STAT 156 Final Project

Yixin Feng, Jaeeun Park, Nikhil Shanbhag

November 2024

1 Introduction

We aim to replicate the findings of the paper "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality" by Maximilian Auffhammer and Ryan Kellogg [1], which analyzed the effectiveness of gasoline regulations aimed at reducing ground-level ozone pollution, focusing on the impact of these regulations on air quality in the United States. Ozone is a harmful pollutant that is linked to respiratory disease and crop damage, formed through chemical reactions between volatile organic compounds (VOCs) and oxides of nitrogen (NOx), largely influenced by human activities. Despite regulation of emissions for the past four decades, there are many regions that still experience ozone concentrations exceeding the Environmental Protection Agency (EPA) standards. Therefore, in order to regulate the chemical composition of gasoline, the US has implemented a set of regulations to target VOC emissions from mobile sources.

1.1 Gasoline regulations

The three main types of regulations that were implemented are:

1. Reid Vapor Pressure (RVP) Regulations

Regulations to reduce ground-level pollution began with the introduction of Reid vapor pressure (RVP) in 1989. This measures the intensity of volatile organic compounds (VOCs) emitted from gasoline, either through vehicle exhaust or evaporation, contributing to ozone formation. RVP regulation was implemented in two phases: phase I (1989-1991) established limits for each state based on EPA modeling of VOC emissions reduction needs, with limits ranging from 9.0 psi - 10.5 psi. Phase II (starting in 1992) set a nationwide summer RVP limit of 9.0 psi and introduced a stricter 7.8 psi limit in southern states with ozone nonattainment areas. Some areas adopted even tighter limits, as low as 7.0 psi.

2. Reformed Gasoline (RFG) Regulations

Federal reformed gasoline (RFG) regulations mandated by the Clean Air Act Amendments of 1990 were enforced by the EPA starting in 1995. It targets areas with severe ozone pollution, and RFG was implemented in two phases: phase I began in 1995 and phase II began in 2000. Phase I RFG required a maximum of 1% benzene, a toxic VOC, and at least 2% oxygen, using additives like MTBE or ethanol. It also mandated a 15% reduction in VOC and toxic air pollutants (TAPs), including benzene, compared to conventional gasoline, and a NOx emissions cap. Phase II RFG increased the VOC reduction to 25% and TAP reduction to 20%, and also called for a 5.5% reduction in NOx emissions year-round.

3. California Air Resources Board (CARB) gasoline

California and Arizona have stricter gasoline standards than RFG, such as California's CARB gasoline, implemented in 1996. This includes a 1% benzene limit, a 7.0 psi RVP limit during summer, and year-round content standards for olefins (6%) and aromatic hydrocarbons (25%), which are more reactive in forming ozone. CARB also requires an 80% sulfur content reduction to lower sulfur dioxide and NOx emissions.

1.2 Key Findings by Auffhammer and Kellogg

Auffhammer and Kellogg investigate how gasoline content regulations, and specifically the introduction of RFG, RVP, and CARB gasoline standards affect air quality, particularly ozone concentrations. They conclude that stricter gasoline regulations such as Federal RFG, and especially CARB, lead to significant reductions in ozone concentrations. Specifically, they found that CARB gasoline standards led to a reduction in log(ozone) of -0.095, which is the equivalent of a 9.1% decrease in absolute ozone concentrations. RVP Phase I and RVP Phase II were also associated with changes in ozone levels, but the effects were smaller and statistically insignificant.

2 Data Collection

The study uses ozone concentration data from the EPA's Air Quality Standards database for 1989-2003 [2], which includes hourly ozone readings from air quality monitors across the US. There are two main daily measures: daily maximum concentration and daily eight-hour maximum. The daily eight-hour maximum was calculated by averaging the ozone levels over all possible eight-hour periods in a day and taking the highest average among them. This aligns with the EPA's air quality standards, which had set a daily maximum of 0.12 ppm until June 2004, after which it shifted to an eight-hour maximum of 0.08 ppm.

Monitors with fewer than nine hours of data between 9 AM and 9 PM and monitor-years missing more than 25% of summer ozone observations between June 1 and August 31 were excluded from the dataset. The final dataset therefore includes 1,144,025 monitor-days, with approximately 80% of monitors being in rural and suburban areas, and 20% in urban areas. Monitored counties were categorized by fuel content regulations, including RVP Phase I, RVP Phase II, federal RFG, and CARB.

In order to control for the influence of weather on ozone concentrations, their analysis uses data from the National Climatic Data Center's Cooperative Station Data (NOAA 2008) [3], which provides daily minimum and maximum temperatures, rains, and snowfall data from over 20,000 weather stations across the US. Any missing weather values were imputed through regression, where the number of stations was iteratively reduced using the nine closest stations, until all missing data was filled. This algorithm was validated by removing a randomly chosen 10% subset of temperature data and comparing the predicted values to actual values, which resulted in a correlation coefficient above 0.95. This means the imputation of data was accurate, allowing for reliable weather controls. A summary of our data points for each covariate is shown below.

Data Cleaning

The code follows a systematic approach for filtering, merging, and transforming multiple datasets in this data cleaning process, to prepare the datasets for empirical analysis of the impacts of gas content regulations on air quality. The first steps pertain to data quality and comparability between observations. The code then first restricts the sample to years before 2004, filters out days and monitor years with fewer than 5 observations per hour, and further limits the sample to summer months (June, July, August) when ozone levels are of greater interest. The dataset can then be cleaned to remove incomplete or underrepresented observations, providing a dataset more reflective of the conditions under which the effects of fuel regulations on ozone metrics can be reliably captured.

Polynomial terms are generated from weather variables to more flexibly model the effects of temperature and precipitation on ozone formation. Additional income and geographic data are merged with the dataset and filtered in a consistent manner such that the important variables like ozone measures and income have meaningful and positive values. Regional dummies help classify states to census regions to capture broad geographic and policy contexts, and time trends and treatment interaction terms provide opportunities to perform difference-in-difference or panel argumentation of the model. The motivation for this cleaning and enrichment is to isolate the effect of gasoline regulations from confounding factors and generate a final dataset amenable to robust statistical inference.

Finally, we prepare the data for the main empirical analysis following data panel creation. With time trends, region classes, and weather interaction, the cleaned dataset is prepared for regression models that can robustly infer the causal effect of gasoline content regulation on air quality measures. Extending beyond the creation of visual representations, we are now prepared to generate figures and graphs that illustrate trends and comparisons between treatment and control groups over time, using a meticulously curated dataset.

Table 1 below gives summary statistics for our dataset, including each year recorded and the number of observations and counties included in that year. Additionally, we focus on the number of active monitors that are in each county, for both urban and suburban counties, with roughly twice as many monitors placed in rural counties. Finally, for each of the four regulations, we have given a total count of the number of monitors. Our dataset perfectly replicates that of Auffhammer and Kellogg.

Table 1—Summary Statistics on Monitors and Regulation for the Summer Ozone Season (June–August)

Year	Observations/	Counts of active monitors			Regulations			
	(counties)	Total	Urban	Rural	RVP1	RVP2	RFG95	CARB
1989	63,076/(418)	720	153	244	371	0	0	0
1990	66,108/(436)	751	157	268	381	0	0	0
1991	69,164/(451)	782	151	297	395	0	0	0
1992	69,848/(452)	789	155	300	0	132	0	0
1993	72,606/(469)	815	167	301	0	140	0	0
1994	74,440/(473)	835	163	316	0	140	0	0
1995	77,007/(477)	865	170	330	0	111	111	0
1996	76,462/(471)	854	165	330	0	76	106	48
1997	78,283/(478)	873	166	336	0	76	108	48
1998	79,544/(487)	889	165	344	0	82	108	49
1999	80,750/(485)	899	168	344	0	87	108	49
2000	82,466/(489)	915	178	346	0	97	107	49
2001	83,781/(490)	929	178	355	0	97	108	47
2002	85,230/(495)	943	177	361	0	100	109	49
2003	85,260/(498)	945	180	362	0	101	108	50
Total	1,144,025/(NA)							
Average	76,268/(471)	854	166	322				

In order to better understand each regulation, we analyze the change in average daily maximum ozone levels over time, as displayed in Figure 3. As done in the study, we averaged the daily ozone maximum concentration for each year from 1989-2003, over the months June, July, and August. Our control group, or "Baseline" counties are the ones with 9.0 psi RVP limits under RVP Phase II while RVP counties are the ones with 7.8 psi limits or lower under Phase II. The remaining two groups are the RFG and CARB counties, which implemented stricter regulations given that they have higher ozone concentrations on average. From Figure 3, it is already evident that CARB counties, which implemented regulations beginning in 1996, saw the steepest decline in average daily ozone maximum levels. However, this time-series plot alone is not enough to conclude that there was a causal effect of these regulations on the ozone levels. It is, however, an indicator of a strong association between the two.

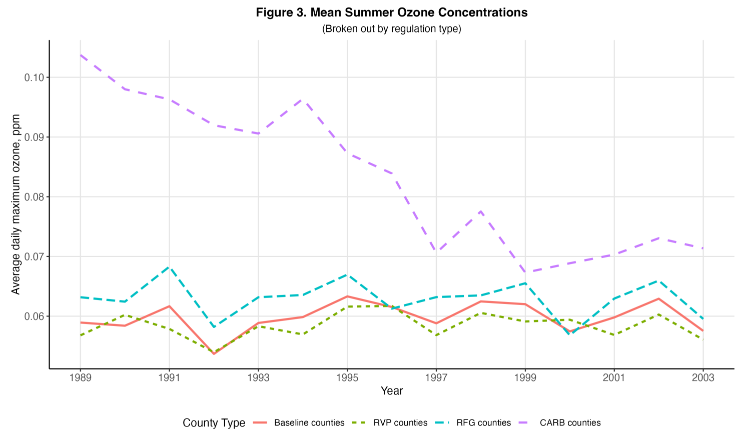


Figure 3. Mean Summer Ozone Concentrations (Broken out by regulation type)

3 Methods

The goal is to identify the causal effect of each type of each content regulation type: (i) Summer RVP for 9.0 psi counties (Baseline); (ii) Summer RVP for 9.5 or 10.5 psi counties (Phase I); (iii) Summer RVP of 7.8 psi or below counties (Phase II); (iv) Federal RFG counties; and (v) CARB counties.

Figure 4 below suggests that we should use other covariates such as weather to improve our causal identification. We plot the average daily maximum temperature as a function of time under the same conditions as Figure 3, and the similarity in the structure of each plot shows that there is a strong positive correlation between temperature and ozone levels, especially for RFG counties. This is why Auffhammer and Kellogg chose to use data related to weather in order to improve the estimators.

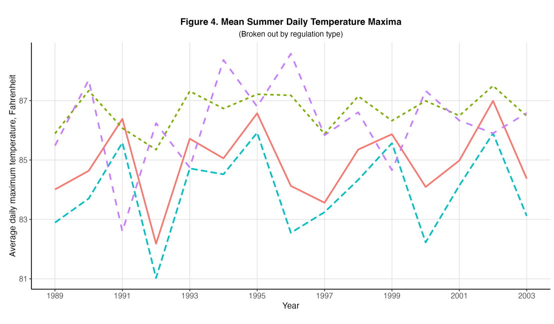


Figure 4. Mean Summer Daily Temperature Maxima (Broken out by regulation type)

3.1 Difference-in-Differences (DiD)

Difference-in-Differences (DiD) models serve to estimate the causal effect of our treatment by comparing the changes in outcomes overtime between the treatment group and control group. This is done by computing the before-and-after differences for both groups and then subtracting the change observed in the control group from the change observed in the treatment group. This way, we can isolate the effect of the treatment from other time-related factors that care capable of influencing the outcome.

The most basic form of the model is

$$\begin{aligned}\log(\text{ozone\_max}_{it}) = {} & \alpha_1 \cdot \text{treat\_rvpI} + \alpha_2 \cdot \text{treat\_rvpII} \\ & + \alpha_3 \cdot \text{treat\_rfg} + \alpha_4 \cdot \text{treat\_CARB} \\ & + \mu_i + \eta_{ry} + \epsilon_{it},\end{aligned}$$

where log(ozone_max) is regressed on the four treatment variables RVPI, RVPII, RFG, and CARB. Each α_i term captures the effect of the corresponding regulation on the log of the maximum daily ozone concentration. The additional term μ_i represents the fixed effect of each monitor, η_{ry} represents the region-by-year fixed effect, and ϵ_{it} represents the error term. Therefore, we run a fixed effects ordinary least squares (FEOLS) regression in order to account for any bias in our estimate, including the fact that treated countries will tend to have higher levels of ozone pollution, both before and after treatment.

3.2 Regression Discontinuity (RD)

The regression discontinuity (RD) design can be used to identify the causal effects of gasoline regulations on ozone concentrations by focusing on the discontinuous changes in ozone levels around the implementation dates of these regulations. This creates a sharp cutoff where emission zones are expected to change right after the policy takes effect.

The model can be written as:

$$\begin{aligned}\log(\text{ozone\_max}_{it}) = {} & \alpha_1 \cdot \text{treat\_rfg}_{ct} + \alpha_2 \cdot \text{treat\_rvpI}_{ct} + \alpha_3 \cdot \text{treat\_rvpII}_{ct} \\ & + \alpha_4 \cdot \text{treat\_CARB}_{ct} + \beta_1 \cdot \text{TempMax}_{it} + \beta_2 \cdot \text{Rain}_{it} \\ & + f(\text{income}_{ct}, \text{other\_controls}) + \mu_i + \epsilon_{it},\end{aligned}$$

where:

- TempMax_{it} is the maximum temperature recorded at location i on day t.
- Rain_{it} is the amount of rainfall at location i on day t.
- $f(\text{income}_{ct}, \text{other\_controls})$ is a function that includes the average income variable and any other potential control variables that can affect ozone concentrations, including population density or air quality indices.

- μ_i is the individual fixed effects for each monitoring station or region, which accounts for the time-invariant characteristics of each monitor.
- ϵ_{it} is the error term capturing random shocks to ozone levels at time t in region i.

3.3 Implementation of DiD and RD Design

In order to measure the impact of confounders such as weather shocks, temperature, and precipitation factors, we implemented the following 8 models, as did the study. Note that the first 5 models use the maximum ozone concentration as the outcome whereas the last 3 models use the daily eight hour ozone maximum as the outcome. The outcomes of their implementation are shown in Section 4 in Tables 2, 3, and 4.

(1) Regress log(ozone_max) on each regulation type and the monitor FEs, accounting for the region-year FEs.

$$\begin{aligned}\log(\text{ozone\_max}_{it}) &= \beta_1 \cdot \text{treat\_rvpI}_{it} + \beta_2 \cdot \text{treat\_rvpII}_{it} \\ &+ \beta_3 \cdot \text{treat\_rfg}_{it} + \beta_4 \cdot \text{treat\_CARB}_{it} \\ &+ \alpha_i + \gamma_{r,t} + \epsilon_{it}.\end{aligned}$$

(2) Regress log(ozone_max) on each regulation type, monitor FEs, region-year FEs, DOY interaction, and weather controls.

$$\begin{aligned}\log(\text{ozone\_max}_{it}) &= \beta_1 \cdot \text{treat\_rvpI}_{it} + \beta_2 \cdot \text{treat\_rvpII}_{it} \\ &+ \beta_3 \cdot \text{treat\_rfg}_{it} + \beta_4 \cdot \text{treat\_CARB}_{it} \\ &+ \alpha_i + \gamma_{r,t} + \gamma_{r,\text{DOW}} + \gamma_{r,\text{DOY}} \\ &+ \delta_1 \cdot \text{TempMax}_{it} + \delta_2 \cdot \text{TempMin}_{it} \\ &+ \delta_3 \cdot \text{Rain}_{it} + \delta_4 \cdot \text{Snow}_{it}. + \epsilon_{it}\end{aligned}$$

(3) Same as model 2, but add the income term.

$$\begin{aligned}\log(\text{ozone\_max}_{it}) &= \beta_1 \cdot \text{treat\_rvpI}_{it} + \beta_2 \cdot \text{treat\_rvpII}_{it} \\ &+ \beta_3 \cdot \text{treat\_rfg}_{it} + \beta_4 \cdot \text{treat\_CARB}_{it} \\ &+ \beta_5 \cdot \text{income}_{it} \\ &+ \alpha_i + \gamma_{r,t} + \gamma_{r,\text{DOW}} + \gamma_{r,\text{DOY}} \\ &+ \delta_1 \cdot \text{TempMax}_{it} + \delta_2 \cdot \text{TempMin}_{it} \\ &+ \delta_3 \cdot \text{Rain}_{it} + \delta_4 \cdot \text{Snow}_{it} + \epsilon_{it}.\end{aligned}$$

(4) Same as model 3, but add regulation-region trends.

$$\begin{aligned}\log(\text{ozone\_max}_{it}) = {} & \sum_{j=1}^{4} \beta_j \cdot \text{treat\_var}_{j,it} + \beta_5 \cdot \text{incomeD}_{it} + \sum_{k=6}^{18} \beta_k \cdot \text{TrendVar}_{k,it} \\ & + \sum_{l=19}^{48} \beta_l \cdot \text{TempVar}_{l,it} + \sum_{m=49}^{77} \beta_m \cdot \text{YearVar}_{m,it} \\ & + \sum_{n=78}^{83} \beta_n \cdot \text{DOWXregVar}_{n,it} + \epsilon_{it}.\end{aligned}$$

(5) Same as model 4, but add regulation-region quadratic trends.

$$\begin{aligned}\log(\text{ozone\_max}_{it}) = {} & \sum_{j=1}^{4} \beta_j \cdot \text{treat\_var}_{j,it} + \beta_5 \cdot \text{incomeD}_{it} + \sum_{k=6}^{16} \beta_k \cdot \text{TrendRVPVar}_{k,it} \\ & + \sum_{l=17}^{18} \beta_l \cdot \text{TrendCARBVar}_{l,it} + \sum_{m=19}^{36} \beta_m \cdot \text{QTrendRVPVar}_{m,it} \\ & + \sum_{n=37}^{48} \beta_n \cdot \text{TempVar}_{n,it} + \sum_{p=49}^{77} \beta_p \cdot \text{YearVar}_{p,it} \\ & + \sum_{q=78}^{83} \beta_q \cdot \text{DOWXregVar}_{q,it} + \epsilon_{it}.\end{aligned}$$

(6) Same as model 1, but we use the eight hour maximum as the outcome.

$$\begin{aligned}\log(\text{8hr\_max}_{it}) = {} & \beta_1 \cdot \text{treat\_rvpI}_{it} + \beta_2 \cdot \text{treat\_rvpII}_{it} + \beta_3 \cdot \text{treat\_rfg}_{it} \\ & + \beta_4 \cdot \text{treat\_CARB}_{it} + \alpha_i + \gamma_{r,t} + \epsilon_{it}.\end{aligned}$$

(7) Same as model 2, but we use the eight hour maximum as the outcome.

$$\begin{aligned}\log(\text{8hr\_max}_{it}) = {} & \beta_1 \cdot \text{treat\_rvpI}_{it} + \beta_2 \cdot \text{treat\_rvpII}_{it} + \beta_3 \cdot \text{treat\_rfg}_{it} \\ & + \beta_4 \cdot \text{treat\_CARB}_{it} + \alpha_i + \gamma_{r,t} + \gamma_{r,\text{DOW}} + \gamma_{r,\text{DOY}} \\ & + \delta_1 \cdot \text{TempMax}_{it} + \delta_2 \cdot \text{TempMin}_{it} \\ & + \delta_3 \cdot \text{Rain}_{it} + \delta_4 \cdot \text{Snow}_{it} + \epsilon_{it}.\end{aligned}$$

(8) Same as model 5, but we use the eight hour maximum as the outcome.

$$
\begin{aligned}
\log(\text{8hr\_max}_{it}) = &\sum_{j=1}^{4} \beta_j \cdot \text{treat\_var}_{j,it} + \beta_5 \cdot \text{incomeD}_{it} + \sum_{k=6}^{19} \beta_k \cdot \text{TrendRVPVar}_{k,it} \\
&+ \sum_{l=20}^{39} \beta_l \cdot \text{QTrendVar}_{l,it} + \sum_{m=40}^{69} \beta_m \cdot \text{TempVar}_{m,it} \\
&+ \sum_{n=70}^{81} \beta_n \cdot \text{DOWXTempVar}_{n,it} + \sum_{p=82}^{94} \beta_p \cdot \text{YearVar}_{p,it} \\
&+ \sum_{q=95}^{99} \beta_q \cdot \text{RWDOWXregVar}_{q,it} + \epsilon_{it}.
\end{aligned}
$$

3.4 Key Model Assumptions

For the DiD model, the key assumption for identifying the causal effect is that county-specific unobserved factors affecting ozone concentrations remain constant over time. This means the following holds:

$$E[\text{treat\_var}_{j,it} \cdot \epsilon_{it} \mid \text{site\_id}, \text{year}, \text{state\_code}] = 0.$$

This may not hold if treated counties experience differential trends in ozone levels compared to control counties, which could be caused by urbanization or pollution control measures, leading to major confounding. As such, the estimate of the treatment effect would be biased, and the model would should add variables to control for these time-varying factors.

Additionally, the DiD model has an identification assumption that unobserved factors are not correlated with the treatment when conditioned on covariates. For example, the following holds:

$$E[\text{treat\_var}_{j,it} \cdot \epsilon_{it} \mid \text{TempMax}, \text{Rain}, \text{urban}, \text{RVPI}, \mu_i, \eta_{ry}] = 0.$$

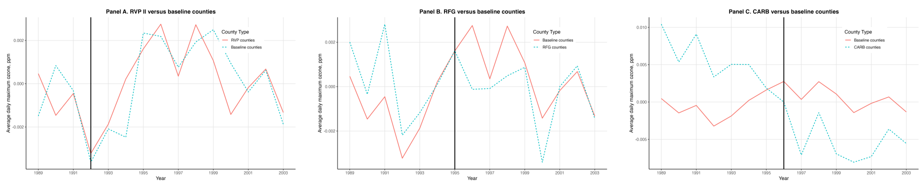
We can condition on any other treatment or temperature variable. This is why we must construct a non-linear model to understand unobserved factors that may affect ozone levels, that may not be captured by covariates such as weather or temperature.

The RD approach assumes that the effect of regulations on ozone concentrations can be identified by the change in ozone within a narrow time window around the phase-in of each regulation. This approach allows unobserved factors to affect ozone non-linearly over time, as long as they do not cause a discontinuity at the point of implementing the regulation. This corrects for the main flaw of the DiD model, which is that unobserved factors follow a linear time trend. Therefore, using the RD model in this study allowed the Auffhammer and Kellogg to account for the non-linear trend of unobserved confounders right after a new regulation is implemented.

4 Replication Results

The plots below in Figure 5 confirm the effectiveness of RFG and especially CARB emissions, along with the ineffectiveness of RVP regulations in lowering daily ozone concentrations.

4.1 DiD Estimation Results



(a) RVP II vs. baseline counties

(b) RFG vs. baseline counties

(c) CARB vs. baseline counties

Figure 5: Summer ozone concentrations for different county comparisons: RVP II, RFG, and CARB versus baseline counties. Values are averaged residuals after weather effects are removed. The panels illustrate annual residuals of ozone concentrations by regulation type.

The implementation of each regression is shown in Table 2, 3, and 4 below, with the table specifying whether or not each covariate was accounted for. Each column represents the model number and each row represents the regressand. The coefficients show the how much each regulation either increased or decreased the log(ozone_max) and log(8hr_max) along with the R-squared value, which reflects the percentage of variance in the log(ozone_max) and log(8hr_max) can be explained by each model, adjusted for the number of predictors.

Our results from Table 2 largely agree with Auffhammer and Kellogg in that CARB gasoline and more generally Federal RFG have consistently negative coefficients that are significant at the 1% level. For CARB regulations, we see an approximate 8 to 9% reduction in ozone levels and for Federal RFG regulations, we see an approximate 3 to 5% reduction in ozone levels. Therefore, it is evident that reformulated gasoline policies are effective in reducing ozone levels, with higher levels of reduction for more strict regulation policies. The inclusion of county income as a control allowed us to understand how socioeconomic factors can impact ozone levels in certain counties, and model (3) showed that counties with higher income have significantly lower ozone levels at the 1% level. This makes sense, given that counties with higher income have more resources to invest in environmentally-friendly technologies, making this variable a confounder.

When broken down by county type (rural, suburban, or urban) in Table 3, we see the impact of RFG remains moderate yet statistically significant, reducing

Table 2—Difference-in-Differences Estimation Results

	Dependent var: ln(daily maximum ozone concentration)					ln(daily max 8 hour concentration)		
Regressand	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RVP Phase I:	0.016	0.018	0.020	0.010	0.009	0.018	0.021	0.011
9.5 or 10.5 psi	(0.010)	(0.012)	(0.012)	(0.013)	(0.015)	(0.010)	(0.013)	(0.017)
RVP Phase II:	-0.007	-0.012	-0.008	-0.014	-0.022	-0.005	-0.010	-0.022
7.8 psi or lower	(0.006)	(0.007)	(0.007)	(0.009)	(0.012)	(0.006)	(0.007)	(0.013)
Federal RFG	-0.029***	-0.030***	-0.018*	-0.046***	-0.046***	-0.028***	-0.029***	-0.051***
	(0.006)	(0.007)	(0.007)	(0.012)	(0.013)	(0.006)	(0.007)	(0.014)
CARB gasoline	-0.095***	-0.089***	-0.077***	-0.081**	-0.089**	-0.090***	-0.086***	-0.090**
	(0.014)	(0.016)	(0.016)	(0.032)	(0.020)	(0.013)	(0.016)	(0.033)
County income			-1.281***	-0.206	-0.213			-0.012
($ billion)			(0.337)	(0.260)	(0.251)			(0.258)
Monitor FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region–year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region–DOW FEs	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Region FE–DOY interaction	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Weather controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Income	No	No	Yes	Yes	Yes	No	No	Yes
Regulation–region trends	No	No	No	Yes	Yes	No	No	Yes
Regulation–region quad trends	No	No	No	No	Yes	No	No	Yes
Observations					1,144,025			
R^2 (within-monitor)	0.315	0.424	0.425	0.258	0.258	0.327	0.433	0.252

* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

ozone concentrations anywhere between 0.2 to 5.1%. Meanwhile, CARB remains the move effective, reducing these concentrations by 6% to 10.5% on average. These results are especially apparent in urban and suburban counties. Even though the impact of RFG and CARB are definitely not as strong for rural areas, the results appear to not be statistically significant.

Finally, when we look at the DiD results for the monitors that are report observations every year from 1989-2003 in Table 4, it becomes apparent that every regulation type causes a decrease in ozone levels to some extent, save for RVP Phase I regulations. The authors suspect that the reason for this is because monitors are being placed in areas that are on average experiencing an increase in ozone levels. We see that the results for RFG regulations are once again significant at the 1% level, leading to the same 3 to 5% reduction in ozone levels suggested by Table 2. However, the effect of the CARB regulations are now magnified, ranging from approximately an 11% to 16% reduction in ozone concentrations, once again significant at the 1% level. This is significantly higher than what Table 2 or 3 would suggest, but once again, the result may be biased because of the monitors deliberately being placed in areas where the ozone levels are rising.

Table 3—Difference-in-Difference Estimation Results: Urban versus Suburban versus Rural

	Dependent var: ln(daily maximum ozone concentration)					
	Urban		Suburban		Rural	
Regressand	(1)	(2)	(3)	(4)	(5)	(6)
RVP Phase I:	0.019	0.019	0.029	0.011	0.020	0.004
9.5 or 10.5 psi	(0.025)	(0.019)	(0.019)	(0.014)	(0.022)	(0.018)
RVP Phase II:	0.008	0.005	-0.009	-0.023*	-0.018	-0.016
7.8 psi or lower	(0.018)	(0.014)	(0.009)	(0.011)	(0.012)	(0.011)
Federal RFG	-0.005	-0.038*	-0.025*	-0.058***	-0.025	-0.045***
	(0.017)	(0.015)	(0.010)	(0.015)	(0.014)	(0.013)
CARB gasoline	-0.063	-0.079**	-0.105***	-0.095**	-0.060**	-0.068*
	(0.032)	(0.029)	(0.026)	(0.033)	(0.022)	(0.034)
County income	-1.307**	0.438	-1.513**	-0.677**	-1.438	0.079
($ billion)	(0.445)	(0.445)	(0.549)	(0.234)	(0.835)	(0.853)
Monitor FEs	Yes	Yes	Yes	Yes	Yes	Yes
Region–year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Region–DOW FEs	Yes	Yes	Yes	Yes	Yes	Yes
Region FE–DOY interaction	Yes	Yes	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes	Yes	Yes
Regulation–region trends	No	Yes	No	Yes	No	Yes
Observations	222,982	222,982	490,539	490,539	430,504	430,504
R^2 (within-monitor)	0.475	0.279	0.420	0.272	0.402	0.236

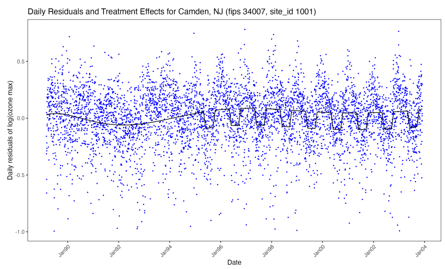
* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

Table 4—Difference-in-Differences Estimation Results: Monitors Recording Data in Every Year

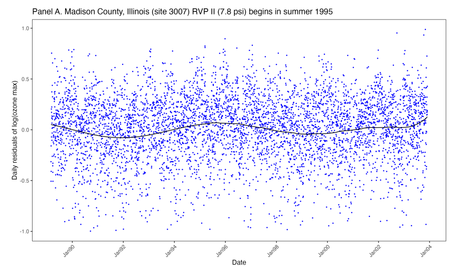
	Dependent var: ln(daily maximum ozone concentration)					ln(daily max 8 hour concentration)		
Regressand	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RVP Phase I:	-0.009	0.000	-0.001	-0.008	-0.007	-0.007	0.003	-0.006
9.5 or 10.5 psi	(0.012)	(0.015)	(0.016)	(0.014)	(0.016)	(0.012)	(0.016)	(0.017)
RVP Phase II:	-0.009	-0.016	-0.011	-0.023*	-0.033*	-0.009	-0.015	-0.033*
7.8 psi or lower	(0.007)	(0.009)	(0.009)	(0.011)	(0.013)	(0.008)	(0.009)	(0.014)
Federal RFG	-0.031***	-0.036***	-0.023*	-0.066*	-0.065***	-0.031***	-0.036***	-0.071***
	(0.007)	(0.008)	(0.010)	(0.014)	(0.016)	(0.008)	(0.009)	(0.017)
CARB gasoline	-0.148***	-0.132***	-0.108***	-0.151***	-0.159***	-0.139***	-0.124***	-0.163***
	(0.022)	(0.027)	(0.027)	(0.035)	(0.037)	(0.021)	(0.027)	(0.039)
County income	–	–	-1.677***	-0.233	-0.252	–	–	-0.042
($ billion)	–	–	(0.439)	(0.278)	(0.251)	–	–	(0.286)
Monitor FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region–year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region–DOW FEs	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Region FE–DOY interaction	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Weather controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Income	No	No	Yes	Yes	Yes	No	No	Yes
Regulation–region trends	No	No	No	Yes	Yes	No	No	Yes
Regulation–region quad trends	No	No	No	No	Yes	No	No	Yes
Observations					455,084			
R^2 (within-monitor)	0.307	0.429	0.430	0.278	0.278	0.308	0.429	0.271

* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

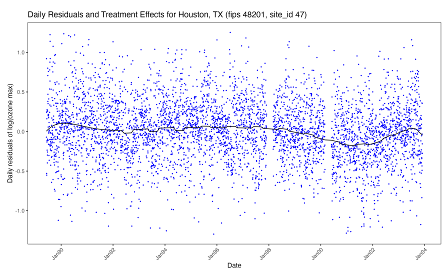
4.2 RD Design Results



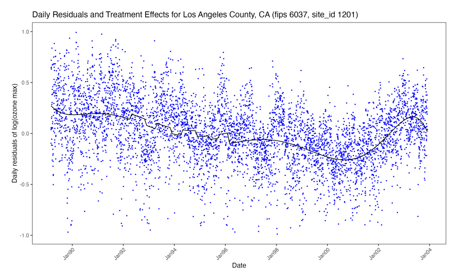
(a) Camden, NJ (fips 34007, site_id 1001)



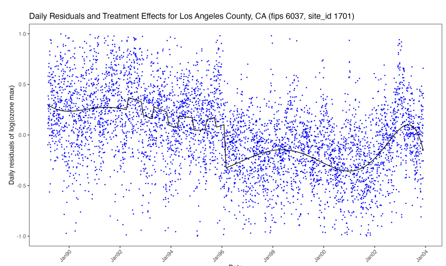
(b) Madison County, IL (site 3007), RVP II



(c) Houston, TX (fips 48201, site_id 47)



(d) Los Angeles County, CA (site 1201)



(e) Los Angeles County, CA (site 1701)

Figure 6: Daily residuals of log(ozone max) with treatment effects across different regions and regulatory implementations. Fitted lines represent predicted values from regression models, including treatment dummies and an eighth-order polynomial time trend.

Our findings in Figure 6 are consistent with those of Auffhammer and Kellogg: Panel A shows no significant effect of the 1995 RVP Phase II regulation on ozone levels at Monitor 3007 in Madison County, Illinois, with an RD estimate of +0.004, indicating no clear shift in residuals or the fitted line. Panel B highlights a significant reduction in ozone levels at Monitor 1001 in Camden County, New

Jersey, following the 1995 introduction of RFG, suggesting a strong seasonal shift post-regulation. There is no obvious shift in residuals for Panel C when RFG was imposed in 1995. All of these are also consistent with the findings of the DiD model. For Panel D and Panel E, which show the trend of Los Angeles County, we focused on the impact of implementing CARB regulations. Panel D, which focuses on coastal monitors, does not show a significant drop in ozone levels. However, Panel E shows a significant drop in the residuals of ozone levels after CARB was implemented in 1996 for the inland monitors, showing that ozone levels decreased by at least 30% for these regions.

5 Robustness Checks

Robustness checks are a critical part of observational studies as they help validate whether the estimated effects from the main analysis are reliable and hold under different assumptions. The main analysis by Auffhammer and Kellogg uses a Difference-in-Differences (DiD) approach to estimate the impact of gasoline regulations on ozone concentrations. However, the results might be influenced by time-varying unobservables—factors such as regional economic activity, technological advances, or changes in other environmental policies that evolve over time and might confound the treatment effect.

In this robustness check, we incorporate linear and quadratic regulation-specific time trends into the original DiD model to assess whether the estimated effects are sensitive to different specifications of time trends. This design is effective because it allows us to control for non-linear trends that may impact ozone levels independently of the gasoline regulations, thereby testing the robustness of the estimated regulatory effects.

5.1 Why the Robustness Check Is Needed

In the original DiD setup, the key identification assumption is that, in the absence of treatment, the treated and untreated counties would follow parallel trends in ozone concentrations over time. However, if treated counties were already experiencing different trends before the implementation of the regulations, our estimates might be biased. Adding regulation-specific linear and quadratic time trends provides a more flexible approach to control for these unobserved temporal dynamics, reducing the risk of omitted variable bias and assessing whether the observed treatment effect is genuinely attributable to the regulations.

5.2 Methodology

To assess the robustness of our findings, we augment our original Difference-in-Differences (DiD) model by incorporating linear and quadratic regulation-specific time trends. This allows us to control for unobserved, time-varying factors that may differentially affect ozone concentrations in treated and control counties over time.

Our original DiD model is specified as:

$$\begin{aligned}\log(\text{ozone\_max}_{it}) = {} & \alpha_1 \cdot \text{treat\_rvpI}_{ct} + \alpha_2 \cdot \text{treat\_rvpII}_{ct} + \alpha_3 \cdot \text{treat\_rfg}_{ct} \\ & + \alpha_4 \cdot \text{treat\_CARB}_{ct} + \mu_i + \eta_{ry} + \epsilon_{it},\end{aligned}$$

where:

- $\log(\text{ozone\_max}_{it})$ is the natural logarithm of the maximum daily ozone concentration at monitor i on day t.
- treat_rvpI_{ct}, treat_rvpII_{ct}, treat_rfg_{ct}, and treat_CARB_{ct} are binary indicators equal to 1 if county c at time t is subject to the respective gasoline regulation, and 0 otherwise.
- μ_i represents monitor fixed effects to control for time-invariant characteristics of each monitor.
- η_{ry} represents region-by-year fixed effects to control for regional trends and shocks over time.
- ϵ_{it} is the error term.

To control for potential nonlinear time trends that may affect the treated counties differently, we incorporate regulation-specific linear and quadratic time trends into the model. The augmented model becomes:

$$\begin{aligned}\log(\text{ozone\_max}_{it}) = {} & \ldots \quad \text{(Original Treatment Terms)} \\ & + \theta_j \cdot (\text{Time}_t \times \text{treat}^j_{ct}) + \phi_j \cdot (\text{Time}^2_t \times \text{treat}^j_{ct}) \\ & + \mu_i + \eta_{ry} + \epsilon_{it},\end{aligned}$$

where:

- Time_t is a continuous variable representing time (e.g., year).
- θ_j and ϕ_j are coefficients for the linear and quadratic time trends specific to each regulation j.
- treat^j_{ct} represents the treatment indicator for regulation j.

By interacting Time_t and Time^2_t with the treatment indicators, we allow the effect of each regulation to vary over time in both linear and nonlinear ways. This helps to account for any regulation-specific trends that could bias our estimates if left unaddressed.

In addition to the time trends, we include weather covariates to control for confounding factors known to influence ozone formation.

5.3 Results and Interpretation

1) Linear-Time Trend:

We found that minor changes to the estimated impact of gasoline regulations were obtained when we also included the linear-time trend. For RVP and CARB, Auffhammer and Kellogg found that their estimated impacts were slightly attenuated from the original model but were still statistically significant. This implies that any linear temporal changes in ozone levels that are not captured by observed factors were controlled for and that the treatment effects for RVP and CARB are relatively robust.

2) Quadratic-Time Trend:

Furthermore the results held greater sensitivity once we added the quadratic time trend. Auffhammer and Kellogg reported that the effect of RFG varies substantially between these specifications: with a linear-time trend, they estimate that RFG reduces log(ozone) by a statistically significant 0.036, but with a quadratic trend, this effect is only -0.019 and is not statistically significant. The estimate of the effect of RFG is sensitive to the use of quadratic time trends implying that the original effect might have been due to natural differential nonlinear trends rather than RFG itself.

5.4 Short Summary

Robustness checks show that effects of RVP and CARB regulation remain both relatively stable and statistically significant in various model specifications, while the effect of the RFG is not. Significantly the inclusion of quadratic time trends in the RFG effect increases its sensitivity. This calls into question the robustness of the original estimated effect. The implication of these results is that policy impacts should be assessed using DiD models that account for the possibility of differential trends.

The robustness checks thus suggest that the estimated impact of RFG on ozone concentrations is sensitive to the model specifications, and therefore does not further strengthen our confidence on validity of all original results. However, the same also emphasizes caution in interpreting the RFG effect, and initiates further investigation in understanding it.

6 Reanalysis using IPW Estimators

We analyzed the main results of the paper using inverse probability weighting (IPW) estimators, which are recognized as a crucial method to improve the performance of the DiD approach. The DiD method is prone to bias caused by confounders, but this can be addressed through re-weighting observations using inverse probability weighting (IPW) estimators, which adjust for confounding and enhance the robustness of treatment effect estimates. Therefore, we employ this as a core method for re-analysis, as IPW estimators correct for confounding bias in DiD models.

6.1 IPW Estimators

The IPW estimator allows us to weight individual observations using the probability of treatment.

IPW weights are defined with the propensity score, $e(X_i)$, which is the probability of receiving the treatment given covariates X_i:

$$w_i = \begin{cases} \frac{1}{e(X_i)}, & \text{if } T_i = 1 \\ \frac{1}{1-e(X_i)}, & \text{if } T_i = 0. \end{cases}$$

We ensure that three key assumptions are satisfied:

(1) Stable Unit Treatment Value Assumption (SUTVA), which means there is no interference between units, and treatment consistency across observations.

(2) Conditional independence, which means given the covariates X, the potential outcomes $Y(0)$ and $Y(1)$ are independent of treatment T.

(3) Positivity: For all covariate values, $0 < e(X_i) < 1$.

Furthermore, the IPW estimator for the treatment effect is expressed as follows:

$$\hat{\mathbb{E}}[Y(1)] = \frac{\sum_{i=1}^{n} T_i \cdot Y_i \cdot w_i}{\sum_{i=1}^{n} T_i \cdot w_i}.$$

$$\hat{\mathbb{E}}[Y(0)] = \frac{\sum_{i=1}^{n} (1 - T_i) \cdot Y_i \cdot w_i}{\sum_{i=1}^{n} (1 - T_i) \cdot w_i}.$$

There are efficiency gains when using IPW, because we can also use the doubly robust estimator [4]:

$$\hat{\tau}_{DR} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{T_i(Y_i - m_1(X_i))}{e(X_i)} + m_1(X_i) - \frac{(1 - T_i)(Y_i - m_0(X_i))}{1 - e(X_i)} - m_0(X_i) \right],$$

where $m_1(X_i)$ and $m_0(X_i)$ are the predicted outcomes for treated and control groups, respectively. This method balances confounding variables between the treatment and control groups by multiplying with the inverse. Combined with the outcome regression models, this ensure reliable inference because $\sqrt{n}(\hat{\tau}_{DR} - \tau) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

6.1.1 Estimating the Propensity Score

We assumed that urbanization introduces confounding bias on our data. To control for this, we calculated the propensity score using urbanization as the treatment variable and included a range of covariates that we had previously used in our regression models. The propensity model is constructed using a multinomial model, as the urban variable takes values of 1, 2, and 3, rather than being binary. The propensity model with multinomial model can be represented as:

$$
\begin{aligned}
\log\left(\frac{P(\text{urban}_i = c)}{P(\text{urban}_i = \text{baseline})}\right) =& \beta_0^{(c)} + \sum_{j=1}^{4} \beta_j^{(c)} \cdot \text{treat\_var}_{j,i} + \beta_5^{(c)} \cdot \text{incomeD}_i \\
&+ \sum_{k=6}^{16} \beta_k^{(c)} \cdot \text{TrendRVPVar}_{k,i} + \sum_{l=17}^{18} \beta_l^{(c)} \cdot \text{TrendCARBVar}_{l,i} \\
&+ \sum_{m=19}^{36} \beta_m^{(c)} \cdot \text{QTrendRVPVar}_{m,i} + \sum_{n=37}^{48} \beta_n^{(c)} \cdot \text{TempVar}_{n,i} \\
&+ \sum_{p=49}^{77} \beta_p^{(c)} \cdot \text{YearVar}_{p,i} + \sum_{q=78}^{83} \beta_q^{(c)} \cdot \text{DOWXregVar}_{q,i}.
\end{aligned}
$$

6.1.2 Weight Adjustment

The potential influence on the treatment assignment is managed through the weight adjustment process in the regression models, allowing for the correction of bias without relying on randomization. We adjusted IPW weights in model 5 to regress log(ozone_max) and in model 7 to regress log(8hr_max). It corrected for confounding biases and provided adjusted estimates of the impact of regulatory treatments on ozone levels. This also allowed for a clearer understanding of the correlation between air quality and regulations by controlling for the interaction between urbanization and other covariates, including temporal trends, socio-economic factors, and day-of-week effects, in this dataset.

6.2 Results of the reanalysis

Table 5—Results of Reanalysis with IPW Estimators: Correcting for Urbanization as a Confounding Factor

ln(Daily Max Ozone Concentration)		
	Model	**IPW**
RVPI	-0.007	-0.008
RVPII	-0.033	-0.031
RFG	-0.065	-0.063
CARB	-0.159	-0.164
Income	-0.252	0.104

ln(daily max 8 hour concentration)		
	Model	**IPW**
RVPI	-0.006	-0.006
RVPII	-0.033	-0.030
RFG	-0.071	-0.069
CARB	-0.163	-0.165
Income	-0.042	0.322

After reanalysis, the income coefficient changed from negative to positive. This change in Table 5 led to the conclusion that there is a positive relationship between ozone levels and income. Correcting for urbanization as a confounding factor revealed that higher income levels are associated with increased ozone concentrations. This suggests that urbanized, higher income areas may experience greater ozone levels due to increased human activities, such as industrial emissions and vehicular traffic. The reanalysis using IPW estimators provided new perspectives on the original research topic.

6.3 Pros and Cons of IPW Estimators

IPW estimators are powerful for addressing confounding variables in our observational study by weighting observations based on their propensity scores. This ensures there is a balance between the treatment and control groups, which allows our estimates to be robust even if there is no randomization. Additionally, we can use IPW to evaluate multinomial treatments and further improve our measure of the estimated effects through doubly robust estimators. In our replication, IPW was able to correct inconsistent results by turning unexpected positive coefficients into credible negative ones, allowing our conclusions to be more valid.

However, it is important to consider the limitations of IPW as a means of reanalysis. Any misspecification of the propensity score model can lead to biased results. Furthermore, low propensity scores in our study could produce large weights, leading to instability in our estimates and potentially high estimates. Therefore, it is important to specify the propensity score model carefully to avoid introducing new biases. Given the dimensions of our data, adding more confounding variables that are correlated with ozone concentrations can make IPW computationally expensive. This is also true if more treatment categories, which are regulations in this case, are added.

7 Conclusion and Limitations

We replicated the the findings of Auffhammer and Kellogg's study "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality," analyzing the impact of gasoline content regulations—Reid Vapor Pressure (RVP), Reformulated Gasoline (RFG), and California Air Resources Board (CARB) gasoline—on ozone concentrations in the United States. Using Difference-in-Differences (DiD) and Regression Discontinuity (RD) models, our findings are consistent with the study, which concluded that stricter gasoline regulations effectively reduce ozone levels, with CARB regulations yielding the most significant impact. Our analysis shows that CARB gasoline reduces ozone concentrations by 8% to 16%, while Federal RFG regulations result in a 3% to 5% decrease, both statistically significant at the 1% level. These results are consistent across urban and suburban areas, though rural areas show less pronounced effects. Robustness checks, including linear and quadratic time trends, confirm the stability of the CARB effect but reveal sensitivity in RFG's effect. Additionally, the inverse probability weighting (IPW) reanalysis provided new perspectives on the research by correcting bias utilizing urbanization data as a confounding variable. The revised conclusion that higher income levels are associated with increased ozone concentrations reinforces the need for further studies.

Despite our findings, there are several limitations that should be considered. While we controlled for weather and socioeconomic factors, there could be many other confounders including technological advancements or changes in environmental policies that our RD model did not account for. The assumptions of DiD and RD should be recognized, particularly the need for parallel trends in the former and abrupt changes to policies in the latter. However, these may not always hold true, thereby introducing bias in our estimates. Furthermore, through our robustness check, we noticed sensitivity in the results of RFG's impact to model specifications, suggesting that non-linear trends may be causing some unobserved effects. It is also important to consider the scope of our dataset. We only cover the years 1989 to 2003 and the summer months June, July, and August. However, this may overlook long-term trends in the data and the impact of recent developments in fuel or vehicle technology, such as the mass introduction of electric vehicles. Therefore, future research on this topic should look to add other confounders to the model and analyze more comprehensive datasets. This way, we will be able to account for the long-term impact of fuel technologies as they continuously evolve and the impact of regulatory policies as they continuously evolve.

References

[1] Auffhammer, Maximilian, and Ryan Kellogg. "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality." *The American Economic Review*, Oct. 2011. `https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.101.6.2687`.

[2] "Air Data: Air Quality Data Collected at Outdoor Monitors Across the US." *EPA, Environmental Protection Agency*. `https://www.epa.gov/outdoor-air-quality-data`.

[3] National Centers for Environmental Information (NCEI). "NCEI Online Store: National Centers for Environmental Information (NCEI)." *NCEI Online Store*, `https://www.ncei.noaa.gov/nespls/olstore.prodspecific?prodnum=4934`.

[4] Funk, Michele Jonsson, et al. "Doubly Robust Estimation of Causal Effects." *American Journal of Epidemiology*, vol. 173, no. 7, 2011, pp. 761–767. `https://pmc.ncbi.nlm.nih.gov/articles/PMC3070495/`. 2011 Mar 8.