties that had a RV but were not assigned into nonattainment. Counties without monitors that are not part of any nonattainment area have no RV. Since every nonattainment area has the RV of the county with the highest RV within its area, we assign those RVs to each nonattainment area using data and detailed discussions from US EPA (2004b, 2005c, 2021), and update them with the supplementary amendments contained in US EPA (2005b, c). We next match areas to more recent, granular measures of particulate matter concentrations.

B. *Pollution Data*

We use annual estimates of ground level $PM_{2.5}$ concentrations from three sources. All three are based on satellite data combined with chemical transport models and calibrated to fit ground level monitor readings. Our main results use pollution data

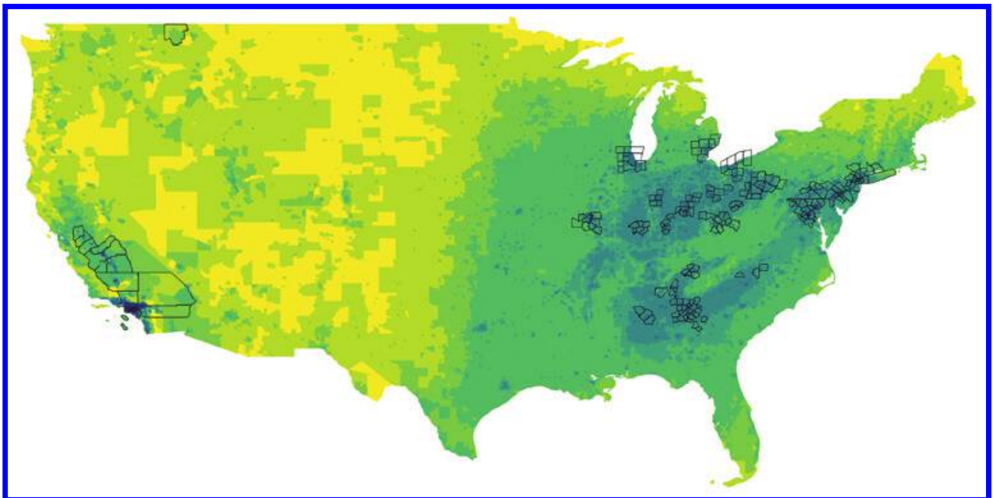
[9] Therefore EPA technical documents often refer to attainment areas as unclassifiable/attainment. Our simple difference-in-differences estimates using satellite-based pollution measures are virtually identical for the whole sample of all counties or the smaller sample of counties with RVs (and monitors).

[10] In the technical EPA documentation, these EPA-registered $PM_{2.5}$ values are referred to as 'Design Values.'

[11] The nine factors that define the appropriate boundaries of areas are emissions, air quality, population density, commuting patterns, expected growth, meteorology, geography, jurisdictional boundaries, and control of emission sources. See US EPA (2004a) for a detailed explanation.

[12] The EPA only groups counties together with the highest RV in nonattainment areas, not attainment areas. In some counties, there are multiple monitors allowing for spatial averaging, and some exceptionally large counties might only be partially included in an area.

Panel A. Contiguous United States



Panel B. Indianapolis – Evansville – Louisville – Cincinnati nonattainment counties

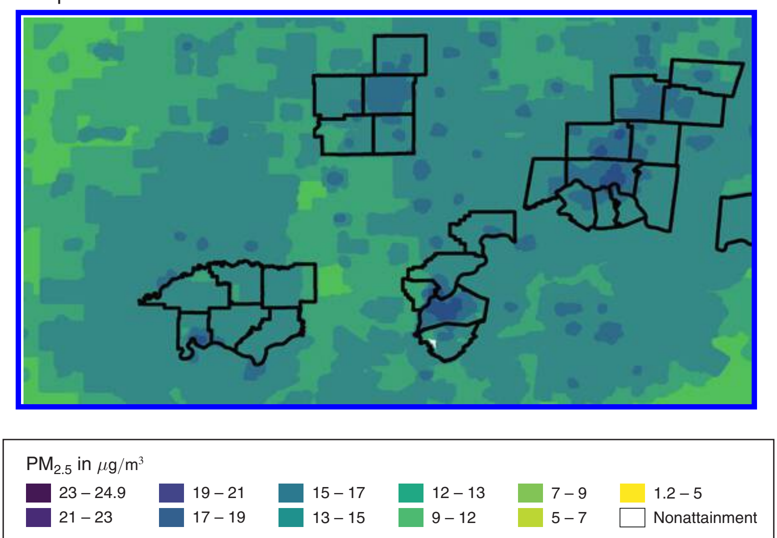


FIGURE 2. BASELINE $PM_{2.5}$ (2001–2003) AND NONATTAINMENT COUNTIES

Notes: The figures show average baseline $PM_{2.5}$ (2001–2003) at the tract level using data from Meng et al. (2019b), and nonattainment counties. Panel A shows the entire contiguous United States, and panel B zooms into the area around Indianapolis (North), Evansville (Southwest), Louisville (South) and Cincinnati (Northeast).

from Meng et al. (2019b), but we show in the Appendix that all our results are similar when using two alternative datasets from Di et al. (2021) or van Donkelaar et al. (2021b).[13]

All three datasets provide $PM_{2.5}$ concentrations at a spatial resolution of $0.01^\circ \times 0.01^\circ$(approximately 1 km × 1 km cells in the United States, depending on

[13] The data by Meng et al. (2019b) extend the furthest back in time. For the van Donkelaar et al. (2021b) data, we use the latest recommended version (V5.GL.03).

latitude). The universal spatial coverage and high resolution of these products allow us to assign pollution levels to all census blocks in the contiguous United States on an annual basis starting from 2000 based on each of the three datasets, and from 1989 for the data based on Meng et al. (2019b).[14] This allows us to calculate $PM_{2.5}$ concentrations also for those counties that do not have RVs Since these data use predicted values there is uncertainty in some of the estimates, especially for those areas that are further away from ground-based monitors. We test robustness to uncertainty in Section IIIC by leveraging information on uncertainty in the underlying predictions and by replicating our analysis using only pollution monitor data from US EPA (2022a).

C. *Census Data and Mapping $PM_{2.5}$ Concentrations*

We use population counts from the 2000, 2010, and 2020 US census and area boundaries from the 2010 US census (Manson et al. 2022). The boundaries allow us to map geocoded $PM_{2.5}$ data into the around 11 million census blocks (sub-divisions of tracts). Since blocks are very small and often do not contain $PM_{2.5}$ grid points at the 1 km resolution, each block is assigned the $PM_{2.5}$ concentration of the grid point closest to the block centroid.[15] We use block level population data as weights to aggregate pollution up to census tracts (of which there are around 70,000). We also use the population data to weight tract level regressions by population and to calculate $PM_{2.5}$ exposure differences between population groups when we turn to the analysis of exposure disparities.

The resulting data provides detailed measures of average $PM_{2.5}$ exposure in each tract and each year. The map in Figure 2, panel A shows tract level $PM_{2.5}$ concentrations averaged between 2001 and 2003 across the contiguous United States. We use this detailed data structure to exploit variation not comprehensively captured by monitoring data, specifically to investigate heterogeneity within counties, visible in Figure 2, panel B, and to measure changes in air quality even in those tracts that are not close to a ground level pollution monitor.

D. *House Price Data*

We demonstrate the implications of our estimates for the implied damages from $PM_{2.5}$ exposure capitalized in house prices. Our measure of changes in house prices relies on data provided by the Federal Housing Finance Agency (US FHFA 2021), specifically, the annual house price index (HPI) at the tract level (further described in Bogin, Doerner, and Larson 2019).

[14] Meng et al. (2019b) add monitor readings of PM_{10} concentrations to help model $PM_{2.5}$ concentrations before 1999.

[15] Note that assigning pollution to census blocks (and their population counts) is equivalent to using restricted individual level data and assigning pollution to individuals based on their census block, as e.g., in Currie, Voorheis, and Walker (2023) (see also Supplemental Appendix A.13A). We complement the census data with information on commuting zones and tract-level characteristics from ChettyDATA (Chetty and Friedman 2019).

E. *Period Choice*

The process of nonattainment designation occurred in multiple stages—with initial state level suggestions for nonattainment designation in February 2004, and final EPA designations in April 2005. There may already have been anticipatory effects of nonattainment designation before 2005 (as shown in Bishop, Ketcham, and Kuminoff 2023, who consider 2004 as the start of the posttreatment period). To avoid any bias from anticipatory effects in 2004, we define our pretreatment period as the three-year average between 2001 and 2003.[16] Taking three-year averages helps to lower the risk of misattributing short-term fluctuations in air quality or measurement error to the CAA rules. To allow for time-varying effects, we report results for two posttreatment periods, respectively, five years (2006–08 average) and ten years (2011–13 average) after the pretreatment period. Supplemental Appendix Table A.1 provides summary statistics for the final analysis sample.[17]

II. Empirical Strategy: The Effect of CAA Nonattainment Designation

Our goal is to estimate the treatment effect of nonattainment designation in 2005 on $PM_{2.5}$ concentrations. To conceptualize our approach, consider the following expression for the level of $PM_{2.5}$ in census tract i during period t:

$$PM_{i,t} \;=\; \beta NA_{i,t} + \delta_i + \lambda_t + \xi_{i,t}, \tag{1}$$

where $NA_{i,t}$ denotes nonattainment status of tract i in period t and β is our treatment effect of interest. Tract level fixed effects δ_i capture factors that affect $PM_{2.5}$ and possibly nonattainment status, but do not vary meaningfully over the relevant time horizon—historical pollution, population density, road infrastructure, topology and the like. Period fixed effects λ_t capture aggregate trends that affect all tracts, such as changes in technologies or federal regulation and policies. Finally, the error term $\xi_{i,t}$ captures tract-period specific fluctuations. For now, we assume that the treatment effect is constant across tracts, an assumption we will relax later.[18]

A. *Likely Bias of Simple Difference-in-Differences* (*DiD*)

Our baseline empirical strategy is a standard DiD approach which compares changes in $PM_{2.5}$ from the pretreatment to the posttreatment period between treated

[16] Note that we provide some descriptive statistics using Meng et al. (2019b) data going back to 1989, but since data before 1999 is less accurate due to the lack of $PM_{2.5}$ ground monitors, we exclude these earlier periods for our empirical analysis.

[17] While our data extends forward to 2016 (in the case of Meng et al. 2019b, Di et al. 2021) and 2020 (in the case of van Donkelaar et al. 2021b), respectively, we avoid measuring effects too long after nonattainment designation in 2005 to avoid using areas that change treatment status. Some nonattainment areas came into attainment, particularly in 2013 and 2014. Furthermore, updates to the threshold rules came into effect in December 2009 and April 2015 placing additional areas into nonattainment.

[18] Following the literature on NAAQS nonattainment designations, we make the stable unit treatment value assumption (SUTVA) that rules out spillover effects from nonattainment into attainment counties (see Hollingsworth et al. (2022) or Walker (2013) for more discussion). We address this issue in a robustness check by excluding counties in attainment that share a border with a nonattainment county in Supplemental Appendix Table A.6.

and untreated units. We can express this taking first differences of equation (1).[19] Simplifying and rearranging terms yields our baseline regression equation using the change in $PM_{2.5}$ for tract i between the pretreatment and the posttreatment periods as outcome:

$$\Delta PM_i = \alpha + \beta \Delta NA_i + \Delta \xi_i, \tag{2}$$

where ΔNA_i is an indicator variable that takes value one for all tracts that become subject to regulatory treatment from 2005 onward.[20] The identifying assumption is that parallel trends between treated and untreated tracts hold: absent regulation, nonattainment and attainment areas would have experienced the same average change in $PM_{2.5}$. In other words, β yields a consistent estimate of the average treatment effect if $\text{cov}\left[\Delta NA_i, \Delta \xi_i\right] = 0$.

While the parallel trends assumption cannot be directly tested, it is common practice to look at pretreatment trends to assess whether the assumption is plausible. Panel A of Figure 3 plots average $PM_{2.5}$ concentrations over time for four groups binned according to their EPA-registered $PM_{2.5}$ values, including two nonattainment groups ($15 < RV \leq 20$ and $RV > 20$) and two groups in attainment ($10 < RV \leq 15$ and $RV < 10$). Nonattainment and attainment areas appear to follow somewhat different trends both before and after 2005.[21] Panel B shows these higher pollution improvements in nonattainment areas before 2005 more explicitly in an event study version of equation (1), plotting the annual difference in $PM_{2.5}$ levels between nonattainment and attainment areas relative to the difference in 2005. This suggests that the parallel trends assumption is violated, as it shows significant differences in pre-trends that are similar to those after treatment. Supplemental Appendix Figures A.21 and A.22 confirm statistically significant pre-trend differences in the two alternative pollution data sources based on Di et al. (2021) and van Donkelaar et al. (2021b).

We conduct several robustness tests for these statistically significant pre-trends. Supplemental Appendix Figure A.1 shows the pre-trends exist when (i) using 2000 census borders and population instead of the 2010 versions and repeating the analysis at the census block level directly without aggregating to tracts, (ii) using interpolated population weights from the 2000, 2010, and 2020 census, allowing for population changes at the census block level, (iii) assigning entire commuting zones into nonattainment beyond the EPA-defined air regions for commuting zones that contain at least one nonattainment area, and when (iv) dropping all attainment counties that border a nonattainment area to address potential spillover effects. We

[19] That is $\Delta PM_i = PM_{i,post} - PM_{i,pre} = \beta\left(NA_{i,post} - NA_{i,pre}\right) + \left(\delta_i - \delta_i\right) + \left(\lambda_{post} - \lambda_{pre}\right) + \left(\xi_{i,post} - \xi_{i,pre}\right)$. All our specifications in changes could equivalently be modeled as panel regressions of levels with two-way fixed effects (TWFE) and various interaction terms. We show that such an approach in a tract-year panel from 2000–2015 produces very similar estimates in Supplemental Appendix Table A.10, which shows a panel equivalent to Table 1.

[20] We weight all regressions by tract population. Because attainment is assigned to counties spanning multiple tracts, we cluster standard errors at the county level, allowing for arbitrary correlation in the errors within counties.

[21] Note that areas with the highest EPA-registered $PM_{2.5}$ values ($RV > 20$) experience an increase in $PM_{2.5}$ concentrations before policy implementation in 2005. One possibility are anticipatory effects in nonattainment areas (Clay et al. 2021). Regardless of the underlying reason, our alternative strategies limit the risk of such confounding trends.

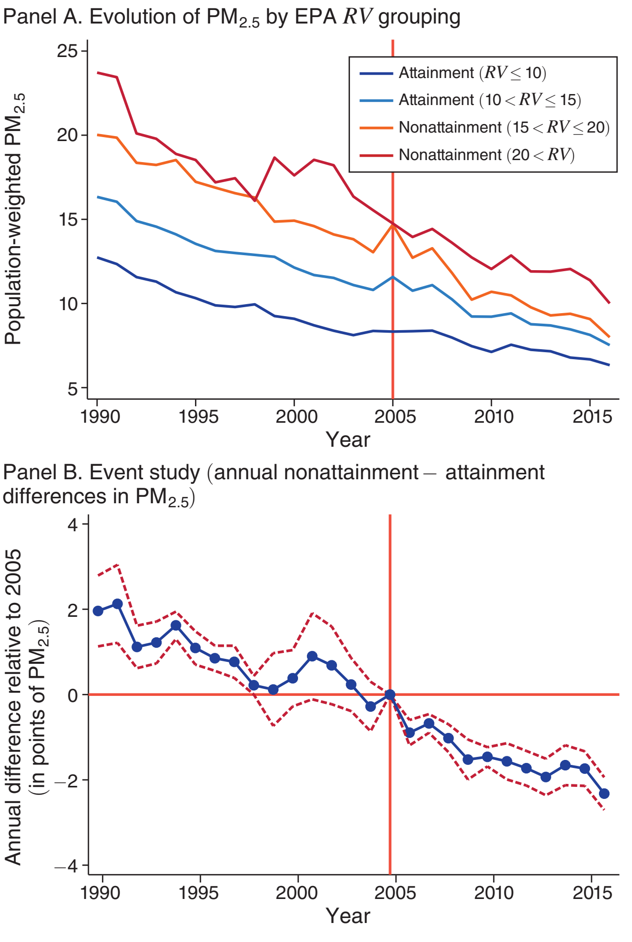


Figure 3. Trends in $PM_{2.5}$ and Event Study Analysis

Notes: Panel A shows the change in $PM_{2.5}$ averages at the tract level (population-weighted) over time. Each line represents a different bin of EPA-registered $PM_{2.5}$ values assigned to each attainment/nonattainment area, each of which usually comprises multiple counties and tracts. Panel B shows coefficient estimates from a regression that includes a treatment dummy interacted with years, controlling for year fixed effects. The dotted blue line shows point estimates and the dashed red lines show 95 percent confidence intervals based on standard errors that are cluster-robust at the level of counties. Both panels are based on data from Meng et al. (2019b).

compare our event study to the event study in Currie, Voorheis, and Walker (2023) along with other differences in detail in Supplemental Appendix A.13.[22] Their event study does not exhibit pre-trends as they assign the subset of nonattainment areas that have previously been treated with 1990 PM_{10} nonattainment into the control group. Intuitively, this subset tends to have higher pollution levels and pre-trends therefore evening out pre-trend differences between treated and controls, but assigning these areas into the control group implicitly assumes no $PM_{2.5}$ treatment effect

[22] We thank the authors of Currie, Voorheis, and Walker (2023), particularly Reed Walker, for a helpful discussion of this comparison.

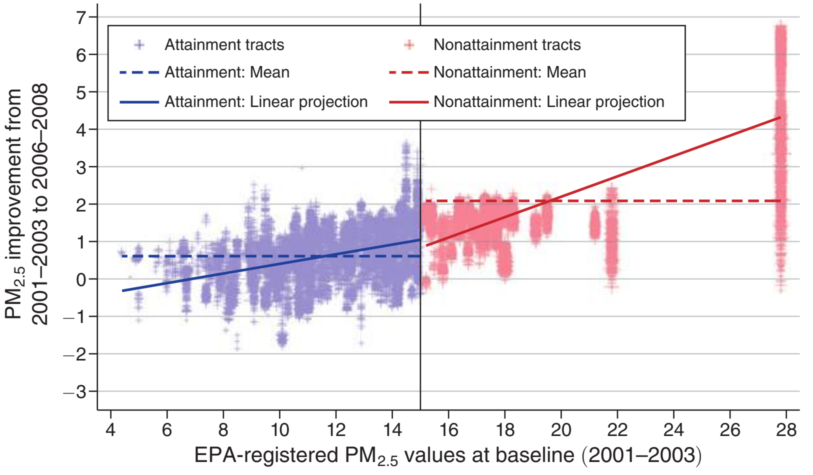


Figure 4. Improvement in Tract $PM_{2.5}$ Averages and EPA-Registered $PM_{2.5}$ Values

Notes: The figure shows the improvement in $PM_{2.5}$ averages at the tract level between two periods: 2001–2003 and 2006–2008. The size of the markers reflect tract level populations. The $PM_{2.5}$ improvements are plotted against the EPA-registered $PM_{2.5}$ values of each attainment/nonattainment area, each of which usually comprises multiple counties and tracts. The dashed line plots the average $PM_{2.5}$ improvement for tracts in nonattainment and attainment areas separately, weighted by tract population, equivalent to the standard DiD estimate. The solid lines plot the linear projection of tract level $PM_{2.5}$ improvements on the RV of the nonattainment and attainment areas separately, weighted by tract population. Based on data from Meng et al. (2019b).

for these areas.[23] We formally test treatment effect heterogeneity by previous 1990 PM_{10} nonattainment status in Section IIF and IIIE, and show evidence that these areas actually tend to exhibit larger $PM_{2.5}$ treatment effects.

The issue becomes even clearer when looking at Figure 4, which plots EPA-registered $PM_{2.5}$ values on the horizontal and tract level negative ΔPM from 2001–03 to 2006–08, i.e., pollution improvements, on the vertical axis. Nonattainment areas are those with a RV higher than the threshold value 15.[24] Crucially, we see a positive association between RV and $-\Delta PM$ on both sides of that cutoff, indicated by the solid linear regression lines. This suggests that nonattainment areas would likely have experienced a larger reduction in $PM_{2.5}$ concentrations also in the absence of nonattainment designation, much like attainment tracts with higher RV have experienced larger reductions than other attainment tracts with lower RV. Since nonattainment designation is a function of RV ($\text{cov}[\Delta NA_i, RV_i] > 0$) and Figure 4 suggests that RV and $\Delta\xi_i$ are correlated ($\text{cov}[-\Delta\xi_i, RV_i] > 0$), it appears highly likely that the identifying assumption for DiD is not satisfied ($\text{cov}[\Delta NA_i, \Delta\xi_i] \neq 0$).[25] Both ΔPM and nonattainment status are correlated with pretreatment pollution levels, confounding the standard DiD estimate.

[23] When dropping this subset of areas instead, the significant pre-trend reappears as shown in Supplemental Appendix Figure A.19—see also Supplemental Appendix Table A.7.

[24] The census tracts at the right end of the figure belong to Los Angeles area, the nonattainment area with the highest RV.

[25] The simple DiD approach in (2) measures the average difference between tracts left and right of the $RV = 15$ cutoff. That is the difference between the horizontal dashed lines.

Supplemental Appendix Figures A.23 and A.24 show almost identical patterns using the two alternative pollution data sources from Di et al. (2021) and van Donkelaar et al. (2021b). In Supplemental Appendix Figure A.4, we use EPA monitor level data instead and show that the pattern is similar at the monitor level.[26] Supplemental Appendix Figure A.2 shows a similar pattern when taking a 10 year difference from 2001–03 to 2011–13.

One reason for the different trends might be that areas that are in nonattainment of $PM_{2.5}$ standards in 2005 are also more likely to have been in nonattainment of previous standards, such as those for PM_{10}, which would explain why they are cleaning up their air already before 2005. We show in the Supplemental Appendix that different trends persist, however, even when dropping all counties that were previously in nonattainment of the PM_{10} standard (Supplemental Appendix Figure A.19 and Supplemental Appendix Table A.7). Another reason for the different absolute trends could be similar relative improvements, for example due to technological change, that translate into bigger absolute improvements in more polluted areas.[27] Supplemental Appendix Figure A.3 shows this is not enough to explain differences in absolute terms, as more polluted areas also experienced greater relative improvements in air quality over time, independent of attainment status.[28] Other reasons for the different trends could be linked to state level policies (and therefore different trends by state), or reasons related to population density (with urban, suburban, and rural areas experiencing different trends). In Supplemental Appendix Tables A.8 and A.9, we provide robustness checks including state level trends or trends by population density, which help explain some of the difference in trends, but not all.

While we can remain agnostic about the particular combination of reasons for these correlated time trends, we need to address the bias they introduce in a standard DiD setting, for which we employ three different strategies. The first is to include baseline pollution controls (DiDwb). This maintains the sample, but introduces a control variable. The other two approaches restrict the sample to observations for which parallel trends are more likely to hold. The second approach is a matching DiD hybrid (MDiD) and the third is a regression discontinuity design (RD).

B. *Difference-in-Differences Estimation With Baseline Controls* (*DiDwb*)

Our first strategy is to explicitly control for the confounding factor suggested by the relationship in Figure 4 using an augmented version of equation (2). In particular, we assume that the error can be decomposed into

$$\Delta\xi_i = \gamma PM_{i,pre} + \Delta\epsilon_i$$

[26] We calculate three year averages for each monitor averaging over various series that are, e.g., certified or not certified.

[27] Consider the example of road traffic. If all regions maintain the same volume of traffic, but newer cars generate 10 percent less emissions per mile driven, we would expect 10 percent less traffic related pollution in all areas, which would be a larger absolute improvement in high-traffic areas.

[28] Colmer et al. (2020) analyse a much longer time period from 1981 to 2016 and find that absolute improvements in $PM_{2.5}$ pollution are much larger for the most polluted census tracts, while they find less difference in relative improvements. Our reported patterns are consistent with their observed reversion to the mean.

so that we can linearly control for baseline pollution $(PM_{i,pre})$, assuming a residual error $\Delta \epsilon_i$:

$$\Delta PM_i = \alpha + \beta NA_i + \gamma PM_{i,pre} + \Delta\epsilon_i. \tag{3}$$

We refer to this DiD approach with baseline controls as DiDwb. Note that including $PM_{i,pre}$ as control in our specification in differences is equivalent to controlling for $PM_{i,pre}$ separately by period (λ_t) in the levels specification in equation (1), where $PM_{i,pre}$ is absorbed by tract fixed effects.[29] This approach absorbs any improvements in air quality over time that are proportional to baseline $PM_{2.5}$ levels (e.g., $\gamma = -0.1$ would indicate a 10 percent reduction for all tracts).

The identifying assumption becomes an augmented version of the parallel trends assumption. Nonattainment and attainment areas would have experienced the same average change in $PM_{2.5}$ over time absent regulation, conditional on a linear association between between baseline $PM_{2.5}$ and ΔPM. Put differently, we require that $\mathrm{cov}\left[\Delta NA_i, \Delta\epsilon_i | PM_{i,pre}\right] = 0$. Supplemental Appendix Figure A.7 shows insignificant pre-trend differences with this augmentation. Notably, this assumes that residual pollution shocks persist across periods. That is, we assume that ϵ_{it} follows an AR1 process such as $\epsilon_{it} = \epsilon_{it-1} + \mu_{it}$ where μ_{it} is uncorrelated with PM_{it-1}.[30] This is satisfied if a shock in the pretreatment period—from, say, new industrial units or infrastructure projects—persists through the posttreatment period (when a new shock can arrive). On the other hand, if a shock in the pretreatment period is only transitory—from, say, unusual weather conditions in a given year—$\Delta\epsilon_i$ would be correlated with $PM_{i,pre}$, introducing bias into equation (3). To mitigate such bias from transitory shocks, we use three-year averages of $PM_{2.5}$ in both the pre- and posttreatment periods such that transitory shocks like weather are unlikely to be captured. We also demonstrate in Supplemental Appendix Table A.3 that results remain unchanged when using higher-order interactions with baseline $PM_{2.5}$ allowing for more flexible nonlinearities.

C. *Matched Difference-in-Differences* (*MDiD*)

Our second approach exploits the fact that our analysis is at the tract level while nonattainment is assigned at the level of the county and/or commuting zone. This means that, even though the RV distributions of nonattainment and attainment areas are disjoint (separated at $RV = 15$), there is overlap for tract level $PM_{2.5}$, allowing us to match nonattainment tracts to attainment tracts with similar baseline $PM_{2.5}$. The map in Figure 2, panel B illustrates this for the region around Indianapolis, showing that there are tracts with low and high baseline $PM_{2.5}$ in both nonattainment and attainment areas. Figure 5 shows the overlap in the distributions of baseline

[29] That is $PM_{2.5_{i,t}} = \beta NA_{it} + \gamma\lambda_t PM_{i,pre} + \delta_i + \lambda_t + \epsilon_{i,t}$ in levels is $\Delta PM_i = \alpha + \beta NA_i + \gamma PM_{i,pre} + \Delta\epsilon_i$ in differences.

[30] In our case the process is a random walk, but using earlier periods than the pretreatment period could correspond to different AR processes.

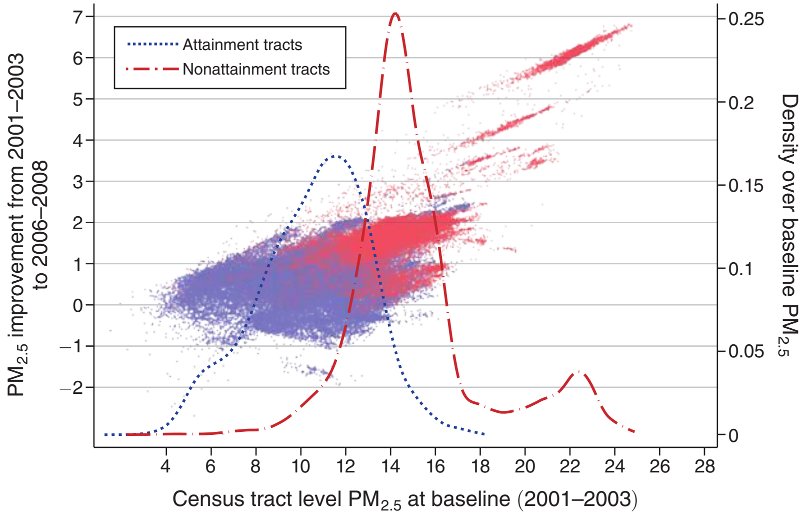


FIGURE 5. IMPROVEMENT IN TRACT $PM_{2.5}$ AVERAGES AND BASELINE $PM_{2.5}$ LEVELS

Notes: The markers in the figure show the improvement in $PM_{2.5}$ averages at the tract level between two periods: 2001–2003 and 2006–2008. The $PM_{2.5}$ improvements are plotted against the baseline $PM_{2.5}$ levels of each tract, using two different colors for tracts in nonattainment and attainment areas. The kernel density (right axis) shows the overlap between the baseline $PM_{2.5}$ distributions of nonattainment and attainment tracts, weighted by tract population. The figure is based on data from Meng et al. (2019b).

$PM_{2.5}$ (2001–2003), plotting $-\Delta PM$ against $PM_{i,pre}$.[31] The density plots in Figure 5 show that there are tracts in attainment areas with average $PM_{2.5}$ values above the EPA threshold of 15, likely because the EPA air pollution ground monitor network has incomplete coverage (Sullivan and Krupnick 2018). There are also many tracts in nonattainment areas with baseline $PM_{2.5}$ values below the cutoff.

We use a one to one matching based on propensity scores with replacement to calculate weights W_i for control tracts.[32] In our main version, we estimate tract propensity scores for treatment based on pretreatment pollution $PM_{i,pre}$ alone, which we call M1DiD. In a second version, which we call M2DiD, we additionally match on pretreatment tract population and population density (both based on the 2000 census). For both M1DiD and M2DiD, we impose a common support

[31] Supplemental Appendix Figures A.25 and A.26 show the same pattern for our two alternative sources of pollution data.

[32] Matching has been used in the literature to evaluate other CAA rules. Usually this is done at the county level instead of the tract level as we do here. In an early example, Greenstone (2004) estimates the effect of the NAAQS for SO_2 between 1975 and 1992. He uses propensity scores to match counties based on lagged pollution levels, income, population, and attainment status for other pollutants. This is similar in spirit to Chay and Greenstone (2005), who compare TSP nonattainment counties of the 24 hour standard to a control group that is in attainment of the 24 hour standard, limited to cases where both groups have similar annual TSP concentrations (and nonattainment is triggered by a daily threshold). We mirror their approach more closely in column 7 of Supplemental Appendix Table A.2 where we only look at a subset of areas that are all in attainment of the 24 hour RV threshold in 2005, but some are in nonattainment of the annual threshold (see Figure 1). The results indicate that such an approach reduces some of the observed bias in DiD, but not all. Another early application is by List et al. (2003) who estimate the effect of Ozone nonattainment status on manufacturing plant births between 1980 and 1990. Sanders, Barreca, and Neidell (2020) match on baseline population and mortality to control for trends in mortality.

condition by dropping all nonattainment tracts with a propensity score that is higher than the maximum in the control group. For M1DiD this corresponds to dropping the rightmost tracts in Figure 5.[33] Tracts that act as matched control for multiple treated tracts get an accordingly higher weight. Unlike the raw sample, the resulting matched sample is balanced between nonattainment and attainment tracts as shown in Supplemental Appendix Table A.4. We use these matching weights to weight our DiD regression,[34] equivalent to

$$\Delta PM_i \sqrt{W_i} \;=\; \alpha \sqrt{W_i} + \beta \Delta NA_i \sqrt{W_i} + \Delta \xi_i \sqrt{W_i}. \tag{4}$$

Our identifying assumption now becomes that nonattainment and their propensity-matched attainment areas would have experienced the same average change in $PM_{2.5}$ over time absent the regulation, i.e., $\mathrm{cov}\left[\Delta NA_i \sqrt{W_i}, \Delta \xi_i \sqrt{W_i}\right] = 0$. Intuitively, this assumption addresses the issue visible in Figures 4 and 5 as it places lower weight on control tracts further to the left that have low baseline pollution and are thus less likely to be matched. The correlation coefficient between ΔNA_i and $PM_{i,pre}$ is 0.64, but only -0.08 between $\Delta NA_i \sqrt{W_i}$ and $PM_{i,pre} \sqrt{W_i}$. We show that the event study graph in Supplemental Appendix Figure A.8 no longer shows significant differences in pre-trends when using the matched sample.

One concern with this approach may be that bias from SUTVA violations due to spillovers could be exacerbated relative to standard DiD, if matched control units tend to be geographically closer to treated units absorbing more spillovers. To address this issue, we exclude all counties in attainment that share a border with a nonattainment county and show that the pattern of our baseline results are robust in Supplemental Appendix Figure A.6 and Supplemental Appendix Table A.6.

D. *Regression Discontinuity Design* (*RD*)

Our third approach exploits the discontinuous assignment rule used for nonattainment designations based on the EPA-registered $PM_{2.5}$ threshold ($RV = 15$). We implement a regression discontinuity (RD) design where we compare nonattainment tracts with a value just above the threshold to attainment tracts just below the threshold.[35] We determine the window of EPA-registered $PM_{2.5}$ values around 15 by using the optimal bandwidth selection procedure for local polynomial regression discontinuity estimation following Calonico, Cattaneo, and Titiunik (2014) and Calonico, Cattaneo, and Farrell (2020).[36]

[33] This effectively limits treated units to those with a baseline $PM_{2.5}$ level of up 18.3. Note that this still includes a subset of tracts in counties with the highest RVs (Los Angeles area) on the right in Figure 4.

[34] Since we weight all regressions by tract population, we take the product of matching weights and population weights for our MDiD approaches.

[35] In an early example, Chay and Greenstone (2005) exploit the discontinuous nature of the 1971 NAAQS for TSP. Specifically, they restrict their DiD sample to a narrow window around the TSP cutoff value, akin to our RD0 approach. See also Sanders and Stoecker (2015).

[36] This is akin to using binary weights in equation (4), set to 1 for treated and untreated observations close to the cutoff.

We estimate two versions of the model based on the restricted sample. One version simply estimates the DiD design around the regression discontinuity, which we call RD0 (since it allows for a polynomial of degree 0). The other version allows for a linear relationship between our outcome ΔPM_i and RV, even in the small window around the threshold, which we call RD1 (since it allows for a polynomial of degree 1). To implement RD1, the RV (recentered around 15) enter as a control variable[37]

$$\Delta PM_i = \alpha + \beta \Delta NA_i + \lambda RV_i + \Delta \xi_i. \tag{5}$$

The identifying assumption of the regression discontinuity approach is that the potential outcomes in ΔPM_i are continuous around the threshold. This assumption includes the usual requirement that there are no discontinuous jumps in factors associated with ΔPM at $RV = 15$, and that there is no manipulation around the threshold that may correlate with ΔPM. In Supplemental Appendix Figure A.11 we illustrate that there does not appear to be a discontinuous jump in tract population and population densities around the treatment cutoff. In Supplemental Appendix Figure A.12 we show density plots for RV, which do not show evidence of manipulation around the treatment cutoff within the optimally chosen bandwidths, and pass the formal sorting around the threshold test (Cattaneo, Frandsen, and Titiunik 2015, McCrary 2008). Figure A.9 shows insignificant pre-trends with our RD design.

We argue that the DiDwb, MDiD and RD approaches address the bias in the simple DiD that stems from correlation of both outcome and treatment with baseline pollution as shown in Figure 4 that violates the parallel trends assumption underlying DiD. However, they differ with respect to the estimand: While the DiD, DiDwb, and MDiD approaches, correctly identified, estimate the average treatment effect on the treated (ATT), the RD approach estimates the local average treatment effect (LATE) of nonattainment designation around the $RV = 15$ annual threshold.

E. *Heterogeneous Treatment Effects by Baseline Pollution Levels*

We have so far assumed that the treatment effect β is homogeneous across all tracts. Heterogeneous treatment effects β_i are potentially important because even if we fail to detect average treatment effects, the policy may be effective in a subset of tracts in nonattainment areas, possibly the most polluted ones. Auffhammer, Bento, and Lowe (2009), for example, find no statistically significant effect of nonattainment designation under the 1990 CAA amendments for PM_{10} at the county level, but find significant reductions in PM_{10} for individual monitors that are in nonattainment. Similarly, Bento, Freedman, and Lang (2015) and Gibson (2019) find larger improvements of air quality near binding pollution monitors that are responsible for assignment into nonattainment of an area compared to less binding monitors in the same areas.

[37] For the empirical implementation, we also interact the values with nonattainment status to allow for different slopes on either side of the cutoff.

To account for potential heterogeneity in treatment effects, we repeat the standard DiD and all three of our approaches with an added interaction term between nonattainment status and baseline levels of $PM_{2.5}$ in 2001–2003. The standard DiD equation (2) becomes

$$\Delta PM_i = \alpha + \beta_1 \Delta NA_i + \beta_2 \Delta NA_i PM_{i,pre} + \Delta\xi_i. \quad (6)$$

Treatment β_i therefore varies along the dimension of pretreatment pollution, or $\beta_i = \beta_1 + \beta_2 PM_{i,pre}$.[38]

F. *Heterogeneous Treatment Effects by Previous PM_{10} Nonattainment Status*

Of the 208 nonattainment counties, 71 counties were in nonattainment of the 1990 NAAQS for PM_{10} in the years leading up to 2005. Since $PM_{2.5}$ and PM_{10} are highly correlated, and indeed often emitted by the same sources, on-going regulation of PM_{10} emissions may well alter the impact of additional $PM_{2.5}$ regulation. Supplemental Appendix Figure A.14 repeats our Figure 4 but shows four groups based on both $PM_{2.5}$ nonattainment and PM_{10} nonattainment. We address previous nonattainment in two ways. Our first approach is to show robustness of our results to dropping all areas in PM_{10} nonattainment (Supplemental Appendix Table A.7). Our second approach is to explicitly allow heterogeneous treatment effects of $PM_{2.5}$ nonattainment based on previous PM_{10} nonattainment. Specifically, we estimate a naive DiD regression (as well as our other models) that allows for such heterogeneous effects:

$$\Delta PM_i = \alpha + \beta_1 \Delta NA_{2005_i}(1 - NA_{1990_i}) + \beta_2 \Delta NA_{2005_i} NA_{1990_i} + \beta_3 NA_{1990_i} + \Delta\xi_i. \quad (7)$$

The coefficients β_1 and β_2 capture the $PM_{2.5}$ nonattainment effects for areas without (β_1) and with previous (β_2) PM_{10} nonattainment status, respectively. Note that we control for differential trends based on PM_{10} nonattainment status separately (captured by β_3), so β_1 and β_2 represent the marginal effect of $PM_{2.5}$ nonattainment for the two groups, respectively. In principle, β_2 could be smaller than β_1, e.g., because switching from no treatment into treatment has the most impact, but β_2 could also be larger than β_1, e.g., if being in nonattainment with both NAAQS has compounding impact. This specification allows us to test the difference between β_1 and β_2. This also tests the validity of a nonattainment "switcher" approach which assigns $PM_{2.5}$ nonattainment areas that are also PM_{10} nonattainment areas into the control group (e.g., Currie, Voorheis, and Walker 2023), as it implicitly assumes that $\beta_2 = 0$.

[38] Note that $PM_{i,pre}$ is absorbed in the fixed effect in the equation in levels from which the above equation has been derived. That is: $PM_{2.5_{i,t}} = \beta_1 NA_{it} + \beta_2 NA_{it} PM_{i,pre} + \delta_i + \lambda_t + \xi_{i,t}$, where the uninteracted effect $PM_{i,pre}$ is co-linear with fixed effect δ_i.

III. Results: The Effect of CAA Nonattainment on $PM_{2.5}$

We now compare estimated effects of 2005 nonattainment designations on subsequent changes in $PM_{2.5}$ concentrations using the four approaches outlined above. Our baseline period is the three-year average over 2001–2003. Our posttreatment periods are five (2006–2008) and ten (2011–2013) years later.

A. *Large Effects Suggested by Difference-in-Differences* (*DiD*)

Standard DiD estimation suggests large and statistically significant reductions of $PM_{2.5}$ concentrations in nonattainment areas. This is shown in column 1 of Table 1. The coefficient estimate ($\hat{\beta}$) in panel A shows that nonattainment tracts experienced a $1.5\,\mu g/m^3$ larger reduction in $PM_{2.5}$ than attainment tracts between 2001–2003 and 2006–2008 (equal to the gap between the red and blue dashed lines in Figure 4). column 5 restricts the sample to only those counties for which RVs are available. Results are virtually the same for this smaller sample, indicating no sample selection issues. All results in Table 1 are also virtually identical if we use interpolated population weights from the 2000, 2010, and 2020 census instead of the 2010 census population weights.[39]

Given the issues regarding the parallel trends assumption underlying these estimates discussed above, we conduct 'placebo tests' shown in panel B. Here, we limit our sample to only tracts in attainment areas, i.e., those areas with $RV \leq 15$, and assign a placebo treatment to all those areas with a RV above the median for that group ($RV \geq 11.5$). We then reestimate the DiD model. As shown in panel B, columns 1 and 5, standard DiD suggests that the placebo treatment was associated with significant improvements in air quality (-0.3 and $-0.5\ \mu g/m^3$), providing further evidence that the DiD approach may be biased.[40]

B. *Smaller but Still Positive Effects with DiDwb, MDiD, and RD*

Results from our three alternative approaches are shown in the remaining columns of Table 1. Column 2 shows the estimates for DiDwb, which adds a control for baseline $PM_{2.5}$ to the DiD regression. The coefficient estimate for $\hat{\beta}$ falls to -0.49 in column 2, which implies that the correlation between time trend and baseline levels accounts for much of the DiD estimate.[41]

Column 3 shows estimates from our matched difference-in-differences approach, using baseline $PM_{2.5}$ as the sole matching variable (M1DiD). Column 4 matches

[39] Results available from authors upon request.

[40] As in panel A, column 1 includes unclassifiable areas (without RV) in attainment as per EPA rules, while column 5 drops all areas without RV. Similar results can be seen in Supplemental Appendix Table A.2, where we reestimate the same DiD model on subsets of areas that are successively closer to the treatment cutoff ($RV = 15$). Treatment effect estimates fall as we narrow the window, indicating that there may be a time trend that is unrelated to treatment status but correlated with EPA-registered $PM_{2.5}$ values. If we only drop the nonattainment area with the highest RV (Los Angeles area), corresponding to the observations in the right of Figure 4, we obtain a DiD estimate of -0.9 instead of the reported -1.5.

[41] Note that our DiDwb estimate is based on the exact same sample with the same weights as in DiD, while our other alternative estimates make sample or weighting restrictions instead of adding controls.

Table 1—Nonattainment Status and Changes in $PM_{2.5}$

	ATT					LATE	
	All tracts				With RV	Optimal bandw.	
	DiD (1)	DiDwb (2)	M1DiD (3)	M2DiD (4)	DiD (5)	RD0 (6)	RD1 (7)
Part A. Effect from 2001–03 to 2006–08							
Panel A. Homogeneous treatment effect: from 2001–03 to 2006–08							
Nonattainment	−1.47	−0.49	−0.41	−0.40	−1.48	−0.36	−0.023
	(0.34)	(0.098)	(0.16)	(0.20)	(0.35)	(0.28)	(0.40)
Observations	72,043	72,043	28,291	28,909	47,962	7,026	10,459
Panel B. Placebo treatment effect: from 2001–03 to 2006–08							
Nonattainment	−0.32	−0.12	−0.060	0.018	−0.49	−0.11	−0.26
	(0.12)	(0.11)	(0.12)	(0.12)	(0.14)	(0.21)	(0.30)
Observations	49,357	49,357	20,388	20,127	25,276	2,143	5,411
Panel C. Heterogeneous treatment effect: from 2001–03 to 2006–08							
Nonattainment	4.82	3.85	1.83	3.39	4.81	3.79	3.73
	(0.81)	(0.83)	(0.30)	(0.66)	(0.82)	(0.76)	(0.62)
NA × Baseline	−0.42	−0.33	−0.16	−0.26	−0.42	−0.29	−0.26
	(0.060)	(0.062)	(0.020)	(0.048)	(0.060)	(0.047)	(0.032)
Observations	72,043	72,043	28,291	28,909	47,962	7,026	10,459
Implied ATE	−1.47	−1.06	−0.55	−0.57	−1.48	−0.57	−0.21
10th pct	−0.32	−0.16	−0.11	0.16	−0.32	0.23	0.52
90th pct	−3.56	−2.70	−1.34	−1.89	−3.57	−2.02	−1.51
Part B. Effect from 2001–03 to 2011–13							
Panel D. Homogeneous treatment effect: from 2001–03 to 2011–13							
Nonattainment	−2.35	−0.56	−0.44	−0.55	−2.44	−1.26	−1.11
	(0.27)	(0.096)	(0.096)	(0.11)	(0.28)	(0.35)	(0.37)
Observations	72,043	72,043	28,291	28,909	47,962	6,137	25,856
Panel E. Placebo treatment effect: from 2001–03 to 2011–13							
Nonattainment	−0.95	0.015	0.19	0.15	−1.57	0.23	0.43
	(0.13)	(0.12)	(0.14)	(0.14)	(0.15)	(0.20)	(0.31)
Observations	49,357	49,357	20,388	20,127	25,276	1,046	4,626
Panel F. Heterogeneous treatment effect: from 2001–03 to 2011–13							
Nonattainment	3.91	−0.24	4.78	4.57	3.83	4.45	3.38
	(0.41)	(0.42)	(0.44)	(0.50)	(0.41)	(0.79)	(0.78)
NA × Baseline	−0.42	−0.024	−0.37	−0.36	−0.42	−0.40	−0.31
	(0.029)	(0.032)	(0.033)	(0.036)	(0.029)	(0.053)	(0.053)
Observations	72,043	72,043	28,291	28,909	47,962	6,137	25,856
Implied ATE	−2.35	−0.61	−0.78	−0.79	−2.44	−1.54	−1.21
10th pct	−1.20	−0.54	0.24	0.19	−1.29	−0.44	−0.37
90th pct	−4.43	−0.73	−2.63	−2.58	−4.52	−3.54	−2.74

Notes: The table shows coefficient estimates for the treatment effect of nonattainment status on the change in $PM_{2.5}$ levels between the pre- and posttreatment periods. Each panel × column combination is from a separate regression as described in the text. Column 1 uses simple DiD; column 2 adds controls for baseline $PM_{2.5}$ (2001–03); column 3 runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, column 4 matches on baseline $PM_{2.5}$, tract population and population density (both 2000); column 5 again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists; columns 6 and 7 use the limited sample based on optimal bandwidth selection in a regression discontinuity framework. Standard errors in parentheses are clustered at the county level. All results based on Meng et al. (2019b).

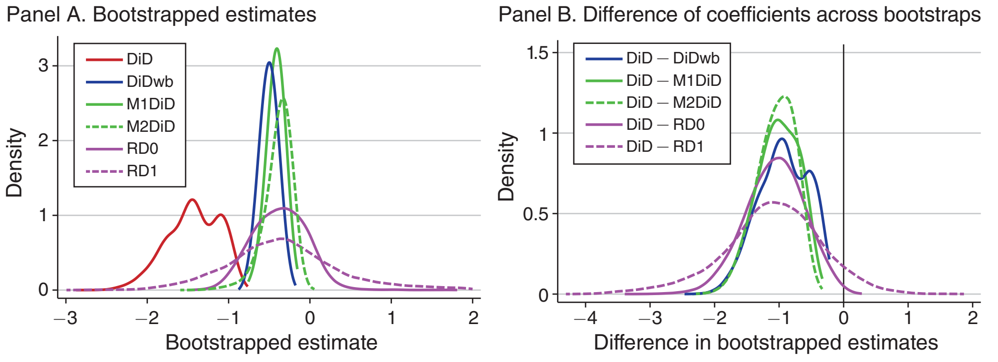


Figure 6. Distribution of Estimates from Bootstraps

Notes: Both panels show distributions of cluster-bootstrapped estimates of our different models corresponding to panel A in Table 1 using a triangle kernel smoother. We draw counties to allow for clustering with replacement based on two strata (attainment and nonattainment), estimate the different models, and repeat the process 10,000 times. Panel A shows the distribution of estimates across the bootstraps for each model. Panel B shows the distribution of the difference between DiD and our alternative models across bootstraps. The area above zero represents the *p*-value of a test of equality of coefficients across models. Supplemental Appendix Table A.5 shows these *p*-values for two-sided tests (i.e., doubling the area in the tail to the right of zero). Based on data from Meng et al. (2019b).

on baseline $PM_{2.5}$, population and population density (M2DiD). Both estimates are substantially smaller then the DiD estimates, with a reduction of about 0.4 $\mu g/m^3$ following nonattainment designation. This effect corresponds to a 3 percent decrease from average concentrations in the pretreatment period.

Columns 6 and 7 show results for our regression discontinuity approaches RD0 and RD1.[42] The point estimate for RD0 is similar to our other strategies, and close to zero for RD1, but both estimates are imprecise. Due to the smaller number of observations around the cutoff, we lack statistical power resulting in larger standard errors. However, effect estimates in both RD0 and RD1 are highly statistically significant when accounting for heterogeneity in panel C, or when considering longer term impacts in panel D, in line with the dynamic effects shown in Supplemental Appendix Figure A.9.

Overall, our preferred specification is M1DiD. It is almost identical to M2DiD, which implies that adding additional matching variables provides little additional benefits to remove bias. M1DiD includes a broader set of tracts and counties than either RD approaches resulting in higher statistical power, but excludes outliers too far from the cutoff that are included in DiDwb, thus presenting a reasonable compromise.

Two points stand out when comparing the four approaches. First, the effect sizes for our three alternative approaches shown in panel A are less than a third the size (around −0.4 to −0.5) of the standard DiD estimates (−1.5) across the board. To statistically test for equality of coefficients across these models, in panel A of Table 1 we use a cluster-bootstrap by drawing counties with replacement by attain-

[42] Graphical representations of these RD approaches are provided in Supplemental Appendix Figure A.10. We provide results for additional bandwidths in Supplemental Appendix Table A.2.

ment and nonattainment strata. Figure 6, panel A shows the resulting distribution of estimates across 10,000 draws, showing clearly that the DiD estimates are centered at a much lower mean with little overlap with our alternative approaches that are centered closer to zero and overlapping with each other. Figure 6, panel B shows the distribution of the differences between the DiD estimate and those of each of our models. The corresponding *p*-values for all pairwise two-sided tests for equality of coefficients are shown in Supplemental Appendix Table A.5. All estimates from our alternative models are significantly different from the DiD estimates at the 1 percent level except the RD1 model due to noisier estimates (see Figure 6, panel A). Conversely, Supplemental Appendix Table A.5 shows that we cannot reject equality of coefficients in all pairwise tests between our alternative models, suggesting that they recover a similar effect.

Second, note that the placebo tests in panel B of Table 1 yield smaller and insignificant coefficients for our three alternative approaches. The pattern is similar when we use the other two sources of pollution data, as we show in Supplemental Appendix Tables A.23 and A.24.

C. *Robustness*

We next discuss robustness to several concerns for our analysis: (i) spillovers, (ii) preceding PM_{10} nonattainment designation, (iii) additional controls for trends, (iv) concurrent air pollution policies, (v) uncertainty of pollution data, and (vi) alternative models for estimation.

First, we exclude all 300 counties in attainment that share a border with a county in nonattainment, to reduce potential bias from spatial spillovers of air quality changes. Supplemental Appendix Table A.6 shows that corresponding estimates are, if anything, slightly higher suggesting that there may be some small spatial spillovers as pollutants can travel across space. However, and importantly, the pattern of much lower estimates compared to DiD is similar to our main results. Second, the pattern also holds when we focus on counties that switch into nonattainment by excluding all areas that were in nonattainment of the NAAQS for PM_{10} in the years leading up to 2005 (71 of 208 $PM_{2.5}$ nonattainment counties in 2001–04, see also Supplemental Appendix Figure A.14), as we show in Supplemental Appendix Table A.7. This implies that the bias of standard DiD cannot be explained by correlations with previous CAA rules. We further explore interaction with PM_{10} nonattainment by explicitly allowing for heterogeneous treatment effects further below.

Third, the findings remain unchanged when we add further controls, which allow for state by period specific time trends in $PM_{2.5}$ and period by quartile-of-tract-population-density specific time trends, as shown in Supplemental Appendix Tables A.8 and A.9, respectively.

Fourth, apart from the nonattainment designations under the NAAQS for $PM_{2.5}$, two separate air quality policies came into effect during our study period: the NO_x Budget Trading Program (NBP) and it's successor, the Clean Air Interstate Rule (CAIR). They target NO_x, SO_2, and Ozone emissions. NO_x and SO_2 are precursors to $PM_{2.5}$, so that overlap with these policies could partially drive our results. To test this, we collect data on regulated facilities under these programs, with details discussed in

Supplemental Appendix A.8. Controlling for NBP and CAIR status does not affect our estimates either for the DiD case or our alternative DiDwb. On the contrary, the estimated effect of those policies depends dramatically on inclusion of $PM_{2.5}$ nonattainment controls.

Fifth, our air pollution data comes from reanalysis models where some predictions may be more uncertain, e.g., due to larger distances to ground-based air pollution monitors. If the measurement error is nonclassical, such that higher $PM_{2.5}$ regions or changes are systematically over- or underestimated, ignoring such uncertainty may introduce bias. We address this concern in three ways. First, we use the data from van Donkelaar et al. (2021b) that also quantifies the uncertainty for each data point from the underlying reanalysis model and raw data. We drop the 30 percent of data points with the highest uncertainty and reestimate our models. Second, we only keep counties if they or any of their neighboring county contain a ground-based monitor. Third, in our most restrictive version with the least observations, we only use monitor data directly from EPA (2022a). We repeat the estimation of the first part of Table 1 and show that our estimates of both naive DiD as well as of our alternative models are robust in Supplemental Appendix Tables A.12 and A.13.

Sixth, we provide results from Synthetic Difference-in-Differences (SDiD) estimation recently proposed by Arkhangelsky et al. (2021). SDiD weights control units (and pretreatment years) to minimize the mean difference in time trends between treated and control groups. Supplemental Appendix A.10 shows that SDiD produces very similar estimates ($-0.41\ \mu g/m^3$) as our three alternatives.

D. *Heterogeneous Treatment Effects Vary with Baseline Pollution Levels*

Our results so far have focused on the average treatment effect of nonattainment designation. We now investigate the possibility of treatment heterogeneity. To do so, we repeat all of the above estimations but add an interaction term between nonattainment status and baseline levels of $PM_{2.5}$ in 2001–2003, following equation (6). The results are shown in panel C of Table 1 and indicate that there is indeed significant treatment heterogeneity. The negative interaction coefficient implies that more polluted tracts experience larger improvements following nonattainment designation. In our M1DiD specification, the improvement in $PM_{2.5}$ concentrations is estimated to be $0.1\,\mu g/m^3$ at the tenth percentile of baseline pollution levels, while it is $1.3\,\mu g/m^3$ at the ninetieth percentile.[43] The heterogeneous treatment effects are in line with previous findings by Auffhammer, Bento, and Lowe (2009) and others discussed above. One possible explanation may be regulatory attention on those areas triggering nonattainment status and where population health is most at risk.

The implied (local) average treatment effects calculated from the two reported coefficients are also shown in the table and, again, are significantly smaller than those produced by standard DiD. Compared to panel A, the coefficients in panel C

[43] For estimating the interaction effects, we only use the units within the sample for each column, e.g., within the RD-chosen window. Note that we use the same overall tenth and ninetieth percentiles of baseline pollution for calculating the corresponding effects at these percentiles across columns for consistency, extrapolating for those models that use a smaller window. While the 90th percentile within the RD0 window is lower than the overall (15.7 versus 20), there is substantial variation in tract level pollution even within the county-based window.

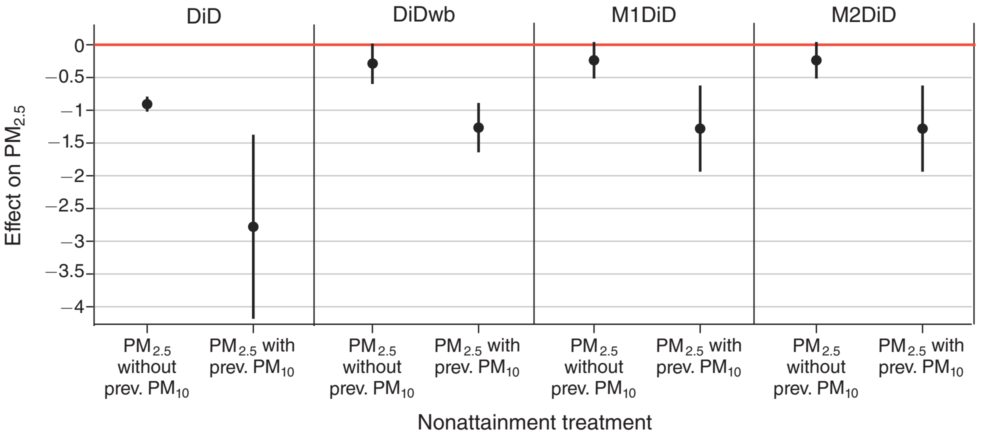


Figure 7. Heterogeneous $PM_{2.5}$ Nonattainment Treatment Effect by Previous PM_{10} Nonattainment Status

Notes: The figure shows the effect of the 2005 nonattainment designation on $PM_{2.5}$ from 2001–03 to 2006–08 for four models. The two estimates for each of the four models show effects of $PM_{2.5}$ nonattainment for those areas that have no previous PM_{10} nonattainment on the left, and for those areas that have previous PM_{10} nonattainment on the right. The estimates come from a single regression with appropriate identifiers for the groups and a control for trends based on previous PM_{10} nonattainment designation alone, so the estimates can be interpreted as the marginal effects of $PM_{2.5}$ nonattainment designation for the two groups. The two estimates are significantly different from each other at the 1 percent level for each model. Based on data from Meng et al. (2019b).

on nonattainment and the interaction are also highly statistically significant for all our strategies, a property we will rely on when employing instrumental variable regressions below. This effect heterogeneity replicates with the other two sources of pollution data, as we show in Supplemental Appendix Tables A.23 and A.24.

E. *Larger Treatment Effects with Previous PM_{10} Nonattainment Status*

We next allow for heterogeneous treatment effects by previous PM_{10} nonattainment status as in equation (7). Figure 7 shows the effect of 2005 $PM_{2.5}$ nonattainment by previous 1990 PM_{10} nonattainment status. Note that we additionally control for flexible trends by previous PM_{10} nonattainment status, so the effect shown is the marginal effect of $PM_{2.5}$ nonattainment status. For both, the naive DiD model, as well as our alternative models, the $PM_{2.5}$ nonattainment effect is significantly larger for those areas that have previously been in PM_{10} nonattainment, with differences significant at the 1 percent level (Supplemental Appendix Table A.14 shows results including the RD models omitted here because they do not always include areas with both treatment groups).[44]

[44] The average treatment effect across both groups in Table 1 is between the two heterogeneous effects shown in Figure 7. Importantly, the effect heterogeneity here does not merely capture heterogeneity from baseline air quality discussed in the previous section. The pattern between the two groups is the same if we additionally allow for heterogeneous treatment effects by baseline air quality as in equation (6), which additionally shows that within the two groups, the initially more polluted tracts see larger air quality improvements.

Importantly, these results show that not only areas that switched from previous PM_{10} *attainment* to $PM_{2.5}$ nonattainment see an effect of $PM_{2.5}$ nonattainment. On the contrary, areas in previous PM_{10} nonattainment see an even larger effect of $PM_{2.5}$ nonattainment. This explains why assigning this latter group into the control group as in Currie, Voorheis, and Walker (2023) flattens pre-trends and lowers estimated effects of $PM_{2.5}$ nonattainment, as areas with the largest treatment effect are added to the control group. Nevertheless, as we show in Supplemental Appendix A.13B, even when assigning these areas into the control group, adjusting for confounding trends is essential, as DiD still significantly overestimates nonattainment effects compared to DiDwb (Supplemental Appendix Table A.18).[45]

F. *Effects over the Longer Time Horizon to 2011–13*

In part B of Table 1, we repeat the analysis of part A but use the years 2011–13 as end point instead of 2006–08. The idea is to test for impacts of nonattainment designation that may take some time to take effect, or that are cumulative. Indeed, all estimates become larger, implying slightly bigger effects of the policy over the ten year period than the five year period. The large difference between DiD ($-2.3\,\mu g/m^3$) and our alternative approaches (-0.4 to $-1.3\,\mu g/m^3$) also persists over this longer horizon.[46]

IV. Implications for Equity and Pollution Damages

We have shown that different estimation strategies yield substantially different estimates for the effect of nonattainment designation on $PM_{2.5}$ concentrations. Difference-in-differences (DiD) estimates suggest the largest improvements, likely due to bias. Our three alternative methods—controlling for baseline pollution (DiDwb), matched difference-in-differences (MDiD), and regression discontinuity (RD)—show substantially smaller, though nonzero effects. In this section, we show how the differences in effect sizes matter for two important applications: one focused on structural pollution exposure disparities and environmental justice, and the other focused on estimating pollution damages as capitalized in house prices.

A. *The Role of the CAA in Shrinking Racial and Urban-Rural Pollution Gaps*

We first focus on disparities in $PM_{2.5}$ exposure in the United States and the contribution of the 2005 CAA NAAQS in reducing these disparities. We begin with the mean pollution exposure gap between Black and White Americans, which has been well documented (Jbaily et al. 2022; Currie, Voorheis, and Walker 2023).[47] Currie, Voorheis, and Walker (2023) show that this Black-White $PM_{2.5}$ gap

[45] These patterns also persist when we instead drop these areas as in Supplemental Appendix Table A.7.

[46] The DiD estimate is equal to the gap between the red and blue dashed lines in Supplemental Appendix Figure A.2. The pattern is similar when using the other two pollution data sources, see Supplemental Appendix Tables A.23 and A.24.

[47] We use our tract level $PM_{2.5}$ concentrations (which are already population weighted by census block populations) and aggregate them up to the national level using tract level Black and White non-Hispanic population

Table 2—Pollution Disparities–Counterfactual Gap Analysis

	$PM_{2.5}$ exposure		Black-White gap							
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
Panel A. Black-White pollution gap										
					Contribution of CAA (in %) [homogeneous effect]					
2001–2003	13.3	11.62	1.69							
2006–2008	12.15	10.57	1.58	−0.11	193	64	53	52	47	3
2011–2013	9.64	8.63	1.00	−0.69	49	12	9	12	26	23
					Contribution of CAA (in %) [heterogeneous effect]					
2001–2003	13.3	11.62	1.69							
2006–2008	12.15	10.57	1.58	−0.11	287	213	108	135	140	86
2011–2013	9.64	8.63	1.00	−0.69	64	14	30	29	47	36
					Contribution of CAA (in %) [+race interactions]					
2001–2003	13.3	11.62	1.69							
2006–2008	12.15	10.57	1.58	−0.11	130	57	86	95	−23	−52
2011–2013	9.64	8.63	1.00	−0.69	68	18	45	41	49	47
	$PM_{2.5}$ exposure		Urban-Rural gap							
	Urban	Rural	(levels)	(change)						
Panel B. Urban-Rural pollution gap										
					Contribution of CAA (in %) [homogeneous effect]					
2001–2003	12.59	10.21	2.38							
2006–2008	11.26	9.62	1.64	−0.74	52	17	14	14	13	1
2011–2013	9.28	7.78	1.49	−0.89	70	17	13	16	37	33
					Contribution of CAA (in %) [heterogeneous effect]					
2001–2003	12.59	10.21	2.38							
2006–2008	11.26	9.62	1.64	−0.74	73	54	28	34	35	20
2011–2013	9.28	7.78	1.49	−0.89	87	19	39	38	63	49
					Contribution of CAA (in %) [+urban interactions]					
2001–2003	12.59	10.21	2.38							
2006–2008	11.26	9.62	1.64	−0.74	71	52	32	37	35	21
2011–2013	9.28	7.78	1.49	−0.89	88	20	39	40	63	52

Notes: Left columns show average $PM_{2.5}$ exposure of Black, White, Urban and Rural populations, and difference between groups, as derived from census block level pollution concentrations and population counts. Right columns show the contribution of CAA nonattainment designations in 2005 based on counterfactual calculations that factor out nonattainment treatment effects as estimated in columns 1–4, 6, and 7 of Table 1. Population data is from the 2000, 2010, and 2020 waves of the US census, linearly interpolated for years in between. Pollution data is from Meng et al. (2019b).

fell by $0.6\,\mu g/m^3$ between 2005 and 2015, and that a substantial portion (61.2 percent) of that narrowing can be attributed to the effects of the 2005 nonattainment designations.

In panel A of Table 2 we conduct a similar counterfactual accounting exercise. Our data shows that the Black-White $PM_{2.5}$ gap fell by $0.69\,\mu g/m^3$ over the ten years from 2001–03 to 2011–13. To measure the potential contribution of the CAA NAAQS, we use coefficient estimates from Table 1. Our DiD estimates suggest that nonattainment designations alone contributed 49 percent to that narrowing, or

counts as weights. In Supplemental Appendix A.12 we show that our $PM_{2.5}$ exposure levels are virtually identical to those in Jbaily et al. (2022), and show the same in Supplemental Appendix A.13A for Currie, Voorheis, and Walker (2023).

64 percent when we allow for heterogeneous treatment effects following panel C of Table 1. When we allow for heterogeneous effects by racial composition of census tracts—by including additional interaction terms with the share of the tract population that was Black in 2000 as well as the interaction between this share and baseline pollution levels—the contribution slightly increases to 68 percent. Importantly, our alternative estimation strategies all show a role for the CAA NAAQS in narrowing the Black-White pollution gap, but the estimated contribution is considerably smaller, often around half the size (between 9 and 26 percent for homogeneous treatment effects, 14 and 47 percent with heterogeneous effects, and 18 and 49 percent with additional race interaction terms). A similar pattern is observed for the shorter five-year period ending in 2006–08.[48]

While we look at slightly different time periods and report main results using data from Meng et al. (2019b) instead of Di et al. (2021), our estimated CAA contribution based on standard DiD with heterogeneous effects (68 percent) until 2011–2013 is broadly in line with the findings in Currie, Voorheis, and Walker (2023) of a contribution of 61.2 percent from 2005–2015. As shown in Supplemental Appendix A.13A, when we follow the approach of Currie, Voorheis, and Walker (2023) based on RIF/Quantile Regressions and their treatment assignment, we recover an almost identical 61.1 percent contribution.[49] However, as we show in Supplemental Appendix A.13B, controlling for confounding trends (i.e., DiDwb) in their approach also reduces the CAA contribution to 18.6 percent (Supplemental Appendix Table A.19). The same pattern holds when we use their RIF/Quantile Regression approach but our treatment assignment, which shows a contribution of 22.5 percent based on DiDwb much in line with our estimated 24 percent using the same data based on Di et al. (2021), in Supplemental Appendix Table A.25.

We next explore spatial pollution gaps between urban and rural residents.[50] In panel B of Table 2, we document a similar role of CAA rules in narrowing the Urban-Rural gap in $PM_{2.5}$. Urban centers, especially those with high population densities and large traffic volumes, are arguably those areas with the highest particulate matter concentrations and tend to have different socio-economic characteristics than rural counterparts. We observe a large Urban-Rural $PM_{2.5}$ gap, even larger than the Black-White gap by around 40 percent. The Urban-Rural gap also narrowed substantially from 2001–2003 to 2011–2013. Again, 2005 nonattainment designations account for some of this narrowing, with DiD estimates suggesting the largest contribution (70–88 percent) while the other approaches yield significantly smaller estimates (13–63 percent).

Overall, our results show that the NAAQS for $PM_{2.5}$ enacted in 2005 significantly contributed towards reducing pollution exposure disparities. Our results also highlight the sensitivity of such analyses to the underlying method of identifying

[48] A contribution of more than 100 percent as is the case in all DiD estimates implies that the counterfactual gap would have increased.

[49] In Supplemental Appendix A.13B we also show that other minor data differences to Currie, Voorheis, and Walker (2023) are negligible.

[50] We do so by calculating weighted average exposure levels using the number of urban and rural residents in each tract as weights. These classifications are based on the 2000 census definition which classifies blocks as urbanized areas (UAs) and urban clusters (UCs) based on population density.

treatment effects, demonstrating that the contribution may have been substantially smaller than suggested by standard DiD estimates. While Table 2 includes changes in population distributions (interpolating linearly between 2000, 2010, and 2020 census waves), we show in Supplemental Appendix Table A.20 that the results hold when population is fixed at 2010 levels switching off any population sorting channels. Supplemental Appendix Tables A.25 and A.26 show that the patterns are similar when using the two alternative pollution data sources.

B. *Instrumenting Pollution with CAA Nonattainment to Estimate Effects on House Prices*

So far, we have focused on air pollution as outcome variable, and the role of the CAA rules in reducing $PM_{2.5}$ concentrations. We now turn to the damages of $PM_{2.5}$ exposure as capitalized in residential real estate values, using nonattainment designations as instrument for pollution. To do so, we estimate the following simple model to describe the change in the log of house prices in tract i:

$$\Delta Y_i \;=\; \alpha + \theta \Delta PM_i + \Delta \mu_i, \tag{8}$$

which is equivalent to estimating the relationship in levels with tract and period fixed effects. We estimate this equation via OLS or IV, using nonattainment as instrument for ΔPM_i using either DiD or our three alternative approaches.

Following the literature that uses nonattainment instruments for pollution, this assumes that nonattainment designations have no direct impact on our outcome, house prices, apart from their impact through pollution reductions. This would be violated if there are, for example, substantial employment effects from regulation (Walker 2013) that also impact house prices, or if nonattainment and attainment areas experience different house price trends for other reasons.[51] In Supplemental Appendix A.16, we show a version of the below analysis with additional commuting zone fixed effects in equation (8) that should capture most of the labor market effects. This changes the interpretation of coefficients and estimates become smaller, but the relative pattern between different IV estimates discussed below are robust.[52]

Two mechanisms could explain why we expect the results to differ between standard DiD and our three alternative estimation strategies. First, variation in the estimates of nonattainment effects in the 'first stage' (Table 1) will mechanically alter the estimated effect of pollution on house prices. Second, there may be differences in house price trends that co-vary with baseline pollution. For example, we could imagine that polluted urban centers experienced a different evolution of house prices

[51] While the exclusion restriction cannot be tested conclusively, we see no significant differences in pre-trends in the house price event study equivalent to Table 3 shown in Supplemental Appendix Figure A.20.

[52] A specification with commuting zone fixed effects uses only variation in $PM_{2.5}$ induced by the interaction of nonattainment designations and baseline $PM_{2.5}$, while binary nonattainment designations are absorbed. If nonattainment designations affect house prices through employment or similar effects at the commuting zone level, those will no longer be a source of bias. But doing so also changes the interpretation of our estimates. We no longer capture house price changes due to different pollution trajectories between commuting zones, but only differential trajectories of tracts within a given commuting zone.

Table 3—Pollution Damages–Instrumental Variable Comparison

	OLS (1)	DiD-IV (2)	DiDwb-IV (3)	M1DiD-IV (4)	M2DiD-IV (5)	RD0-IV (6)	RD1-IV (7)
Panel A. Effect of $PM_{2.5}$ ***increases*** *on house price index growth 2001–03 to 2006–08*							
ΔPM2.5	−0.040	−0.064	−0.15	−0.12	−0.10	−0.16	−0.17
	(0.017)	(0.0080)	(0.011)	(0.029)	(0.011)	(0.11)	(0.048)
Observations	54,529	54,529	54,529	21,152	21,693	5,087	7,937
K-P F statistic		72.8	22.8	25.0	26.3	47.5	55.1
Elasticity	−0.48	−0.77	−1.81	−1.44	−1.26	−1.98	−2.00
Panel B. Effect of $PM_{2.5}$ ***increases*** *on house price index growth 2001–03 to 2011–13*							
ΔPM2.5	−0.012	−0.016	−0.035	−0.033	−0.045	−0.032	−0.022
	(0.0092)	(0.012)	(0.041)	(0.019)	(0.017)	(0.040)	(0.031)
Observations	54,378	54,378	54,378	21,062	21,608	4,496	19,035
K-P F statistic		305.0	25.1	135.8	114.9	146.7	145.9
Elasticity	−0.14	−0.19	−0.42	−0.39	−0.54	−0.39	−0.26

Notes: The dependent variable is the change in the logarithm of the house price index. ΔPM2.5 is the change in $PM_{2.5}$ since 2001–03 in $\mu g/m^3$, instrumented by CAA nonattainment status for $PM_{2.5}$, allowing for heterogeneous effects in the instrument by previous PM_{10} nonattainment status and by baseline $PM_{2.5}$ levels in 2001–03. First-stage specifications in Columns 2–7 correspond to Columns 1–4, 6, and 7 in Table 1. Standard errors in parentheses are clustered at the county level. Pollution data is from Meng et al. (2019b).

over time.[53] Such biases in the reduced-form relationship between nonattainment designations and house price growth could work in both directions. Our three estimation strategies also address this second bias. DiDwb directly controls for such trends in house prices, while MDiD and RD both compare treated with control units that have similar baseline pollution levels and thus similar associated trends. As we show in Supplemental Appendix Table A.22, there are only small differences in the reduced-form relationships across empirical strategies, suggesting that the bias mainly operates through the first mechanism linked to the first stage. To increase instrument power, we include the set of instruments that exploit the two types of treatment effect heterogeneity: the heterogeneity in panels C and F of Table 1, as well as the heterogeneity based on previous PM_{10} nonattainment treatment status as in Figure 7.

Column 1 in Table 3 shows results when running OLS without instruments, and implies that a one unit increase in $PM_{2.5}$ is associated with a reduction in house prices by 4 percent $(\exp(-0.04) - 1)$. Instrumenting $PM_{2.5}$ with nonattainment status corresponding to the simple DiD approach in column 2 shows an effect that is larger implying a semielasticity of around 6 percent. This is expected as pollution may exhibit classical measurement error and is correlated with desirable factors such as economic activity, introducing attenuation and upward bias. The remainder of Table 3 shows corresponding estimates from our three approaches that address the time trend that is correlated with baseline $PM_{2.5}$. Column 3 shows estimates that include baseline $PM_{2.5}$ as a control (DiDwb-IV), columns 4 and 5 are based on matched DiD (MDiD-IV), and in columns 6 and 7 we use the regression discontinuity strategy (RD-IV).

[53] See also Sanders and Stoecker (2015), Sanders, Barreca, and Neidell (2020) who address differential trends in their health outcome variables when estimating the impact of pollution.

The IV estimates based on our three alternative approaches yield larger pollution damages, around 50 percent to 150 percent larger than those based on the standard DiD-IV. Our preferred approach for this setting is M1DiD-IV, shown in column 4, which implies that a one unit increase in $PM_{2.5}$ lowers house prices by 11 percent. This effect is almost twice that in the standard DiD-IV. Our house price effects are also larger than those found for previous NAAQS targeting coarser categories of particles. While this could in part be due to the finest particles mattering more or that house prices have become more sensitive to pollution over time, our results show that it could also be due to the downward bias in the standard DiD-IV estimate, which is more in line with previous results.[54] This implies that while simple DiD may *overestimate* the effect of nonattainment on $PM_{2.5}$, it may *underestimate* the effect of $PM_{2.5}$ on house prices when nonattainment status is used as an instrument for $PM_{2.5}$. A similar pattern holds when we extend the posttreatment period to 2011–13. Again, the DiDwb-IV, MDiD-IV and RD-IV yield larger estimates of pollution damages as capitalized by house prices. The pattern is similar when we use the other two sources of pollution data, as we show in Supplemental Appendix Tables A.27 and A.28.

Finally, when we estimate the effect of nonattainment designation on house prices directly (reduced form), the results show that house prices in nonattainment areas gained an additional 6 percent on average due to being designated into nonattainment.[55]

V. External Validity

Our focus so far has been on the $PM_{2.5}$ rules and we demonstrated the importance of accounting for trends in pollution that correlate with baseline pollution and assignment into treatment. We next examine how likely it is that this insight extends to NAAQS beyond the 2005 $PM_{2.5}$ rules.

The forerunner of the 2005 $PM_{2.5}$ regulation was the 1990 PM_{10} regulation, widely studied in the literature (e.g., Bento, Freedman, and Lang 2015; Auffhammer, Bento, and Lowe 2009). To gauge the issue of confounding trends for this older regulation, we use the historic $PM_{2.5}$ data from Meng et al. (2019b) going back to the 1980s together with the 1990 PM_{10} nonattainment areas.[56] First, Supplemental Appendix Figure A.5 shows that there is indeed a similar pattern where $PM_{2.5}$ improvement is clearly associated with 1987–89 baseline $PM_{2.5}$ concentrations even in the absence of 1990 PM_{10} nonattainment. Second, we estimate the impact of PM_{10} nonattainment comparing

[54] The implied elasticity of -1.4 is larger than the elasticity of -0.6 in Bento, Freedman, and Lang (2015) who study the effects of PM_{10} on house prices, or the elasticity of around -0.2 to -0.35 reported for TSP (PM_{100}) in Chay and Greenstone (2005). Note that the elasticity for the endline 2011–13 is around -0.4, and thus more in line with previous estimates, but also 100 percent larger than the elasticity based on simple DiD-IV. Graff Zivin and Singer (2023) explore differential capitalization rates by racial groups using micro data, but find similar overall effects on house prices based on our proposed approaches.

[55] This policy effect is based on the average "reduced form" effect estimated in Supplemental Appendix Table A.22. Alternatively, we can calculate an approximation by multiplying the $-0.55\ \mu g/m^3$ reduction in $PM_{2.5}$ from Table 1 with the house price effect of -11 percent per $\mu g/m^3$ from Table 3, which yields an increase of around 7 percent $(\exp(-0.55 \times -0.12) - 1)$.

[56] Note that we use $PM_{2.5}$ concentrations instead of PM_{10} because of much better spatial coverage due to Meng et al. (2019b). $PM_{2.5}$ is highly correlated with PM_{10} as it is a subset of PM_{10}.

Table 4—The Effect of 1990 PM_{10} Nonattainment Designation on $PM_{2.5}$ Concentrations

	DiD (1)	DiDwb (2)	M1DiD (3)	M2DiD (4)
Homogeneous treatment effect: from 1987–89 to 1991–93				
Nonattainment	−0.75	−0.37	−0.27	−0.46
	(0.29)	(0.056)	(0.26)	(0.31)
Observations	72,043	72,043	20,174	22,094

Notes: The table shows coefficient estimates for the treatment effect of nonattainment status with the 1990 PM_{10} NAAQS (instead of the 2005 $PM_{2.5}$ NAAQS) on the change in $PM_{2.5}$ levels between the pre- and posttreatment periods of 1987-89 and 1991-93, respectively. Each column is from a separate regression, where column 1 uses simple DiD, column 2 adds controls for baseline $PM_{2.5}$ (1987–89), column 3 runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, and column 4 matches on baseline $PM_{2.5}$, tract population and population density (both 2000). Standard errors in parentheses are clustered at the county level. Pollution data is from Meng et al. (2019b).

1987–89 and 1991–93 analogous to our main analysis for the $PM_{2.5}$ rules. Table 4 shows that naive DiD has a similar upward bias (column 1), while DiDwb, M1DiD and M2DiD have a lower estimated nonattainment impact of around half the size.[57] This suggests that our insights are likely just as relevant for the earlier 1990 PM_{10} standards.

Apart from the closely related 1990 PM_{10} rules, the problem of correlated trends may apply more broadly to NAAQS and related policies. Indeed, Greenstone (2004) mentions possible "mean reversion" going back to the SO rules in the 1970s and Clay et al. (2021) show that to-be-treated units were on different trends for the original CAA in 1970. In our robustness Section IIIC, we briefly discuss the NBP and CAIR to rule them out as possible confounding concurrent air quality policies. We can, however, also use the data on NBP and CAIR treatment to evaluate whether controlling for trends based on baseline pollution alters the estimated effect of NBP and CAIR designation per se. Supplemental Appendix Table A.11, panel C and D show that, in contrast to simple DiD, a DiDwb approach produces a much smaller effect of NBP or CAIR treatment on subsequent $PM_{2.5}$ levels. Controlling for baseline trends in Ozone has little effect on estimated effects on Ozone levels, however, suggesting that confounding trends for $PM_{2.5}$ may be particularly problematic (panel E).[58] Finally, we use the comprehensive EPA data on all NAAQS nonattainment areas (EPA 2022b) to focus on those areas which have consistently been in attainment, i.e., were never subject to any NAAQS nonattainment regulation in history and also not subject to the NBP or CAIR. Even in this subset of "never treated" areas, we document that there are differential trends in air quality improvements by baseline pollution. Using our $PM_{2.5}$ data from 1981 (Meng et al. 2019b), Figure 8, panels A, B, and C show that baseline $PM_{2.5}$ on the horizontal axes predicts 10-year improvements ($-\Delta PM_{2.5}$) on the vertical axes, akin to Figure 5. Panel D plots coefficients from a

[57] We use the years 1987–89 as baseline here. We do not use an RD framework here due to lack of access to EPA-registered PM_{10} values for the 1990 regulation.

[58] Panels D and E also replicate the results from Deschênes, Greenstone, and Shapiro (2017), see Supplemental Appendix A.8.

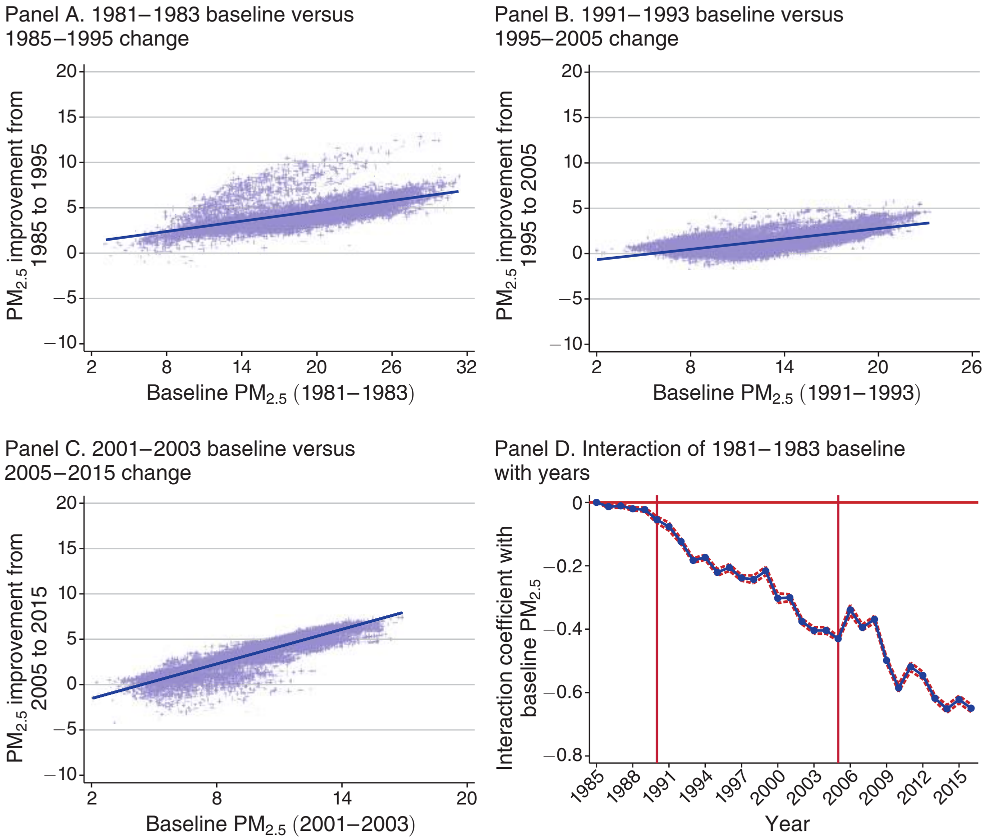


Figure 8. Long-Running Correlation between Baseline Pollution and Pollution Changes

Notes: Panels A, B, and C plot tract level mean $PM_{2.5}$ concentrations in 1981–83, 1991–93, and 2001–03, respectively, on the horizontal axes, and 1985–1995, 1995–2005, and 2005–2015 improvements in $PM_{2.5}$ concentrations on the vertical axes. Panel D shows interaction coefficients estimated in a tract-year panel regression with $PM_{2.5}$ concentrations as dependent variable. The plotted estimates are for tract level baseline $PM_{2.5}$ (1981–83 average) interacted with year dummies with 95 percent confidence bands based on standard errors clustered at the county level. The figure is based on data from Meng et al. (2019b).

regression of annual $PM_{2.5}$ levels on 1981–83 baseline $PM_{2.5}$, and shows a general trend correlated with baseline pollution in the 'never treated' group. This suggests that the issue of differential trends that we identity is relevant beyond our focus on the 2005 rules, as the "never treated" group is likely to be a control group in most analyses of CAA policies.[59]

The issue of correlated trends is often not accounted for in the literature. There are few exceptions that address possible confounding trends which, however, have no explicit discussion of bias (Greenstone 2004; Chay and Greenstone 2005; Auffhammer, Bento, and Lowe 2009; Bishop, Ketcham, and Kuminoff 2023). Greenstone 2004 controls for and matches on baseline levels for analyzing the

[59] Colmer et al. (2020) show a convergence of pollution concentrations, but for the entire United States, not just the "never treated" group.

1970s SO regulation. Chay and Greenstone (2005) use a variant of regression discontinuity with manual window selection to study TSP rules in the 1970s–80s (see also Sanders and Stoecker (2015). Auffhammer, Bento, and Lowe (2009) include monitor-specific time trends in their analysis of the 1990 rules for PM_{10}, and Bishop, Ketcham, and Kuminoff (2023) control for baseline $PM_{2.5}$ when exploiting nonattainment designations to estimate $PM_{2.5}$ effects on dementia prevalence in a cross-sectional analysis. However, it remains common to estimate nonattainment effects without adjusting for confounding trends by baseline pollution, including in the growing literature focusing on $PM_{2.5}$ nonattainment or the previous PM_{10} or TSP nonattainment designations (e.g., Grainger 2012, Isen, Rossin-Slater, and Walker 2017, Sanders, Barreca, and Neidell 2020, Colmer and Voorheis 2021, Colmer et al. 2022, Hollingsworth et al. 2022, Currie, Voorheis, and Walker 2023).

For practitioners, our findings show that it is important to take into account trends based on baseline pollution. This implies adding controls (or matching on) baseline pollution levels when using differenced outcomes, or allowing for interactions between baseline levels and year dummies in a panel fixed effect settings. While it may depend on context, our findings also imply that nonattainment areas that have previously been in nonattainment should either be kept in the treated group (possibly with a heterogeneous treatment effect) or dropped, but not assigned into the control group.

VI. Conclusion

Did the National Ambient Air Quality Standards for fine particulate matter pollution introduced in 2005 trigger air quality improvements? Our results show that areas in nonattainment of the standards indeed experienced faster reductions in $PM_{2.5}$ levels following regulation. This is in line with the empirical literature evaluating earlier iterations of CAA rules (Currie and Walker 2019; Aldy et al. 2022).

We find, however, that difference-in-differences (DiD) estimation tends to overstate the achieved pollution reductions. This bias is driven by a correlation between baseline levels and changes of pollution, even in the absence of nonattainment designations. We propose three alternative approaches that address this source of bias: DiD with added controls for baseline pollution trends (DiDwb), matched DiD (MDiD), and regression discontinuity designs (RD). All three produce similar estimates which are less than half the size of those produced by standard DiD. The strategies are easy to implement and our results imply that it may be worth including them in assessments of CAA nonattainment rules, or when using CAA nonattainment designations as instrument for air pollution.

We further show that the choice of estimation strategy can have important implications for the role of the CAA with regards to pollution exposure disparities and environmental justice. We find the 2005 CAA rules likely contributed to the narrowing of the Urban-Rural and Black-White gaps in $PM_{2.5}$ exposure, but less so than DiD estimates would suggest. Similarly, the choice of empirical strategy matters when estimating pollution damages with nonattainment instruments. As we show for the case of house prices, while standard DiD overstates the impact of the regulation on pollution, it understates the impact of pollution when nonattainment is used as instrument. Similar differences likely hold in other settings where nonattainment designations are used as instruments, including estimates of health or productivity losses.

Our findings provide a cautionary tale when it comes to estimating the effects of nonattainment designations which are a central element of Clean Air Act rules. We find that nonattainment designations in 2005 cannot be considered random and that nonattainment areas likely followed a different time trend than attainment areas. Similar time trends are apparent going back to at least the 1980s, suggesting possible confounding bias for analyses of previous NAAQS.

REFERENCES

Aguilar-Gomez, Sandra, Holt Dwyer, Joshua Graff Zivin, and Matthew Neidell. 2022. "This is Air: The "Nonhealth" Effects of Air Pollution." *Annual Review of Resource Economics* 14: 403–25.

Aldy, Joseph E., Maximilian Auffhammer, Maureen Cropper, Arthur Fraas, and Richard Morgenstern. 2022. "Looking Back at 50 Years of the Clean Air Act." *Journal of Economic Literature* 60 (1): 179–232.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. "Synthetic Difference-In-Differences." *American Economic Review* 111 (12): 4088–4118.

Auffhammer, Maximilian, Antonio M. Bento, and Scott E. Lowe. 2009. "Measuring the Effects of the Clean Air Act Amendments on Ambient Concentrations: The Critcal Importance of a Spatially Disaggregated Analysis." *Journal of Environmental Economics and Management* 58 (1): 15–26.

Banzhaf, Spencer, Lala Ma, and Christopher Timmins. 2019. "Environmental Justice: The Economics of Race, Place, and Pollution." *Journal of Economic Perspectives* 33 (1): 185–208.

Bento, Antonio, Matthew Freedman, and Corey Lang. 2015. "Who Benefits from Environmental Regulation? Evidence from the Clean Air Act Amendments." *Review of Economics and Statistics* 97 (3): 610–22.

Bishop, Kelly C., Jonathan D. Ketcham, and Nicolai V. Kuminoff. 2023. "Hazed and Confused: The Effect of Air Pollution on Dementia." *Review of Economic Studies* 90 (5): 2188–2214.

Bogin, Alexander, William Doerner, and William Larson. 2019. "Local House Price Dynamics: New Indices and Stylized Facts." *Real Estate Economics* 47 (2): 365–98.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell. 2020. "Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs." *Econometrics Journal* 23 (2): 192–210.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Cattaneo, Matias D., Brigham R. Frandsen, and Rocio Titiunik. 2015, "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the US Senate." *Journal of Causal Inference* 3 (1): 1–24.

Chay, Kenneth Y., and Michael Greenstone. 2005. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy* 113 (2): 376–424.

Chetty, Raj, and John N. Friedman. 2019. "Replication Data for: A Practical Method to Reduce Privacy Loss when Disclosing Statistics Based on Small Samples." Inter-University Consortium for Political and Social Research. http://doi.org/10.3886/E116494V1.

Clay, Karen, Akshaya Jha, Joshua A. Lewis, and Edson R. Severnini. 2021. "Impacts of the Clean Air Act on the Power Sector from 1938-1994: Anticipation and Adaptation." NBER working paper 28962.

Colmer, Jonathan, and John Voorheis. 2021. "The Grandkids Aren't Alright: The Intergenerational Effects of Prenatal Pollution Exposure." Unpublished.

Colmer, Jonathan, John Voorheis, and Brennan Williams. 2022. "Air Pollution and Economic Opportunity in the United States." Unpublished.

Colmer, Jonathan, Ian Hardman, Jay Shimshack, and John Voorheis. 2020. "Disparities in PM2.5 Air Pollution in the United States." *Science* 369 (6503): 575–78.

Currie, Janet, and Reed Walker. 2019. "What do Economists have to Say About the Clean Air Act 50 Years after the Establishment of the Environmental Protection Agency?." *Journal of Economic Perspectives* 33 (4): 3–26.

Currie, Janet, John Voorheis, and Reed Walker. 2023. "What Caused Racial Disparities in Particulate Exposure to Fall? New Evidence from the Clean Air Act and Satellite-Based Measures of Air Quality." *American Economic Review* 113 (1): 71–97.

Deschênes, Olivier, Michael Greenstone, and Joseph S. Shapiro. 2017. "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program." *American Economic Review* 107 (10): 2958–89.

Di, Qian, Heresh Amini, Liuhua Shi, Itai Kloog, Rachel Silvern, James Kelly, M. Benjamin Sabath, et al. 2019. "An Ensemble-Based Model of PM2. 5 Concentration across the Contiguous United States with High Spatiotemporal Resolution." *Environment International* 130: 104909.

Di, Qian, Y. Wei, A. Shtein, C. Hultquist, X. Xing, H. Amini, L. Shi, et al. 2021. "Daily and Annual PM2.5 Concentrations for the Contiguous United States, 1-km grids, v1 (2000–2016)." NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/0rvr-4538.

Drupp, Moritz A., Ulrike Kornek, Jasper N. Meya, and Lutz Sager. 2021. "Inequality and the Environment: The Economics of a Two-Headed Hydra." CESifo Working Paper 9447.

Fowlie, Meredith, Edward Rubin, and Reed Walker. 2019. "Bringing Satellite-Based Air Quality Estimates Down to Earth." *AEA Papers and Proceedings* 109: 283–88.

Gibson, Matthew. 2019. "Regulation-Induced Pollution Substitution." *Review of Economics and Statistics* 101 (5): 827–40.

Graff Zivin, Joshua, and Matthew Neidell. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review* 102 (7): 3652–73.

Graff Zivin, Joshua, and Gregor Singer. 2023. "Disparities in Pollution Capitalization Rates: The Role of Direct and Systemic Discrimination." NBER working paper 30814.

Grainger, Corbett A. 2012. "The Distributional Effects of Pollution Regulations: Do Renters Fully Pay for Cleaner Air?" *Journal of Public Economics* 96: 840–52.

Greenstone, Michael. 2002. "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures." *Journal of Political Economy* 110 (6): 1175–1219.

Greenstone, Michael. 2004. "Did the Clean Air Act Cause the Remarkable Decline in Sulfur Dioxide Concentrations?." *Journal of Environmental Economics and Management* 47 (3): 585–611.

Henderson, J. Vernon. 1996. "Effects of Air Quality Regulation." *American Economic Review* 86 (4): 789–813.

Hollingsworth, Alex, Taylor Jaworski, Carl Kitchens, and Ivan J. Rudik. 2022. "Economic Geography and the Efficiency of Environmental Regulation." NBER Working Paper 29845.

Isen, Adam, Maya Rossin-Slater, and W. Reed Walker. 2017. "Every Breath You Take—Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970." *Journal of Political Economy* 125 (3): 848–902.

Jbaily, Abdulraham, Xiaodan Zhou, Jie Liu, Ting-Hwan Lee, Leila Kamareddine, Stéphane Verguet, and Francesca Dominici. 2022. "Air Pollution Exposure Disparities across US Population and Income Groups." *Nature* 601 (7892): 228–33.

Jha, Akshaya, Peter H. Matthews, and Nicholas Z. Muller. 2019. "Does Environmental Policy Affect Income Inequality? Evidence from the Clean Air Act." *AEA Papers and Proceedings* 109: 271–76.

Kahn, Matthew E., and Erin T. Mansur. 2013. "Do Local Energy Prices and Regulation Affect the Geographic Concentration of Employment?" *Journal of Public Economics* 101: 105–14.

Landrigan, Philip J., Richard Fuller, Nereus J. R. Acosta, Olusoji Adeyi, Robert Arnold, Niladri Basu, Abdoulaye Bibi Baldé, et al. 2018. "The Lancet Commission on Pollution and Health." *The Lancet* 391 (10119): 462–512.

List, John A., Daniel L. Millimet, Per G. Fredriksson, and W. Warren McHone. 2003. "Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Matching Estimator." *Review of Economics and Statistics* 85 (4): 944–52.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2022. "Ipums National Historical Geographic Information System: Version 17.0." IPUMS. http://doi.org/10.18128/D050.V17.0. (accessed February 2, 2022).

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714.

Meng, Jun, Chi Li, Randall V. Martin, Aaron van Donkelaar, Perry Hystad, and Michael Brauer. 2019a. "Estimated Long-Term (1981–2016) Concentrations of Ambient Fine Particulate Matter across North America from Chemical Transport Modeling, Satellite Remote Sensing, and Ground-Based Measurements." *Environmetal Science & Technology* 53 (9): 5071–79.

Meng, Jun, Chi Li, Randall V. Martin, Aaron van Donkelaar, Perry Hystad, and Michael Brauer. 2019b. "Historical PM2.5 across North America." Washington University of St. Louis Atmospheric Composition Analysis Group. https://sites.wustl.edu/acag/datasets/historical-pm2-5-across-north-america. (accessed February 2, 2022).

Sager, Lutz, and Gregor Singer. 2025. *Data and Code for: "Clean Identification? The Effects of the Clean Air Act on Air Pollution, Exposure Disparities and House Prices."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E192280V1.

Sanders, Nicholas J., and Charles Stoecker. 2015. "Where Have All the Young Men Gone? Using Sex Ratios to Measure Fetal Death Rates." *Journal of Health Economics* 41: 30–45.

Sanders, Nicholas J., Alan I. Barreca, and Matthew J. Neidell. 2020. "Estimating Causal Effects of Particulate Matter Regulation on Mortality." *Epidemiology* 31 (2): 160–67.

Shapiro, Joseph S., and Reed Walker. 2018. "Why is Pollution from US Manufacturing Declining? The Roles of Environmental Regulation, Productivity, and Trade." *American Economic Review* 108 (12): 3814–54.

Sullivan, Daniel M., and Alan Krupnick. 2018. "Using Satellite Data to Fill the Gaps in the US Air Pollution Monitoring Network." Resources for the Future Working Paper 18–21.

US Environmental Protection Agency (EPA). 2004a. *Technical Support for State and Tribal Air Quality Fine Particle (PM2.5) Designations*, *Chapter 5: An Explanation of EPA's Nine-Factor Analysis*. Research Triangle Park, NC: EPA.

US Environmental Protection Agency (EPA). 2004b. *Technical Support for State and Tribal Air Quality Fine Particle (PM2.5) Designations, Chapter 6: Nine Factor Analyses of Individual Nonattainment Areas*. Research Triangle Park, NC: EPA.

US Environmental Protection Agency (EPA). 2005a. "Air Quality Designations and Classifications for the Fine Particles (PM2.5) National Ambient Air Quality Standards; Final Rule." *Federal Register* 70 (3).

US Environmental Protection Agency (EPA). 2005b. "Air Quality Designations for the Fine Particles (PM2.5) National Ambient Air Quality Standards – Supplemental Amendments; Final Rule." *Federal Register* 70 (3).

US Environmental Protection Agency (EPA). 2005c. *Technical Support Document for PM2.5 Designations - Supplemental Notice*. EPA.

US Environmental Protection Agency (EPA). 2016. "Fine Particulate Matter National Ambient Air Quality Standards: State Implementation Plan Requirements; Final Rule." *Federal Register* 81 (164).

US Environmental Protection Agency (EPA). 2018a. "County PM2.5 24h Design Values, 1999-2017." Particulate Matter NAAQS Review - Analyses and Data Sets. https://www.epa.gov/sites/default/files/2020-03/county_level_pm2.5_24hr_design_values_1999_2017.xlsx. (accessed February 1, 2022).

US Environmental Protection Agency (EPA). 2018b. "County PM2.5 annual design values, 1999-2017." Particulate Matter NAAQS Review - Analyses and Data Sets. https://www.epa.gov/sites/default/files/2020-03/county_level_pm2.5_annual_design_values_1999_2017.xlsx. (accessed February 1, 2022).

US Environmental Protection Agency (EPA). 2021. "Green Book PM-2.5 1997, Designated Area Design Values." EPA Green Book. https://www3.epa.gov/airquality/greenbook/qdtc.html (accessed September 24, 2021).

US Environmental Protection Agency (EPA). 2022a. "AirData Annual Summary Data." EPA. https://www.epa.gov/outdoor-air-quality-data. (accessed February 1, 2022).

US Environmental Protection Agency (EPA). 2022b. "Green Book National Area and County-Level Multi-Pollutant Information." EPA. https://www.epa.gov/green-book/green-book-national-area-and-county-level-multi-pollutant-information. (accessed February 1, 2022).

US Federal Housing Finance Agency (FHFA). 2021. "Annual House Price Indexes, Census Tracts (version: February 23, 2021)." Federal Housing Finance Agency (FHFA). https://www.fhfa.gov/sites/default/files/documents/bdl_faqs_local_hpis.pdf (accessed February 1, 2022).

van Donkelaar, Aaron, Melanie S. Hammer, Liam Bindle, Michael Brauer, Jeffery R. Brook, Michael J. Garay, N. Christian Hsu, et al. 2021a. "Monthly Global Estimates of Fine Particulate Matter and their Uncertainty." *Environmental Science & Technology* 55 (22): 15287– 15300.

van Donkelaar, Aaron, Melanie S. Hammer, Liam Bindle, Michael Brauer, Jeffery R. Brook, Michael J. Garay, N. Christian Hsu, et al. 2021b. "Surface PM2.5 data (v5.gl.03)." Washington University of St. Louis Atmospheric Composition Analysis Group. https://sites.wustl.edu/acag/datasets/surface-pm2-5. (accessed February 2, 2022).

Walker, W. Reed. 2013. "The Transitional Costs of Sectoral Reallocation: Evidence from the Clean 2107 Air Act and the Workforce." *Quarterly Journal of Economics* 128 (4): 1787–1835.