Grantham Research Institute on Climate Change and the Environment

Clean identification? The effects of the Clean Air Act on air pollution, exposure disparities and house prices

Lutz Sager and Gregor Singer

September 2023

Grantham Research Institute on Climate Change and the Environment Working Paper No. 376

ISSN 2515-5717 (Online)

Grantham Foundation for the Protection of the Environment

The London School of Economics and Political Science

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and a number of other sources. It has 13 broad research areas:

1. Biodiversity
2. Climate change adaptation and resilience
3. Climate change governance, legislation and litigation
4. Climate, health and environment
5. Environmental behaviour
6. Environmental economic theory
7. Environmental policy evaluation
8. International climate politics
9. Science and impacts of climate change
10. Sustainable public and private finance
11. Sustainable natural resources
12. Transition to zero emissions growth
13. UK national and local climate policies

More information about the Grantham Research Institute is available at: www.lse.ac.uk/GranthamInstitute

Suggested citation:
Sager L and Singer G (2023) *Clean identification? The effects of the Clean Air Act on air pollution, exposure disparities and house prices*. Grantham Research Institute on Climate Change and the Environment Working Paper 376. London: London School of Economics and Political Science

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

Lutz Sager* Gregor Singer*

September 11th, 2023

LINK TO LATEST VERSION

Abstract

We assess the US Clean Air Act standards for fine particulate matter $(\text{PM}_{2.5})$. Using high-resolution data, we find that the 2005 regulation reduced $\text{PM}_{2.5}$ 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 $\text{PM}_{2.5}$ exposure disparities, but less than difference-in-differences suggest. Pollution damages capitalized into house prices, on the other hand, appear larger than previously thought when leveraging regulatory variation.

Keywords: air pollution, clean air act, environmental justice, regulation, house prices

JEL codes: Q52, Q53, Q58

*Corresponding authors: Lutz Sager and Gregor Singer. Lutz Sager: McCourt School of Public Policy, Georgetown University, 3700 O St. NW, Washington, DC 20057, USA, E-mail: Lutz.Sager@georgetown.edu. Gregor Singer: London School of Economics, Houghton St, London WC2A 2AE, United Kingdom, E-mail: g.a.singer@lse.ac.uk. We thank Raphael Calel, Roger Fouquet, Josh Graff Zivin, Matthew Neidell, Arik Levinson, 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. G.S. acknowledges support from the Grantham Research Institute on Climate Change and the Environment at the London School of Economics.

The 1970 Clean Air Act (CAA) and subsequent amendments are the cornerstone of air quality regulation in the United States (US). 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 & Neidell 2012) and other non-health 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.

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 US. We use high-resolution data from three leading reanalysis projects (Meng et al. 2019*a*, Di et al. 2019, van Donkelaar et al. 2021*a*) that estimate $PM_{2.5}$ concentrations by combining ground monitors, satellite data and chemical transport models for the entire contiguous US. 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 et al. 2023, Jha et al. 2019, Sanders et al. 2020, Currie et al. 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} stan-

[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.

[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.

dard.[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 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% 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 et al. (2023), and that these gaps narrowed, in part due to NAAQS nonattainment areas (Currie et al. 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

[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 et al. (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.

[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.

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% 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 et al. 2015) and up to four times the elasticity for Total Suspended Particles (TSP or PM_{100}) (Chay & 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 et al. 2009, Bento et al. 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 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 & 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 & Greenstone 2005), and PM_{10} concentrations (Auffhammer et al. 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

[5] Nonattainment designation has also been linked to changes in industrial activity (Henderson 1996, Greenstone 2002), within-product improvements in emission intensity (Shapiro & Walker 2018), and employment (Kahn & Mansur 2013, Walker 2013). Deschenes et al. (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 & Walker (2019).

pollution gaps and environmental justice (Currie et al. 2023, Jbaily et al. 2022, Banzhaf et al. 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 & Greenstone 2005, Grainger 2012, Bento et al. 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 et al. 2017, Sanders & Stoecker 2015, Sanders et al. 2020, Colmer & Voorheis 2021, Bishop et al. 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.

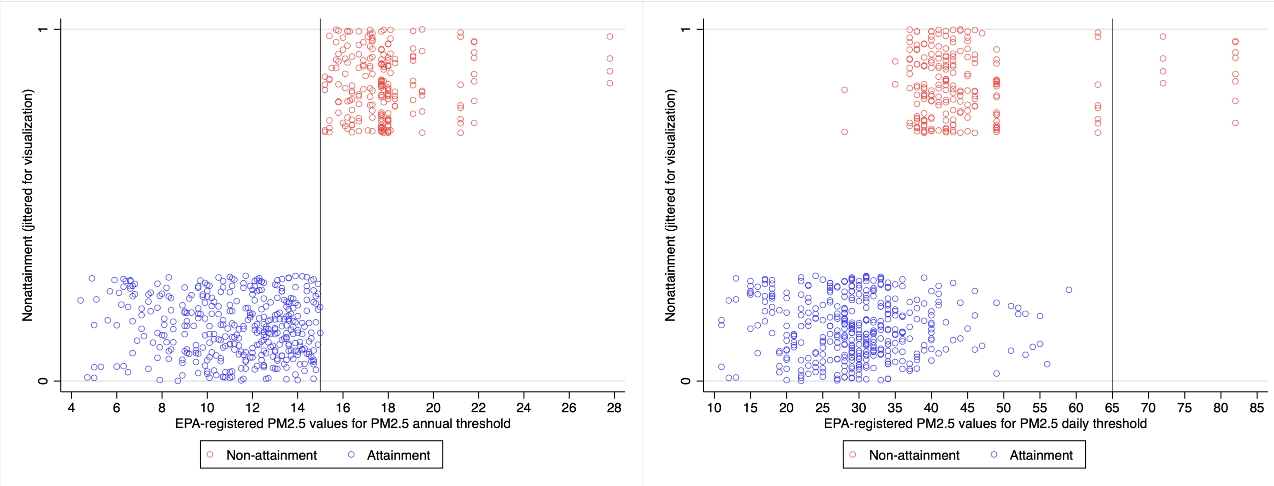
I. Data and Regulatory Context

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 $2.5\mu m$ in diameter, came into effect.[6] The EPA introduced 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 98^{th} 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

[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 EPA (2005*a*, 2016). Previous NAAQS regulated coarser particulate matter PM_{10} and TSP, equivalent to PM_{100}.

[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.



(a) Annual threshold (15 $\mu g/m^3$)

(b) Daily threshold (65 $\mu g/m^3$)

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 2001-2003. Panel (b) shows the EPA-registered $PM_{2.5}$ values for the three-year average of the 98th percentile daily threshold of 65 $\mu g/m^3$ from 2001-2003. 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 requirment 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).

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 & 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 (EPA 2005*a*). While these monitors were usually placed in more populous counties, they only covered around 20% of counties, possibly missing counties that would otherwise be regulated (Sullivan & Krupnick 2018, Fowlie et al. 2019).[9] The EPA took the three year averages of monitor readings from 2001-2003 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

[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.

[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).

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 2a 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 (EPA 2005*a*,*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 (EPA 2018*b*,*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 EPA (2004*b*, 2005*c*, 2021), and update them with the supplementary amendments contained in EPA (2005*b*,*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 from Meng et al. (2019*b*), but we show in the Appendix that all our results are similar when using two alternative data sets from Di et al. (2021) or van Donkelaar et al. (2021*b*).[13]

[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 EPA (2004*a*) 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.

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

All three data sets provide $PM_{2.5}$ concentrations at a spatial resolution of $0.01° \times 0.01°$ (approximately 1km $\times$ 1km cells in the US, depending on latitude). The universal spatial coverage and high resolution of these products allow us to assign pollution levels to all Census blocks in the contiguous US on an annual basis starting from 2000 based on each of the three data sets, and from 1989 for the data based on Meng et al. (2019*b*).[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 III.C. by leveraging information on uncertainty in the underlying predictions and by replicating our analysis using only pollution monitor data from EPA (2022*a*).

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 1km 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.

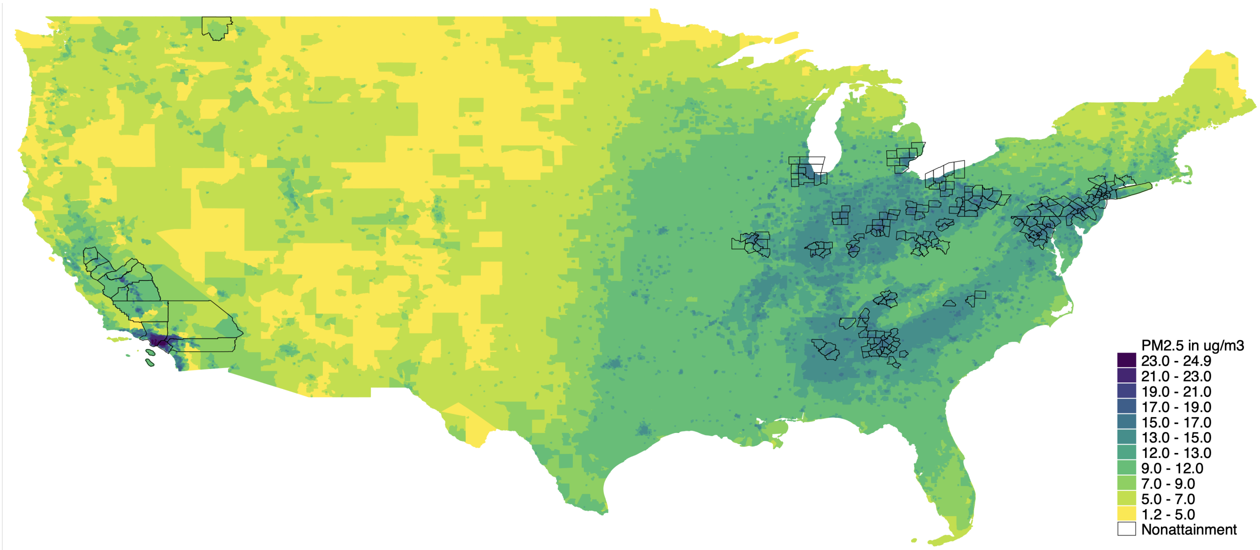
The resulting data provides detailed measures of average $PM_{2.5}$ exposure in each tract and each year. The map in Figure 2a shows tract level $PM_{2.5}$ concentrations averaged between 2001-2003 across the contiguous US. We use this detailed data structure to exploit variation not comprehensively captured by monitoring data, specifically to investigate heterogeneity within counties, visible in Figure 2b, 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

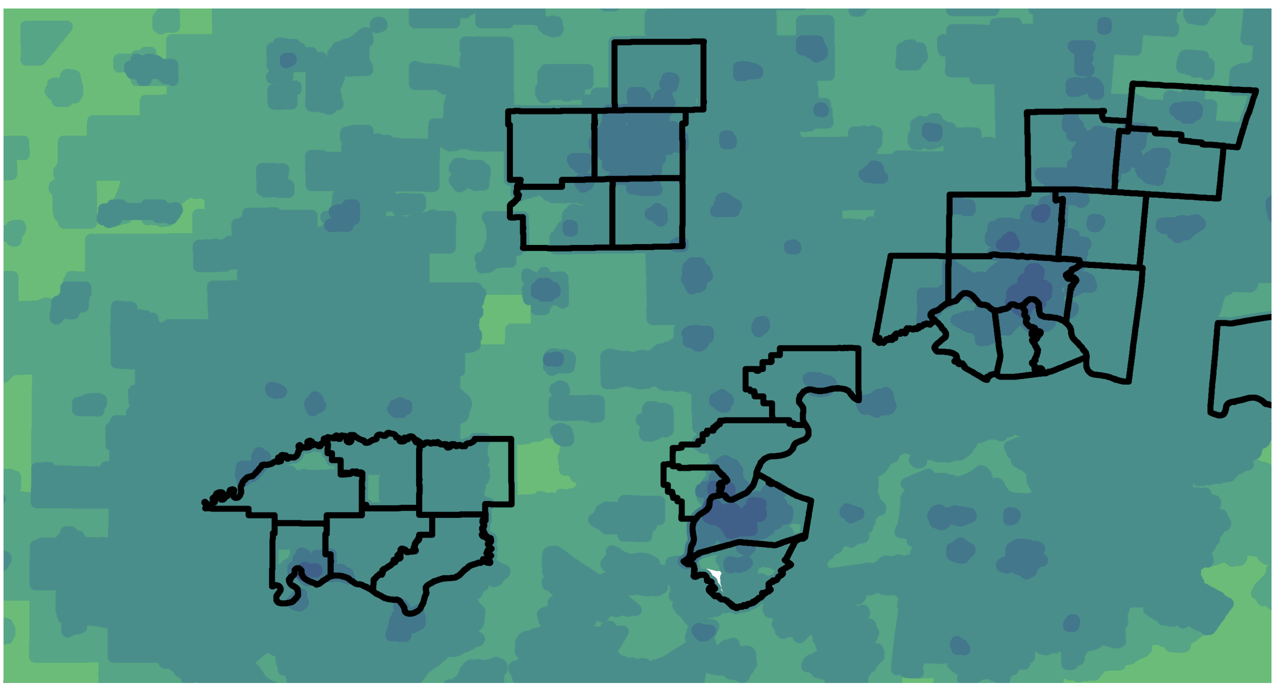
We demonstrate the implications of our estimates for the implied damages from $PM_{2.5}$ exposure capitalized in house prices. Our measure of changes in house prices relies on data provided by the

[14] Meng et al. (2019*b*) 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 et al. (2023) (see also Appendix A.13A.). We complement the Census data with information on commuting zones and tract-level characteristics from Chetty & Friedman (2019).



(a) Contiguous United States



(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. (2019*b*), and nonattainment counties. Panel (a) shows the entire contiguous US, and Panel (b) zooms into the area around Indianapolis (North), Evansville (South-West), Louisville (South) and Cincinatti (North-East).

Federal Housing Finance Agency (FHFA 2021), specifically, the annual house price index (HPI) at the tract level (further described in Bogin et al. 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 et al. 2023, who consider 2004 as the start of the post-treatment period). To avoid any bias from anticipatory effects in 2004, we define our pre-treatment period as the three-year average between 2001-2003.[16] Taking three-year averages helps to lower the risk of mis-attributing short-term fluctuations in air quality or measurement error to the CAA rules. To allow for time-varying effects, we report results for two post-treatment periods, respectively five years (2006-08 average) and ten years (2011-13 average) after the pre-treatment period. 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:

$$\mathrm{PM}_{i,t} = \beta \mathbf{NA}_{i,t} + \delta_i + \lambda_t + \xi_{i,t} \tag{1}$$

where $\mathbf{NA}_{i,t}$ 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]

[16]Note that we provide some descriptive statistics using Meng et al. (2019*b*) 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. 2019*b*, Di et al. 2021) and 2020 (in the case of van Donkelaar et al. 2021*b*) 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

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 pre-treatment to the post-treatment period between treated 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 pre-treatment and the post-treatment periods as outcome:

$$\Delta \mathrm{PM}_i = \alpha + \beta \Delta \mathbf{NA}_i + \Delta \xi_i \tag{2}$$

where $\Delta \mathbf{NA}_i$ 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 $\mathrm{cov}(\Delta \mathbf{NA}_i, \Delta \xi_i) = 0$.

While the parallel trends assumption cannot be directly tested, it is common practice to look at pre-treatment 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. 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. (2021*b*).

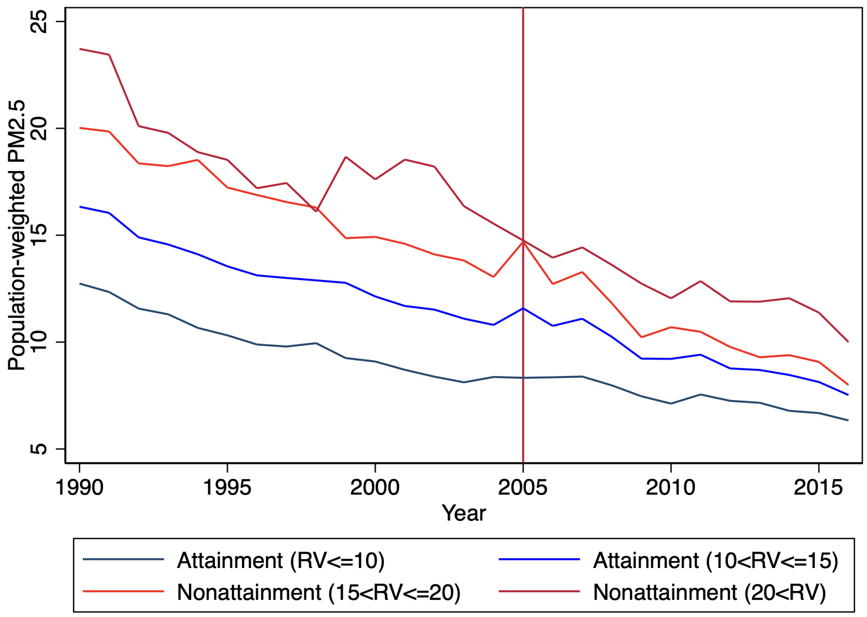
We conduct several robustness tests for these statistically significant pre-trends. Appendix Figure A.1 shows the pre-trends exist when (i) using 2000 Census borders and population instead of

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 Appendix Table A.6.

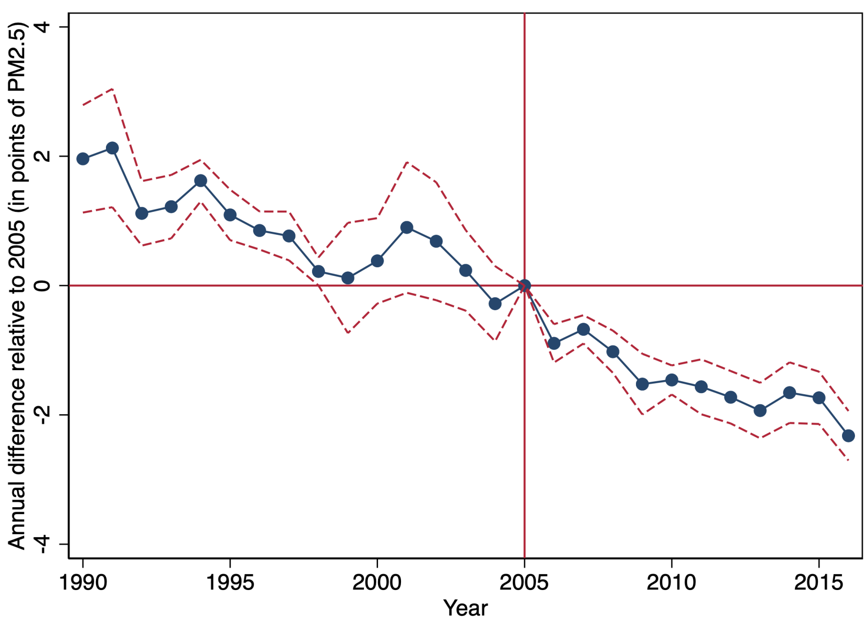
[19] That is $\Delta \mathrm{PM}_i = \mathrm{PM}_{i,post} - \mathrm{PM}_{i,pre} = \beta(\mathbf{NA}_{i,post} - \mathbf{NA}_{i,pre}) + (\delta_i - \delta_i) + (\lambda_{post} - \lambda_{pre}) + (\xi_{i,post} - \xi_{i,pre})$. All our specifications in changes could equivalently be modeled as panel regressions 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 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.



(a) Evolution of $PM_{2.5}$ by EPA RV grouping



(b) Event study (annual nonattainment-attainment differences in $PM_{2.5}$)

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% 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. (2019*b*).

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 compare our event study to the event study in Currie et al. (2023) along with other differences in detail in 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 for these areas.[23] We formally test treatment effect heterogeneity by previous 1990 PM_{10} nonattainment status in Section II.F. and III.E., 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 ($\text{cov}(\Delta\mathbf{NA}_i, RV_i) > 0$) 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 ($\text{cov}(\Delta\mathbf{NA}_i, \Delta\xi_i) \neq 0$).[25] Both ΔPM and nonattainment status are correlated with pre-treatment pollution levels, confounding the standard DiD estimate.

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. (2021*b*). In Appendix Figure A.4, we use EPA monitor level data instead and show that the pattern is similar at the monitor level.[26] Ap-

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

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

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

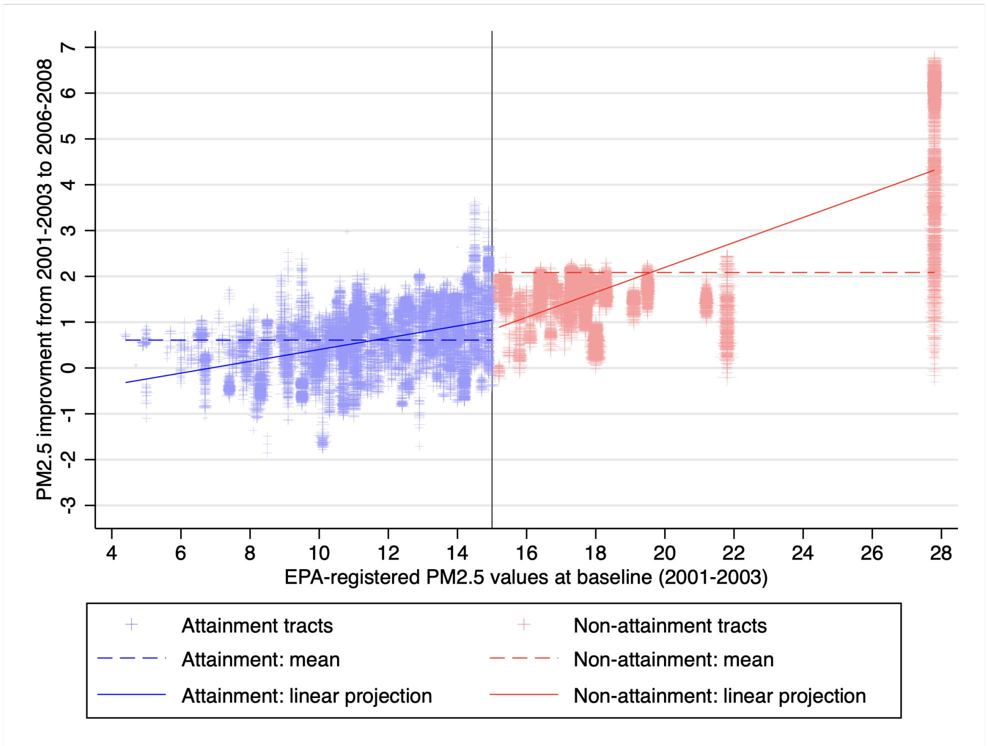


Figure 4: Improvement in tract $PM_{2.5}$ averages and EPA-registered $PM_{2.5}$ values

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. (2019*b*).

pendix 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 Appendix that different trends persist, however, even when dropping all counties that were previously in nonattainment of the PM_{10} standard (Figure A.19 and 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] 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

[27]Consider the example of road traffic. If all regions maintain the same volume of traffic, but newer cars generate 10% less emissions per mile driven, we would expect 10% 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 $PM_{2.5}$

different trends by state), or reasons related to population density (with urban, suburban, and rural areas experiencing different trends). In 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:

$$\Delta\xi_i = \gamma \text{PM}_{i,pre} + \Delta\epsilon_i$$

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

$$\Delta\text{PM}_i = \alpha + \beta\mathbf{NA}_i + \gamma\text{PM}_{i,pre} + \Delta\epsilon_i \tag{3}$$

We refer to this DiD approach with baseline controls as DiDwb. Note that including $\text{PM}_{i,pre}$ as control in our specification in differences is equivalent to controlling for $\text{PM}_{i,pre}$ separately by period (λ_t) in the levels specification in Equation (1), where $\text{PM}_{i,pre}$ is absorbed by tract fixed effects.[29] This approach absorbs any improvements in air quality over time that are proportional to baseline $\text{PM}_{2.5}$ levels (e.g. $\gamma = -0.1$ would indicate a 10% 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 $\text{PM}_{2.5}$ over time absent regulation, conditional on a linear association between between baseline $\text{PM}_{2.5}$ and ΔPM. Put differently, we require that $\text{cov}(\Delta\mathbf{NA}_i, \Delta\epsilon_i|\text{PM}_{i,pre}) = 0$. 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 $\epsilon_{it} = \epsilon_{it-1} + \mu_{it}$

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.

[29]That is $\text{PM2.5}_{i,t} = \beta\mathbf{NA}_{it} + \gamma\lambda_t\text{PM}_{i,pre} + \delta_i + \lambda_t + \epsilon_{i,t}$ in levels is $\Delta\text{PM}_i = \alpha + \beta\mathbf{NA}_i + \gamma\text{PM}_{i,pre} + \Delta\epsilon_i$ in differences.

where μ_{it} is uncorrelated with PM_{it-1}.[30] This is satisfied if a shock in the pre-treatment period—from, say, new industrial units or infrastructure projects—persists through the post-treatment period (when a new shock can arrive). On the other hand, if a shock in the pre-treatment period is only transitory—from, say, unusual weather conditions in a given year—$\Delta\epsilon_i$ would be correlated with $\text{PM}_{i,pre}$, introducing bias into equation (3). To mitigate such bias from transitory shocks, we use three-year averages of $\text{PM}_{2.5}$ in both the pre- and post-treatment periods such that transitory shocks like weather are unlikely to be captured. We also demonstrate in Appendix Table A.3 that results remain unchanged when using higher-order interactions with baseline $\text{PM}_{2.5}$ 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 $\text{PM}_{2.5}$, allowing us to match nonattainment tracts to attainment tracts with similar baseline $\text{PM}_{2.5}$. The map in Figure 2b illustrates this for the region around Indianapolis, showing that there are tracts with low and high baseline $\text{PM}_{2.5}$ in both nonattainment and attainment areas. Figure 5 shows the overlap in the distributions of baseline $\text{PM}_{2.5}$ (2001-2003), plotting $-\Delta\text{PM}$ against $\text{PM}_{i,pre}$.[31] The density plots in Figure 5 show that there are tracts in attainment areas with average $\text{PM}_{2.5}$ values above the EPA threshold of 15, likely because the EPA air pollution ground monitor network has incomplete coverage (Sullivan & Krupnick 2018). There are also many tracts in nonattainment areas with baseline $\text{PM}_{2.5}$ 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 pre-treatment pollution $\text{PM}_{i,pre}$ alone, which we call M1DiD. In a second version, which we

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

[31] 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 1975-1992. 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 & 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 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 1980-1990. Sanders et al. (2020) match on baseline population and mortality to control for trends in mortality.

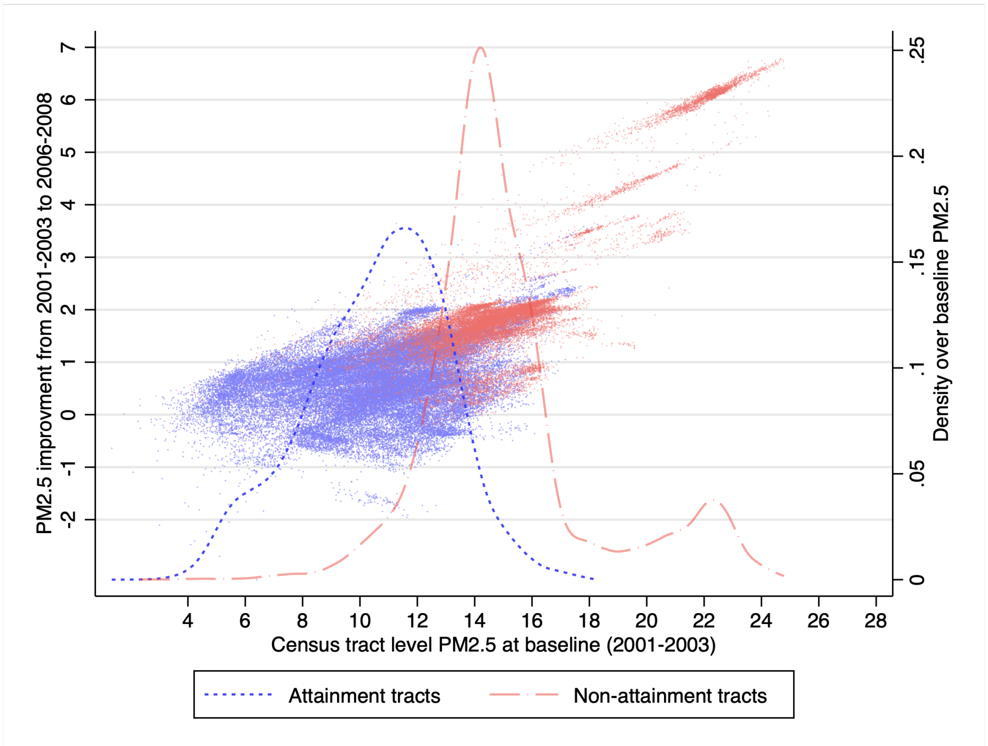


Figure 5: Improvement in tract $PM_{2.5}$ averages and baseline $PM_{2.5}$ levels

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. (2019*b*).

call M2DiD, we additionally match on pre-treatment tract population and population density (both based on the 2000 Census). For both M1DiD and M2DiD, we impose a common support 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 Appendix Table A.4. We use these matching weights to weight our DiD regression[34], equivalent to:

$$\Delta \text{PM}_i \sqrt{W_i} = \alpha \sqrt{W_i} + \beta \Delta \mathbf{NA}_i \sqrt{W_i} + \Delta \xi_i \sqrt{W_i} \tag{4}$$

Our identifying assumption now becomes that nonattainment and their propensity-matched at-

[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.

tainment areas would have experienced the same average change in $PM_{2.5}$ over time absent the regulation, i.e. $\text{cov}(\Delta\mathbf{NA}_i\sqrt{W_i}, \Delta\xi_i\sqrt{W_i}) = 0$. 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 $\Delta\mathbf{NA}_i$ and $\text{PM}_{i,pre}$ is 0.64, but only -0.08 between $\Delta\mathbf{NA}_i\sqrt{W_i}$ and $\text{PM}_{i,pre}\sqrt{W_i}$. We show that the event study graph in 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 Appendix Figure A.6 and 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 et al. (2014) and Calonico et al. (2020).[36]

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]:

$$\Delta\text{PM}_i = \alpha + \beta\Delta\mathbf{NA}_i + \lambda\text{RV}_i + \Delta\xi_i \tag{5}$$

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 require-

[35]In an early example, Chay & 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 & 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.

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

ment 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 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 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 et al. 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 et al. (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 et al. (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.

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:

$$\Delta \text{PM}_i = \alpha + \beta_1 \Delta \mathbf{NA}_i + \beta_2 \Delta \mathbf{NA}_i \text{PM}_{i,pre} + \Delta \xi_i \tag{6}$$

Treatment β_i therefore varies along the dimension of pre-treatment pollution, or $\beta_i = \beta_1 + \beta_2 \text{PM}_{i,pre}$.[38]

[38]Note that $\text{PM}_{i,pre}$ is absorbed in the fixed effect in the equation in levels from which the above equation has been derived. That is: $\text{PM2.5}_{i,t} = \beta_1 \mathbf{NA}_{it} + \beta_2 \mathbf{NA}_{it} \text{PM}_{i,pre} + \delta_i + \lambda_t + \xi_{i,t}$, where the uninteracted effect $\text{PM}_{i,pre}$ is co-linear with fixed effect δ_i.

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. 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 (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:

$$\Delta \text{PM}_i = \alpha + \beta_1 \Delta \textbf{NA2005}_i (1 - \text{NA1990}_i) + \beta_2 \Delta \textbf{NA2005}_i \text{NA1990}_i + \beta_3 \text{NA1990}_i + \Delta \xi_i \quad (7)$$

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 et al. 2023), as it implicitly assumes that $\beta_2 = 0$.

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 2001-03. Our post-treatment periods are five (2006-08) and ten (2011-13) 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 at-

Table 1: Nonattainment status and changes in $PM_{2.5}$

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(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	72043	72043	28291	28909	47962	7026	10459
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	49357	49357	20388	20127	25276	2143	5411
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(x)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	72043	72043	28291	28909	47962	7026	10459
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	72043	72043	28291	28909	47962	6137	25856
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	49357	49357	20388	20127	25276	1046	4626
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(x)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	72043	72043	28291	28909	47962	6137	25856
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 post-treatment periods. Each panel(x)column combination is from a separate regression as described in the text. (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists, (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. (2019*b*).

tainment 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 re-estimate 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 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% decrease from average concentrations in the pre-treatment period.

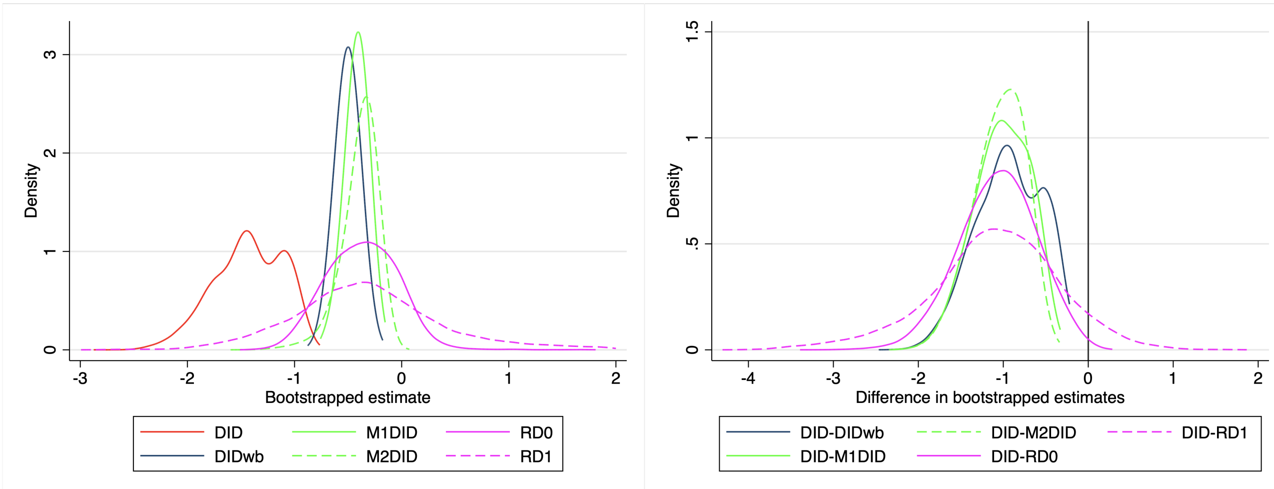
Columns 6 and 7 show results for our regression discontinuity approaches RD0 and RD1.[42] The point estimate for RD0 is similar to our other strategies, and close to zero for RD1, but both estimates are imprecise. Due to the smaller number of observations around the cutoff, we lack statistical power resulting in larger standard errors. However, effect estimates in both RD0 and RD1 are highly statis-

[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 Appendix Table A.2, where we re-estimate 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.

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



(a) Bootstrapped estimates

(b) Difference of coefficients across bootstraps

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. 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. (2019*b*).

tically 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 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 attainment and nonattainment strata. Figure 6a 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 6b 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 Appendix Table A.5. All estimates from our alternative models are significantly different from the DiD estimates at the 1% level except the RD1 model due to noisier estimates (see Figure 6a). Conversely, 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 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. 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 Figure A.14), as we show in 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 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 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 non-classical, 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. (2021*b*) that also quantifies the uncertainty for each data point from the underlying reanalysis model and raw data. We drop the 30% of data points with the highest uncertainty and re-estimate 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 (2022*a*). 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 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 pre-treatment years) to minimize the mean difference in time trends between treated and control groups. 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 10th percentile of baseline pollution levels, while it is 1.3 $\mu g/m^3$ at the 90th percentile.[43] The heterogeneous treatment effects are in line with previous findings by Auffhammer et al. (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) on nonattainment and the interaction are also

[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 10th and 90th 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 vs 20), there is substantial variation in tract level pollution even within the county-based window.

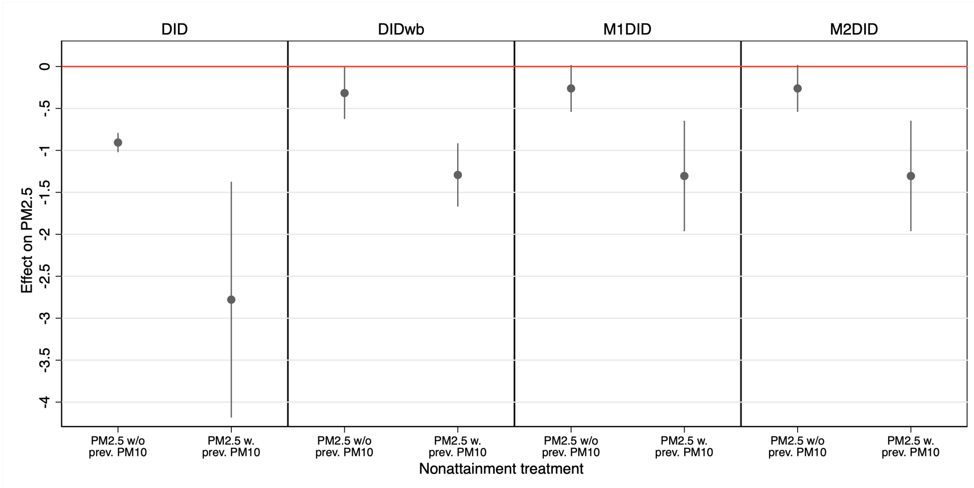


Figure 7: Heterogeneous $PM_{2.5}$ nonattainment treatment effect by previous PM_{10} nonattainment status

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% level for each model. Based on data from Meng et al. (2019*b*).

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 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% level (Appendix Table A.14 shows results including the RD models omitted here because they do not always include areas with both treatment groups).[44]

Importantly, these results show that not only areas that switched from previous PM_{10} *attainment*

[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.

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 et al. (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 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 (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 US 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 et al. 2023).[47]

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

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

Currie et al. (2023) show that this Black-White $PM_{2.5}$ gap fell by 0.6 $\mu g/m^3$ between 2005 and 2015, and that a substantial portion (61.2%) 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% to that narrowing, or 64% 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%. 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 9-26% for homogeneous treatment effects, 14-47% with heterogeneous effects, and 18-49% 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. (2019*b*) instead of Di et al. (2021), our estimated CAA contribution based on standard DiD with heterogeneous effects (68%) until 2011-2013 is broadly in line with the findings in Currie et al. (2023) of a contribution of 61.2% from 2005-2015. As shown in Appendix A.13A., when we follow the approach of Currie et al. (2023) based on RIF/Quantile Regressions and their treatment assignment, we recover an almost identical 61.1% contribution.[49] However, as we show in Appendix A.13B., controlling for confounding trends (i.e. DiDwb) in their approach also reduces the CAA contribution to 18.6% percent (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% based on DiDwb much in line with our estimated 24% using the same data based on Di et al. (2021) in 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

In 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 Appendix A.13A. for Currie et al. (2023).

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

[49]In Appendix A.13B. we also show that other minor data differences to Currie et al. (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.

Table 2: Pollution Disparities - Counterfactual Gap Analysis

Panel (a): Black-White Pollution Gap

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [homogeneous effect]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
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

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [heterogeneous effect]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
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

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [+race interactions]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
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

Panel (b): Urban-Rural Pollution Gap

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [homogeneous effect]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
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

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [heterogeneous effect]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
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

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [+urban interactions]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
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. (2019*b*).

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%. 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%) while the other approaches yield significantly smaller estimates (13-63%).

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 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 Appendix Table A.20 that the results hold when population is fixed at 2010 levels switching off any population sorting channels. 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:

$$\Delta \mathrm{Y}_i = \alpha + \theta \Delta \mathrm{PM}_i + \Delta \mu_i \tag{8}$$

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 $\Delta \mathrm{PM}_i$ 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 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

[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 Appendix Figure A.20.

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 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 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% $(\exp(-0.04) - 1)$. Instrumenting $PM_{2.5}$ with nonattainment status corresponding to the simple DiD approach in Column 2 shows an effect that is larger implying a semi-elasticity of around 6%. 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).

The IV estimates based on our three alternative approaches yield larger pollution damages, around 50% to 150% larger than those based on the standard DiD-IV. Our preferred approach for

[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.

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

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	54529	54529	54529	21152	21693	5087	7937
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	54378	54378	54378	21062	21608	4496	19035
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. (2019*b*).

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%. 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 post-treatment 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 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% on average due to being designated into nonattainment.[55]

[54] The implied elasticity of -1.4 is larger than the elasticity of -0.6 in Bento et al. (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 & 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% larger than the elasticity based on simple DiD-IV. Graff Zivin & 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 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% per $\mu g/m^3$ from Table 3, which yields an increase of around 7% ($\exp(-0.55 * -.12) - 1$).

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 et al. 2015, Auffhammer et al. 2009). To gauge the issue of confounding trends for this older regulation, we use the historic $PM_{2.5}$ data from Meng et al. (2019*b*) going back to the 1980s together with the 1990 PM_{10} nonattainment areas.[56] First, 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 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_2 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 III.C., 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. 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 2022*b*) 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. 2019*b*), Figure 8 Panels (a), (b) and (c) show

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

[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 Deschenes et al. (2017), see Appendix A.8.

Table 4: The effect of 1990 PM_{10} nonattainment designation on $PM_{2.5}$ concentrations

	DiD	DiDwb	M1DiD	M2DiD
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	72043	72043	20174	22094

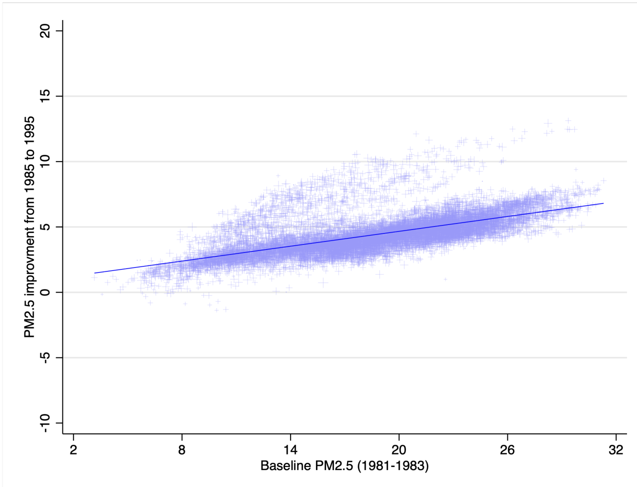
Notes: The table shows coefficient estimates for the treatment effect of nonattainment status with the 1990 PM_{10} NAAQS (instead of the 2005 $PM_{2.5}$ NAAQS) on the change in $PM_{2.5}$ levels between the pre- and post-treatment periods of 1987-89 and 1991-93 respectively. Each column is from a separate regression, where (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (1987-89), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (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. (2019*b*).

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 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]

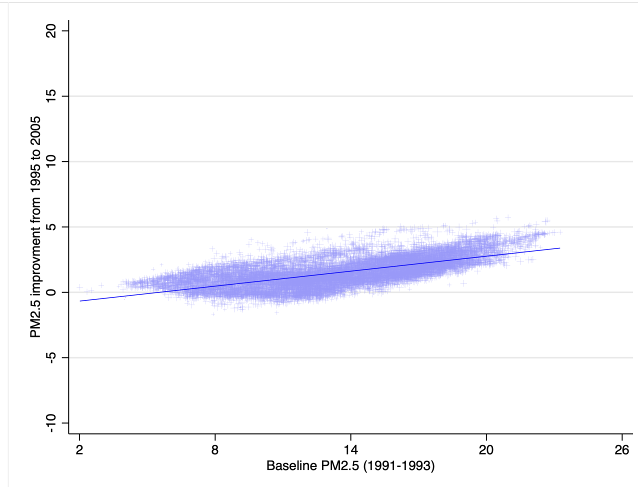
The issue of correlated trends is often not accounted for in the literature. There are few exceptions that address possible confounding trends which, however, have no explicit discussion of bias (Greenstone 2004, Chay & Greenstone 2005, Auffhammer et al. 2009, Bishop et al. 2023). Greenstone (2004) controls for and matches on baseline levels for analyzing the 1970s SO_2 regulation. Chay & Greenstone (2005) use a variant of regression discontinuity with manual window selection to study TSP rules in the 1970s-80s (see also Sanders & Stoecker (2015). Auffhammer et al. (2009) include monitor-specific time trends in their analysis of the 1990 rules for PM_{10}, and Bishop et al. (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 et al. 2017, Sanders et al. 2020, Colmer & Voorheis 2021, Colmer et al. 2022, Hollingsworth et al. 2022, Currie et al. 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

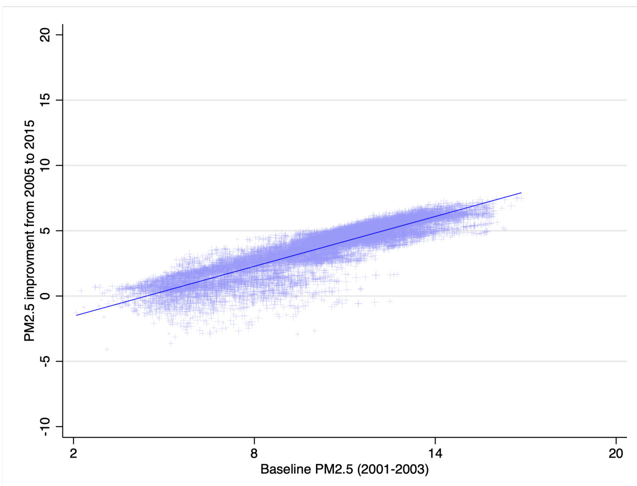
[59] Colmer et al. (2020) show a convergence of pollution concentrations, but for the entire US, not just the 'never treated' group.



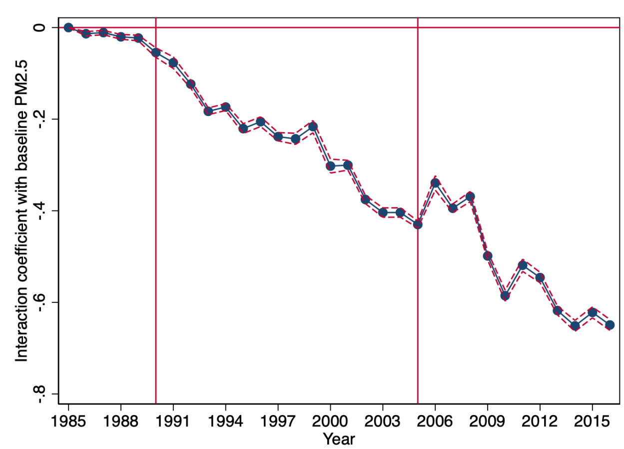
(a) 1981-83 Baseline vs. 1985-1995 Change



(b) 1991-93 Baseline vs. 1995-2005 Change



(c) 2001-03 Baseline vs. 2005-2015 Change



(d) Interaction of 1981-83 baseline with years

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% confidence bands based on standard errors clustered at the county level. The figure is based on data from Meng et al. (2019*b*).

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 & 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, S., Dwyer, H., Graff Zivin, J. & Neidell, M. (2022), 'This is air: The "nonhealth" effects of air pollution', *Annual Review of Resource Economics* **14**, 403–425.

Aldy, J. E., Auffhammer, M., Cropper, M., Fraas, A. & Morgenstern, R. (2022), 'Looking back at 50 years of the clean air act', *Journal of Economic Literature* **60**(1), 179–232.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W. & Wager, S. (2021), 'Synthetic difference-in-differences', *American Economic Review* **111**(12), 4088–4118.

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

Banzhaf, S., Ma, L. & Timmins, C. (2019), 'Environmental justice: The economics of race, place, and pollution', *Journal of Economic Perspectives* **33**(1), 185–208.

Bento, A., Freedman, M. & Lang, C. (2015), 'Who benefits from environmental regulation? evidence from the clean air act amendments', *Review of Economics and Statistics* **97**(3), 610–622.

Bishop, K. C., Ketcham, J. D. & Kuminoff, N. V. (2023), 'Hazed and Confused: The Effect of Air Pollution on Dementia', *The Review of Economic Studies* **00**(00), 1–27.

Bogin, A., Doerner, W. & Larson, W. (2019), 'Local house price dynamics: New indices and stylized facts', *Real Estate Economics* **47**(2), 365–398.

Calonico, S., Cattaneo, M. D. & Farrell, M. H. (2020), 'Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs', *The Econometrics Journal* **23**(2), 192–210.

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014), 'Robust nonparametric confidence intervals for regression-discontinuity designs', *Econometrica* **82**(6), 2295–2326.

Cattaneo, M. D., Frandsen, B. R. & Titiunik, R. (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, K. Y. & Greenstone, M. (2005), 'Does air quality matter? evidence from the housing market', *Journal of Political Economy* **113(2)**, 376–424.

Chetty, R. & Friedman, J. N. (2019), 'Replication data for: A practical method to reduce privacy loss when disclosing statistics based on small samples', American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor], `http://doi.org/10.3886/E116494V1`.

Clay, K., Jha, A., Lewis, J. A. & Severnini, E. R. (2021), 'Impacts of the clean air act on the power sector from 1938-1994: Anticipation and adaptation', *NBER working paper* (28962).

Colmer, J., Hardman, I., Shimshack, J. & Voorheis, J. (2020), 'Disparities in PM2.5 air pollution in the united states', *Science* **369**(6503), 575–578.

Colmer, J. & Voorheis, J. (2021), 'The grandkids aren't alright: the intergenerational effects of prenatal pollution exposure', *Mimeo* .

Colmer, J., Voorheis, J. & Williams, B. (2022), 'Air pollution and economic opportunity in the united states', *Mimeo* .

Currie, J., Voorheis, J. & Walker, R. (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.

Currie, J. & Walker, R. (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.

Deschenes, O., Greenstone, M. & Shapiro, J. S. (2017), 'Defensive investments and the demand for air quality: Evidence from the NOx budget program', *American Economic Review* **107**(10), 2958–2989.

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M. B., Choirat, C., Koutrakis, P., Lyapustin, A. 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, Q., Wei, Y., Shtein, A., Hultquist, C., Xing, X., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J. et al. (2021), 'Daily and annual PM2.5 concentrations for the contiguous United States, 1-km grids, v1 (2000–2016) [dataset]', Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC), `https://doi.org/10.7927/0rvr-4538`, accessed February 2, 2022.

Drupp, M. A., Kornek, U., Meya, J. & Sager, L. (2021), Inequality and the environment: The economics of a two-headed hydra, CESifo Working Paper 9447, Center for Economic Studies and the ifo Institute.

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

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

EPA (2005*a*), 'Air quality designations and classifications for the fine particles (PM2.5) national ambient air quality standards; final rule', *Federal Register* **70**(3).

EPA (2005*b*), 'Air quality designations for the fine particles (PM2.5) national ambient air quality standards – supplemental amendments; final rule', *Federal Register* **70**(3).

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

EPA (2016), 'Fine particulate matter national ambient air quality standards: State implementation plan requirements; final rule', *Federal Register* **81**(164).

EPA (2018*a*), 'County PM2.5 24h design values, 1999-2017 [dataset]', U.S. Environmental Protection Agency (EPA), `https://www.epa.gov/sites/default/files/2020-03/county_level_pm2.5_24hr_design_values_1999_2017.xlsx`, accessed February 1, 2022.

EPA (2018*b*), 'County PM2.5 annual design values, 1999-2017 [dataset]', U.S. Environmental Protection Agency (EPA), `https://www.epa.gov/sites/default/files/2020-03/county_level_pm2.5_annual_design_values_1999_2017.xlsx`, accessed February 1, 2022.

EPA (2021), *Green Book. PM-2.5 (1997) Designated Area Design Values*, U.S. Environmental Protection Agency.

EPA (2022*a*), *AirData Annual Summary Data*, U.S. Environmental Protection Agency. Available via https://www.epa.gov/outdoor-air-quality-data. Accessed February 1, 2022.

EPA (2022*b*), *Green Book. Green Book National Area and County-Level Multi-Pollutant Information*, U.S. Environmental Protection Agency. Available from EPA's Air Markets Program Data web site: https://www.epa.gov/green-book. Accessed February 1, 2022.

FHFA (2021), 'Annual house price indexes, census tracts (version: February 23, 2021) [dataset]', Federal Housing Finance Agency (FHFA), `https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_AT_BDL_tract.csv`, accessed February 1, 2022.

Fowlie, M., Rubin, E. & Walker, W. R. (2019), 'Bringing satellite-based air quality estimates down to earth', *AEA Papers and Proceedings* **109**, 283–88.

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

Graff Zivin, J. & Neidell, M. (2012), 'The impact of pollution on worker productivity', *American Economic Review* **102**(7), 3652–73.

Graff Zivin, J. & Singer, G. (2023), 'Disparities in pollution capitalization rates: The role of direct and systemic discrimination', *NBER working paper* (30814).

Grainger, C. A. (2012), 'The distributional effects of pollution regulations: Do renters fully pay for cleaner air?', *Journal of Public Economics* **96**, 840–852.

Greenstone, M. (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, M. (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. V. (1996), 'Effects of air quality regulation', *American Economic Review* **86(4)**, 789–813.

Hollingsworth, A., Jaworski, T., Kitchens, C. & Rudik, I. J. (2022), Economic geography and the efficiency of environmental regulation, Technical report, National Bureau of Economic Research.

Isen, A., Rossin-Slater, M. & Walker, W. R. (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, A., Zhou, X., Liu, J., Lee, T.-H., Kamareddine, L., Verguet, S. & Dominici, F. (2022), 'Air pollution exposure disparities across us population and income groups', *Nature* **601**(7892), 228–233.

Jha, A., Matthews, P. H. & Muller, N. Z. (2019), 'Does environmental policy affect income inequality? evidence from the clean air act', *AEA Papers and Proceedings* **109**, 271–76.

Kahn, M. & Mansur, E. T. (2013), 'Do local energy prices and regulation affect the geographic concentration of employment?', *Journal of Public Economics* **101**, 105–114.

Landrigan, P. J., Fuller, R., Acosta, N. J., Adeyi, O., Arnold, R., Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J. I., Breysse, P. N. et al. (2018), 'The lancet commission on pollution and health', *The lancet* **391**(10119), 462–512.

List, J. A., Millimet, D. L., Fredriksson, P. G. & McHone, W. W. (2003), 'Effects of environmental regulations on manufacturing plant births: evidence from a propensity score matching estimator', *Review of Economics and Statistics* **85**(4), 944–952.

Manson, S., Schroeder, J., Van Riper, D., Kugler, T. & Ruggles, S. (2022), 'Ipums national historical geographic information system: Version 17.0 [dataset]', `http://doi.org/10.18128/D050.V17.0`, accessed February 2, 2022.

McCrary, J. (2008), 'Manipulation of the running variable in the regression discontinuity design: A density test', *Journal of econometrics* **142**(2), 698–714.

Meng, J., Li, C., Martin, R. V., van Donkelaar, A., Hystad, P. & Brauer, M. (2019*a*), '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', *Environmental science & technology* **53**(9), 5071–5079.

Meng, J., Li, C., Martin, R. V., van Donkelaar, A., Hystad, P. & Brauer, M. (2019*b*), 'Historical PM2.5 across North America [dataset]', 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.

Sanders, N. J., Barreca, A. I. & Neidell, M. J. (2020), 'Estimating causal effects of particulate matter regulation on mortality', *Epidemiology (Cambridge, Mass.)* **31**(2), 160.

Sanders, N. J. & Stoecker, C. (2015), 'Where have all the young men gone? using sex ratios to measure fetal death rates', *Journal of health economics* **41**, 30–45.

Shapiro, J. S. & Walker, R. (2018), 'Why is pollution from us manufacturing declining? the roles of environmental regulation, productivity, and trade', *American Economic Review* **108**(12), 3814–54.

Sullivan, D. M. & Krupnick, A. (2018), 'Using satellite data to fill the gaps in the us air pollution monitoring network', *Resources for the Future Working Paper* pp. 18–21.

van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., Hsu, N. C., Kalashnikova, O. V., Kahn, R. A., Lee, C. et al. (2021*a*), 'Monthly global estimates of fine particulate matter and their uncertainty', *Environmental Science & Technology* **55**(22), 15287–15300.

van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., Hsu, N. C., Kalashnikova, O. V., Kahn, R. A., Lee, C. et al. (2021*b*), 'Surface pm2.5 data (v5.gl.03) [dataset]', Washington University of St. Louis Atmospheric Composition Analysis Group, `https://sites.wustl.edu/acag/datasets/surface-pm2-5/`, accessed February 2, 2022.

Walker, W. R. (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.

APPENDIX FOR ONLINE PUBLICATION

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

by Lutz Sager[1,*], Gregor Singer[2,*]

[1]Georgetown University

[2]London School of Economics

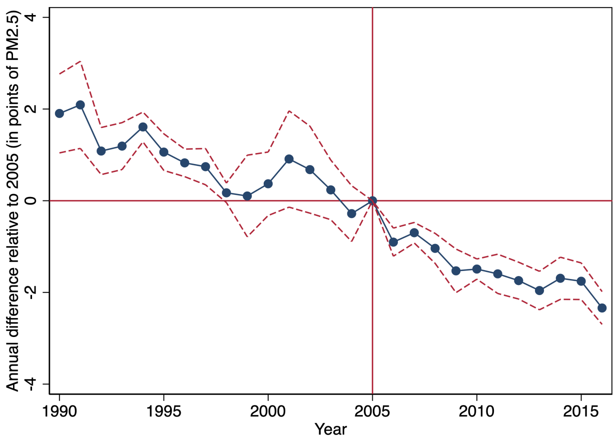
[*]Correspondence to: Lutz Sager: Lutz.Sager@georgetown.edu, Gregor Singer: g.a.singer@lse.ac.uk

A.1 Descriptive Statistics & Graphs

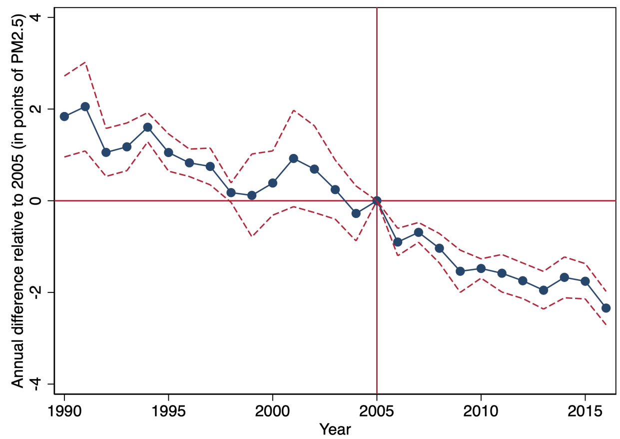
Table A.1: Descriptive Statistics

	mean	sd	min	max
Panel (a): Baseline Period (2001-2003)				
PM25 (Meng et al.)	11.98	3.33	1.35	24.79
PM25 (Di et al.)	12.58	3.31	2.51	29.58
PM25 (Van Donkelaar et al.)	12.19	3.23	3.68	27.30
Observations	72043			
Panel (b): Five Year Post Period (2006-2008)				
PM25 (Meng et al.)	10.91	2.57	1.75	18.57
PM25 (Di et al.)	11.12	2.36	2.19	21.18
PM25 (Van Donkelaar et al.)	10.99	2.50	3.67	22.45
Observations	72043			
Panel (c): Ten Year Post Period (2011-2013)				
PM25 (Meng et al.)	8.99	1.96	1.22	16.96
PM25 (Di et al.)	9.20	1.80	1.91	18.63
PM25 (Van Donkelaar et al.)	8.99	1.74	3.52	18.19
Observations	72043			

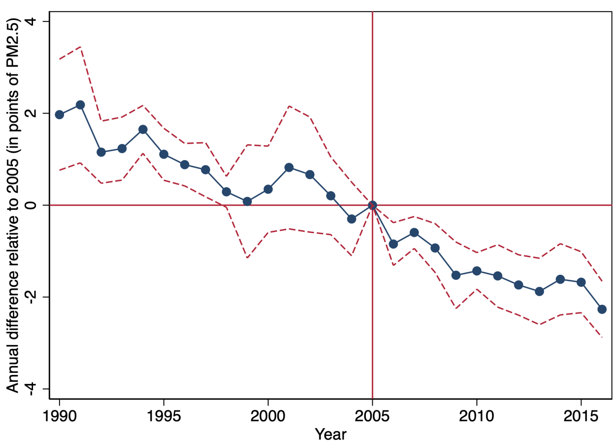
Notes: Tract level summary statistics, averaged over the respective 3-year periods, weighted by population weights accounting for population differences within tracts as well as across tracts. Pollution data is from Meng et al. (2019*b*), Di et al. (2021) and van Donkelaar et al. (2021*b*).



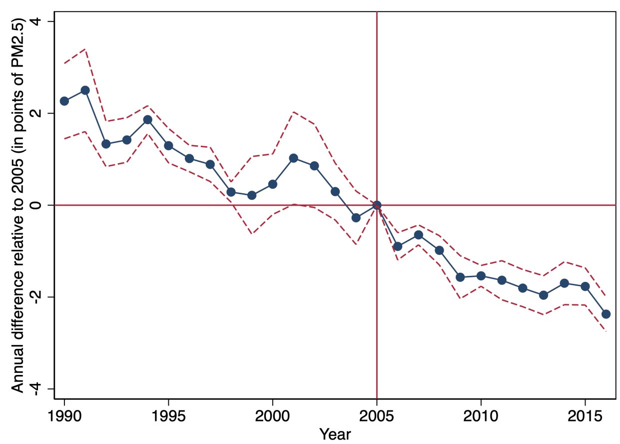
(a) 2000 block borders and block level analysis



(b) Interpolated population weights 2000, 2010, 2020



(c) Nonattainment based on commuting zones



(d) Dropping neighboring controls

Figure A.1: Robustness of differences in pre-trends in event study

Notes: The figure replicates the event study graph from Panel (b) of Figure 3. Panel (a) uses borders and population counts from the 2000 Census instead of the 2010 Census. In addition, the analysis is at the Census block level, rather than pre-aggregating to the Census tract level using Census block weights as in our main analysis (the results are equivalent using either). Panel (b) uses population weights that are interpolated between the 2000, 2010 and 2020 Census, using the IPUMS NHGIS crosswalk, instead of constant population weights at the 2010 level. Panel (c) assigns all counties in a commuting zone into nonattainment, as long as a single county in that commuting zone is in nonattainment, resulting in 428 nonattainment counties compared to the 208 actual nonattainment counties based on EPA air regions. Panel (d) drops all attainment counties that border a nonattainment county allowing for possible spatial spillovers (dropping 300 counties). All results based on Meng et al. (2019*b*), but the same patterns hold for data from Di et al. (2021) or van Donkelaar et al. (2021*b*). Standard errors are clustered at the county level except for Panel (c) where we cluster at the commuting zone level.

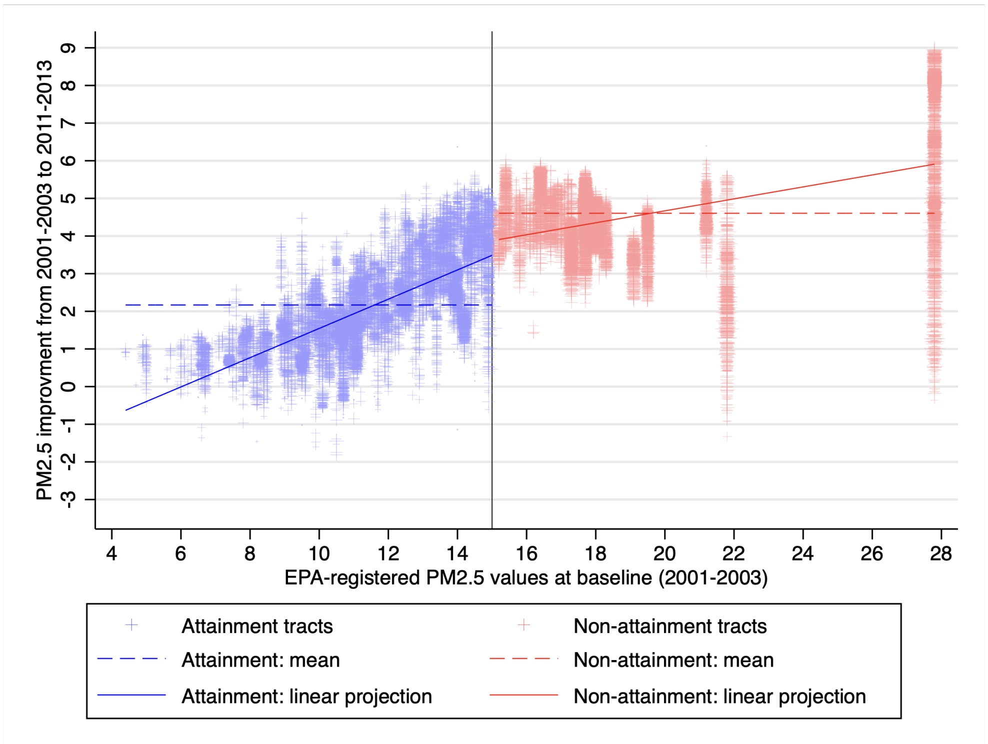


Figure A.2: Ten year improvement in tract $PM_{2.5}$ averages and EPA-registered $PM_{2.5}$ values

Notes: The figure shows the improvement in $PM_{2.5}$ averages at the tract level between two periods, 2001-2003 and 2011-2013. 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. The solid lines plot the linear projection of tract level $PM_{2.5}$ improvements on the EPA-registered $PM_{2.5}$ values of the nonattainment and attainment areas separately, weighted by tract population. Based on data from Meng et al. (2019*b*).

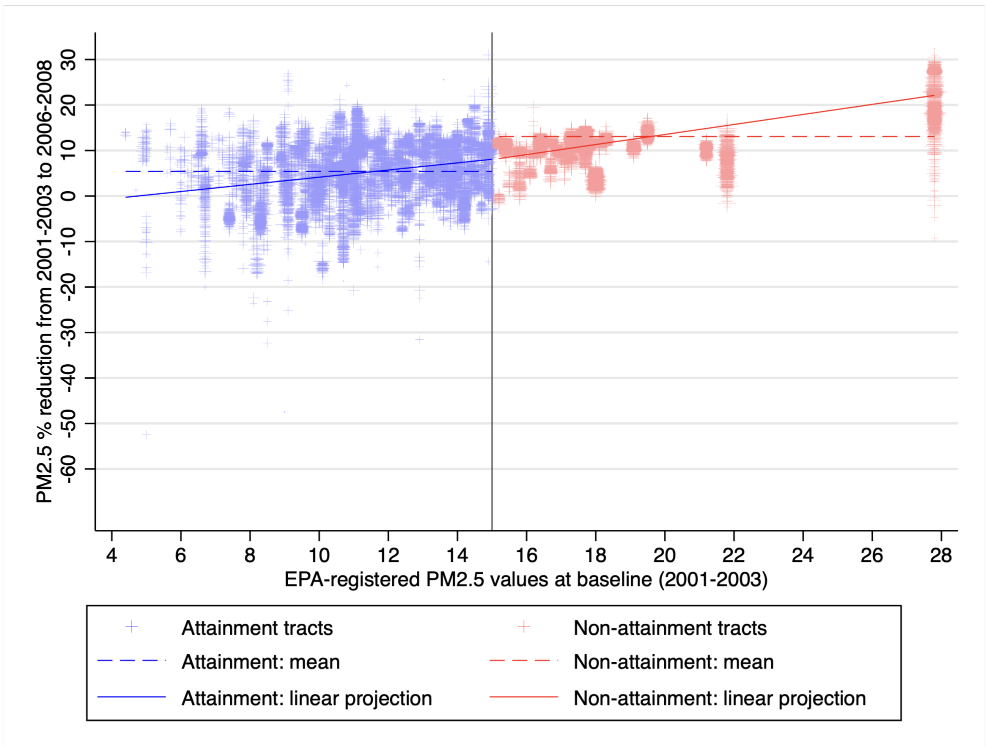


Figure A.3: Relative improvement in tract $PM_{2.5}$ averages and EPA-registered $PM_{2.5}$ values

Notes: The figure replicates Figure 4 but converts the vertical axis to percentage changes. It shows the improvement in $PM_{2.5}$ averages at the tract level between two periods, 2001-2003 and 2006-2008, expressed in percent of the 2001-2003 values. 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. The solid lines plot the linear projection of tract level $PM_{2.5}$ improvements on the EPA-registered $PM_{2.5}$ values of the nonattainment and attainment areas separately, weighted by tract population. Based on data from Meng et al. (2019*b*).

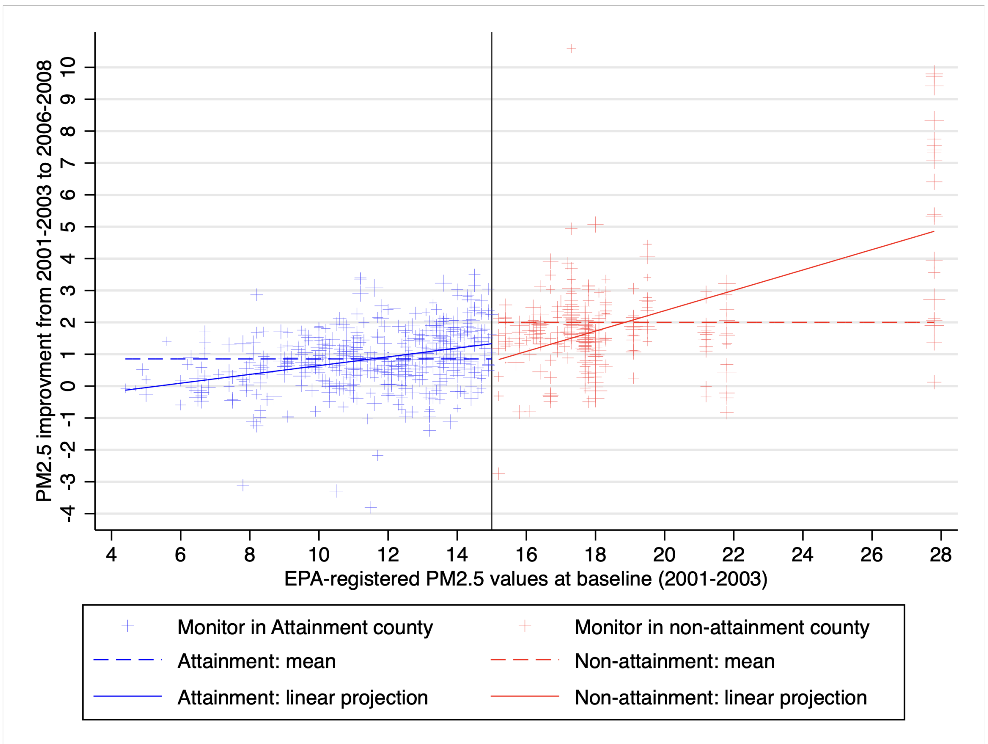


Figure A.4: Improvement in EPA monitor $PM_{2.5}$ averages and EPA-registered $PM_{2.5}$ values

Notes: The figure shows the improvement in $PM_{2.5}$ averages at the EPA monitor level between two periods, 2001-2003 and 2006-2008, taking the average $PM_{2.5}$ for each monitor. The size of the markers reflect tract level populations in which the monitor is situated. 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 monitors in nonattainment and attainment areas separately, weighted by tract population. The solid lines plot the linear projection of monitor level $PM_{2.5}$ improvements on the EPA-registered $PM_{2.5}$ values of the nonattainment and attainment areas separately, weighted by tract population. Based on data from EPA (2022*a*).

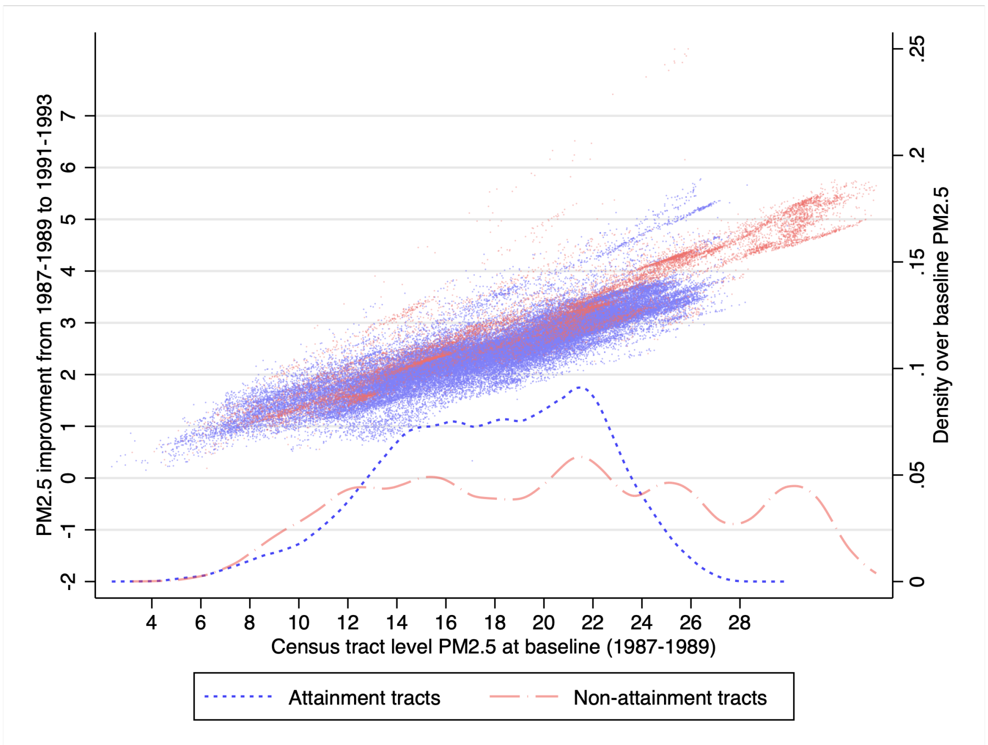


Figure A.5: Improvement in tract $PM_{2.5}$ averages and baseline $PM_{2.5}$ levels for the 1987 PM_{10} rules coming into effect in 1990

Notes: The markers in the figure show the improvement in $PM_{2.5}$ averages at the tract level between two periods, 1987-1989 and 1991-1993. 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, based on the 1987 PM_{10} NAAQS EPA designations. 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. (2019*b*).

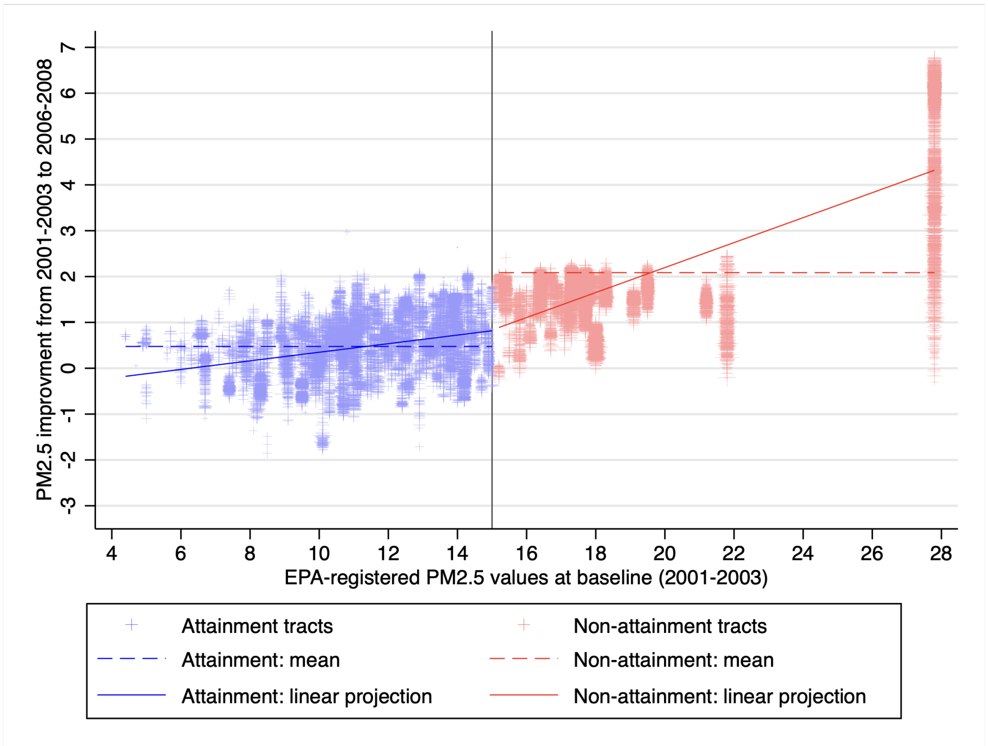


Figure A.6: Improvement in tract $PM_{2.5}$ averages and EPA-registered $PM_{2.5}$ values excluding attainment counties that share a border with a nonattainment county

Notes: The figure shows the improvement in $PM_{2.5}$ averages at the tract level between two periods, 2001-2003 and 2006-2008. Tracts in attainment counties that share a border with a nonattainment county are dropped. 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. The solid lines plot the linear projection of tract level $PM_{2.5}$ improvements on the EPA-registered $PM_{2.5}$ values of the nonattainment and attainment areas separately, weighted by tract population. Based on data from Meng et al. (2019*b*).

A.2 Difference in differences with different bandwidths or subsamples

Table A.2: Difference-in-differences estimates using different bandwidths or sub-samples using Chay & Greenstone (2005) approach.

	All tracts	Tracts with RV: with RV	10-20	13-17	OBW	Binding w/ RV: 13-17	Daily RV: 28-63
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a): Change from 2001-03 to 2006-08							
Nonattainment	-1.47	-1.48	-0.83	-0.64	-0.36	-0.62	-0.72
	(0.34)	(0.35)	(0.093)	(0.18)	(0.28)	(0.22)	(0.094)
Observations	72043	47962	37366	12738	7026	10388	35820
Panel (b): Change from 2001-03 to 2011-13							
Nonattainment	-2.35	-2.44	-1.85	-1.48	-1.26	-1.45	-1.78
	(0.27)	(0.28)	(0.12)	(0.22)	(0.35)	(0.29)	(0.11)
Observations	72043	47962	37366	12738	6137	10388	35820

Notes: The table shows coefficient estimates from a simple difference-in-differences estimation following equation 2. Panel (a) uses average $PM_{2.5}$ across years 2006-2008 as post-treatment outcome. Panel (b) uses average $PM_{2.5}$ across years 2011-2013 as post-treatment outcome. Both use 2001-2003 as pre-treatment period. Column 1 uses full sample of tracts, Column 2 only those tracts for which EPA-registered $PM_{2.5}$ values are available, Column 3 only those tracts in a narrow window of these values around treatment cutoff $(10 < RV < 20)$, Column 4 an even narrower window $(13 < RV < 17)$, and Column 5 an optimal bandwidth as discussed in the section on regression discontinuity. Column 6 is the same as Column 4 but additionally restricts the treated counties to only contain those counties that have the highest EPA pollution readings within each nonattainment area and are therefore the binding counties that assign an area into nonattainment. Column 7 follows a strategy similar to Chay & Greenstone (2005), restricting the sample to areas in attainment of the daily standard and in the overlapping range of daily RV (28-63) shown in Figure 1b. Data from Meng et al. (2019*b*). Standard errors clustered at the county level in parentheses.

A.3 Difference in differences with baseline controls

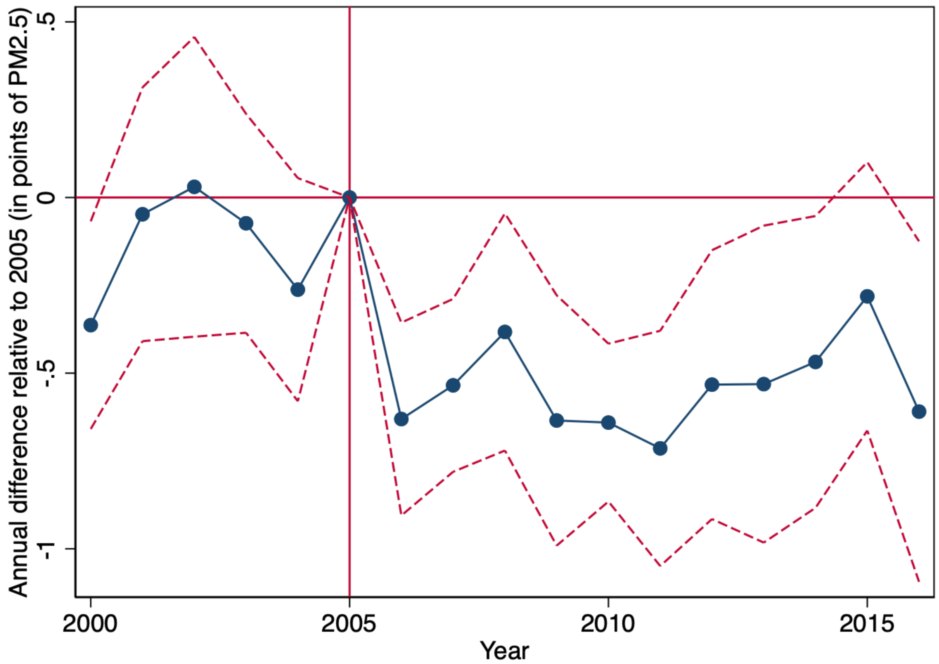


Figure A.7: Event study analysis with controls for baseline pollution

Notes: The Figure replicates the event study graph from Panel (b) of Figure 3 but controls for an interaction between time dummies and baseline pollution equivalent to Column 2 Table 1. All results based on Meng et al. (2019*b*).

Table A.3: Difference-in-differences estimates using different polynomials of baseline $PM_{2.5}$.

	Linear (1)	Quadratic (2)	Cubic (3)	Quartic (4)
Change from 2001-03 to 2006-08				
Nonattainment	-0.49 (0.098)	-0.41 (0.070)	-0.52 (0.072)	-0.51 (0.071)
Observations	72043	72043	72043	72043
Panel (b): Change from 2001-03 to 2011-13				
Nonattainment	-0.56 (0.096)	-0.55 (0.094)	-0.52 (0.096)	-0.53 (0.094)
Observations	72043	72043	72043	72043

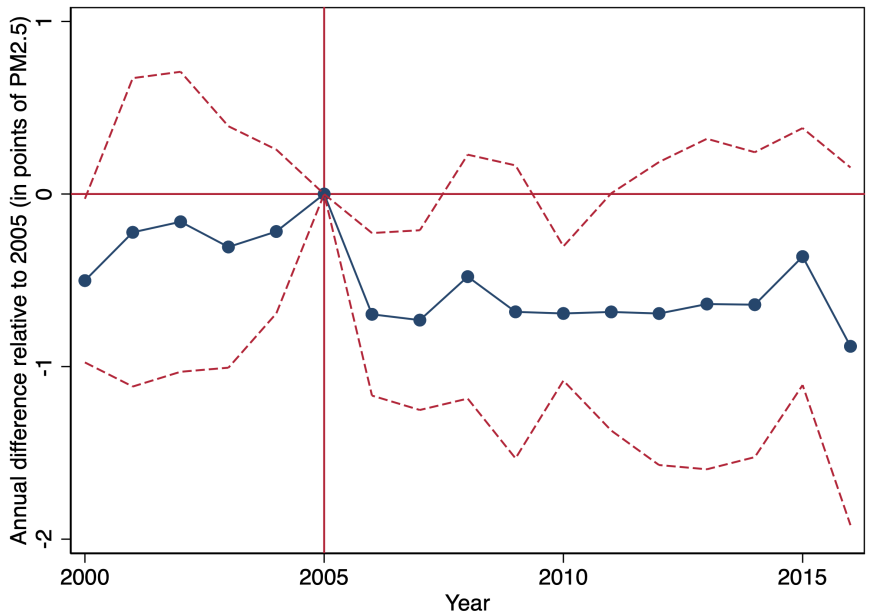
Notes: The table shows coefficient estimates from specifications with control for baseline $PM_{2.5}$ (DiDwb). Column 1 uses linear control and is identical to Column 2 of Table 1. Columns 2, 3 and 4 successively add quadratic, cubic and quartic terms. Data from Meng et al. (2019*b*). Standard errors clustered at the county level in parentheses.

A.4 Matching

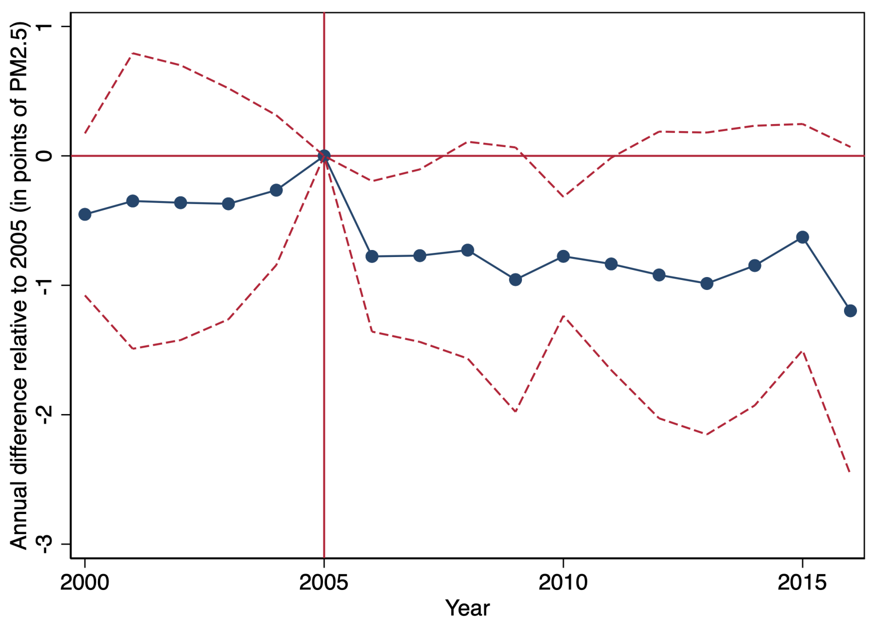
Table A.4: Matched samples - balance tests

	All Tracts		All Tracts with RV	
	unmatched	matched	unmatched	matched
	(1)	(2)	(3)	(4)
M1: Matching on baseline $PM_{2.5}$				
Baseline PM2.5	0.95	-0.064	1.12	-0.063
	(0.20)	(0.27)	(0.25)	(0.31)
Population	182.7	229.0	195.0	267.9
	(88.4)	(111.8)	(102.7)	(137.5)
Pop. Density	7148.4	6582.5	5992.9	5684.0
	(2605.3)	(2645.6)	(2617.7)	(2654.7)
Observations	28291	28291	26647	26647
M2: Matching on baseline $PM_{2.5}$, population, density				
Baseline PM2.5	1.25	0.054	1.15	-0.17
	(0.24)	(0.38)	(0.26)	(0.35)
Population	142.4	113.5	86.4	127.6
	(81.2)	(117.7)	(119.0)	(132.0)
Pop. Density	5530.8	-1821.7	4791.2	3854.1
	(2242.0)	(5803.9)	(2396.8)	(2463.8)
Observations	28909	28909	26637	26637

Notes: The table shows comparisons of average pre-treatment differences before and after our matching procedure. Shown are population-weighted average differences between nonattainment tracts and attainment tracts for baseline $PM_{2.5}$ (2001-03), baseline population (2000), and baseline population density (2000), without using matching weights (*unmatched*) and with using matching weights (*matched*). Matching approach M1 is the same as used in Columns 3 and 7 of Table 1, M2 is the same as in Columns 4 and 8. Standard errors in parentheses are cluster-robust at the level of counties. All results based on Meng et al. (2019*b*).



(a) M1: Matching on baseline $PM_{2.5}$



(b) M2: Matching on baseline $PM_{2.5}$, population, density

Figure A.8: Matched samples - event study analysis

Notes: The figure replicates the event study graph from Panel (b) of Figure 3 with the matched sample and weights underlying the matched difference-in-differences estimation shown in Columns 3 and 4 of Table 1. All results based on Meng et al. (2019*b*).

A.5 Regression discontinuity

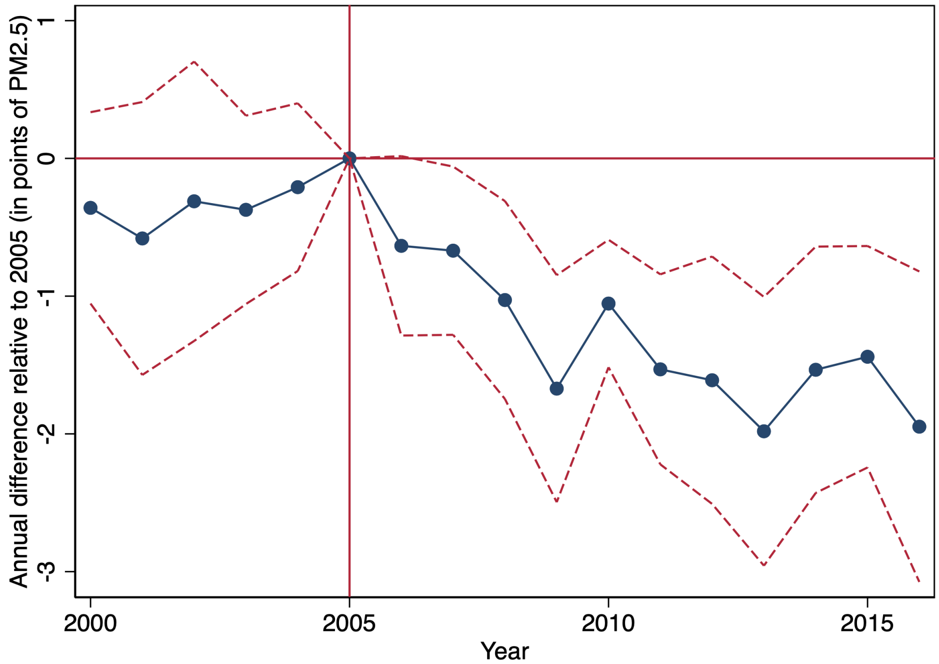
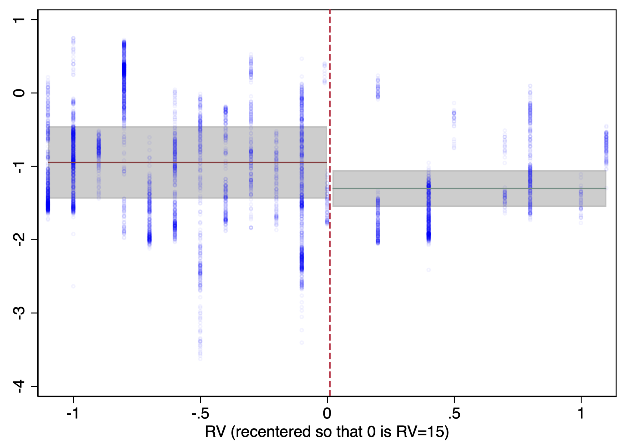
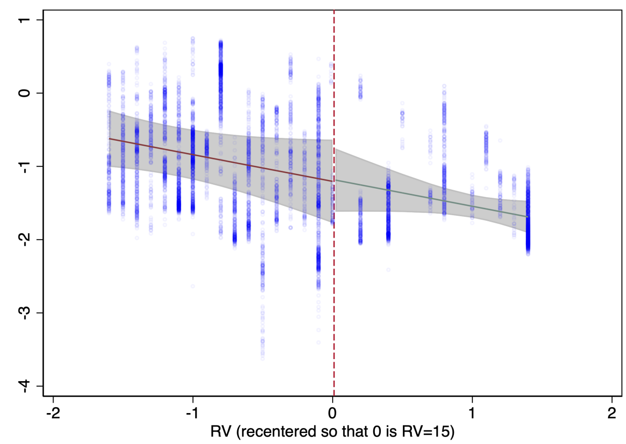


Figure A.9: Regression discontinuity analysis (RD0) - event study analysis

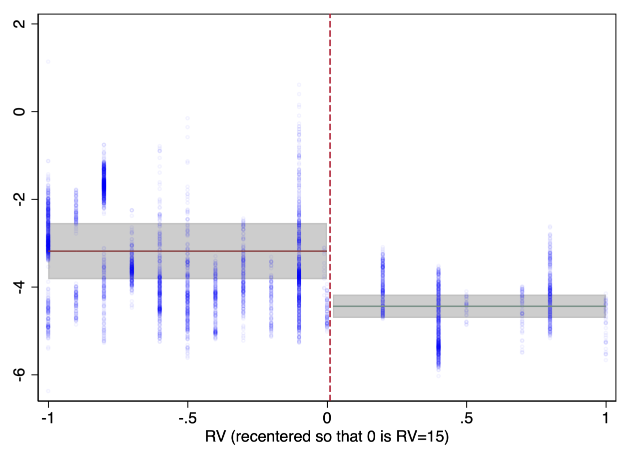
Notes: The figure replicates the event study graph from Panel (b) of Figure 3 with the restricted sample underlying the regression discontinuity estimation shown in Column 6 Table 1. All results based on Meng et al. (2019*b*).



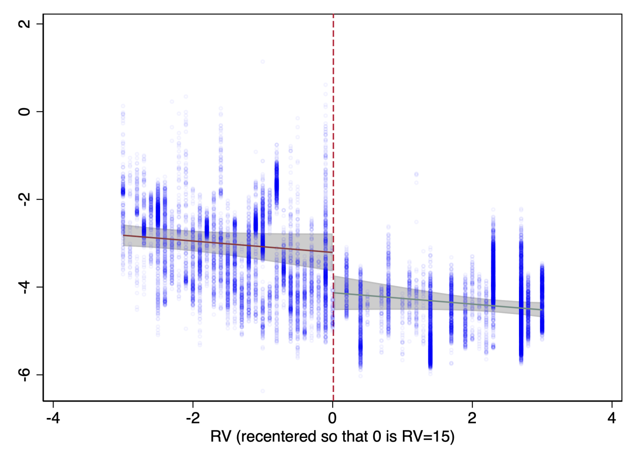
(a) RD0: No slopes (2001-03 vs. 2006-08)



(b) RD1: Linear slopes (2001-03 vs. 2006-08)



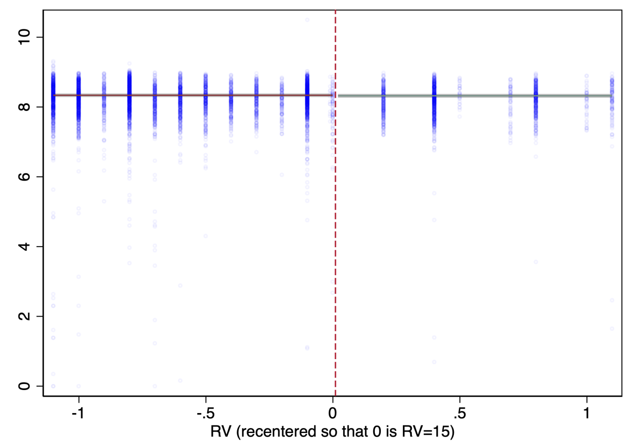
(c) RD0: No slopes (2001-03 vs. 2011-13)



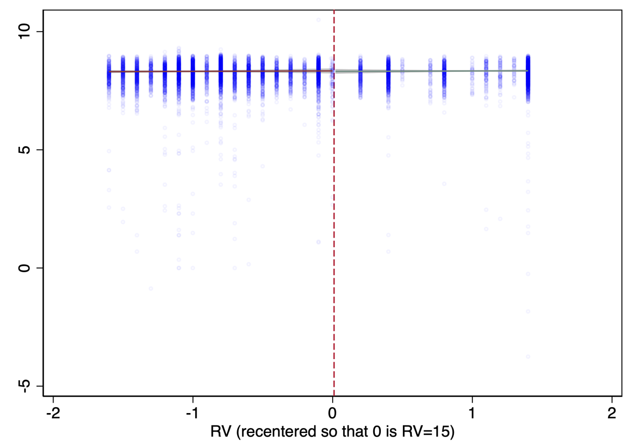
(d) RD1: Linear slopes (2001-03 vs. 2011-13)

Figure A.10: Regression discontinuity – change in $PM_{2.5}$

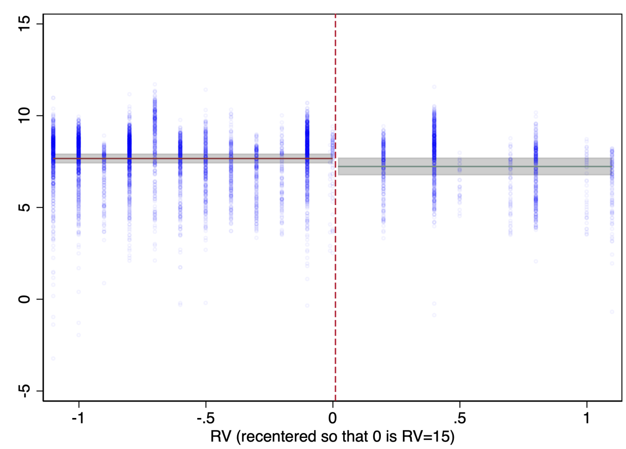
Notes: The figures show visual representations of the regression discontinuity approaches without slopes (RD0) and with linear trends (RD1). Lines are model predictions, shaded areas represent 95% confidence intervals. Horizontal axis shows EPA-registered $PM_{2.5}$ values, re-centered so that 0 is the cutoff (RV=15), and narrowed to the optimal bandwidth. Vertical axis shows the change in $PM_{2.5}$ from 2001-03 to 2006-08 in Panels (a) and (b) or from 2001-03 to 2011-13 in Panels (c) and (d). The jump between the fitted lines at the cutoff is equal to the coefficient estimate in Columns 9 and 10 of Table 1. The figure is based on data from Meng et al. (2019*b*).



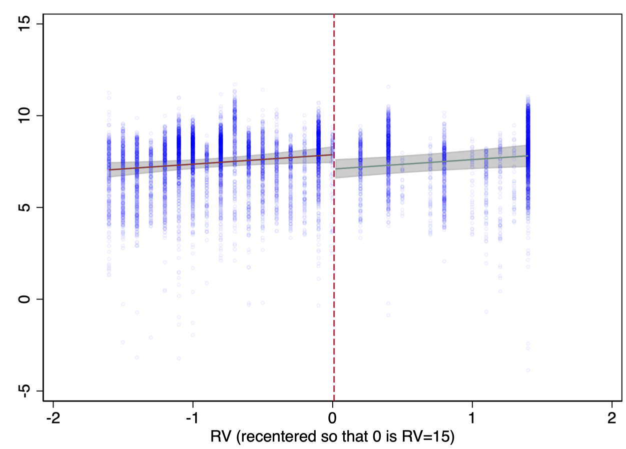
(a) RD0: No slopes (log population)



(b) RD1: Linear slopes (log population)



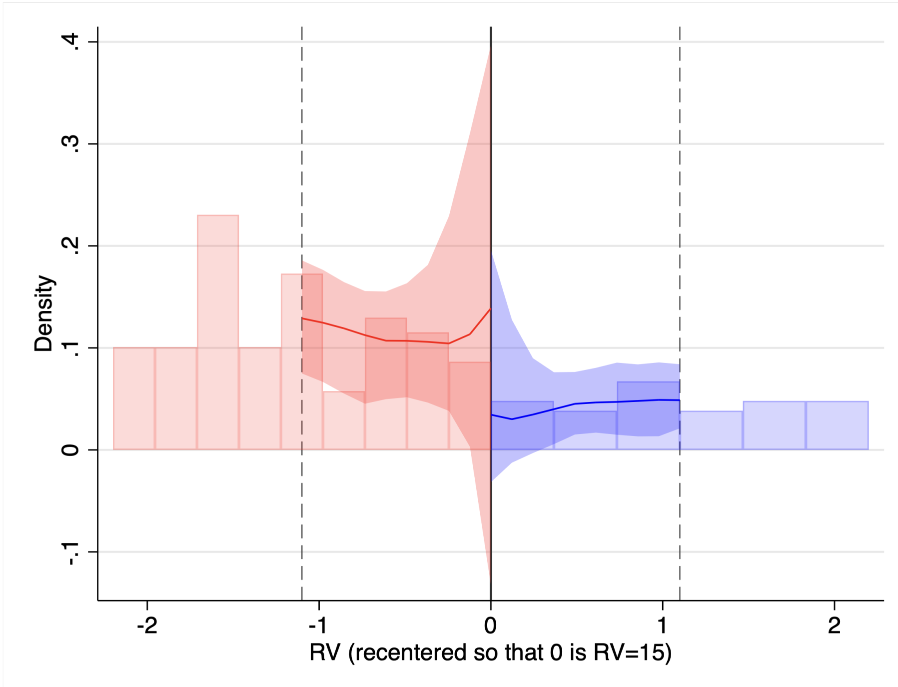
(c) RD0: No slopes (log population density)



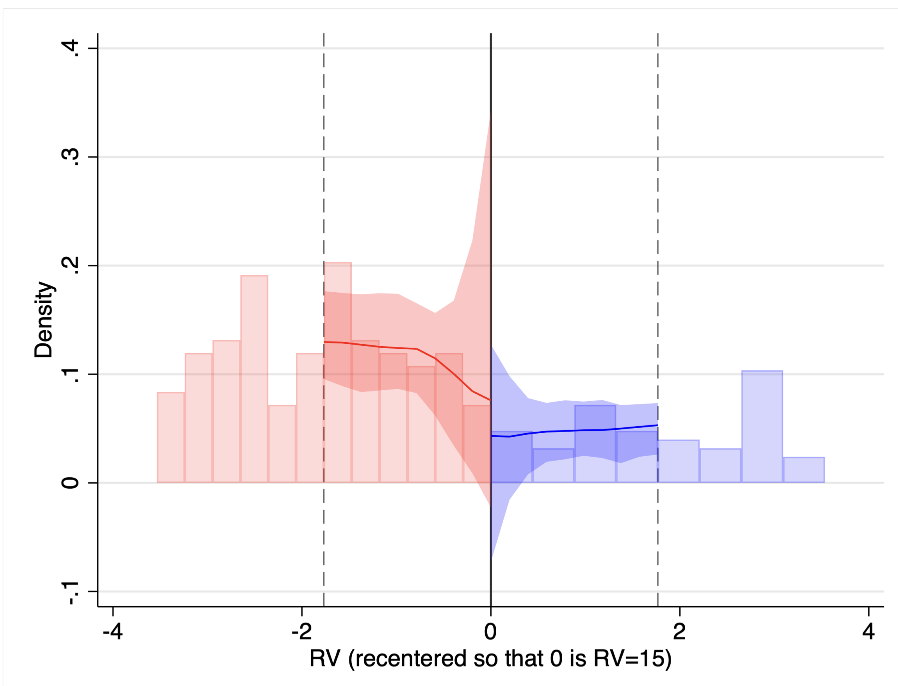
(d) RD1: Linear slopes (log population density)

Figure A.11: Continuity of predetermined population and population density

Notes: The figures show tests for continuity of predetermined covariates applied to the regression discontinuity approaches without slopes (RD0) and with linear trends (RD1). Panels (a) and (b) use 2000 population (log) and Panels (b) and (c) use 2000 population density (log). Lines are model predictions, shaded areas represent 95% confidence intervals. The horizontal axis shows EPA-registered $PM_{2.5}$ values, re-centered so that 0 is the cutoff (RV=15), and narrowed to the optimal bandwidth. Bandwidth selection in this figure is based on data from Meng et al. (2019*b*).



(a) RD0 optimal bandwidth



(b) RD1 optimal bandwidth

Figure A.12: Regression discontinuity – manipulation around threshold

Notes: The figures show tests for manipulation around threshold for regression discontinuity approaches without slopes (RD0) and with linear trends (RD1). The horizontal axis shows EPA-registered $PM_{2.5}$ values at the level of commuting zones, re-centered so that 0 is the cutoff (RV=15), and narrowed to twice the optimal bandwidth. Vertical axis shows density of RV values, in absolute (histogram) and polynomial approximation within the optimal bandwidth (dashed lines) following Calonico et al. (2014) and Calonico et al. (2020). A large discontinuous and significant jump between the fitted lines at the cutoff would indicate manipulation around the threshold. The figure is based on data from Meng et al. (2019*b*).

A.6 Testing for differences in coefficients across models

Table A.5: Testing for equality of coefficients across models

	DiD	DiDwb	M1DiD	M2DiD	RD0
DiDwb	0				
M1	0	0.410			
M2	0	0.424	0.463		
RD0	0.005	0.686	0.867	0.968	
RD1	0.127	0.790	0.919	0.967	0.937

Notes: The table shows p-values from two-sided tests of pairwise equality of coefficients corresponding to Panel (a) in Table 1. To calculate p-values, we cluster-bootstrap estimates for our different estimators. We draw counties (allowing for clustering) with replacement based on two strata (attainment and nonattainment), estimate the different models, and repeat the process 10,000 times and calculate the difference in coefficients for each bootstrap. The p-values correspond to the share of runs where the difference has the opposite sign as the difference in Table 1, multiplied by two to allow for two-sided testing.

A.7 Nonattainment status and changes in $PM_{2.5}$ – Robustness

Table A.6: Nonattainment status and changes in $PM_{2.5}$: *Dropping attainment counties with neighboring county in nonattainment*

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Part A: Effect from 2001-03 to 2006-08							
Panel (a): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-1.57	-0.50	-0.72	-0.67	-1.61	-0.64	-0.77
	(0.34)	(0.15)	(0.080)	(0.100)	(0.35)	(0.29)	(0.40)
Observations	64516	64516	26130	26664	43523	5441	10194
Panel (b): Placebo Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.22	-0.10	-0.028	0.037	-0.39	-0.037	-0.25
	(0.12)	(0.12)	(0.12)	(0.13)	(0.14)	(0.21)	(0.34)
Observations	41830	41830	15411	15145	20837	2341	6123
Panel (c): Heterogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	4.72	4.13	1.08	1.93	4.68	3.51	2.80
	(0.81)	(0.83)	(0.23)	(0.33)	(0.82)	(0.76)	(0.61)
NA(x)Baseline	-0.42	-0.36	-0.13	-0.18	-0.42	-0.29	-0.25
	(0.060)	(0.061)	(0.016)	(0.025)	(0.060)	(0.047)	(0.030)
Observations	64516	64516	26130	26664	43523	5441	10194
Implied ATE	-1.57	-1.29	-0.85	-0.83	-1.61	-0.85	-0.98
10th pct	-0.41	-0.30	-0.50	-0.32	-0.46	-0.052	-0.28
90th pct	-3.66	-3.10	-1.50	-1.75	-3.70	-2.30	-2.23
Part B: Effect from 2001-03 to 2011-13							
Panel (d): Homogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-2.52	-0.64	-0.43	-0.50	-2.60	-1.51	-1.78
	(0.27)	(0.11)	(0.13)	(0.14)	(0.28)	(0.44)	(0.61)
Observations	64516	64516	26130	26664	43523	4562	13442
Panel (e): Placebo Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-0.98	-0.050	0.17	0.15	-1.53	0.15	0.26
	(0.15)	(0.14)	(0.16)	(0.16)	(0.16)	(0.19)	(0.28)
Observations	41830	41830	15411	15145	20837	1204	3375
Panel (f): Heterogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	3.74	-0.13	5.06	4.71	3.66	4.20	3.18
	(0.41)	(0.43)	(0.51)	(0.44)	(0.41)	(0.83)	(1.01)
NA(x)Baseline	-0.42	-0.039	-0.39	-0.37	-0.42	-0.40	-0.34
	(0.029)	(0.032)	(0.038)	(0.032)	(0.029)	(0.053)	(0.053)
Observations	64516	64516	26130	26664	43523	4562	13442
Implied ATE	-2.52	-0.72	-0.82	-0.82	-2.60	-1.80	-1.95
10th pct	-1.37	-0.61	0.26	0.20	-1.45	-0.70	-1.01
90th pct	-4.61	-0.92	-2.77	-2.66	-4.69	-3.79	-3.66

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 post-treatment periods. All estimations exclude counties that were in attainment but had a neighboring county in nonattainment. Each panel(x)column combination is from a separate regression as described in the text. (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists, (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. (2019*b*).

Table A.7: Nonattainment status and changes in $PM_{2.5}$: *Dropping PM_{10} nonattainment counties*

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Part A: Effect from 2001-03 to 2006-08							
Panel (a): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.91	-0.59	-0.39	-0.27	-0.90	-0.64	0.048
	(0.059)	(0.083)	(0.12)	(0.16)	(0.092)	(0.22)	(0.39)
Observations	60978	60978	22889	22763	37507	9920	9559
Panel (b): Placebo Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.33	-0.13	-0.056	0.026	-0.51	-0.053	-0.22
	(0.13)	(0.12)	(0.13)	(0.13)	(0.16)	(0.18)	(0.31)
Observations	45114	45114	18904	18662	21643	3229	4447
Panel (c): Heterogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	0.65	-0.33	1.16	1.29	0.65	3.39	3.94
	(0.28)	(0.32)	(0.30)	(0.31)	(0.29)	(0.59)	(0.62)
NA(x)Baseline	-0.11	-0.019	-0.11	-0.11	-0.11	-0.28	-0.27
	(0.020)	(0.026)	(0.020)	(0.020)	(0.020)	(0.036)	(0.032)
Observations	60978	60978	22889	22763	37507	9920	9559
Implied ATE	-1.01	-0.61	-0.49	-0.36	-1.00	-0.78	-0.14
10th pct	-0.70	-0.56	-0.18	-0.061	-0.70	-0.010	0.61
90th pct	-1.56	-0.71	-1.04	-0.91	-1.55	-2.16	-1.50
Part B: Effect from 2001-03 to 2011-13							
Panel (d): Homogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-2.02	-0.70	-0.55	-0.57	-2.07	-0.79	-1.03
	(0.078)	(0.085)	(0.088)	(0.094)	(0.11)	(0.27)	(0.39)
Observations	60978	60978	22889	22763	37507	3712	12664
Panel (e): Placebo Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-0.93	0.039	0.22	0.15	-1.59	0.090	0.67
	(0.14)	(0.13)	(0.14)	(0.15)	(0.16)	(0.16)	(0.34)
Observations	45114	45114	18904	18662	21643	2133	3932
Panel (f): Heterogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	3.20	-0.98	4.68	4.66	3.16	5.58	4.58
	(0.33)	(0.35)	(0.33)	(0.33)	(0.34)	(0.80)	(0.71)
NA(x)Baseline	-0.37	0.020	-0.37	-0.37	-0.37	-0.43	-0.39
	(0.024)	(0.028)	(0.024)	(0.024)	(0.024)	(0.058)	(0.042)
Observations	60978	60978	22889	22763	37507	3712	12664
Implied ATE	-2.36	-0.68	-0.88	-0.90	-2.40	-0.95	-1.29
10th pct	-1.34	-0.73	0.14	0.12	-1.38	0.25	-0.21
90th pct	-4.21	-0.58	-2.74	-2.75	-4.25	-3.12	-3.24

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 post-treatment periods. All estimations exclude areas that were in nonattainment of the PM_{10} NAAQS between 2001-04. Each panel(x)column combination is from a separate regression as described in the text. (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ values exists, (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. (2019*b*).

Table A.8: Nonattainment status and changes in $PM_{2.5}$: *with flexible state time trends*

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Part A: Effect from 2001-03 to 2006-08							
Panel (a): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-1.11	-0.22	-0.13	-0.31	-1.36	-0.073	0.074
	(0.26)	(0.065)	(0.080)	(0.14)	(0.37)	(0.048)	(0.16)
Observations	72043	72043	28290	28908	47962	7026	10459
Panel (b): Placebo Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.20	0.0054	-0.018	0.020	-0.47	0.038	0.076
	(0.098)	(0.087)	(0.073)	(0.083)	(0.14)	(0.15)	(0.20)
Observations	49357	49357	20388	20127	25276	2143	5411
Panel (c): Heterogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	4.76	3.01	2.67	3.45	4.65	2.08	2.57
	(0.59)	(0.57)	(0.22)	(0.40)	(0.54)	(0.18)	(0.19)
NA(x)Baseline	-0.39	-0.24	-0.20	-0.26	-0.39	-0.16	-0.18
	(0.043)	(0.042)	(0.016)	(0.031)	(0.041)	(0.014)	(0.0091)
Observations	72043	72043	28290	28908	47962	7026	10459
Implied ATE	-1.09	-0.61	-0.29	-0.40	-1.16	-0.26	-0.21
10th pct	-0.016	0.053	0.25	0.31	-0.089	0.17	0.30
90th pct	-3.04	-1.82	-1.28	-1.68	-3.09	-1.04	-1.13
Part B: Effect from 2001-03 to 2011-13							
Panel (d): Homogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-1.56	-0.26	-0.097	-0.27	-1.76	0.18	-0.31
	(0.27)	(0.058)	(0.087)	(0.15)	(0.40)	(0.069)	(0.15)
Observations	72043	72043	28290	28908	47962	6137	25856
Panel (e): Placebo Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-0.43	0.017	0.0046	0.040	-0.78	-0.058	0.43
	(0.065)	(0.050)	(0.055)	(0.065)	(0.12)	(0.052)	(0.23)
Observations	49357	49357	20388	20127	25276	1046	4626
Panel (f): Heterogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	5.57	1.83	5.12	5.42	5.68	5.51	4.47
	(0.45)	(0.41)	(0.23)	(0.38)	(0.41)	(0.39)	(0.31)
NA(x)Baseline	-0.47	-0.16	-0.37	-0.39	-0.48	-0.39	-0.34
	(0.034)	(0.031)	(0.016)	(0.030)	(0.031)	(0.030)	(0.020)
Observations	72043	72043	28290	28908	47962	6137	25856
Implied ATE	-1.54	-0.51	-0.40	-0.41	-1.51	-0.28	-0.58
10th pct	-0.23	-0.084	0.61	0.67	-0.19	0.78	0.35
90th pct	-3.90	-1.29	-2.24	-2.34	-3.90	-2.21	-2.27

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 post-treatment periods. All estimations include state fixed effects. Each panel(x)column combination is from a separate regression as described in the text. (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists, (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. (2019*b*).

Table A.9: Nonattainment status and changes in $PM_{2.5}$: *with flexible state time trends and quartile of density time trends*

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Part A: Effect from 2001-03 to 2006-08							
Panel (a): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.97	-0.23	-0.14	-0.35	-1.27	-0.11	-0.042
	(0.24)	(0.062)	(0.079)	(0.14)	(0.34)	(0.047)	(0.14)
Observations	71951	71951	28264	28882	47881	7021	10451
Panel (b): Placebo Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.12	0.00093	0.026	0.049	-0.44	0.057	0.076
	(0.097)	(0.088)	(0.072)	(0.083)	(0.14)	(0.15)	(0.19)
Observations	49289	49289	20373	20127	25219	2141	5396
Panel (c): Heterogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	4.74	2.81	2.36	3.41	4.66	1.49	1.63
	(0.61)	(0.51)	(0.30)	(0.51)	(0.57)	(0.32)	(0.32)
NA(x)Baseline	-0.39	-0.23	-0.18	-0.25	-0.39	-0.12	-0.12
	(0.046)	(0.038)	(0.022)	(0.039)	(0.043)	(0.023)	(0.020)
Observations	71951	71951	28264	28882	47881	7021	10451
Implied ATE	-1.06	-0.59	-0.27	-0.39	-1.15	-0.24	-0.20
10th pct	0.0036	0.032	0.21	0.30	-0.086	0.078	0.14
90th pct	-2.99	-1.72	-1.15	-1.66	-3.09	-0.82	-0.80
Part B: Effect from 2001-03 to 2011-13							
Panel (d): Homogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-1.30	-0.27	-0.11	-0.34	-1.61	0.089	-0.34
	(0.25)	(0.056)	(0.082)	(0.15)	(0.35)	(0.074)	(0.14)
Observations	71951	71951	28264	28882	47881	6133	25831
Panel (e): Placebo Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-0.29	-0.0060	0.093	0.098	-0.72	-0.15	0.45
	(0.066)	(0.050)	(0.055)	(0.065)	(0.12)	(0.084)	(0.25)
Observations	49289	49289	20373	20127	25219	1044	4612
Panel (f): Heterogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	5.38	1.61	4.52	5.04	5.39	4.01	3.37
	(0.52)	(0.36)	(0.29)	(0.49)	(0.48)	(0.68)	(0.34)
NA(x)Baseline	-0.45	-0.14	-0.33	-0.36	-0.46	-0.28	-0.26
	(0.039)	(0.027)	(0.021)	(0.037)	(0.036)	(0.049)	(0.022)
Observations	71951	71951	28264	28882	47881	6133	25831
Implied ATE	-1.41	-0.49	-0.37	-0.41	-1.48	-0.23	-0.54
10th pct	-0.16	-0.10	0.53	0.59	-0.22	0.55	0.18
90th pct	-3.67	-1.19	-1.99	-2.22	-3.77	-1.64	-1.84

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 post-treatment periods. All estimations include state fixed effects and tract population density quartile fixed effects. Each panel(x)column combination is from a separate regression as described in the text. (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists, (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. (2019*b*).

Table A.10: TWFE estimates in tract-year panel (2000-2015)

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
NA Effect	-1.675	-0.544	-0.469	-0.855	-1.719	-0.794	-0.727
	(0.254)	(0.069)	(0.073)	(0.262)	(0.260)	(0.268)	(0.208)
Observations	1152688	1152688	488400	516464	767392	98192	180448

Notes: The table shows results from a panel regression with two-way fixed effects (TWFE) with a homogeneous treatment effect from nonattainment designations, equivalent to Panels (a) and (d) of Table 1. Data used is a tract-year panel from 2000 to 2015. Estimation includes tract and year fixed effects. Standard errors clustered at the county level in parentheses. All results based on Meng et al. (2019*b*).

A.8 NBP and CAIR as potential confounders

Since nonattainment designations under the NAAQS for $PM_{2.5}$ are not the only air quality policies during our sample period, we may worry about mis-attributing changes in air quality to $PM_{2.5}$ nonattainment designations if those other air quality policies are correlated with nonattainment designations. Specifically, during our sample period, the NO_x Budget Trading Program (NBP) (discussed in e.g. Deschenes et al. 2017, Curtis 2018) and its' successor, the Clean Air Interstate Rule (CAIR) were implemented. The NBP was a cap-and-trade program enacted by twenty eastern states plus DC in 2003/2004 that targeted NO_x emissions from power plants and other large stationary sources. In 2009, the NBP was replaced by the CAIR which expanded geographic coverage and targeted SO_2 and Ozone emissions in addition to NO_x.

As NO_x and SO_2 are precursors to $PM_{2.5}$, and these policies coincide with our study period, we next verify that these are not partially driving our results. To do so, we collect data on all facilities subject to regulation under the NBP and CAIR from the EPA's Clean Air Markets Data Program Facility Attributes Table (EPA 2023). We generate a binary variable indicating if a county contained a facility that became subject to NBP (for the 5-year period ending in 2006-08) or CAIR (for the 10-year period ending in 2011-13), and include this indicator as a control variable in our DiD and DiDwb regressions. Part A of Table A.11 shows the results for NBP and CAIR respectively in Panel (a) and (b). Column 1 and 3 reproduce our baseline coefficients from Table 1. Columns 2 and 4 show that controlling for these programs leaves our $PM_{2.5}$ nonattainment coefficients from Column 1 and 3 largely unchanged, indicating that any potential bias from these other policies is likely minimal.[60]

Instead of testing robustness of our effects to NBP/CAIR controls, we next analyze in Part B of Table A.11 whether the NBP/CAIR estimates suffer from similar bias when ignoring possibly confounding trends in $PM_{2.5}$. First, in Panel (c), we estimate the effect of NBP or CAIR using our $PM_{2.5}$ data. While the estimates in Columns 1 and 2 suggest substantial $PM_{2.5}$ reduction effects, at least for NBP, this effect disappears when controlling for baseline levels of pollution.[61] Second, in Panel (d), we use the data and code from Deschenes et al. (2017*b*) to first replicate their results in Column 1 and 2 (corresponding to their Table 2b Row 9 Columns 4 and 5 – see table notes for details). Once we control for trends based on baseline $PM_{2.5}$, the estimated coefficients fall in corresponding Columns 3 and 4, mirroring the pattern using our data in Panel (c). Deschenes et al. (2017) note that their effects on $PM_{2.5}$ are inconclusive. Once we control for trends, the results are even closer to zero, in line with our main findings. Finally, in Panel (e) we again replicate the results from

[60] The results look almost identical when we control for NBP/CAIR participation at the state level to account for possible spillovers. These results are available from the authors upon request.

[61] Note also that the NBP and CAIR effects from Panel (c) Columns 1 and 2 also disappear when we control for $PM_{2.5}$ nonattainment in Panel (a) and (b) Column 2.

Deschenes et al. (2017) in Columns 1 and 2 but focus on Ozone concentrations instead of $PM_{2.5}$ (corresponding to their Table 2b Row 4 Columns 4 and 5). Interestingly, controlling for trends based on baseline Ozone reduces estimated coefficients only by a small amount, largely confirming the results in Deschenes et al. (2017). This suggest that confounding trends in $PM_{2.5}$ may be particularly severe.

Table A.11: Controlling for potential confounding by contemporaneous policy changes and exploring NBP/CAIR effects

	(DiD) (1)	(DiD) (2)	(DiDwb) (3)	(DiDwb) (4)
Part A: Robustness controlling for NBP or CAIR				
Panel (a): NBP enacted 2003/04 (period ending 2006-08)				
Nonattainment	-1.47	-1.55	-0.49	-0.56
	(0.34)	(0.48)	(0.098)	(0.095)
NBP		0.18		0.17
		(0.36)		(0.22)
Observations	72043	72043	72043	72043
Panel (b): CAIR enacted 2009 (period ending 2011-13)				
Nonattainment	-2.35	-2.39	-0.56	-0.60
	(0.27)	(0.29)	(0.096)	(0.097)
CAIR		0.22		0.18
		(0.20)		(0.078)
Observations	72043	72043	72043	72043
Part B: Focusing on NBP or CAIR without/with controlling for trends				
Panel (c): NBP or CAIR				
NBP	-0.49		0.0034	
	(0.22)		(0.21)	
CAIR		-0.20		0.11
		(0.28)		(0.085)
Observations	72043	72043	72043	72043
Panel (d): $PM_{2.5}$ – replicating Deschenes et al. (2017)				
NBP x Post x Summer	-0.45	-1.03	0.19	-0.55
	(0.32)	(0.27)	(0.31)	(0.36)
Observations	4172	4172	4172	4172
Panel (e): Ozone 8-hour value (ppb) – replicating Deschenes et al. (2017)				
NBP x Post x Summer	-3.38	-3.37	-3.01	-3.06
	(0.56)	(0.54)	(0.54)	(0.56)
Observations	2352	2352	2352	2352

Notes: Part A tests robustness of our nonattainment estimates to controlling for NBP or CAIR status. Column 1 shows the baseline DiD results identical to Column 1 in Table 1. Column 3 shows the baseline DiDwb results identical to Column 2 in Table 1. Columns 2 and 4 add indicators for counties containing at least one facility that was subject to regulation under NBP or CAIR as control variables in Panel (a) and (b) respectively. Part B focuses on NBP and CAIR and tests robustness to controlling for trends based on baseline pollution (DiDwb). Panel (c) uses our approach and $PM_{2.5}$ data from Meng et al. (2019*b*), and shows results for DiD in Columns 1 and 2 for NBP and CAIR, where the endlines are 2006-08 and 2011-13 respectively. Columns 3 and 4 add the controls (DiDwb) to Columns 1 and 2. In Panel (d) and (e), we use the data and code from Deschenes et al. (2017*b*) to first replicate their results in Columns 1 and 2, focusing on the NBP with data from 2001-2007. Panel (d) and (e) Columns 1 and 2 correspond to their Table 2b, Row 9 and Row 4, Columns 4 and 5 respectively. The analysis is for panel data at the county-by-year-by-season level and both columns include county-by-season, summer-by-year, and county-by-year fixed effects as well as detailed weather controls. Column 1 is weighted by emission/pollution monitors, and Column 2 by population. Columns 3 and 4 (DiDwb) add year dummies interacted with baseline pollution and seasons to Columns 1 and 2 respectively. Panel (d) uses their $PM_{2.5}$ data as outcome and control for baseline trend, and Panel (e) uses their Ozone data as outcome and control for baseline trend. Standard errors in parentheses are clustered at the county level for Panels (a) to (c) and clustered at the state by season level for Panel (d) and (e).

A.9 Addressing uncertainty in pollution data

To address possible non-classical measurement error in the $PM_{2.5}$ reanalysis data, we show three robustness test.

The first tests relies on the uncertainty data in van Donkelaar et al. (2021*b*). For each grid-point, this data not only contains the estimated $PM_{2.5}$ concentration, but also information of the uncertainty around this estimate due to local geo-physical characteristics or distance to monitors. We drop 30% of Census tracts with the largest uncertainty average over our sample period and normalized by its mean.[62] To avoid mixing $PM_{2.5}$ data sources, we rely on pollution data from van Donkelaar et al. (2021*b*) exclusively for this first exercise, so the most comparable baseline table that retains all observations is Appendix Table A.24, which corresponds to main Table 1 based on data from Meng et al. (2019*b*). Part A of Table A.12 shows that our estimates are robust to dropping 30% of tracts that have the most uncertain $PM_{2.5}$ data estimates.

Second, since areas with ground-based air pollution monitors likely have less uncertainty in the $PM_{2.5}$ data, we drop all counties that neither contain a monitor themselves nor have a neighboring county with a monitor, using data from Meng et al. (2019*b*). Part B of Table A.12 shows that our results are robust to dropping these counties.

Third, in our most restrictive approach in Table A.13, we rely exclusively on data from ground-based pollution monitors from EPA (2022*a*). This severely reduces our observations, but even in this restrictive version, the patters of our main results and the bias of naive DiD are robust.

[62]That is we first take the average uncertainty and the average $PM_{2.5}$ estimate for each Census tract across our sample period (where we use tract data from population-weighted Census block estimates as in our main paper). We then calculate a normalized measure of uncertainty by dividing the uncertainty from the previous step by the average from the previous step for each tract. We then drop the 30% of tracts with the highest value of normalized uncertainty.

Table A.12: Dropping areas with higher uncertainty in pollution measurements

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Part A: Dropping 30% of tracts with highest uncertainty measures							
Panel (a): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-1.46	-0.29	-0.40	-0.40	-1.47	-0.59	-0.76
	(0.36)	(0.12)	(0.18)	(0.21)	(0.38)	(0.36)	(0.41)
Observations	50452	50452	25499	26061	33582	4499	6166
Panel (b): Placebo Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.24	-0.086	-0.053	0.022	-0.58	-0.21	-0.35
	(0.12)	(0.11)	(0.12)	(0.12)	(0.18)	(0.30)	(0.47)
Observations	29672	29672	17636	17314	13547	1117	2233
Panel (c): Heterogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	5.27	3.50	2.13	3.79	5.37	3.77	3.02
	(0.83)	(0.86)	(0.41)	(0.81)	(0.83)	(0.87)	(0.94)
NA(x)Baseline	-0.44	-0.29	-0.18	-0.29	-0.45	-0.30	-0.25
	(0.059)	(0.063)	(0.028)	(0.058)	(0.059)	(0.052)	(0.052)
Observations	50452	50452	25499	26061	33582	4499	6166
Implied ATE	-1.39	-0.92	-0.54	-0.56	-1.37	-0.81	-0.73
10th pct	-0.17	-0.10	-0.047	0.24	-0.13	0.033	-0.042
90th pct	-3.61	-2.39	-1.43	-2.01	-3.61	-2.33	-1.98
Part B: Only counties (incl. neighboring counties) with monitor							
Panel (d): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-1.53	-0.28	-0.41	-0.32	-1.56	-0.34	-0.016
	(0.37)	(0.16)	(0.20)	(0.21)	(0.37)	(0.32)	(0.42)
Observations	47821	47821	24267	24197	44008	6180	9928
Panel (e): Placebo Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.39	-0.096	0.0027	-0.048	-0.50	-0.14	-0.29
	(0.15)	(0.13)	(0.15)	(0.16)	(0.15)	(0.21)	(0.29)
Observations	26991	26991	15636	15634	23178	2113	5121
Panel (f): Heterogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	4.82	3.58	1.80	1.97	4.79	4.11	3.70
	(0.80)	(0.84)	(0.32)	(0.34)	(0.81)	(0.87)	(0.68)
NA(x)Baseline	-0.42	-0.30	-0.16	-0.16	-0.42	-0.31	-0.26
	(0.059)	(0.064)	(0.020)	(0.021)	(0.059)	(0.054)	(0.035)
Observations	47821	47821	24267	24197	44008	6180	9928
Implied ATE	-1.47	-0.97	-0.54	-0.46	-1.50	-0.54	-0.20
10th pct	-0.32	-0.14	-0.11	-0.014	-0.34	0.32	0.51
90th pct	-3.57	-2.49	-1.32	-1.27	-3.59	-2.08	-1.51

Notes: Part A is equivalent to Appendix Table A.24 based on pollution data from van Donkelaar et al. (2021*b*), after dropping 30% of tracts that have the most uncertain $PM_{2.5}$ predictions. Part B is equivalent to Table 1 based on data from Meng et al. (2019*b*), after dropping all counties that do not contain a monitor and also do not have a neighboring county with a monitor. The table shows coefficient estimates for the treatment effect of nonattainment status on the change in $PM_{2.5}$ levels between the pre- and post-treatment periods. Each panel(x)column combination is from a separate regression as described in the text. (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists, (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.

Table A.13: Nonattainment status and changes in $PM_{2.5}$ using EPA monitor data

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-1.25	-0.012	-0.10	-0.12	-1.22	-0.15	-0.084
	(0.24)	(0.16)	(0.24)	(0.31)	(0.24)	(0.26)	(0.48)
Observations	667	667	279	268	596	36	86
Panel (b): Placebo Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.53	-0.072	-0.37	-0.058	-0.56	-0.63	-0.53
	(0.11)	(0.16)	(0.25)	(0.25)	(0.11)	(0.57)	(0.52)
Observations	431	431	239	244	360	23	113
Panel (c): Heterogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	5.29	3.70	2.07	2.09	5.32	4.54	6.82
	(1.22)	(1.24)	(0.61)	(0.64)	(1.22)	(2.04)	(2.12)
NA(x)Baseline	-0.42	-0.27	-0.15	-0.15	-0.42	-0.31	-0.46
	(0.084)	(0.087)	(0.039)	(0.038)	(0.084)	(0.13)	(0.14)
Observations	667	667	279	268	596	36	86
Implied ATE	-1.25	-0.59	-0.25	-0.29	-1.22	-0.36	-0.43
10th pct	-0.015	0.22	0.19	0.16	0.019	0.57	0.94
90th pct	-2.90	-1.68	-0.84	-0.88	-2.86	-1.59	-2.25

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 post-treatment periods. Each panel(x)column combination is from a separate regression as described in the text. (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists, (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 monitor data from EPA (2022*a*).

A.10 Synthetic Control Estimates

Instead of our MDiD approach combining matching and DiD, an alternative approach can be based on Synthetic Controls (SC). The traditional SC method is designed for a context with few treated units and a small 'donor pool' of control units (Abadie et al. 2010). In our context, however, we have many treated and control units, and therefore use MDiD as on of our primary alternatives. Nevertheless, we extend the SC methods to our setting and provide two sets of results based on synthetic counterfactuals for additional robustness analysis.

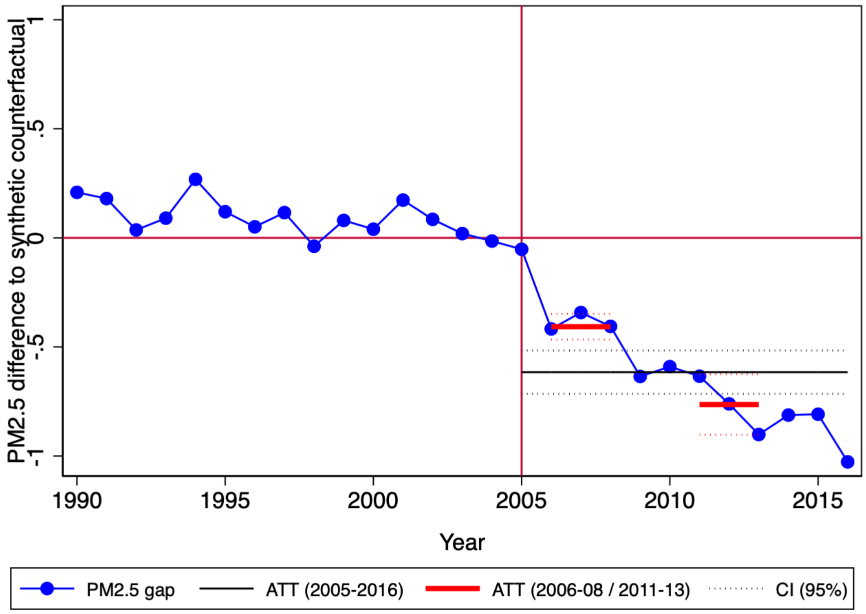
First, we use the recently proposed Synthetic Difference-in-Differences (SDiD) estimator to estimate average treatment effects (Arkhangelsky et al. 2021). In addition to unit weights chosen to closely replicate the average treated unit before treatment as in MDiD, SDiD uses time weights in the pre-treatment period to reduce variation in time trends among control units. SDiD is implemented as a weighted DiD with unit fixed effects and has been shown to perform equally well or better than traditional SC or DiD in common settings (Arkhangelsky et al. 2021). We show county level results based on SDiD in Figure A.13a using annual $PM_{2.5}$ concentrations based on Meng et al. (2019*b*) during the 1990-2004 pre-treatment period as predictor variables.[63] The blue line shows that the gap between the average nonattainment county and the weighted control units is small until 2005, but diverges in the expected direction after 2005. The estimated ATT for the full post-treatment period (2005-2016) shows a 0.62 $\mu g/m^3$ reduction in $PM_{2.5}$ (black line). To compare the SDiD estimates to our main analysis in Table 1 we focus on the same post-treatment time periods. The two red lines in Figure A.13a show that the SDiD estimates are very similar to our main estimates, with an ATT of 0.41 until 2006-08 and an ATT of 0.76 until 2011-13, confirming robustness of our main results.

While traditional SC estimation is not suitable for estimating average treatment effects in our setting, it offers another approach to heterogeneity analysis. We construct synthetic counterfactuals for each of the 208 nonattainment counties based on 1990-2004 $PM_{2.5}$ levels, each time limiting the 'donor pool' to attainment counties from the same Census division.[64] We then compare the change between 2001-03 and 2006-08 between each of the 208 nonattainment counties and its' synthetic counterfactual visualized by the red markers in Figure A.13b. This shows a population weighted average effect of 0.62 (red dotted line), and shows heterogeneity that increases the treatment effect with baseline pollution as captured by EPA-registered $PM_{2.5}$ values, in line with our main results. For a placebo exercise, we also run a SC analysis for each of the 339 attainment counties that have an EPA-registered $PM_{2.5}$ value below 15. The blue markers in Figure A.13b show the results. Reassuringly,

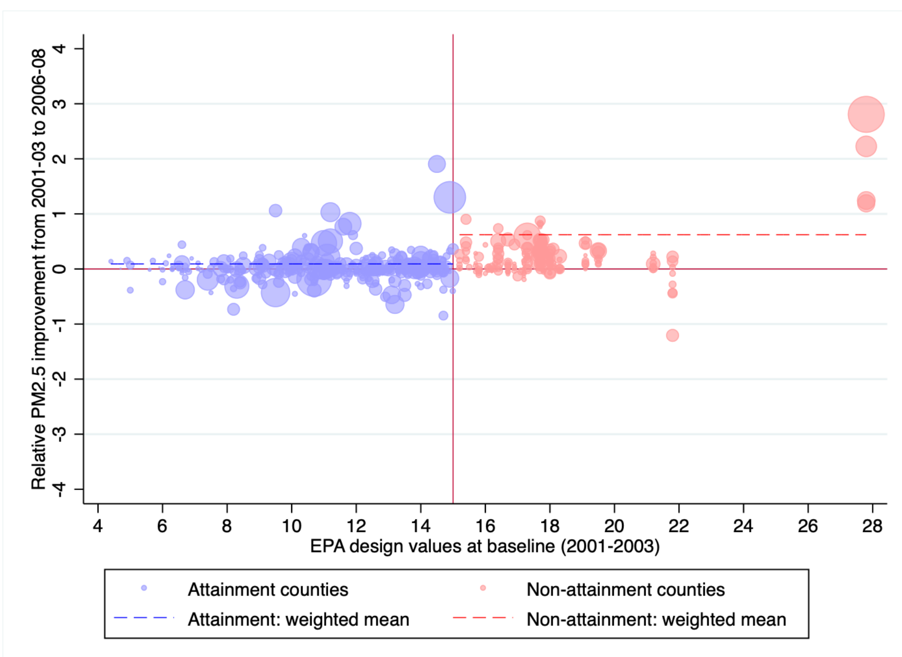
[63] Standard error calculations proposed by Arkhangelsky et al. (2021) are cluster-robust at the level of treated units, i.e. counties in our setting. County $PM_{2.5}$ levels are population-weighted averages across tracts.

[64] The nine Census Divisions are New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain Division and Pacific Division.

and in contrast to Figure 4, there is no visible association between $PM_{2.5}$ improvements and EPA-registered $PM_{2.5}$ values for these placebo attainment counties, as these are evaluated against their own synthetic counterfactuals.



(a) Synthetic Difference-in-Differences Estimation



(b) Unit-wise Synthetic Control Estimation (2001-03 vs. 2006-08)

Figure A.13: Synthetic Control Estimates

Notes: These figures are based on county level synthetic control estimates for 3,109 counties, 208 of which are in nonattainment of the $PM_{2.5}$ NAAQS from 2005. Panel (a) implements Synthetic Difference-in-Differences estimation following Arkhangelsky et al. (2021). The black line shows the estimated average treatment effect on the treated (ATT) for the full post-treatment period (2005-16). The red lines show the ATT for the three-year average periods used in the main analysis (2006-08 and 2011-13). Dashed lines indicate 95% confidence intervals based on standard errors that allow for correlation within county clusters. Panel (b) shows unit-wise Synthetic Control estimates of the $PM_{2.5}$ improvement between 2001-03 and 2006-08 for the 208 nonattainment counties evaluated against their synthetic counterfactual in red, and 339 attainment counties that have an EPA-registered $PM_{2.5}$ value (below 15), each relative to their unit-wise synthetic counterfactual in blue. Bubble size indicates county population in 2010 and dashed lines are population-weighted means. Based on data from Meng et al. (2019*b*).

A.11 Heterogeneous $PM_{2.5}$ nonattainment treatment effect by previous PM_{10} nonattainment status

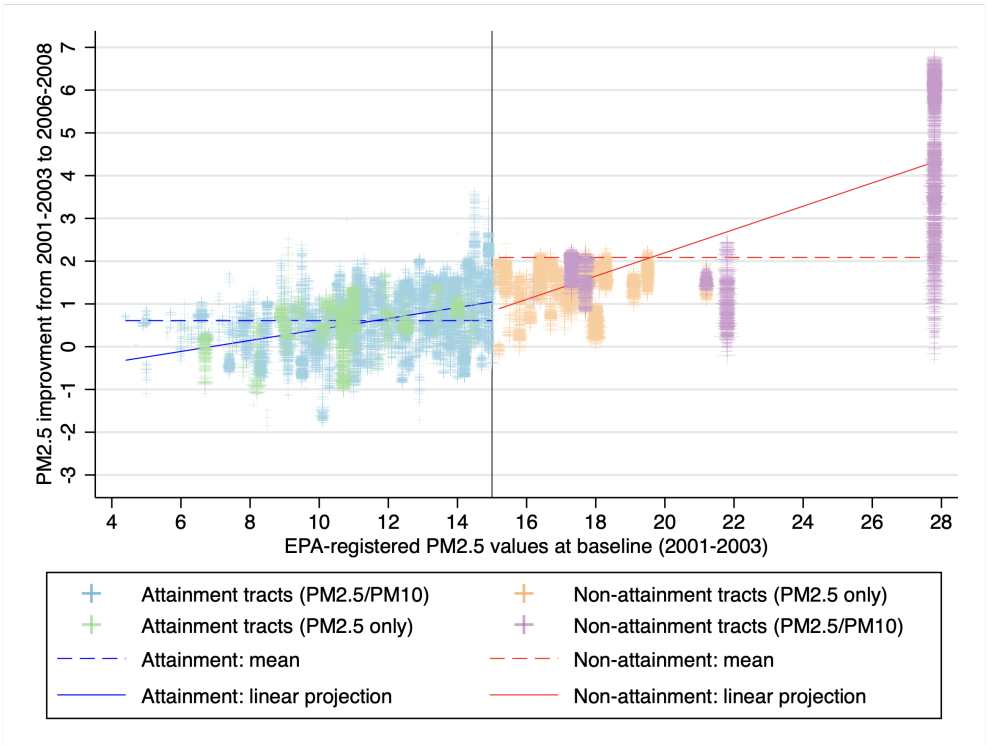


Figure A.14: Improvement in tract $PM_{2.5}$ averages and $PM_{2.5}/PM_{10}$ nonattainment status

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 EPA-registered $PM_{2.5}$ values of the nonattainment and attainment areas separately, weighted by tract population. In addition, the purple markers indicate nonattainment areas that are also in nonattainment of the previous PM_{10} regulation, and the green markers indicate attainment areas that are in nonattainment of the previous PM_{10} regulation. Based on data from Meng et al. (2019*b*).

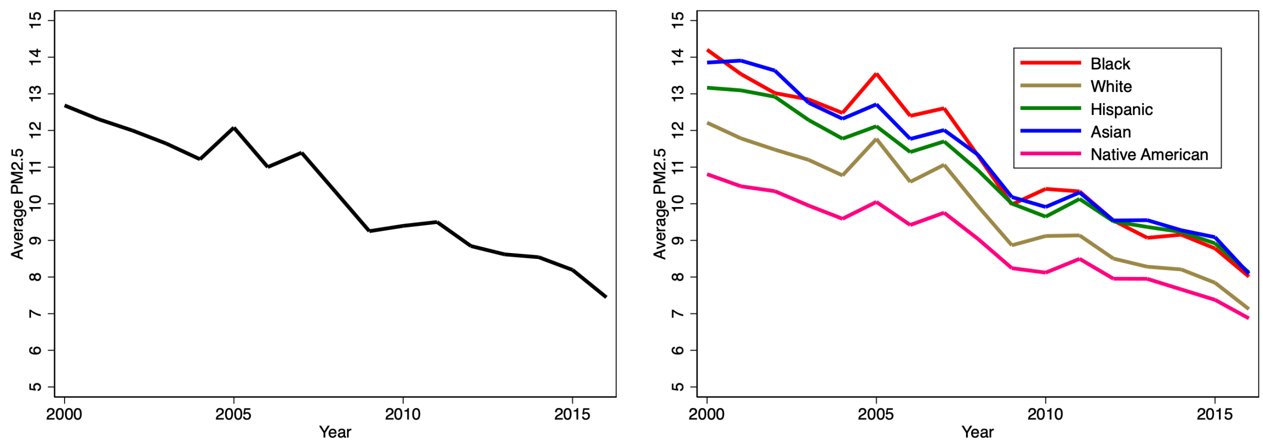
Table A.14: Heterogeneous treatment effects by previous PM_{10} status

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
PM2.5 NA	-0.91	-0.26	-0.32	-0.19	-0.90	-0.35	-0.022
w/o prev. PM10 NA	(0.059)	(0.14)	(0.16)	(0.18)	(0.092)	(0.29)	(0.40)
PM2.5 NA	-2.78	-1.30	-1.29	-1.63	-2.79	0	0.63
w. prev. PM10 NA	(0.72)	(0.34)	(0.19)	(0.31)	(0.72)	(.)	(0.65)
Observations	72043	72043	28291	28909	47962	7026	10459
Panel (b): Homogeneous Treatment Effect: from 2001-03 to 2011-13							
PM2.5 NA	-2.02	-0.63	-0.56	-0.60	-2.07	-1.20	-1.16
w/o prev. PM10 NA	(0.078)	(0.093)	(0.091)	(0.098)	(0.11)	(0.38)	(0.36)
PM2.5 NA	-3.76	-0.57	-1.71	-2.11	-3.72	0	0.69
w. prev. PM10 NA	(0.71)	(0.25)	(0.26)	(0.30)	(0.71)	(.)	(2.17)
Observations	72043	72043	28291	28909	47962	6137	25856

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 post-treatment periods. We allow heterogeneous treatment effects by previous PM_{10} nonattainment status, as in Equation (7). All regressions control for trends based on PM_{10} nonattainment status, so the shown coefficients are the heterogeneous marginal effects of $PM_{2.5}$ nonattainment status, as in Figure 7. Each panel(x)column combination is from a separate regression as indicated: (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists, (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. (2019*b*).

A.12 Replication of $PM_{2.5}$ exposure levels in Jbaily et al. (2022)

Our main analysis uses the same $PM_{2.5}$ data (Meng et al. 2019*b*) that is also used by Jbaily et al. (2022) to document pollution exposure disparities across income and racial groups in the US. We replicate the relevant average exposure levels from their paper in Figure A.15. Panel (a) shows the population-weighted $PM_{2.5}$ exposure across all residents and panel (b) shows averages by racial groups. Reassuringly, both the levels and changes over time are virtually identical to those shown in panels (a) and (b) of Extended Data Fig. 1 in Jbaily et al. (2022).



(a) All residents

(b) By racial group

Figure A.15: Replication of $PM_{2.5}$ levels in Jbaily et al. (2022)

Notes: The Figure replicates population-weighted $PM_{2.5}$ exposure levels by population groups as shown in Extended Data Fig. 1 in Jbaily et al. (2022). Results are based on Meng et al. (2019*b*) and tract level population counts.

A.13 Comparison to recent analysis in Currie et al. (2023)

Part of our analysis of 2005 nonattainment effects, particularly the regulation's impact on $PM_{2.5}$ exposure gaps between Black and White residents, is closely related to the recent contribution by Currie et al. (2023) — henceforth CVW.[65] In the below, we first show that we can replicate some of the main findings of CVW despite using only our publicly available data.[66] Thereafter, we highlight the main differences between our and their approaches and discrepancies in data.

A. Replication of headline results in CVW (Currie et al. 2023)

To replicate the results of CVW, we rely on pollution data from Di et al. (2021) as in their analysis. While CVW use samples of individuals from the long form 2000 Census and the American Community Survey (ACS) from 2001, we use the population counts that account for the universe of individuals from Census records, and linearly interpolate between the 2000, 2010, and 2020 Census records. As in CVW, we assign pollution to individuals based on Census blocks. We run the entire replication analysis at the Census block level. Importantly, we also use their assignment into treatment status for the purposes of replication, which we discuss in more detail in the next section.

Currie et al. (2023) begin by showing that Black Americans are exposed to substantially higher levels of $PM_{2.5}$ than White Americans, and that this gap has narrowed over time. We show in Table A.15 that the average $PM_{2.5}$ exposure levels and the Black-White gap closely, although not exactly, replicate the numbers reported in Table 2 of CVW. The numbers in Column 1 and 3 are particularly similar, as these are both based on the 2000 Census without including data from the American Community Survey (ACS).[67]

Turning to the 2005 NAAQS nonattainment designations, CVW first show an event study of the regulation on $PM_{2.5}$ concentrations. Figure A.16 shows that we can replicate their event study almost exactly. Note that there is no pre-trend in this Figure, which is due to CVW assigning a subset of treated units into the control group as we discuss in more detail in the next section. We next replicate their baseline average treatment effects in Table A.16. With year and county fixed effects, we estimate an ATE of -1.2 $\mu g/m^3$, almost identical to the estimate of -1.230 reported by CVW. Similarly, when adding state-year fixed effects, our estimate falls to -0.76 (compared to -0.737 in CVW).

Finally, CVW ask how much of the reduction in the Black-White exposure gap between 2005 and 2015 can be accounted for by the 2005 nonattainment designations. To account for effect heterogene-

[65]We are grateful for the authors of CVW for helpful discussions, especially Reed Walker.

[66]Their individual level data is indispensable for their analysis of contributions of individual level income to exposure gaps. We only focus on their main results here.

[67]CVW use the 2000 Census long form which is a subset of the 2000 Census.

Table A.15: Replication of Table 2 in CVW

	Actual 2000 Exposure	Actual 2015 Exposure	Counterfactual 2015 using 2000 locations
Panel (a): Original numbers reported in CVW			
White	12.96	8.25	8.22
Black	14.52	8.79	8.89
B-W Difference	1.56	0.54	0.67
Chg. in B-W Diff	0.00	-1.02	-0.89
Panel (b): Replication using our data			
White	12.90	8.12	8.17
Black	14.53	8.80	8.90
B-W Difference	1.63	0.68	0.73
Chg. in B-W Diff	0.00	-0.95	-0.90

Notes: Panel (a) restates Table 2 from Currie et al. (2023). Panel (b) replicates those numbers using our data. Columns 1 and 2 report average $PM_{2.5}$ exposure levels, using block level population weights that are linearly interpolated between 2010 and 2020. Column 3 uses constant 2000 population weights instead. Pollution data is from Di et al. (2021).

ity, they estimate RIF-Quantile treatment effects in 19 pollution 'vigintiles', separately for Black and White residents. We replicate the RIF-Quantile regressions in Figure A.17, which again closely resembles Figure 8 in CVW. We replicate the counterfactual gap accounting based on these regression results in Table A.17. Again, the results closely resemble those reported in Table 4 of CVW. Our overall actual change in the gap (0.47 vs. 0.59 $\mu g/m^3$) and counterfactual change in the gap (0.18 vs 0.23 $\mu g/m^3$) are similar but slightly smaller. Yet, using our publicly available Census based data recovers virtually the same contribution of the CAA nonattainment areas to the reduction in the Black-White pollution gap (61.1%) as the ACS-based individual level sample in CVW (61.2%).

Throughout this section, we used the same classification of treated and control units as in CVW. The small differences in results are due to differences between our (interpolated) block level population counts and the individual level survey sample used in CVW. We next discuss these differences further.

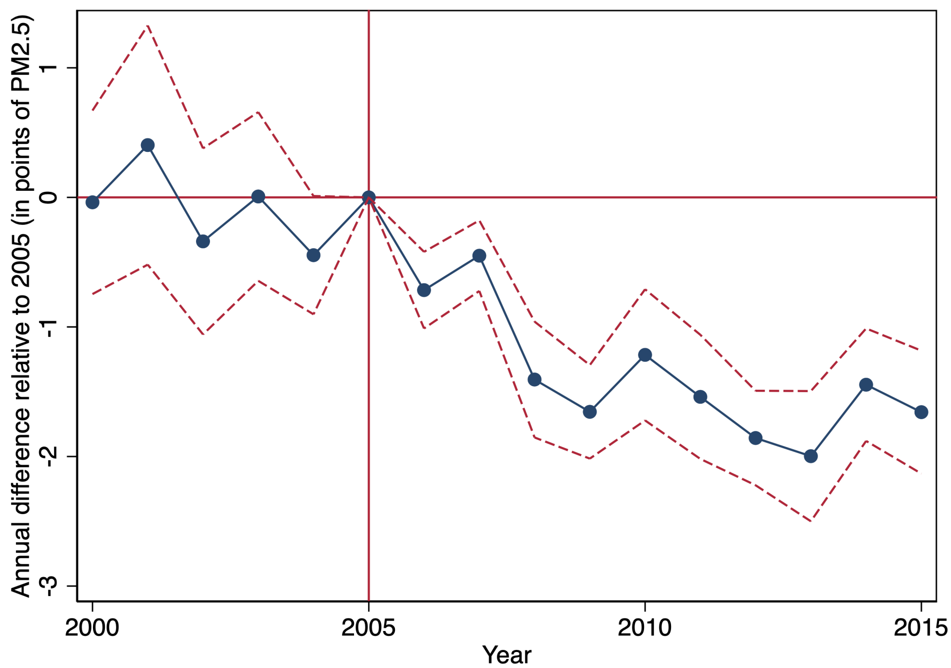


Figure A.16: Replication of Figure 6 in CVW

Notes: This figure replicates Figure 6 in Currie et al. (2023) using the block level data from our paper. The graph shows an event study plotting the coefficients from nonattainment areas as defined by CVW interacted with year dummies. The regression model controls for county fixed effects and year fixed effects. The regression is weighted by block level population counts, linearly interpolated between 2000, 2010 and 2020, and errors are clustered by commuting zone. Pollution data is from Di et al. (2021).

Table A.16: Replication of Table 3 in CVW

	(1)	(2)
Panel (a): Original numbers reported in CVW		
$PM_{2.5}$ NA	-1.230	-0.727
	(0.335)	(0.080)
Observations	32,360,000	32,360,000
Panel (b): Our data		
PM2.5 NA	-1.20	-0.76
	(0.40)	(0.078)
Observations	108,583,670	108,583,670
County FE	Yes	Yes
Year FE	Yes	No
State-Year FE	No	Yes

Notes: The table replicates difference-in-differences estimates of the average treatment effect from nonattainment designations shown in Table 3 in Currie et al. (2023) using our data at the block level with block population weights. Column 1 replicates original Column 1, Column 2 replicates original Column 5. Pollution data is from Di et al. (2021).

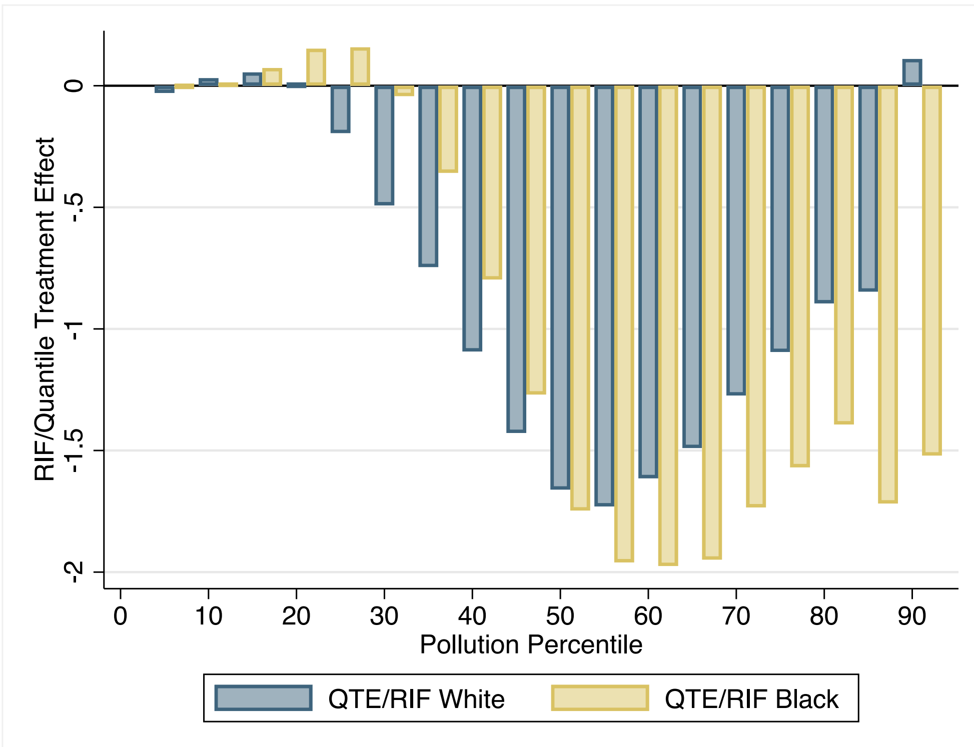


Figure A.17: Replication of Figure 8 in CVW

Notes: This figure replicates Figure 8 in Currie et al. (2023) using the block level data from our paper. It plots regression coefficients from 38 separate regressions, 19 for each race, where the dependent variable consists of the RIF-Quantile transformation of the respective $PM_{2.5}$ vigintile (indicated by the x-axis). The regression model controls for county fixed effects and state-by-year fixed effects. Regressions are weighted by block level population counts, linearly interpolated between 2000, 2010 and 2020, and errors are clustered by commuting zone. Pollution data from Di et al. (2021).

Table A.17: Replication of Table 4 in CVW

$PM_{2.5}$ Quantile Bin	Actual $PM_{2.5}$ in 2005	Actual $PM_{2.5}$ in 2015	White Counterfactual $PM_{2.5}$ in 2015 Without CAA	Black Counterfactual $PM_{2.5}$ in 2015 Without CAA
5	5.38	4.22	4.22	4.22
10	7.94	5.58	5.58	5.58
15	8.97	6.22	6.21	6.22
20	9.7	6.71	6.7	6.7
25	10.36	7.11	7.11	7.09
30	10.91	7.45	7.49	7.42
35	11.43	7.75	7.85	7.76
40	11.92	8.01	8.19	8.11
45	12.36	8.25	8.52	8.48
50	12.74	8.47	8.88	8.87
55	13.1	8.69	9.21	9.3
60	13.46	8.89	9.53	9.76
65	13.82	9.09	9.81	10.05
70	14.18	9.29	10.04	10.39
75	14.55	9.52	10.2	10.55
80	14.95	9.78	10.4	10.82
85	15.34	10.13	10.62	11.12
90	15.85	10.71	11.19	11.85
95	17.35	12.55	12.51	12.97

Main Counterfactual incl. 2005-2015 Mobility Responses		
	Original numbers reported in CVW	**Our data**
2005 Actual B-W Gap	1.20	1.16
2015 Counterfactual B-W Gap	0.97	0.98
Counterfactual Chg in B-W Gap	-0.23	-0.18
Actual Chg in B-W Gap	-0.59	-0.47
% Attributable to CAA	61.2	61.1

Notes: The table replicates Table 4 from Currie et al. (2023) using our block level data. Population counts are linearly interpolated between 2000, 2010 and 2020 to approximate the approach in Currie et al. (2023), who follow individuals in their data as they move across locations. Counterfactuals are calculated as the actual $PM_{2.5}$ levels in 2015 minus the RIF-Quantile treatment effects of nonattainment (applied in proportion to the population share living in nonattainment areas), separately for each vigintile and for each racial group. Pollution data is from Di et al. (2021).

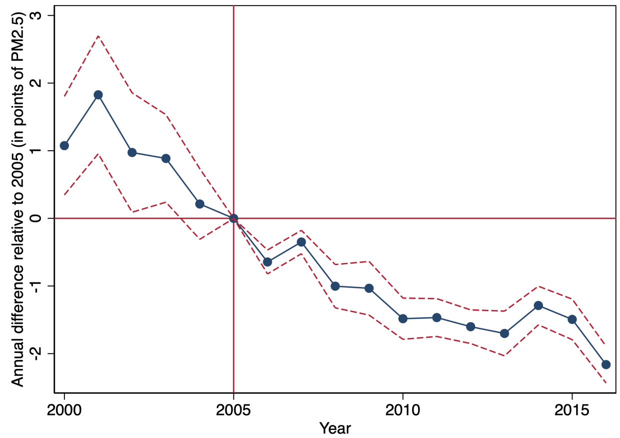
B. *Explaining differences compared to CVW* (*Currie et al. 2023*)

There are several differences between our approach and that of CVW, yet there are only two differences that are important: treatment assignment and controlling for baseline trends. We first briefly discuss minor data discrepancies that make no difference for the main findings before we turn to the two important differences.

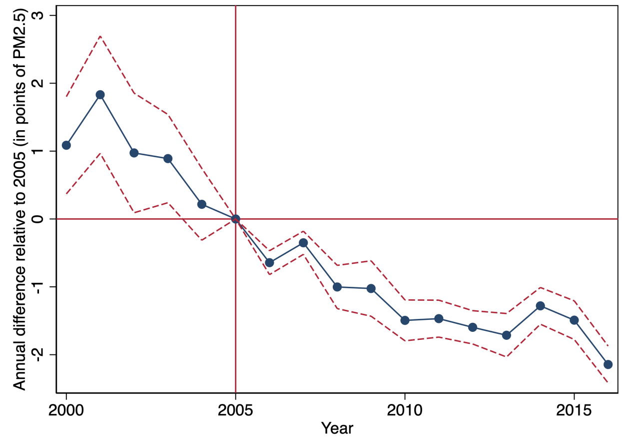
We have shown in the previous section that using our publicly available Census data recovers virtually the same estimated nonattainment effects on pollution and contribution of the CAA nonattainment areas to narrowing the Black-White exposure gap. This is reassuring and shows that any differences due to using publicly available data vs. individual level American Community Survey (ACS) samples are negligible, especially because pollution is assigned to individuals at the Census block level in both approaches. Nevertheless, we briefly list some of the data differences and use the event study to illustrate that they do not matter for this analysis.[68] First, CVW use a sample based on the Census long form as well as the 1% ACS sample. Our data is constructed from the full Census population. If samples are random, we should recover the same estimates in expectation. Second, CVW incorporate year-to-year mobility through the annual ACS samples while we use fixed 2010 location in our main analysis, or interpolated block populations by race using the 2000, 2010 and 2020 Census in the preceding section or for robustness in Figure A.1, for example. The event study graph is virtually indistinguishable when using interpolated vs. constant 2010 population as shown in Figure A.18a, compared to Figure A.21. Third, CVW use 2000 block boundaries while we use 2010 block boundaries. We also aggregate to the tract level using block population weights, which should, however, be equivalent to running the regression at the block level for the purposes of the event study. Figure A.18b shows that the event study is virtually unchanged if we use 2000 block boundaries and run the analysis at the block level, compared to Figure A.21.

While the differences between our data and that used in CVW appear negligible, there are two important differences. The first key difference is the assignment into treatment or control of those areas that are in $PM_{2.5}$ nonattainment, but have also been in PM_{10} nonattainment previously. In CVW, all such areas are assigned into the control group, sometimes known as switcher approach. As these areas do not switch into nonattainment from being in attainment of a previous NAAQS, so are 'merely' in nonattainment of an additional NAAQS ($PM_{2.5}$), one may expect that these areas experience a lower treatment effect from the additional nonattainment assignment. A switcher approach that assigns these areas into the control group assumes a treatment effect of zero for these areas. In Figure 7 we test this assumption and show, however, that the treatment effect for these areas

[68] Note that individual data would be required to assess the contribution of individual level factors to the exposure gap, which, however explain little as shown in CVW.



(a) Interpolated population weights 2000, 2010, 2020



(b) 2000 block borders and block level analysis

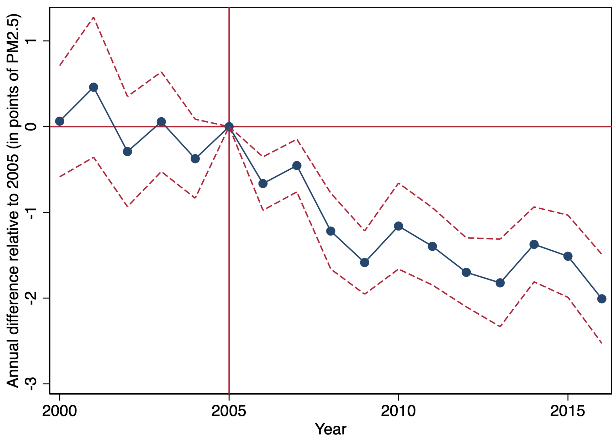
Figure A.18: Robustness of differences in pre-trends in event study

Notes: The figure replicates the event study graph from Panel (b) of Figure A.21. Panel (a) uses population weights that are interpolated between the 2000, 2010 and 2020 Census, using the IPUMS NHGIS crosswalk, instead of constant population weights at the 2010 level. Panel (b) uses borders and population counts from the 2000 Census instead of the 2010 Census. In addition, the analysis is at the Census block level, rather than pre-aggregating to the Census tract level using Census block weights as in our main analysis (the results are equivalent using either). Results are based on Di et al. (2021). Standard errors are clustered at the county level and 95% confidence intervals are shown.

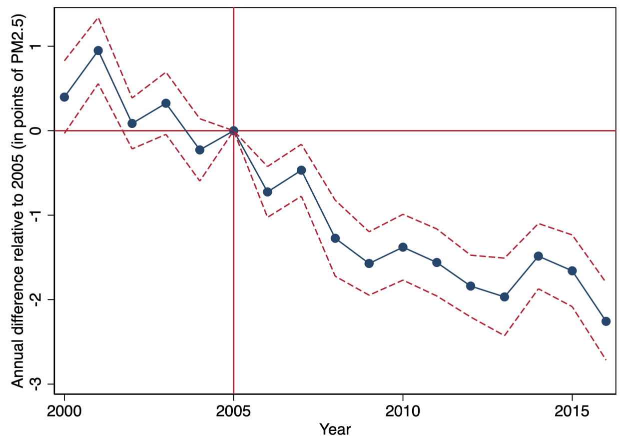
is – if anything – larger than for those areas that switched from PM_{10} attainment to $PM_{2.5}$ nonattainment. Using the treatment assignment of CVW, we can replicate their event study with insignificant pre-trends as shown in Figure A.19a (see also Figure A.16 above). This is intuitive, as $PM_{2.5}$ nonattainment areas that were also in nonattainment for PM_{10} tend to be more polluted and, as we show in our Figure 3a, also likely to exhibit the largest pre-trends, thus assigning them into the control group eliminates the pre-trends on average. If we instead drop these areas entirely, the pre-trends reappear (Figure A.19b). Note that a second, but minor difference in treatment assignment is that CVW assign entire commuting zones (CZ) into nonattainment treatment as long as a county within the CZ is in nonattainment, while we use EPA defined nonattainment areas based on air regions (i.e. the nonattainment counties). Figures A.19c and A.19d show that using CZ instead of counties based on EPA air regions have no discernible implications for the event study.

The second key difference is that we control for pollution trends based on baseline pollution as discussed in detail in our main paper. We next rerun some of the main estimations in CVW but additionally controlling for baseline pollution, similar to our DIDwb approach.[69] First, the estimated nonattainment effect is much smaller (even zero in one specification) as shown in Columns 3 and 4 of Table A.18 (Column 1 and 2 replicate the results in Table A.16 Panel b). This is in line with our main findings that when ignoring such trends, a naive DiD approach overestimates the nonat-

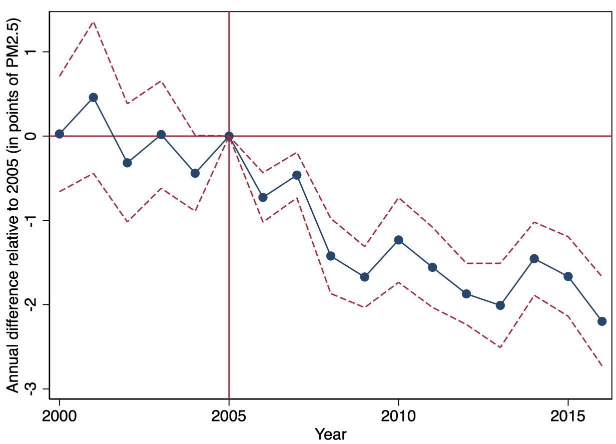
[69]That is we control for baseline pollution in 2000 interacted with year dummies in all of their panel regressions.



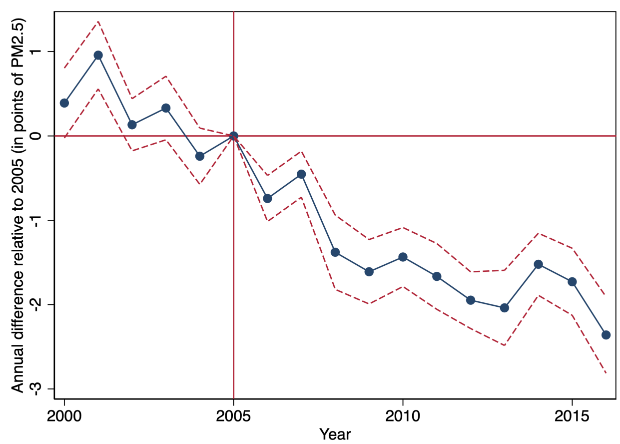
(a) Assigning double treated into control (county level)



(b) Dropping double treated (county level)



(c) Assigning double treated into control (CZ level)



(d) Dropping double treated (CZ level)

Figure A.19: Event study assigning previously treated units into the control group or dropping them

Notes: The figure replicates the event study graph from Panel (b) of Figure A.21 to facilitate comparison with the event study in Currie et al. (2023). Panel (a) assigns all nonattainment counties that were also in nonattainment with the earlier 1990 PM_{10} status in 2001-2004 into the control group (20 counties). Panel (b) instead drops these 20 counties. Panel (c) and (d) repeat these analysis of (a) and (b) respectively, but at the commuting zone level. Commuting zones in nonattainment, where all counties were previously also in nonattainment (i.e. did not switch into nonattainment), are assigned into the control group in Panel (c). These are 6 commuting zones, including, e.g. Los Angeles. In Panel (d), these commuting zones are instead dropped. Panel (b) and (d) look similar when we additionally drop counties from the control group that are in nonattainment with the PM_{10} standard, but in attainment with the $PM_{2.5}$ standard (this drops 71 counties instead of 20 counties, and is the sample we use for Table A.7). Standard errors are clustered at the commuting zone level in all four panels. Pollution data based on Di et al. (2021).

tainment effects significantly. When we use our treatment assignment instead (Columns 4-8), and control for baseline pollution, we recover effects similar as in our main analysis. These are naturally all slightly larger than the corresponding effects based on the CVW treatment assignment, where the most polluting areas with the largest effect are assigned into the control group as discussed in the previous paragraph. Second, turning to the estimation of the contribution of CAA nonattainment designations to narrowing the Black-White exposure gap, Table A.19 shows that controlling for trends based on baseline pollution lowers the estimated contribution from 61.1% (Column 1) to 18.6% (Column 2) using the same RIF analysis and CVW treatment assignment as in the previous replication section. When we use our treatment assignment instead, we find an overestimated contribution of 115.1% (Column 3) versus 22.5% with controls for trends (Column 4), similar to the pattern in our main paper. Our estimated effect in Column 4 aligns closely with our main findings in Table A.25 (i.e. the version of Table 2 that uses Di et al. (2021) data).[70] The main insight is that irrespective of using CVW or our treatment assignment, controlling for secular trends based on baseline pollution significantly reduces the estimated CAA contribution to the narrowing Black-White exposure gap, in this case by a factor of around 3-4. The large upward bias from ignoring such trends dominates the downward bias from using an approach that assigns some treated units already in nonattainment into the control group.

Table A.18: Extended replication of Table 3 in CVW

	CVW treatment assignment				Our treatment assignment			
	DiD		DiDwb		DiD		DiDwb	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PM2.5 NA CVW	-1.20	-0.76	-0.17	0.0064				
	(0.40)	(0.078)	(0.41)	(0.12)				
PM2.5 NA SS					-2.12	-1.49	-0.66	-0.20
					(0.54)	(0.45)	(0.16)	(0.10)
Observations	108,583,670				108,583,670			
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No	Yes	No	Yes	No
State-Year FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The table replicates difference-in-differences estimates of the average treatment effect from nonattainment designations shown in Table 3 in Currie et al. (2023) using our data and a panel regression at the block-by-year level. Column 1 replicates original Column 1, Column 2 replicates original Column 5. Columns 3 and 4 control for baseline $PM_{2.5}$ separately in each year. Columns 5-8 repeat the analysis but use our treatment assignment instead of the CVW treatment assignment. Pollution data is from Di et al. (2021).

[70] For our main findings we allow for simple linear heterogeneity by baseline pollution and race, rather than the RIF approach used here. In Table A.25 based on Di et al. (2021) and DiDwb, we find a contribution of 24% versus the 22.5% estimated using RIF and a slightly different time window.

Table A.19: Extended replication of Table 4b in CVW

Main Counterfactual incl. 2005-2015 Mobility Responses				
	CVW	CVW-wb	SS	SS-wb
	(1)	(2)	(3)	(4)
2005 Actual B-W Gap	1.16	1.16	1.16	1.16
2015 Counterfactual B-W Gap	.98	.78	1.23	.8
Counterfactual Chg in B-W Gap	-.18	-.38	.07	-.36
Actual Chg in B-W Gap	-.47	-.47	-.47	-.47
% Attributable to CAA	61.1	18.6	115.1	22.5
2005 NA Treatment	Switcher	Switcher	All	All
Baseline Control (DiDwb)	No	Yes	No	Yes

Notes: The table shows an extended replication of Table 4 from Currie et al. (2023) using our block level data. Column 1 shows the same RIF-based replication as in Table A.17. Column 2 adds controls for baseline $PM_{2.5}$ in each RIF-Quantile regression. Population counts are linearly interpolated between 2000, 2010 and 2020 to approximate the approach in Currie et al. (2023), who follow individuals in their data as they move across locations. Counterfactuals are calculated as the actual $PM_{2.5}$ levels in 2015 minus the RIF-Quantile treatment effects of nonattainment (applied in proportion to the population share living in nonattainment areas), separately for each vigintile and for each racial group. Pollution data is from Di et al. (2021).

A.14 Counterfactual pollution disparities with constant 2010 population

Table A.20: Pollution disparities - counterfactual gap analysis with constant 2010 population

Panel (a): Black-White Pollution Gap

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [homogeneous effect]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	13.14	11.49	1.65							
2006-2008	12.11	10.53	1.58	-0.07	282	93	78	76	68	4
2011-2013	9.65	8.64	1.01	-0.64	52	13	10	12	28	25

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [heterogeneous effect]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	13.14	11.49	1.65							
2006-2008	12.11	10.53	1.58	-0.07	413	307	155	193	201	122
2011-2013	9.65	8.64	1.01	-0.64	67	14	31	31	49	38

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [+race interactions]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	13.14	11.49	1.65							
2006-2008	12.11	10.53	1.58	-0.07	216	110	124	128	26	-24
2011-2013	9.65	8.64	1.01	-0.64	71	18	43	41	52	47

Panel (b): Urban-Rural Pollution Gap

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [homogeneous effect]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	12.41	10.18	2.23							
2006-2008	11.22	9.61	1.60	-0.63	61	20	17	17	15	1
2011-2013	9.28	7.78	1.49	-0.74	83	20	16	20	44	39

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [heterogeneous effect]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	12.41	10.18	2.23							
2006-2008	11.22	9.61	1.60	-0.63	85	63	32	39	40	23
2011-2013	9.28	7.78	1.49	-0.74	103	23	45	45	74	57

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [+urban interactions]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	12.41	10.18	2.23							
2006-2008	11.22	9.61	1.60	-0.63	82	60	37	42	40	24
2011-2013	9.28	7.78	1.49	-0.74	104	24	46	47	74	61

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 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 2010 Census, held fixed across all years. Pollution data is from Meng et al. (2019*b*).

A.15 Event study for house price trends

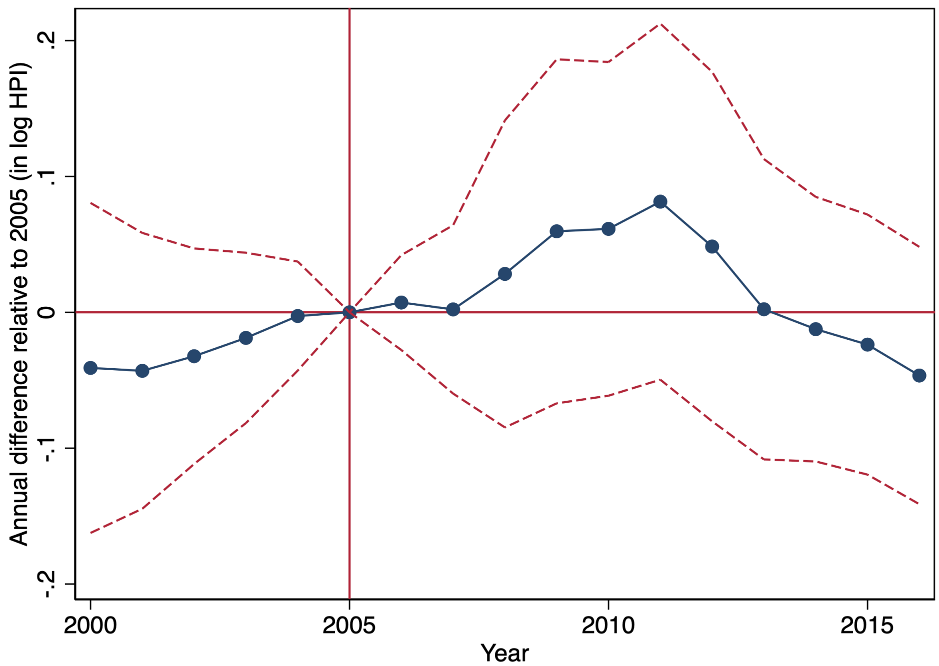


Figure A.20: Event study of house price growth nonattainment vs. attainment areas

Notes: The figure shows an event study plotting the average difference in (log) house prices between $PM_{2.5}$ nonattainment and attainment areas, normalized to 0 in 2005, as predicted from a DiDwb specification that allows for heterogeneous treatment effects by previous PM_{10} nonattainment status and baseline $PM_{2.5}$ levels in 2001-03 as in Table 3. Shown are the average treatment effects at the mean, and 95% confidence intervals are based on standard errors clustered at the county level. Pollution data is from Meng et al. (2019*b*).

A.16 Results for house price changes with commuting zone fixed effects

Table A.21: Pollution damages - instrumental variable comparison (with CZ FE)

	OLS	DiD-IV	DiDwb-IV	M1DiD-IV	M2DiD-IV	RD0-IV	RD1-IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a): Effect of $PM_{2.5}$ ***increases*** *on house price index growth 2001-03 to 2006-08*							
ΔPM2.5	-0.019	-0.025	-0.040	-0.037	-0.039	-0.053	-0.086
	(0.0053)	(0.0076)	(0.013)	(0.013)	(0.0097)	(0.024)	(0.034)
Observations	54483	54483	54483	21080	21602	5086	7936
K-P F statistic		107.2	9.67	92.7	82.9	37.4	98.3
Elasticity	-0.23	-0.30	-0.48	-0.44	-0.46	-0.64	-1.03
Panel (b): Effect of $PM_{2.5}$ ***increases*** *on house price index growth 2001-03 to 2011-13*							
ΔPM2.5	-0.022	-0.039	-0.14	-0.029	-0.042	-0.0060	-0.050
	(0.0099)	(0.0090)	(0.028)	(0.014)	(0.014)	(0.010)	(0.012)
Observations	54332	54332	54332	20990	21517	4495	19034
K-P F statistic		278.4	17.0	296.2	276.1	303.8	288.1
Elasticity	-0.27	-0.47	-1.73	-0.35	-0.50	-0.071	-0.60

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, with commuting zone fixed effects added. Standard errors in parentheses are clustered at the county level. Pollution data is from Meng et al. (2019*b*).

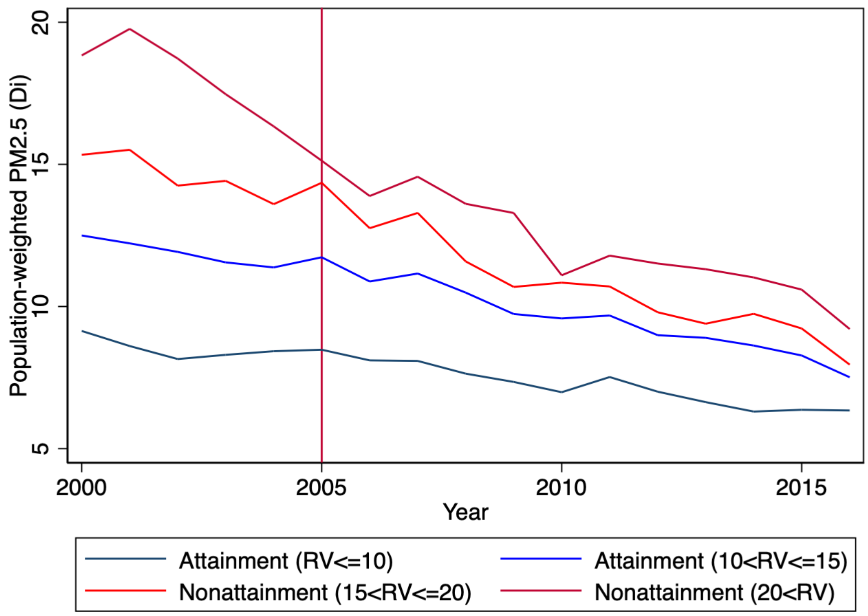
A.17 Reduced form results for house price changes

Table A.22: Reduced form effect of NA on HPI - instrumental variable comparison

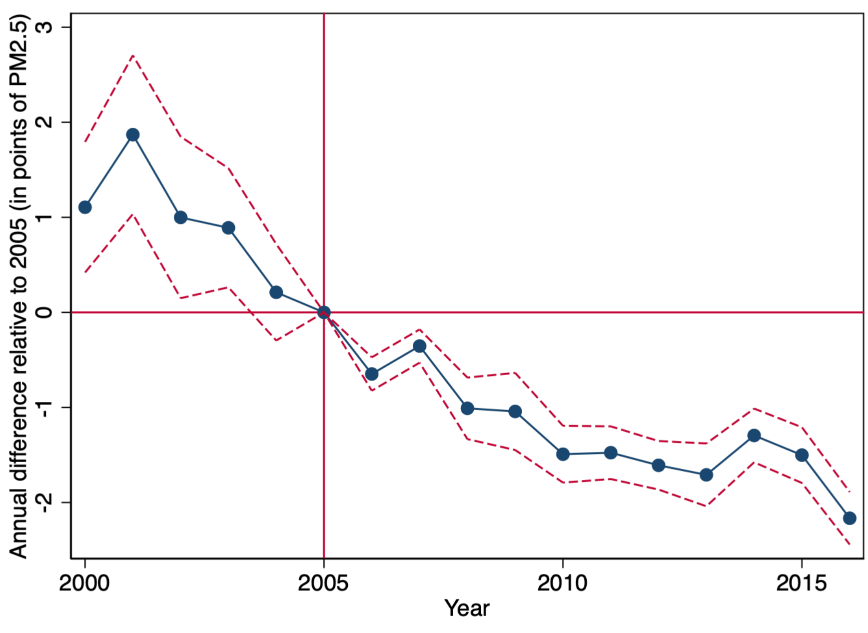
	DiD-RF	DiDwb-RF	M1DiD-RF	M2DiD-RF	RD0-RF	RD1-RF
	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a): Effect of NA on house price index growth 2001-03 to 2006-08						
NA Effect	0.057	0.142	0.057	0.045	0.087	0.008
	(0.025)	(0.030)	(0.038)	(0.044)	(0.066)	(0.120)
Observations	54529	54529	21152	21693	5087	7937
Panel (b): Effect of NA on house price index growth 2001-03 to 2011-13						
NA Effect	-0.018	0.017	-0.027	-0.023	0.062	0.118
	(0.022)	(0.024)	(0.025)	(0.026)	(0.052)	(0.057)
Observations	54378	54378	21062	21608	4496	19035

Notes: The table shows reduced form estimates from a simplified version of our instruments that only includes $PM_{2.5}$ nonattainment (NA) and the interaction with baseline $PM_{2.5}$. Average treatment effects are calculated as linear combination of coefficient estimates for the NA dummy and NA interacted with baseline $PM_{2.5}$, evaluated at the mean. Standard errors in parentheses are clustered at the county level. Pollution data is from Meng et al. (2019*b*).

A.18 Replication of Tables 1-3 and Figures 3-5 in main paper with alternative $PM_{2.5}$ data



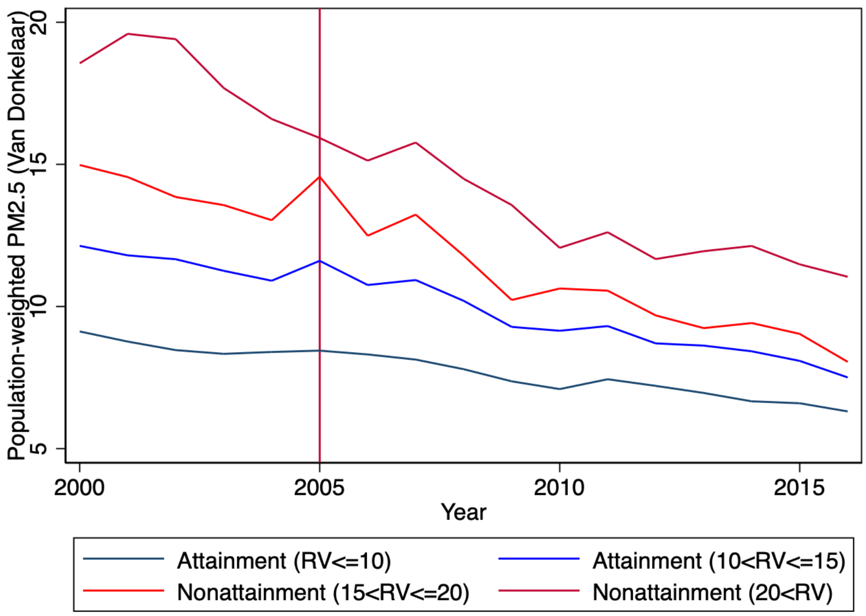
(a) Evolution of $PM_{2.5}$ grouped by EPA RV grouping



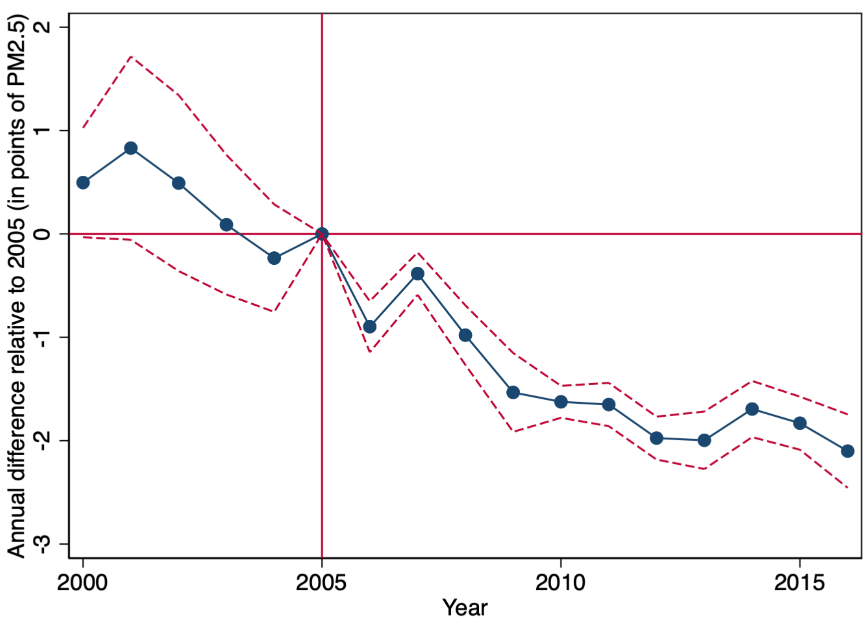
(b) Event study (annual nonattainment-attainment differences in $PM_{2.5}$)

Figure A.21: Trends in $PM_{2.5}$ and event study analysis using Di et al. (2021)

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% confidence intervals based on standard errors that are cluster-robust at the level of counties. Both Panels are based on data from Di et al. (2021).



(a) Evolution of $PM_{2.5}$ grouped by EPA RV grouping



(b) Event study (annual nonattainment-attainment differences in $PM_{2.5}$)

Figure A.22: Trends in $PM_{2.5}$ and event study analysis using van Donkelaar et al. (2021*b*)

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% confidence intervals based on standard errors that are cluster-robust at the level of counties. Both Panels are based on data from van Donkelaar et al. (2021*b*).

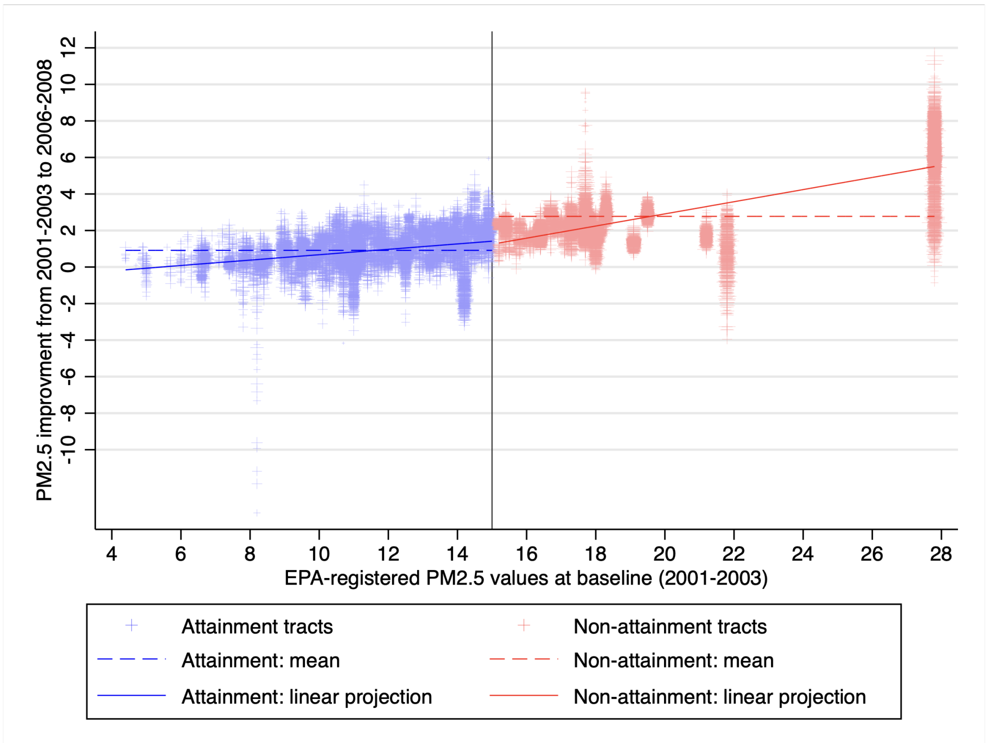


Figure A.23: Improvement in tract $PM_{2.5}$ averages and EPA-registered $PM_{2.5}$ values using Di et al. (2021)

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. The solid lines plot the linear projection of tract level $PM_{2.5}$ improvements on the EPA-registered $PM_{2.5}$ values of the nonattainment and attainment areas separately, weighted by tract population. Based on data from Di et al. (2021).

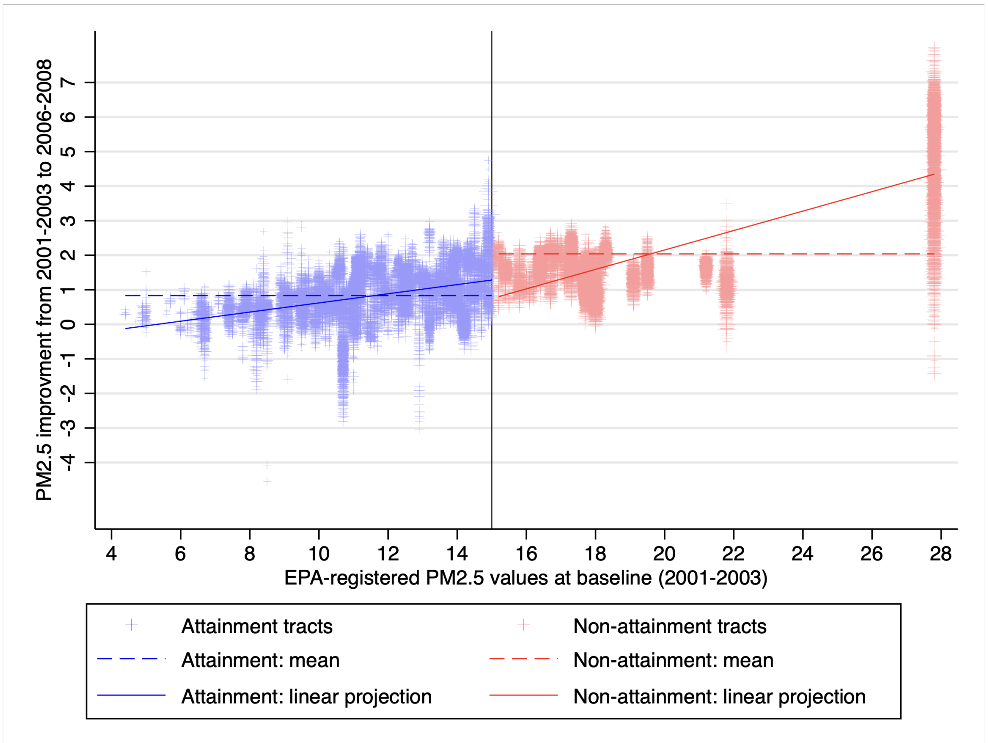


Figure A.24: Improvement in tract $PM_{2.5}$ averages and EPA-registered $PM_{2.5}$ values using van Donkelaar et al. (2021*b*)

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. The solid lines plot the linear projection of tract level $PM_{2.5}$ improvements on the EPA-registered $PM_{2.5}$ values of the nonattainment and attainment areas separately, weighted by tract population. Based on data from van Donkelaar et al. (2021*b*).

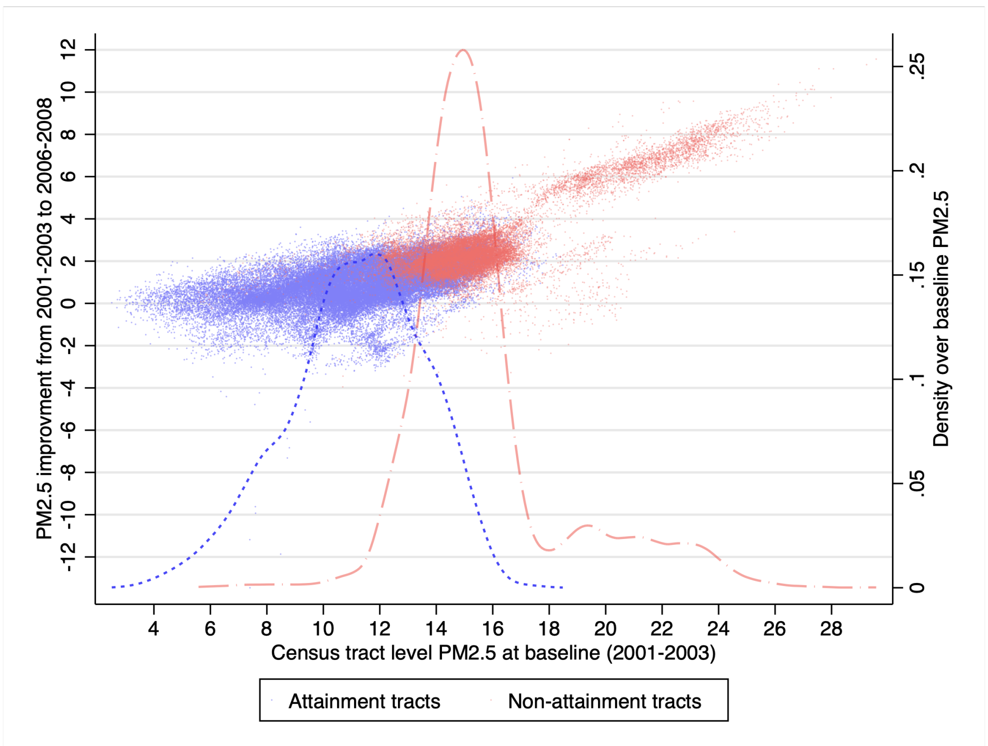


Figure A.25: Improvement in tract $PM_{2.5}$ averages and baseline $PM_{2.5}$ levels using Di et al. (2021)

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 nonattainment and attainment tracts in terms of baseline $PM_{2.5}$, weighted by tract population. The figure is based on data from Di et al. (2021).

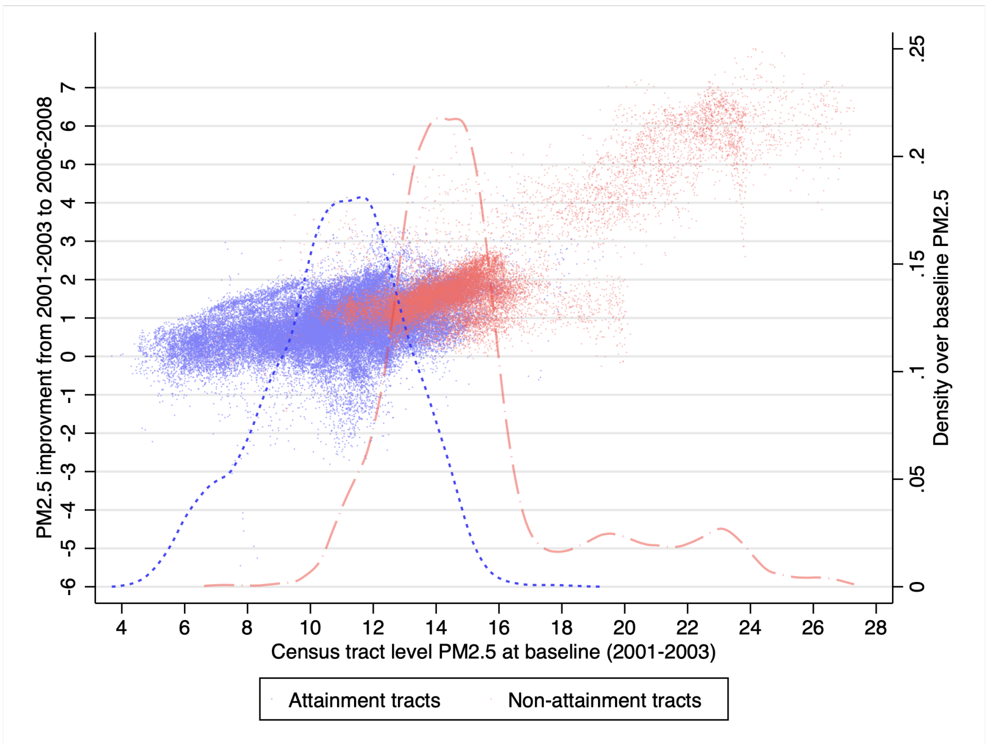


Figure A.26: Improvement in tract $PM_{2.5}$ averages and baseline $PM_{2.5}$ levels using van Donkelaar et al. (2021*b*)

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 nonattainment and attainment tracts in terms of baseline $PM_{2.5}$, weighted by tract population. The figure is based on data from van Donkelaar et al. (2021*b*).

Table A.23: Nonattainment status and changes in $PM_{2.5}$ using Di et al. (2021)

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Part A: Effect from 2001-03 to 2006-08							
Panel (a): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-1.92	-0.49	-0.32	-0.74	-1.86	-0.84	-0.31
	(0.38)	(0.13)	(0.12)	(0.40)	(0.39)	(0.58)	(0.52)
Observations	72043	72043	27827	29932	47962	5234	12738
Panel (b): Placebo Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.46	-0.0090	-0.014	0.15	-0.56	-0.12	0.077
	(0.20)	(0.20)	(0.20)	(0.20)	(0.21)	(0.23)	(0.63)
Observations	49357	49357	20460	20068	25276	3280	4626
Panel (c): Heterogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	6.27	4.11	2.42	7.41	6.33	4.03	4.76
	(0.83)	(0.85)	(0.75)	(0.87)	(0.83)	(1.20)	(1.06)
NA(x)Baseline	-0.52	-0.33	-0.19	-0.52	-0.52	-0.32	-0.34
	(0.059)	(0.062)	(0.052)	(0.061)	(0.059)	(0.068)	(0.061)
Observations	72043	72043	27827	29932	47962	5234	12738
Implied ATE	-1.92	-1.05	-0.50	-0.75	-1.86	-1.04	-0.65
10th pct	-0.61	-0.22	-0.027	0.56	-0.55	-0.23	0.22
90th pct	-4.19	-2.48	-1.30	-3.01	-4.13	-2.45	-2.15
Part B: Effect from 2001-03 to 2011-13							
Panel (d): Homogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-2.85	-0.50	-0.23	-0.93	-2.94	-0.52	-0.73
	(0.39)	(0.11)	(0.11)	(0.42)	(0.41)	(0.22)	(0.37)
Observations	72043	72043	27827	29932	47962	3743	10459
Panel (e): Placebo Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-0.91	0.26	0.33	0.45	-1.50	0.32	1.14
	(0.18)	(0.15)	(0.18)	(0.19)	(0.21)	(0.37)	(0.63)
Observations	49357	49357	20460	20068	25276	2143	4807
Panel (f): Heterogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	6.32	0.98	4.17	8.20	6.24	6.12	4.87
	(0.83)	(0.84)	(0.70)	(0.87)	(0.83)	(1.63)	(1.06)
NA(x)Baseline	-0.59	-0.11	-0.30	-0.58	-0.59	-0.44	-0.37
	(0.058)	(0.060)	(0.049)	(0.060)	(0.058)	(0.11)	(0.067)
Observations	72043	72043	27827	29932	47962	3743	10459
Implied ATE	-2.85	-0.69	-0.52	-0.94	-2.94	-0.77	-0.87
10th pct	-1.38	-0.42	0.23	0.53	-1.46	0.34	0.052
90th pct	-5.39	-1.15	-1.82	-3.48	-5.48	-2.68	-2.46

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 post-treatment periods. Each panel(x)column combination is from a separate regression as described in the text. (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists, (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 Di et al. (2021).

Table A.24: Nonattainment status and changes in $PM_{2.5}$ using van Donkelaar et al. (2021*b*)

	ATT					LATE	
	All Tracts				with RV	Optimal Bandw.	
	DiD	DiDwb	M1DiD	M2DiD	DiD	RD0	RD1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Part A: Effect from 2001-03 to 2006-08							
Panel (a): Homogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-1.22	-0.13	-0.20	-0.63	-1.21	-0.26	0.043
	(0.33)	(0.092)	(0.16)	(0.36)	(0.34)	(0.26)	(0.35)
Observations	72043	72043	28311	29808	47962	7026	15683
Panel (b): Placebo Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	-0.38	-0.15	-0.15	-0.071	-0.52	-0.071	0.029
	(0.12)	(0.13)	(0.13)	(0.14)	(0.17)	(0.44)	(0.50)
Observations	49357	49357	20285	20056	25276	1046	4626
Panel (c): Heterogeneous Treatment Effect: from 2001-03 to 2006-08							
Nonattainment	4.76	3.47	2.52	5.34	4.78	2.04	2.89
	(0.58)	(0.59)	(0.61)	(0.61)	(0.59)	(1.33)	(0.70)
NA(x)Baseline	-0.39	-0.28	-0.19	-0.39	-0.39	-0.16	-0.20
	(0.044)	(0.045)	(0.045)	(0.044)	(0.044)	(0.093)	(0.042)
Observations	72043	72043	28311	29808	47962	7026	15683
Implied ATE	-1.22	-0.70	-0.38	-0.64	-1.21	-0.39	-0.17
10th pct	-0.12	0.063	0.15	0.46	-0.11	0.054	0.39
90th pct	-3.18	-2.07	-1.33	-2.60	-3.16	-1.19	-1.17
Part B: Effect from 2001-03 to 2011-13							
Panel (d): Homogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-2.34	-0.15	-0.072	-0.92	-2.42	-0.29	-0.54
	(0.39)	(0.090)	(0.14)	(0.45)	(0.40)	(0.100)	(0.35)
Observations	72043	72043	28311	29808	47962	3743	12997
Panel (e): Placebo Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	-0.98	0.048	0.095	0.15	-1.58	0.97	1.09
	(0.16)	(0.14)	(0.18)	(0.19)	(0.18)	(0.43)	(0.58)
Observations	49357	49357	20285	20056	25276	676	3280
Panel (f): Heterogeneous Treatment Effect: from 2001-03 to 2011-13							
Nonattainment	5.75	0.60	4.95	7.15	5.67	1.10	3.10
	(0.50)	(0.51)	(0.69)	(0.56)	(0.51)	(0.53)	(0.60)
NA(x)Baseline	-0.53	-0.058	-0.35	-0.53	-0.53	-0.098	-0.26
	(0.037)	(0.039)	(0.051)	(0.038)	(0.037)	(0.037)	(0.035)
Observations	72043	72043	28311	29808	47962	3743	12997
Implied ATE	-2.34	-0.27	-0.41	-0.93	-2.42	-0.39	-0.87
10th pct	-0.86	-0.11	0.57	0.56	-0.93	-0.11	-0.14
90th pct	-4.99	-0.56	-2.17	-3.57	-5.06	-0.87	-2.17

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 post-treatment periods. Each panel(x)column combination is from a separate regression as described in the text. (1) uses simple DiD, (2) adds controls for baseline $PM_{2.5}$ (2001-03), (3) runs DiD using a sample matched (1-to-1) on baseline $PM_{2.5}$, (4) matches on baseline $PM_{2.5}$, tract population and population density (both 2000), (5) again uses simple DiD but with the limited sample of areas for which an EPA-registered $PM_{2.5}$ value exists, (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 van Donkelaar et al. (2021*b*).

Table A.25: Pollution disparities - counterfactual gap analysis using Di et al. (2021)

Panel (a): Black-White Pollution Gap

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [homogeneous effect]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	13.48	12.32	1.16							
2006-2008	11.91	10.91	1.00	-0.16	172	44	28	66	75	28
2011-2013	9.63	9	0.63	-0.53	77	14	6	25	14	20

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [heterogeneous effect]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	13.48	12.32	1.16							
2006-2008	11.91	10.91	1.00	-0.16	171	93	44	65	92	57
2011-2013	9.63	9	0.63	-0.53	77	18	14	25	20	23

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [+race interactions]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	13.48	12.32	1.16							
2006-2008	11.91	10.91	1.00	-0.16	140	62	54	35	-55	-46
2011-2013	9.63	9	0.63	-0.53	83	24	31	31	-6	-2

Panel (b): Urban-Rural Pollution Gap

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [homogeneous effect]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	12.94	11.51	1.43							
2006-2008	11.27	10.54	0.73	-0.70	72	18	12	28	31	12
2011-2013	9.33	8.64	0.69	-0.74	101	18	8	33	19	26

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [heterogeneous effect]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	12.94	11.51	1.43							
2006-2008	11.27	10.54	0.73	-0.70	85	47	23	41	47	33
2011-2013	9.33	8.64	0.69	-0.74	115	27	25	47	37	39

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [+urban interactions]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	12.94	11.51	1.43							
2006-2008	11.27	10.54	0.73	-0.70	91	53	30	47	49	33
2011-2013	9.33	8.64	0.69	-0.74	119	31	30	51	38	42

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 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 Di et al. (2021).

Table A.26: Pollution disparities - counterfactual gap analysis using van Donkelaar et al. (2021*b*)

Panel (a): Black-White Pollution Gap

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [homogeneous effect]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	13.03	11.88	1.15							
2006-2008	11.8	10.73	1.07	-0.08	228	25	37	118	48	-8
2011-2013	9.38	8.74	0.64	-0.51	66	4	2	26	8	15

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [heterogeneous effect]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	13.03	11.88	1.15							
2006-2008	11.8	10.73	1.07	-0.08	218	124	67	110	69	27
2011-2013	9.38	8.74	0.64	-0.51	64	8	10	24	11	24

	$PM_{2.5}$ exposure		Black-White Gap		Contribution of CAA (in %) [+race interactions]					
Period	Black	White	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	13.03	11.88	1.15							
2006-2008	11.8	10.73	1.07	-0.08	174	80	51	64	191	21
2011-2013	9.38	8.74	0.64	-0.51	68	12	14	28	5	39

Panel (b): Urban-Rural Pollution Gap

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [homogeneous effect]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	12.53	11.15	1.38							
2006-2008	11.19	10.23	0.96	-0.42	77	8	12	40	16	-3
2011-2013	9.17	8.26	0.91	-0.47	132	9	4	52	16	30

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [heterogeneous effect]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	12.53	11.15	1.38							
2006-2008	11.19	10.23	0.96	-0.42	92	55	31	55	31	18
2011-2013	9.17	8.26	0.91	-0.47	151	18	35	71	25	58

	$PM_{2.5}$ exposure		Urban-Rural Gap		Contribution of CAA (in %) [+urban interactions]					
Period	Urban	Rural	(levels)	(change)	DiD	DIDwb	M1DiD	M2DiD	RD0	RD1
2001-2003	12.53	11.15	1.38							
2006-2008	11.19	10.23	0.96	-0.42	95	58	36	59	33	24
2011-2013	9.17	8.26	0.91	-0.47	148	15	33	68	27	56

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 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 van Donkelaar et al. (2021*b*).

Table A.27: Pollution damages - instrumental variable comparison using Di et al. (2021)

	OLS (1)	DiD-IV (2)	DiDwb-IV (3)	M1DiD-IV (4)	M2DiD-IV (5)	R0-IV (6)	R1-IV (7)
*Panel (a): Effect of $PM_{2.5}$ **increases** on house price index growth 2001-03 to 2006-08*							
ΔPM2.5	-0.028	-0.048	-0.21	-0.048	-0.081	-0.083	0.25
	(0.014)	(0.0099)	(0.024)	(0.042)	(0.0075)	(0.093)	(0.060)
Observations	54529	54529	54529	20959	22631	3882	9729
K-P F statistic		90.1	17.6	11.9	48.6	8.51	47.5
Elasticity	-0.35	-0.60	-2.62	-0.60	-1.03	-1.04	3.21
*Panel (b): Effect of $PM_{2.5}$ **increases** on house price index growth 2001-03 to 2011-13*							
ΔPM2.5	-0.0037	-0.010	-0.15	0.033	-0.034	-0.025	0.055
	(0.0087)	(0.011)	(0.023)	(0.037)	(0.012)	(0.11)	(0.013)
Observations	54378	54378	54378	20867	22557	2965	7911
K-P F statistic		189.9	17.3	17.5	71.7	40.8	258.2
Elasticity	-0.046	-0.13	-1.89	0.42	-0.43	-0.31	0.69

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 A.23. Standard errors in parentheses are clustered at the county level. Pollution data is from Di et al. (2021).

Table A.28: Pollution damages - instrumental variable comparison using van Donkelaar et al. (2021*b*)

	OLS (1)	DiD-IV (2)	DiDwb-IV (3)	M1DiD-IV (4)	M2DiD-IV (5)	R0-IV (6)	R1-IV (7)
*Panel (a): Effect of $PM_{2.5}$ **increases** on house price index growth 2001-03 to 2006-08*							
ΔPM2.5	-0.039	-0.067	-0.19	-0.032	-0.089	0.13	0.090
	(0.019)	(0.012)	(0.039)	(0.056)	(0.0086)	(0.044)	(0.089)
Observations	54529	54529	54529	21287	22442	5087	11963
K-P F statistic		96.4	23.9	264.9	152.8	41.7	42.9
Elasticity	-0.47	-0.81	-2.31	-0.39	-1.09	1.55	1.09
*Panel (b): Effect of $PM_{2.5}$ **increases** on house price index growth 2001-03 to 2011-13*							
ΔPM2.5	-0.0062	-0.012	-0.18	0.033	-0.032	-0.10	0.054
	(0.010)	(0.012)	(0.054)	(0.031)	(0.013)	(0.099)	(0.014)
Observations	54378	54378	54378	21199	22363	2965	9902
K-P F statistic		408.9	18.1	55.0	202.3	908.5	1038.9
Elasticity	-0.075	-0.14	-2.24	0.40	-0.39	-1.26	0.66

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 A.24. Standard errors in parentheses are clustered at the county level. Pollution data is from van Donkelaar et al. (2021*b*).

REFERENCES

Abadie, A., Diamond, A. & Hainmueller, J. (2010), 'Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program', *Journal of the American statistical Association* **105**(490), 493–505.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W. & Wager, S. (2021), 'Synthetic difference-in-differences', *American Economic Review* **111**(12), 4088–4118.

Calonico, S., Cattaneo, M. D. & Farrell, M. H. (2020), 'Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs', *The Econometrics Journal* **23**(2), 192–210.

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014), 'Robust nonparametric confidence intervals for regression-discontinuity designs', *Econometrica* **82**(6), 2295–2326.

Chay, K. Y. & Greenstone, M. (2005), 'Does air quality matter? evidence from the housing market', *Journal of Political Economy* **113(2)**, 376–424.

Currie, J., Voorheis, J. & Walker, R. (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.

Curtis, E. M. (2018), 'Who loses under cap-and-trade programs? the labor market effects of the NOx budget trading program', *Review of Economics and Statistics* **100**(1), 151–166.

Deschenes, O., Greenstone, M. & Shapiro, J. S. (2017*a*), 'Defensive investments and the demand for air quality: Evidence from the NOx budget program', *American Economic Review* **107**(10), 2958–2989.

Deschenes, O., Greenstone, M. & Shapiro, J. S. (2017*b*), 'Replication data for: Defensive investments and the demand for air quality: Evidence from the NOx budget program', American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor], `http://doi.org/10.3886/E112938V1`.

Di, Q., Wei, Y., Shtein, A., Hultquist, C., Xing, X., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J. et al. (2021), 'Daily and annual PM2.5 concentrations for the contiguous United States, 1-km grids, v1 (2000–2016) [dataset]', Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC), `https://doi.org/10.7927/0rvr-4538`, accessed February 2, 2022.

EPA (2022), *AirData Annual Summary Data*, U.S. Environmental Protection Agency. Available via https://www.epa.gov/outdoor-air-quality-data. Accessed February 1, 2022.

EPA (2023), *Clean Air Markets Program Data*, U.S. Environmental Protection Agency. Washington, DC: Office of Atmospheric Protection, Clean Air Markets Division. Available from EPA's Air Markets Program Data web site: https://campd.epa.gov. Accessed February 2, 2023.

Jbaily, A., Zhou, X., Liu, J., Lee, T.-H., Kamareddine, L., Verguet, S. & Dominici, F. (2022), 'Air pollution exposure disparities across us population and income groups', *Nature* **601**(7892), 228–233.

Meng, J., Li, C., Martin, R. V., van Donkelaar, A., Hystad, P. & Brauer, M. (2019), 'Historical PM2.5 across North America [dataset]', 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.

van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., Hsu, N. C., Kalashnikova, O. V., Kahn, R. A., Lee, C. et al. (2021), 'Surface pm2.5 data (v5.gl.03) [dataset]', Washington University of St. Louis Atmospheric Composition Analysis Group, `https://sites.wustl.edu/acag/datasets/surface-pm2-5/`, accessed February 2, 2022.