8-2020

Three Essays in Environmental Economics

Sangwon Lee
University of Tennessee

To the Graduate Council:

I am submitting herewith a dissertation written by Sangwon Lee entitled "Three Essays in Environmental Economics." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Economics.

J.Scott Holladay, Major Professor

We have read this dissertation and recommend its acceptance:

Marianne Wanamaker, Maria Padilla-Romo, Seong-Hoon Cho

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Three Essays in Environmental Economics

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Sangwon Lee

August 2020

I dedicate my dissertation to my beloved family who is always on my side and supports me.

Acknowledgments

I would like to express my deepest appreciation to my supervisor Dr. J. Scott Holladay for his dedicated support and guidance in my research. He continuously provided encouragement and was always willing to assist in any way he could throughout my entire work.

I am indebted to my dissertation committee of Dr. Marianne Wanamaker, Dr. Maria Padilla-Romo, and Dr. Seong-Hoon Cho for their thoughtful comments and recommendations. I would not have been able to complete this dissertation without their help.

I would also like to express my thanks to my colleagues, Bongkyun Kim, Jinseong Park, Eunsik Chang, Bokseong Jeong, and Dulal Hossen, for generously sharing their time and ideas. I learned much through our conversations.

Finally, to my wife, Sooji, and my son, Gunyul: all your unconditional love and support motivated me to go the extra mile. Without your patience and understanding, I never would have made it. You earned this degree right along with me.

Abstract

This dissertation consists of three essays in environmental economics. The first essay examines the effectiveness of air quality information designed to reduce the public health risks associated with air pollution exposure. Using daily bike-share trip data for the metro DC area, I estimate the causal effect of air quality alerts on avoidance behavior in a regression discontinuity analysis, assigning a cutoff for treatment that triggers air quality alerts. Air quality alerts cause less bike-share trip counts and duration. Results for heterogeneous treatment effects indicate that air quality alerts mainly reduce weekend trips in the central DC, which implies that bike share-reducing effects are driven by leisure trips rather than commuting trips. The second essay investigates whether a gasoline tax can be a useful policy tool to reduce air pollutants emitted from automobiles. Using a difference-in-differences and synthetic control method, I explore variation differences in air quality across New Jersey and the other states, before and after New Jersey's gasoline tax increase in 2016. Although estimates suggest that New Jersey's air pollution levels were lower than those of the other states with the gasoline tax increase, none of these differences are statistically significant. Moreover, the gasoline tax increase was not successful in reducing gasoline consumption and vehicle miles traveled. The third essay examines how the effect of temperature on crime varies across urban and rural areas. Using a 10-year panel of monthly crime and temperature data for California cities, I identify the impact of higher temperatures on violent and property crimes in urban and rural areas. Results show that higher temperatures are correlated with more violent crimes. Urban areas have a higher number of violent crimes than rural areas, holding temperature constant. The number of violent crimes tends to increase in proportion to temperature across both areas, but the marginal effects of temperature are smaller in urban than in rural areas.

Table of Contents

List of Tables

List of Figures

Chapter 1

Air Quality Information and Avoidance Behavior: Evidence from Metro DC's Bike Share

1.1 Introduction

Provision and dissemination of information about ambient air quality are crucial to environmental and health policy. An information-based program holds that individuals respond appropriately to information by reducing air pollution and health risk exposure. The U.S. Environmental Protection Agency (EPA) has developed the air quality index (AQI) that determines air quality. Local air quality control agencies also voluntarily issue free air quality alerts to the public in many U.S. cities when the AQI forecast exceeds a certain level. Thus, the public can easily access local air quality information. Understanding changes in people's activities for informational provision campaign is an important empirical question that aims to determine the effectiveness of this alert program.

In this study, I examine the effect of air quality alerts on avoidance behaviors, focusing on bike-share trips in the Washington DC Metropolitan Area (metro DC area). Exploring a bike-share program in the metro DC area provides an ideal research setting due to three reasons. First, a bike ride is one of the best options for outdoor physical activity as it observes behavioral responses to poor ambient air quality. Second, the bike share system of the metro DC area has extensive real-time data for bike trips as it is the oldest and third largest system in the U.S. The data set provides information about people's time of bike ride, length of ride, most popular stations, and most popular times of bike ride. Lastly, the metro DC area ranks eighth out of 277 metropolitan areas for high-ozone days according to the 2014 "State of the Air" report of the American Lung Association [4].

The study analysis uses a large data set, including air quality alert readings, issuance date of alerts by a local agency, and extensive real-time bike trips recorded by Capital Bikeshare for 2011−2018. I employ the panel structure of data to estimate the alert effects over time across bike stations. To deal with the exogenous characteristic of this intervention, I use a regression discontinuity design (RDD) that elicits the causal effect of air quality alerts on bike-share usage. Alerts are issued above a cutoff AQI forecast of 100. Hence, the probability of treatment changes from 0 to 1 at the cutoff. Bike-share trips on both sides of the cutoff are compared to estimate the average treatment effect. A range of bandwidths and functional forms in this RDD are used to choose an appropriate model specification and to obtain unbiased estimates of the treatment effect.

The main results show that air quality alerts reduce bike-share trips in the metro DC area. In my preferred model, point estimates suggest 1.6−2.7 percent reduction in the number of bike-share trips and 3.4−5.9 percent reduction in the trip duration on air quality alert days. Heterogeneous responses to alerts have an important policy implication in this study. Although trip purpose is not directly observed, day-of-week and time-of-day observations of bike trips shed light on different responses to alerts between commuting and leisure trips. Air quality alerts reduce the number and duration of trips on weekends. However, weekday peak periods have no significant effects. Thus, leisure or recreational trips may be more flexible than commuting trips. In other words, the effect of air quality alerts may rely on the discretionary of an outdoor activity. Considering different locations in DC, I also find that air quality alerts have a significant effect on weekend bike trips, particularly in the central DC areas.

Bike-share users are sensitive to consecutive alerts based on a simple dynamic model with two consecutive alerts. Hence, bike-share users may miss the first alert information. However, they are more likely to be exposed to the information and acknowledge the seriousness of pollution over time. They may be willing to avoid outdoor activities with consecutive alerts considering that increased pollution exposure raises healthcare costs. Using not only code orange but also code red alerts, which are a more severe air pollution level, is reasonable given that an RDD can have multiple cutoffs. Although larger reductions in bike-share trips are expected with code red alerts, this analysis is not feasible due to few code red alerts issued in the sample period.[1]

A growing body of literature has documented the importance of air quality warnings and information. This line of research has found that individuals adjust their time on outdoor activities to reduce air pollution exposure. Graff Zivin and Neidell [90] and Neidell [64] use zoo and observatory attendance to measure outdoor activity and examine behavioral changes in response to smog alerts. However, visiting the zoo is not a frequent outdoor activity and is mainly an attractive option only for children and parents. Thus, their sample may not be representative of the population in their empirical analyses. Moreover, the observatory is known for nighttime activities, such as stargazing. Hence, smog alerts may not be effective as ozone levels drop at night.[2]

In addition, Noonan [67] uses small sample data to observe passerby at Piedmont Park in Atlanta for 35 days. In his study, ozone alerts are issued seven times during the period. He suggests mixed findings, that is, the alerts do not affect aggregate park usage but reduce usage by the elderly and joggers. Among the existing literature, my study is most closely related to Saberian, Heyes, and Rivers [80], who investigate the effect of air quality warnings on people's cycling behavior in Sydney, Australia. In addition to studying a different country, I use a different empirical strategy. They use bushfires as an instrument for air quality in an instrumental variable (IV) analysis, whereas I focus on exogenous variations from a

[1] There were only two code red alerts in 2011−2018. When adding code red alerts to my model specification, I find that code red alerts cause 52.9 percent decrease in the number of trips and 39.4 percent decrease in the trip duration. Although the results are consistent with previous studies in that individuals spend less time outside as air pollution levels increase, I need to wary about reading into these estimates due to a lack of code red alerts in this study.

[2] The Griffith Observatory is open Tue−Fri 2:00 PM to 10:00 PM and Sat−Sun 12:30 PM to 10:00 PM. Hence, it is very busy in the late afternoon, particularly during astronomical events (https://www.griffithobservatory.org/).

discontinuity in treatment assignment.[3] I also extend their research by using a larger sample size to obtain consistent estimates.[4]

Several studies indirectly address avoidance behavior by evaluating the relationship between air pollution and health outcomes. Neidell [63] and Janke [49] provide indirect evidence of air pollution exposure avoidance by examining the differences in hospital admissions for respiratory diseases on poor air quality days. They suggest that air quality information is used appropriately to measure the health costs of air pollution. Currie et al. [23] note that avoidance behavior would be an endogeneity source in empirical studies. Thus, ignoring this behavior could lead to biased health effect estimates of air pollution. This study provides similar evidence of avoidance behavior, though behavioral changes are directly observed by comparing bike-share trips between alert and non-alert days.

This study contributes to the existing literature in several ways. I attempt to identify the short-term effects of air pollution through a direct behavioral response, whereas most studies focused on air pollution effects on health outcomes (e.g., Chay and Greenstone [16]; Currie and Neidell [22]; Schlenker and Walker [83]). My results suggest that information-based programs can effectively reduce health and economic costs due to air pollution. This study supports the government's role in disseminating public information about health hazards. Furthermore, I carefully implement an empirical strategy to capture the effects of air quality alerts. Most related studies use a simple RDD using all their observations, whereas this study provides consistent estimates based on various approaches to the RDD, including different bandwidths and functional forms. In addition, the RDD in this study relies on panel data context, though previous studies are based on cross-sectional and pooled cross-sectional data sets. Examining the effects over time across bike stations helps eliminate much of the error variance, such as the between-station variation.

The remainder of the paper proceeds as follows. Section 2 provides background on air quality information and bike-share program. Section 3 describes data sources used in the

[3] They assume that a bushfire is a quasi-random event, but it is more likely to be correlated with hot and dry weather condition.

[4] I use daily bike trip data from 507 bike share stations across the metro DC for 2011−2018, while Saberian, Heyes, and Rivers focus on daily cycle counts passing only at 26 counters in Sydney cycling paths for 2008−2013.

analysis. Section 4 discusses my identification strategy and empirical model. Section 5 presents the estimation results and Section 6 concludes.

1.2 Background

1.2.1 Air Quality Information

Ambient air pollution is one of the main health problems worldwide. Since the UK's introduction of a national air quality monitoring program in the early 1900s, local most national government agencies have implemented air quality alert programs to manage public health risks due to poor air quality. Air quality forecasts and real-time air quality information are important to sensitive groups, such as children, the elderly, and people with asthma and other lung diseases. During and the following days with high ozone levels, people with respiratory diseases are more likely admitted to emergency rooms and hospitals. Thus, doctors and other health-care providers recommend people to reduce their air pollution exposure, particularly those who are more susceptible to the effects of exposure (US EPA, 2017). Although certain sensitive groups may experience more serious health effects, everyone is encouraged to take precautions to protect his or her health.

An air quality forecast determines local air quality alerts in the metro DC area, which is based on the national AQI. AQI is divided into six categories with levels of health concern and color codes: 0−50 (good, green), 51−100 (moderate, yellow), 101−150 (unhealthy for sensitive groups, orange), 151−200 (unhealthy, red), 201−300 (very unhealthy, purple), 301−500 (hazardous, maroon). When the AQI is predicted to reach 101 or higher in the metro DC area, Metropolitan Washington Council of Governments (MWCOG) issues an air quality alert. Subsequently, various channels, such as TV and radio stations, the internet, mobile applications, and phone text messages, disseminate the information to the public to reduce their exposure to the pollution.[5]

[5]MWCOG is in partnership with the District Department of Energy & Environment, Maryland Department of the Environment, and Virginia Department of Environment Quality. It provides daily air quality forecasts for the metro DC area (https://www.mwcog.org/environment/planning-areas/air-quality/).

US EPA reports AQI for five major air pollutants under the Clean Air Act.[6] The AQI for ozone is metro DC area's primary concern because this area has been in non-attainment of the National Ambient Air Quality Standard (NAAQS) for ozone since 2012 (see Figure A1.1 in the Appendix). Ozone is the biggest air pollution challenge the area faces. Hence, controlling emissions from industrial facilities and the mobile sector is crucial to reducing public health risks. Adding to policies and regulations for air pollution abatement, local governments utilize an air quality alert program to reduce the adverse health effects of ozone exposure.

Local air control agencies provide alert information a day in advance to help individuals avoid pollution exposure and reduce health risks. It is important to note that when air quality alerts are set a day before, they are not amended to correct forecast errors. Thus, actual air quality should not affect the alert issued, which is helpful in constructing my empirical model. If air quality alerts are amended after being set, an endogenous problem arises and thus biased estimates of the alert effect. This condition may violate the assumption that avoidance behavior is driven by alerts measured by air quality forecasts. Behavioral responses to alerts measured by forecasts and actual air quality will understate or overstate avoidance behavior degree. Therefore, forecasting mistakes do not significantly affect the RDD treatment effect in this study.

1.2.2 Bike-Share Program

A bike-share program is a short-term bicycle rental service available in many urban areas as an alternative form of green transportation. The first bike-share projects were designed to reduce air pollution and traffic congestion in Europe in the 1960s. Hamilton and Wichman [41] study the Capital Bikeshare in the metro DC area and show that the bike-share program reduces traffic congestion by 4 percent. This finding implies an annual benefit of approximately $1.28 million from reducing congestion-induced CO_2 emissions. The bike-share program also promotes healthy exercise and outdoor recreation in a city. According

[6]Five criteria pollutants regulated by the Clean Air Act are ground-level ozone (O_3), particulate matter (PM), carbon monoxide (CO), sulfur dioxide (SO_2), and nitrogen dioxide (NO_2). I have found only two $PM_{2.5}$ alerts in my sample period.

to the 2013 Capital Bikeshare user survey, 31.5 and 18.4 percent of 2,830 responses reported reduced stress and weight loss, respectively, as a result of the bike-share program.

Capital Bikeshare is a privately owned public bike-sharing program that serves the metro DC area.[7] SmartBike DC, a former program of Capital Bikeshare, began with 120 bikes at 10 stations in downtown Washington, DC, in August 2008. The program was the first bike-share system in the U.S. In September 2010, the District of Columbia; Alexandria, VA; Arlington, VA; and Montgomery County, MD jointly funded and launched this new regional bike-share program with 400 bikes at 49 stations. Unlike the recent scooter-share or dockless bike-share system, Capital Bikeshare is still the station-based system, where people can use a bike at any station and return it to another station. As of April 2018, this program offers more than 4,300 bikes at 507 stations across six jurisdictions.[8] [9]

1.3 Data

Daily AQI forecast data for the metro DC area comes from EPA's AirNow. Similar to other local air control agencies across many US cities, the local agency in the metro DC area predicts the daily AQI based on weather conditions and current and previous air quality levels and trends and then provides the records to the EPA. Only one local agency issues air quality alerts in the metro DC area and forecasts for ozone only during the summer months, from April to September, as ozone is worse in hotter weather. This study covers 92 orange and only 2 red alert days in 1,464 days. Therefore, 6.4 and 0.1 percent of code orange and red alerts occurred during this period.[10]

[7] Capital Bikeshare offers several pricing options: as of January 2019, annual membership ($85/year), 30-day membership ($28/30 days), 3-day pass ($17/3 days), 24-hour pass ($8/day), and single trip ($2). All bike trips under 30 minutes are included in the prices but there are additional time fees after 30 minutes.

[8] https://www.capitalbikeshare.com/about/. Figure A1.2 in the Appendix reports the number of Capital Bikeshare stations by year.

[9] According to the 2017 annual report of the National Association of City Transportation Officials (NACTO), Capital Bikeshare is the third largest station-based bike-sharing system in the U.S. after Citi Bike in NYC and Divvy in Chicago.

[10] Figure A1.3 in the Appendix reports the number of alert days by year. The number of air quality alerts are large particularly in 2011 and 2012. I estimate the effect of air quality alert on bike-share trip by dropping data from the periods because it might be a potential problem in the analysis (see Table A1.1 in the Appendix).

Table 1.1: Summary Statistics for 2011−2018, April−September

	All days	Normal days	Alert days
Average total daily trips per station (counts)			
Full sample	32.2	32.7	31.0
	[37.5]	[38.1]	[35.9]
Weekday	32.0	32.6	30.8
	[35.2]	[35.8]	[33.8]
Weekend	32.6	33.2	31.4
	[42.2]	[43.0]	[40.3]
Average total daily trip duration per station (minutes)			
Full sample	631.2	650.8	585.2
	[1,076.6]	[1,107.2]	[999.8]
Weekday	547.3	566.6	501.2
	[834.5]	[866.8]	[750.1]
Weekend	828.3	852.1	774.8
	[1,479.9]	[1,517.3]	[1,390.7]
Covariates			
AQI forecast for ozone	57.5	54.0	108.8
	[21.8]	[17.6]	[10.6]
Temperature (°F)	70.8	73.2	65.0
	[10.2]	[8.5]	[11.3]
Wind speed (knots)	7.5	7.4	7.7
	[3.1]	[2.9]	[3.6]
Relative humidity (%)	66.6	66.9	65.9
	[13.4]	[13.2]	[14.0]

Notes: This table shows means and standard deviations (in square brackets) of the number of trips and trip duration, air quality alerts, and weather variables by alert/non-alert days. The total mean values of the number of trips and trip duration are based on the station-day level. This study uses data for the sample period of 2011 to 2018. The months from October to March are dropped here because local air control agencies typically provide air quality forecasts only during the summer months.

I also collected historical real-time bike trip data from Capital Bikeshare for 2011−2018. This data set includes trips' duration, start/end date and time, and starting/ending station location.[11] After collecting the real-time data, I constructed two measures of bike-share trips at the station-day level, namely, the average total daily number of trips and average total

[11] Capital Bikeshare staffs have dropped data associated with test trips at its warehouse and any trips lasting less than 60 seconds. This work eliminates potential data errors in using bikes such as re-docking a bike to ensure its secure.

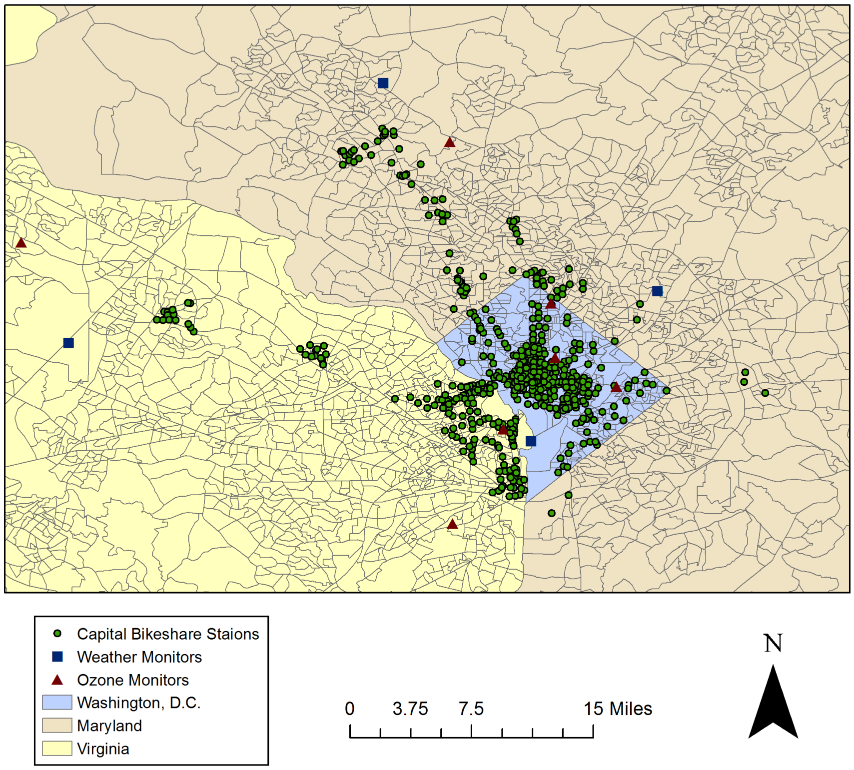


Figure 1.1: Capital Bikeshare Stations in the Metro DC Area

Notes: This figure shows locations of Capital Bikeshare stations, weather monitors, and ozone monitors in the metro DC area. This study uses data from 507 bike stations, 4 weather monitors, and 7 ozone monitors.

daily trip duration by bike station. Table 1.1 shows that bike-share's average total daily trip and trip duration are 32.2 times and 631.2 minutes, respectively.

The trip duration during weekends is longer than during weekdays, whereas weekday and weekend trip counts have no difference. Table 1.1 presents 828.3 minutes of the average total daily trip duration on weekends compared with 547.3 minutes on weekdays. Leisure or recreational bike users may enjoy relatively long-distance trips on weekends, whereas commuters traveling to work and school by bike may comprise the daily short-distance riders.[12] In addition, the bike-share usage on alert days is smaller than on non-alert days.

[12] According to the 2014 report of the US Census Bureau, the median commuting time for those who travel to work by bike is only 19 minutes indicating that an average trip distance is approximately 3 miles.

Table 1.1 shows that air quality alerts have reduced the average total daily trip duration by approximately 10 percent. Summary statistics give a preview of the negative relationship between air quality alerts and bike-share trips that would be obtained by a regression analysis.

I control for meteorological factors in the analysis not only because ozone levels increase with heat and sunlight but also because weather conditions directly affect biking. In other words, the impact of air quality alerts from other unobservable factors of behavioral changes must be separated to quantify the impact of air quality alerts on avoidance behavior. Weather data comes from the Global Historical Climatology Network daily database maintained by the National Oceanic and Atmospheric Administration's National Climate Data Center. Four weather monitors in the metro DC area provide daily measures of average temperature, wind speed, and relative humidity, as shown in Figure 1.1. I identified the closest weather monitor for each bike-share station using ArcGIS Pro. I also collected observed ozone levels, which will be used as an alternative assignment variable in specification tests. Although the metro DC area has eight ozone-monitoring stations, I end up using data from only seven ozone monitors after identifying the nearest monitor for each bike-share station.

1.4 Empirical Strategy

1.4.1 Baseline Specification

This study aims to determine whether air quality alerts reduce outdoor activity and examine the extent to which individuals respond to the alerts. I estimate the effect of air quality alerts on biking to identify direct behavioral response. A causal effect can be estimated as the difference in the outcome between bike trips on alert days and non-alert days. Specifically, I use a sharp RDD to evaluate the causal effect of a binary intervention or treatment. That is, this design uses expected outcome variation (bike-share trips) with and without the treatment (air quality alerts).

Following a generalized RDD model introduced by Imbens and Lemieux [46] and Lee and Lemieux [53], I estimate the natural log of bike-share trips as follows:

$$ln(Trip_{it}) = \alpha + \tau Alert_t + f(AQIforecast_t; \lambda) + \delta X_{it} + \rho_i + \sigma_t + \epsilon_{it}, \tag{1.1}$$

where $Trip_{it}$ denotes a measure of bike-share trips, that is, the number of trips or trip duration, as recorded at start station i on date t.[13] Variations in the number of trips and trip duration could indicate variations in the extensive and intensive margins of bike-share demand. In this study, the extensive margin refers to how many trips, whereas the intensive margin refers to how long to bike per trip. $Alert_t$ is a dummy variable that is set to 1 if ozone alerts are issued on date t.[14]

$$Alert_t = \begin{cases} 1 & \text{if } AQIforecast_t \geq 101 \ ; \\ 0 & \text{else} \end{cases}$$

where $AQIforecast$ is the assignment variable in this RDD model. Thus, the coefficient τ shows the RDD treatment effect in this study. $f(\cdot)$ is a flexible functional form that could be linear, quadratic, or cubic terms of $AQIforecast$, whose slope may vary at the cutoff.[15] X_{it} is a vector of daily weather and pollution variables, that is, a linear form of temperature, wind speed, relative humidity, carbon monoxide (CO), nitrogen dioxide (NO_2), and particulate matter ($PM_{2.5}$). ρ_i is station fixed effects. σ_t is time dummies including year, month, day of the week, and holiday dummies to account for outdoor activity changes. ϵ_{it} is an error term and is clustered on stations to account for within-station error correlations.

The baseline specification in Equation (1.1) assumes that the relationship between the outcome and assignment variables can be nonlinear and linear. Meanwhile, recent literature notes that adding high-order polynomials is often a poor choice in RDD analysis. Gelman

[13] As the outcome variable is logged, a one-unit change in $Alert_t$ would imply a percentage change in bike trips of $100 \cdot (e^{\tau} - 1)$.

[14] I assume that there is no variation in air alert alerts across my study areas because there is only one local agency issuing the alerts in the metro DC area.

[15] I center the assignment variable on the cutoff point by creating a new variable ($=AQIforecast - 101$). Adding interaction terms between assignment variable and treatment variable to the model can be useful for explaining that the treatment affects not only the intercept but also the slope of the regression line (Jacob et al. [48]).

and Imbens [32] describe three main issues of controlling for high-order polynomial terms in a RDD model: noisy estimates, sensitivity to the polynomial order, and a narrow confidence interval. They recommend the use of local low-order polynomial approaches, particularly local linear or local quadratic, as discussed by Hahn, Todd, and van der Klaauw [40] and Porter [75]. Gelman and Zelizer [33] also demonstrate that RDD treatment effects are highly sensitive to high-order polynomial regressions by examining two published research papers. Focusing on the estimates using local linear, quadratic, and cubic regressions in this study, I find that linear regression models are my preferred specification because the results from the second- and third-order polynomial terms are very noisy.

The implementation of the RDD model is based on few observations in a neighborhood around a cutoff point. Recent RDD studies prefer restricting bandwidth to small observations to reduce bias because using all available observations increases the potential for bias. However, an important issue of this approach is to choose the neighborhood size, referred to as a bandwidth. Although researchers commonly select the optimal bandwidth in a nonparametric estimation, I focus on a parametric estimation because the assignment variable in this study is discrete, not continuous. Thus, bandwidth choices can be simplified with the discrete assignment variable in this study (Lee and Lemieux [53]). Lee and Card [52] study the discrete case in RDD design and show that units within narrow bandwidths cannot be compared.[16] They suggest instead modeling parametric regressions from the selected functional form.

1.4.2 Validity of the Regression Discontinuity Design

The validity of the sharp RDD relies on the assumption that the conditional expectations of the observed outcome are continuous in the assignment variable. Variation in treatment around a cutoff point can be as good as random if the assignment variable is randomized. Treated and control groups remain similar in their observed covariates to test whether the RDD of this study meets the local randomization assumption. A possible discontinuity in the value of weather and pollution variables at the cutoff point can be easily identified.

[16] With an irreducible gap between units just below the cutoff and units just above, the treatment effect is not identified in the absence of a parametric assumption (Lee and Card [52]).

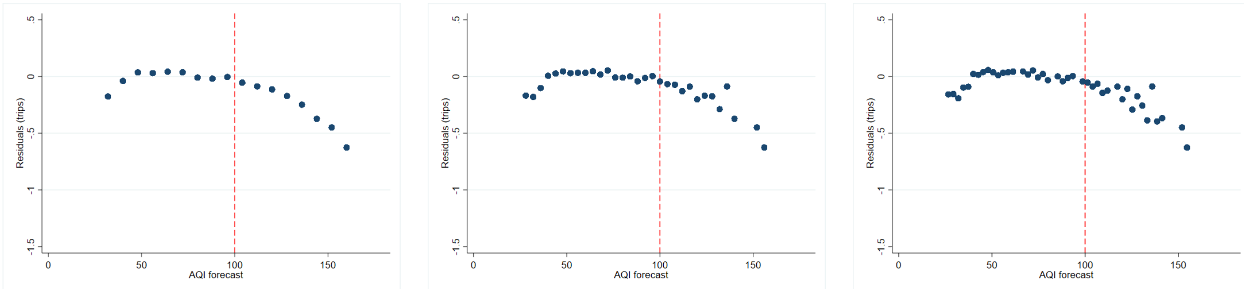
For example, if the amount of precipitation has a discontinuity at the AQI forecast 101, then it is unlikely to be the result of air quality alerts. In Figure 1.4 in the Appendix, no discontinuities of weather and pollution are found at the same cutoff used to give the alerts.

A complementary and more direct approach is used to test the validity of the RDD, that is, examination of the density of the assignment variable itself (McCrary [60]). When the density of the assignment variable for each bike trip is discontinuous, it may provoke skepticism about the manipulation of the assignment variable. For example, in this context, if a local air control agency's director is the only person who forecasts the AQI for ozone, he/she may completely determine whether to issue air quality alerts by setting the AQI accordingly. Therefore, assignment variables can be manipulated to determine the treatment status. Using a conventional McCrary [60] test, I checked for the validity of this continuity assumption, as shown in Figure A1.5 in the Appendix.[17] No discontinuity is found in the density of the AQI around the cutoff, which implies that the assignment variable is not manipulated.

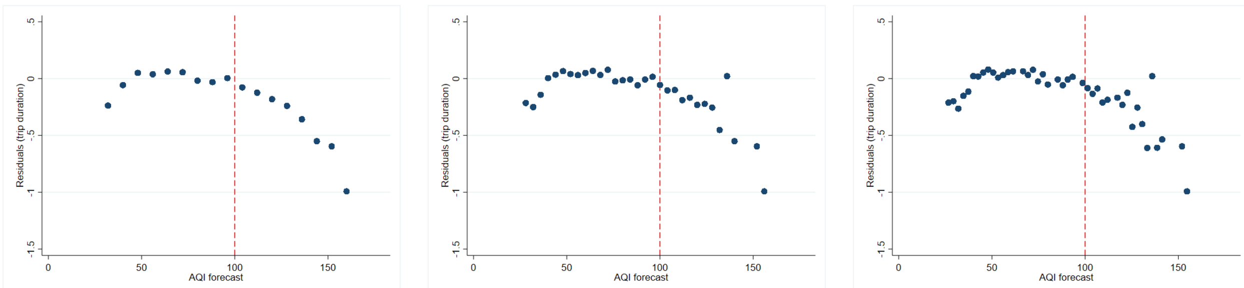
In local randomization, the assignment to treatment is independent of the covariates because the outcome variable and the assignment variable determine a standard RD estimation. Thus, it is not necessary to include additional covariates in a standard RDD model (Lee and Lemieux [53]). However, in practice, applied researchers often include them in their specifications to obtain consistent estimates of the treatment effect (Calonico et al. [12]). Additionally, my RDD differs from many cross-sectional RDD as it uses panel data. Panel data model is identified using variations in time and cross-sectional units. Thus, I can take account of covariates that are discontinuous in time, such as day-of-week effects. For example, recreational bike trips on weekends have longer trips than commuting trips to work on weekdays (Xie and Wang [89]).[18]

[17] McCrary [60] estimates a binned histogram of the assignment variable and then employs a local linear regression to examine a discontinuity at the cutoff point. The test shows the estimated log difference in heights at the cutoff point. I can implement this test using the Stata package *DCdensity*.

[18] I also estimate my RDD model excluding covariates in specification tests and the main results are unchanged.



(a) Trips (20, 40, 60 Bins)



(b) Trip duration (20, 40, 60 Bins)

Figure 1.2: Residual Plots

Notes: Using 20, 40, and 60 bins, points plotted present average residuals by the AQI obtained after estimating Equation (1.1) including weather and pollution conditions, time dummies, and station fixed effects.

1.5 Results

1.5.1 Baseline Results

A RDD analysis typically begins with a graphical presentation to provide visual evidence of the relationship between assignment and outcome variables. It allows researchers to check whether a jump or discontinuity exists around the cutoff. However, a scatter plot of individual data points is quite noisy and less useful using large sample sizes, making it difficult to determine any pattern along the distribution.[19] A binned scatter plot can be a good alternative approach by diving the assignment variable into several equal-sized

[19]Points plotted of the trip duration are based on the individual trip level (over twelve million of trips) while those of the number of trips are based on the aggregate trip counts at the station-day level (over three hundred thousand of trip counts). Scatter plots with the huge observations yielded a solid black cloud and an obscure relationship between the bike trip and the AQI forecast.

intervals, hence making the pattern look clean. Figure 1.2 presents binned residual plots by dividing the AQI forecast into 20, 40, and 60 bins based on their fitted values and plotting the average residuals for each bin.[20] [21] Following Holladay, Price, and Wanamaker's [44] approach, I calculated the residuals after estimating a baseline RDD including weather and pollution controls, time dummies, and station fixed effects. I can see the differences in the unexplained variation in bike-share trips across alert and non-alert days by excluding the alert dummy and the assignment variables.[22] Thus, the residuals arise from the differences that are associated with air quality alerts.

Figure A1.2 shows a downward-sloping outcome with the treatment, suggesting that air quality alerts reduce bike-share trips, though no discernible reduction is observed above the cutoff. One may be concerned about a positive slope of the residuals below the AQI 50. I have found that the low bike share usage is closely related to rainy days. Figure A1.4 in the Appendix shows the negative relationship between humidity and AQI forecast. Rain positively influences air quality but negatively influences biking. Furthermore, the observations far from the cutoff do not significantly affect my RDD estimates because I will focus more on small observations around the cutoff by following the local randomization assumption in RDD.[23]

To assess the statistical significance of the differences in bike trips, I estimate Equation (1.1) using different bandwidths as well as the full sample. Column 1 of Table 1.2 shows the results from my baseline RDD model using all observations in the sample period. Both point estimates are negative and statistically significant, which suggests that air quality alerts reduce the number of bike trips and trip duration by 7 and 10.6 percent, respectively. However, as shown in the figure above, the point estimates may be biased downwards due to a steep negative slope above the cutoff. Using observations far from the cutoff is likely to

[20]Figure A1.6 in the Appendix presents the binned scatter residual plots using 80 and 100 bins.

[21]The vertical dashed red line of the plots shows the cutoff value, above which people receive alert information and below which they do not receive.

[22]To measure the effect of media coverage on energy production, Holladay, Price, and Wanamaker [44] begin by examining residuals from a regression including the set of controls but excluding the media and press release indicators. The residuals show the underlying differences in the unexplained variation in energy production across media and non-media days.

[23]Several exceptional points exist around the AQI 130. I found that a couple of alerts were issued on Independence day weekend. Thus, the exceptional points may be driven by a large number of bike-share trips at that time. This is not a problem when I restrict the sample around the cutoff.

Table 1.2: Baseline Results

	Full sample (1)	2-Bandwidth (2)	5-Bandwidth (3)	10-Bandwidth (4)
Panel A. Log(trips)				
Alert	−.070***	−.016*	−.027***	−.021***
	(.004)	(.009)	(.009)	(.007)
Panel B. Log(duration)				
Alert	−.106***	−.034**	−.059***	−.051***
	(.007)	(.015)	(.016)	(.012)
N	374,862	14,703	25,387	35,985

Notes: This table shows the results of separate RD regressions using different bandwidths as well as the full sample. All regressions include weather and pollution variables, day of the week, holidays, year, month, and station fixed effects. Standard errors presented in parentheses are clustered at station level.
*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

produce an inaccurate approximation near the cutoff and lead to biased estimates (Gelman and Imbens [32]). The results of column 1 demonstrate that graphing the data before running a regression has an important advantage in estimating the treatment effect.

Using a narrow bandwidth helps reduce the bias but weakens the statistical power with a small sample size. For a 2-bandwidth model, point estimates indicate that air quality alerts significant affect bike trips. This finding suggests a 1.6 and 3.4 percent decrease in the number of trips and trip duration. As expected, standard errors are relatively larger. Using 5- and 10-bandwidths models, I find larger effects that air quality alerts reduce the number of trips and trip duration by 2.1−2.7 and 5.1−5.9 percent, respectively. My point estimates are consistent across different bandwidths (see Table A1.2 in the Appendix). These consistent estimates support local governments' air quality alert program that encourage people to reduce air pollution exposure.

Next, I explore the effect of air quality alerts on bike-share trips by employing a nonparametric approach. Numerous RDD studies use nonparametric methods to obtain optimal bandwidth or reduce bias in RDD estimates. Unlike a parametric estimation, a nonparametric estimation relies on a kernel regression, which is a fundamentally local

Table 1.3: Local Linear Regression: Nonparametric Estimation

	(1)	(2)	(3)	(4)
Panel A. Log(trips)				
Alert	−.021**	−.025***	−.023***	−.025**
	(.009)	(.010)	(.008)	(.010)
Bandwidth size	2	5	10	4.672
Polynomial	linear	linear	linear	linear
N	14,703	25,387	35,985	21,299
Panel B. Log(duration)				
Alert	−.043***	−.045***	−.053***	−.040***
	(.015)	(.017)	(.013)	(.015)
Bandwidth size	2	5	10	2.588
Polynomial	linear	linear	linear	linear
N	14,703	25,387	35,985	14,703

Notes: This table shows the results of separate local linear RDD regressions using different bandwidths. Based on a nonparametric approach, the estimates rely on a kernel regression. Column 4 uses the optimal bandwidth which minimizes mean squared error. All regressions include weather and pollution variables, day of the week, holidays, year, month, and station fixed effects. Standard errors presented in parentheses are clustered at station level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

method, hence putting more weight on observations near the cutoff (Lee and Lemieux [53]). Table 1.3 shows the results from nonparametric regressions based on a triangular kernel. Particularly, column 4 of Table 1.3 presents the estimates from the optimal IK bandwidth designed to minimize mean squared error (Imbens and Kalyanaraman [45]).[24]

Table 1.3 illustrates that point estimates are generally consistent with those obtained from a parametric estimation above. In column 4, the optimal bandwidths based on the IK implementation are 4.672 and 2.588, respectively. The number of samples within a bandwidth 2.588 is the same as the number within a bandwidth 2 because this study uses a discrete assignment variable. In the case of a discrete assignment variable, comparing outcomes within narrow bandwidths of the cutoff may be difficult because the treatment effect is not identified in a nonparametric assumption (Lee and Card [52]). Lee and Card suggest instead choosing a particular functional form for the parametric model.

[24] I estimate nonparametric models of the RDD using the Stata package *rd*.

Table 1.4: Low-Order Polynomials

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Log(trips)						
Alert	−.027***	−.023	−.029**	−.021***	−.017*	−.023
	(.009)	(.019)	(.013)	(.007)	(.010)	(.016)
AIC	28619	28621	28621	40521	40524	40526
Panel B. Log(duration)						
Alert	−.059***	−.034	−.086***	−.051***	−.023	−.064**
	(.016)	(.031)	(.023)	(.012)	(.018)	(.027)
AIC	52443	52444	52444	74111	74109	74108
Bandwidth size	5	5	5	10	10	10
Polynomial	linear	quadratic	cubic	linear	quadratic	cubic
N	25,387	25,387	25,387	35,985	35,985	35,985

Notes: This table shows the results of separate local RDD regressions using different polynomial specifications (first-, second-, and third-order polynomials) and 5- and 10- bandwidths. All regressions include weather and pollution variables, day of the week, holidays, year, month, and station fixed effects. Standard errors presented in parentheses are clustered at station level.

*** Statistical significance at the 1 percent level.

** Statistical significance at the 5 percent level.

* Statistical significance at the 10 percent level.

The function $f(\cdot)$ in Equation (1.1) represents the relationship between the outcome and assignment variables. Various functional forms can be tested to determine which model best describes the data and minimizes bias. One of challenges for the parametric estimation is the choice of the polynomial regression order. Table 1.4 explores local low-order polynomial approximations, including first-, second-, and third-order polynomials, which are flexible functional forms. To select the appropriate functional form, I use the Akaike information criterion (AIC), which is a well-known measure of model fit.

Table 1.4 shows the negative relationship between air quality alerts and bike-share trips across three polynomial specifications. Point estimates of linear models are consistent with those of cubic models, but the estimates are somewhat smaller and not statistically significant in the quadratic models. Linear models have the smallest AIC value with 5-bandwidth, whereas a cubic model for the trip duration has the smallest AIC value with 10-bandwidth. A linear regression model can be a preferred specification in this analysis because the estimates with the second- and third-order polynomials are quite noisy.[25] This finding is consistent with Gelman and Imbens [32], suggesting that using high-order polynomials for the assignment variable is not good for RDD. Thus, Gelman and Imbens recommend instead the use of local linear or quadratic polynomials in RDD analysis.

1.5.2 Heterogeneous Treatment Effects

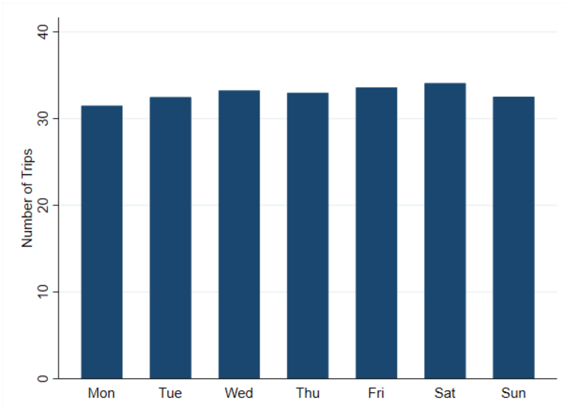
The purpose a bike was used for is difficult to capture in this study because the data set does not include bike user information. Alternatively, day-of-week and time-of-day examinations of bike-share trip patterns may help understand different trip purposes. A bike-share program is popular with commuters during weekday peak hours, whereas leisure users generally take a bike on weekends for fun and exercise (O'Brien, Cheshire, and Batty [70]). Figure 1.3 (a) and (b) show the daily distribution of the average number of trips and trip duration by day-of-week at a bike station level. The average trip duration of bike share is higher on weekends than on weekdays as shown previously in summary statistics. In contrast, no discernible difference is observed in the number of bike-share trips on each day of the week. Such a pattern is logical if leisure riders tend to spend more time with bike share compared with those who commute to work and school by bike share. Figure 1.3 (c) and (d) show that weekday morning and afternoon peaks have higher bike-share demand.

Table 1.5 shows weekday and weekend differential responses to air quality alerts by adding an interaction term of the alert and weekend indicators.[26] [27] Particularly, I examine the effects of air quality alerts at peak and off-peak hours by using disaggregated hourly data. District Department of Transportation defines morning peak from 7:00 AM to 9:00 AM and afternoon peak from 5:00 PM to 7:00 PM. The midday off-peak is from 9:00 AM to 5:00 PM and the night off-peak is from 7:00 PM to 7:00 AM. Weekends and holidays have no peak

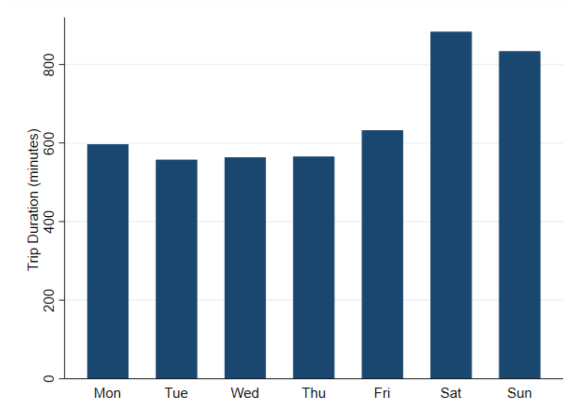
[25] This corresponds to residual plots in Figure 1.2 which show a linear or quadratic relationship.

[26] The weekends include national holidays in this analysis.

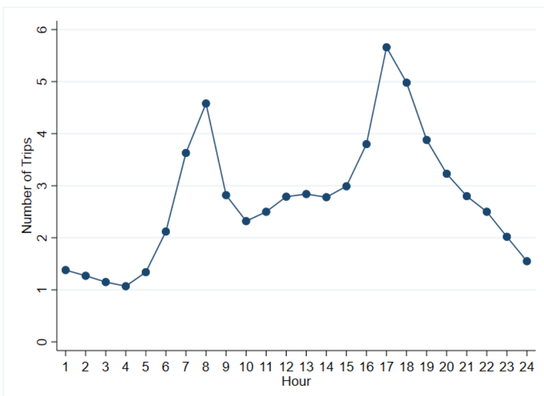
[27] For weekday trips, the coefficient on the alert dummy itself would be interpreted as the alert effect on bike trips because of no interaction term. However, for weekend trips, the effect of the alert on bike trips depends on the coefficients on both the alert dummy and the interaction term. Exploring this heterogeneous treatment effect, I use an interaction approach rather than a separate sample analysis for each subgroup, as splitting the sample causes loss of substantive information and efficiency (Becker, Egger, and Von Ehrlich [8]).



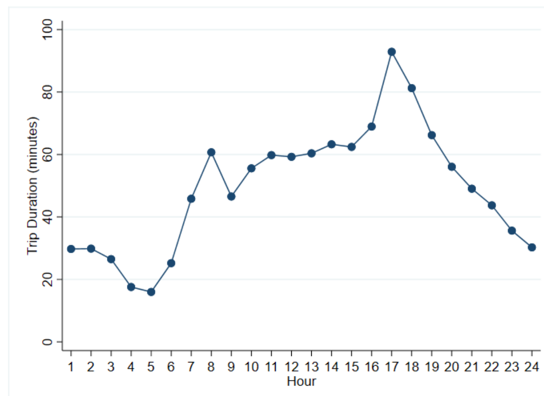
(a) Day-of-week (number of trips)



(b) Day-of-week (trip duration)



(c) Time-of-day (number of trips)



(d) Time-of-day (trip duration)

Figure 1.3: Travel Patterns of Bike Share

Notes: This figure shows the average number of trips and trip duration by day-of-week (upper panel) and by time-of-day on weekdays (lower panel) at a bike station level.
Sources: Author's calculation from the Capital Bikeshare database.

hours. Hence, I do not explore the heterogeneity of the alert effects across various times of the weekend.

This analysis sheds light on the different behavior responses to alerts across different days of the week and times of the day. Table 1.5 presents a relatively greater response to alerts on weekends. This finding suggests that alerts reduce the number of trips and trip duration on weekends by 2.8 and 3.7 percent. For weekday trips, air quality alerts positively correlate with bike-share usage in morning and afternoon peaks, but the point estimates are not

Table 1.5: Weekday and Weekend Trips

	Weekday				Weekend
	Morning peak (1)	Afternoon peak (2)	Midday off-peak (3)	Night off-peak (4)	(5)
Panel A. Log(trips)					
Alert	.004	.015***	−.005	−.007*	−.028***
	(.006)	(.005)	(.003)	(.004)	(.008)
Panel B. Log(duration)					
Alert	.002	.013	−.028***	−.011*	−.037***
	(.009)	(.009)	(.006)	(.006)	(.013)

Notes: This table shows the results of separate local RDD regressions using 5-bandwidth by weekday and weekend trips. All regressions include weather and pollution variables, day of the week, year, month, hour, and station fixed effects. Standard errors presented in parentheses are clustered at station level.
*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

statistically significant in columns 1 and 2. However, they reduce bike-share usage during the off-peak hours. The results are consistent with the findings of Saberian, Heyes, and Rivers [80]. They show that recreational cyclists and cyclists commuting to work differently respond to air quality warnings considering people may easily delay recreational cycling or replace it with other activities on poor air quality days.[28]

The nature of these two trips, a discretion degree, may be helpful in understanding my results. Bike trips could be either mandatory if bike users have a daily schedule, such as trips to work, or optional if their trips are flexible, such as recreational activities.[29] For mandatory trips that are carried out routinely, the cost of changing trip conditions may be high. Mandatory trip users could afford at the expense of poor air quality. By contrast, for flexible trip users, the cost of postponing or cancelling a bike trip may be low. Therefore,

[28]To examine heterogeneous responses, they attempt to separate cycle routes into commuting and leisure routes based on more heavily used on weekdays or on weekends. However, their categorization seems unclear and arbitrary.

[29]If a commuter takes a bike ride every day, it makes more financial sense to buy his/her own bike rather than to purchase a bike-share membership. However, depending on where a commuter live, he/she may benefit from purchasing a membership. Bike share is a good option for those who would not able to ride a bike, commuting long distances by bus, subway, or train. It could be a convenient and effective way to move between home or office and transport hub.

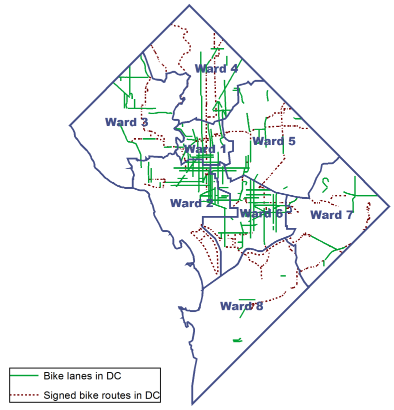


Figure 1.4: Eight Wards in Washington, DC

Notes: This figure shows a map of eight wards including bike lanes and signed bike routes in Washington, DC.

responses to air quality alerts may vary between two trip purposes, depending on the cost of avoiding poor air quality.

Thus far, I have examined the effects of air quality alerts at different times of the day considering the bike usage patterns for commuting and leisure. However, it is likely that these effects also vary even in different locations of the metro DC area. Vogel [87] finds that location characteristics are germane to bike-share trip production and attraction. According to a study of the Capital Bikeshare (2017), numerous Capital Bikeshare stations are located in the central DC area because tourist attractions and office areas are the driving force behind bike-share usage. Bike-share stations are strategically placed around subway stations and in business districts to encourage people to use public transportation and/or bikes for commuting trips. Additionally, several studies have addressed a range of socioeconomic and demographic determinants as well as bike infrastructure and accessibility that influence bike usage in commuting trips. For example, bike commuters are generally more likely to be younger people, who enjoy physical activities. On average, men bike more than women, that is, men account for two-thirds of its members, according to a 2017 Citi Bike NYC report. Therefore, exploring the location characteristics may offer further evidence of differential behavioral responses to air quality alerts in different bike trip purposes.

Table 1.6: Factors Influencing Bike-Share Usage by Each Ward

	Ward 1	Ward 2	Ward 3	Ward 4	Ward 5	Ward 6	Ward 7	Ward 8
population density (people per sq. mi.)	32267.6	8730.4	7575	9152.8	7729.3	13198.1	7961.8	6595.6
housing units	37638	43635	41792	32675	35429	42473	33232	33924
median household income (dollars)	80794	99422	109909	71545	55063	90903	39828	31642
% unemployed	7.1	3.9	3.9	10.7	16.5	7.4	20	25.4
% male	49	50.4	45.1	48	47.3	47.9	46	45
median age	31.3	30.9	36.2	39.2	35.5	34.2	38.1	28.9
% non-white	43.2	22.5	14.4	72.2	80.2	43.4	96.9	95.2
% bachelor's degree or higher	63.7	82.7	85.4	45.3	36.3	65.5	16.1	13.6
% no vehicles available	44.8	47.9	23.1	23.8	31.9	33.4	38.8	46.8
% commuting to work by bike	8.5	5.4	3.6	3.8	5.3	6.5	1.9	1.7
bike share stations	30	91	20	16	23	62	16	19
Metro stations	2	11	5	2	3	11	4	2
bike lanes (mi.)	13.9	22.6	10.4	9.5	10.7	26.3	8.3	3
signed bike routes (mi.)	5.7	11.7	11.7	11.7	12.7	17.9	12.5	8.5

Notes: This table shows the socioeconomic, demographic, and built environment factors that affect bike-share usage for eight wards in DC.
Sources: American Community Survey (2010−2014), District Department of Transportation

Table 1.7: Subgroup Analyses by Ward and Time of Day

	Weekday				Weekend
	Morning peak	Afternoon peak	Midday off-peak	Night off-peak	
	(1)	(2)	(3)	(4)	(5)
Ward 1	.017	.005	.027**	.004	−.014
	(.019)	(.017)	(.011)	(.011)	(.026)
Ward 2	−.004	.004	−.021***	−.012*	−.043***
	(.011)	(.010)	(.007)	(.007)	(.016)
Ward 3	.034	.007	.007	−.028**	.033
	(.028)	(.030)	(.014)	(.010)	(.035)
Ward 4	.004	.032	−.021	−.010	−.005
	(.038)	(.031)	(.012)	(.030)	(.043)
Ward 5	−.018	.047*	.010	.006	−.039**
	(.026)	(.025)	(.015)	(.013)	(.019)
Ward 6	.007	.014	−.001	−.008	−.034*
	(.013)	(.011)	(.008)	(.008)	(.019)
Ward 7	.052	.032**	−.000	.036	.018
	(.030)	(.013)	(.016)	(.023)	(.052)
Ward 8	.011	.054	.018	.027	−.051
	(.054)	(.033)	(.022)	(.020)	(.067)

Notes: This table shows the results of separate local RDD regressions by ward and time of day using 5-bandwidth. Outcome variable is the natural log of the number of bike-share trips. All regressions include weather and pollution variables, day of the week, year, month, hour, and station fixed effects.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

Washington, DC includes diverse places with different socioeconomic situations. To explore variations in behavioral responses to air quality alerts across neighborhoods and over time, I split my estimation sample into the eight wards that comprise Washington, DC, each with approximately 80,000 residents. Figure 1.4 shows a map with eight wards, and Table 1.6 lists some socioeconomic, demographic, and built environmental factors that affect each ward's bike-share usage. Table 1.6 also indicates that the proportion of commuting by bike is higher within downtown neighborhoods, such as wards 1, 2, 5, and 6. Commuting by bike is particularly attractive to downtown residents given traffic congestion, short commuting distance, and protected bike lanes.[30]

To investigate potential heterogeneous responses across the eight wards and over time, Table 1.7 reports results from subgroup analyses showing the estimated effect of air quality alerts on bike share trips.[31] Broadly speaking, these point estimates are statistically insignificant during weekday peak hours. These results provide no evidence that behavioral responses vary by neighborhoods. However, I find that the bike trip-reducing effects during the weekends are concentrated especially in wards 2, 5, and 6, which include busy areas of downtown DC. This finding reaffirms the results in Table 1.5, suggesting that bike trip-reducing effects are driven by recreational and leisure trips, not by commuting trips. In particular, there appears to be strong evidence of the effects in the central area, where people can have more alternative indoor activities, such as museums and major retails.

1.5.3 Specification Tests

I conduct several specification tests to explore alternative possibilities that may explain my results. Each of the following specifications is based on a preferred specification, which is a linear RDD model with 5-bandwidth. First, I examine the existence of a significant effect at a placebo cutoff point, where no effect should be observed. My identification strategy is based on the assumption that AQI forecast values below 101 are a good comparison to isolate bike-share trips from the effect of air quality alerts. However, if bike-share usage has behavioral changes below the AQI forecast 101 or an unobserved pattern exists, which is not controlled for, then the RDD treatment effect in this study would be biased. Assuming that bike users receive air quality alerts when the AQI forecast is greater than or equal to 91, column 1 of Table 1.8 reports the result from this placebo test. No significant treatment effect occurs at the placebo cutoff point as expected.

Lee and Lemieux [53] note that the inclusion of covariates reduces the sampling variability in the RDD estimation, if the RDD design is valid. However, including them in a local

[30] Figure A1.7 in the Appendix shows that bike-share usage is much higher in downtown neighborhoods than in non-downtown neighborhoods.

[31] Table A1.3 in the Appendix also presents the effect of air quality alerts on bike share trip duration, separated both by time of day and by ward.

Table 1.8: Specification Tests

	Placebo cutoff (1)	No covariates (2)	Observed AQI (3)	Lagged dep. var. (4)	Jun-Aug (5)
Panel A. Log(trips)					
Alert	.009	−.020**	−.026**	−.024***	−.038***
	(.016)	(.009)	(.012)	(.009)	(.011)
Panel B. Log(duration)					
Alert	−.005	−.039***	−.062***	−.058***	−.068***
	(.029)	(.015)	(.021)	(.015)	(.018)
N	21,235	25,387	17,676	24,489	19,018

Notes: This table shows the results of different model specifications. Column 1 indicates a placebo test using a cutoff value 91. Column 2 excludes covariates in the baseline model. Column 3 uses an actual observed AQI rather than the AQI forecast as the assignment variable. Column 4 uses a lagged dependent variable as another covariate. Column 5 uses samples only for warmer months, June and August. All regressions are based on a 5-bandwidth model. Standard errors presented in parentheses are clustered at station level.
*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

randomization in the RDD design is unnecessary. Although the covariates are used to improve the precision in the baseline RDD model, the presence of the covariates should not change the identification strategy (Imbens and Lemieux [46]). In column 2, I explore the sensitivity of the results by excluding weather and pollution variables. I can observe significant reductions in bike-share trips on alert days, though point estimates are somewhat smaller because the estimate for *Alert* suffers from a downward bias with the omitted variables.[32]

AQI forecast determines air quality alerts. For this reason, I have defined the AQI forecast as an assignment variable in my RDD analysis. As the AQI forecast value and an actual AQI value are more likely to be correlated, assuming that the actual AQI value can be an alternative assignment variable in specification testing is sensible.[33] Point estimates

[32]Weather conditions and pollution levels are likely to be correlated with both biking decisions and air quality alerts.

[33]In a 5-bandwidth model, alerts were issued 48 days in 85 days. However, there were 21 days of forecasts error that actual AQI values were not above 100. On the other hand, 6 days of forecast error had actual AQI values of above 100 on non-alert days.

in column 3 remain unchanged, which implies that my analysis is free from forecast errors. Following Lee and Lemieux [53], I also consider the lagged outcome variable $Trip_{it-1}$ as simply another control variable in period t. If $Trip_{it}$ is highly persistent over time, $Trip_{it-1}$ will be a good predictor in reducing the sampling error in the panel data set. In column 4, the effects of air quality alerts have similar magnitude to the baseline results. This finding suggests that the main results are not sensitive to the lagged outcome variable. Furthermore, I test for effects using samples for summer months, from June to August, as shown in column 5. The bike trip-reducing effects are stronger in warmer months because air quality alerts are issued more often during hot weather.[34]

1.5.4 Consecutive Alerts

An extension of my empirical strategy is a dynamic model that allows for the effects of air quality alerts issued on consecutive days on bike-share usage. In the summer months, poor ozone pollution sometimes lasts for a few consecutive days. In the sample period, one-time alerts account only for 31 percent of all alerts but more than 60 percent of alerts are issued on consecutive days. Individuals' different responses to consecutive alerts may be important for an air quality alert program. To see the behavioral dynamics of consecutive alerts, I modify my baseline model following a two-day model introduced by Graff Zivin and Neidell [90].[35] I express the dynamic version of Equation (1.1) as follows:

$$\begin{aligned} ln(Trip_{it}) = \alpha + \beta_1 Alert_t + \beta_2 Alert_{t-1} + \beta_3 Alert_t \times Alert_{t-1} + f(AQI\,forecast_t; \lambda_1) \\ + g(AQI\,forecast_{t-1}; \lambda_2) + \delta_1 X_{it} + \delta_2 X_{it-1} + \rho_i + \sigma_t + \sigma_{t-1} + \epsilon_{it}, \end{aligned} \tag{1.2}$$

$$Alert_t = \begin{cases} 1 & \text{if } AQI\,forecast_t \geq 101 \text{ ;} \\ 0 & \text{else} \end{cases}$$

[34] Air quality alerts issued from June to August account for approximately 80 percent of all alerts in my study.

[35] Using attendance data at the Los Angeles Zoo and the Griffith Park Observatory, Graff Zivin and Neidell [90] show the possibility of 'alert fatigue' that individuals are less likely to be sensitive to consecutive alerts. When alerts are issued on consecutive days, they show evidence of a rebound effect because visiting a zoo is an infrequent event or not easy to postpone.

Table 1.9: Consecutive Alerts

	(1)	(2)	(3)
Panel A. Log(trips)			
First day response	.002	−.014	−.014*
	(.010)	(.010)	(.008)
Second day response	−.033*	−.010	−.037***
	(.017)	(.013)	(.010)
p-value of difference	.053	.736	.013
Panel B. Log(duration)			
First day response	−.017	−.044***	−.044***
	(.018)	(.017)	(.013)
Second day response	−.050*	−.016	−.068***
	(.028)	(.022)	(.017)
p-value of difference	.283	.189	.143
Bandwidth size	2	5	10
Polynomial	linear	linear	linear
N	14,162	24,489	34,800

Notes: This table shows the results of separate local RDD regressions using different bandwidths and two consecutive alerts. A p-value comes from an F-test for whether the coefficients of two different responses are the same. All regressions include weather and pollution variables, day of the week, holidays, year, month, and station fixed effects. Standard errors presented in parentheses are clustered at station level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

$$Alert_{t-1} = \begin{cases} 1 & \text{if } AQI\ forecast_{t-1} \geq 101 \ ; \\ 0 & \text{else} \end{cases}$$

where $Alert_{t-1}$ is a lagged alert dummy variable. I add an interaction term between contemporaneous and lagged alerts to allow the effect of the alert issued on date t to vary depending on the alert issued on date $t-1$. The effect of a one-time alert is still β_1 because $Alert_{t-1}$ is zero. However, when the alert is issued on two consecutive days, the effect of the alert is $\beta_1 + \beta_3$ on the second day.

Table 1.9 reports the results for two consecutive alerts.[36] The results suggest that bike-share users are more sensitive to second-day alerts than to first-day alerts in 2- and 10-bandwidth models, though the two responses have no statistically significant difference.[37] Bike-share users may miss the first alert information, but they are more likely to be exposed to the information and acknowledge the seriousness of pollution over time. As increased pollution exposure raises healthcare costs, they may be willing to avoid outdoor activities with consecutive alerts. In light of policy implications, local governments and the media may increase their efforts to publicize first-day alerts. As global warming and climate change make poor air quality longer and more frequent, understanding these dynamic responses is important (Graff Zivin and Neidell [90]).

1.6 Conclusion

I examine the link between air quality information and avoidance behavior in the analysis using a panel data set of bike-share trips. Employing a range of bandwidths and functional forms in RDD, I provide strong empirical evidence that bike-share users respond to air quality alerts, reducing the number of trips and trip duration in the metro DC area. The results support government's role in improving public health by disseminating air quality information about health risks. In terms of policy implications, information disclosure and air pollution exposure avoidance can be cost-effective health and environmental policies. Furthermore, understanding heterogeneous treatment effects has important implications for an air quality alert program, which suggests that alert responses vary across different bike trip purposes.

This study has several limitations that future research should consider. First, as this empirical analysis focuses on a reduced form model identifying a treatment effect, I am unable to determine the extent to which individuals pay attention to air quality information.[38] A

[36] When I use the dynamic analysis with three or four consecutive alerts, the results are not feasible due to a small number of them.

[37] Large p-values of difference indicate no statistically significant difference between the two estimated coefficients.

[38] Future research would use data on the intensity of media coverage or Google Trends data to explore the public's interest and concern for air quality alerts.

statistically significant treatment effect indicates that reduced bike-share trips result from avoidance behavior. However, this finding is insufficient to show evidence of individual avoidance behavior in response to air quality alerts. Potential avoidance behavior may be greater among sensitive groups, such as the elderly and children. Second, a bike ride is more discretionary than other types of outdoor activities, such as farming and landscaping. These outdoor activities are less discretionary and may not respond to air quality alerts because the cost of reducing air pollution exposure is relatively high. Lastly, information on bike users riding their private bikes is not available. They may react in different ways to air quality alerts.

Chapter 2

Effect of Gasoline Tax on Air Quality: Lesson from New Jersey's Gasoline Tax Increase

2.1 Introduction

High sales taxes affect consumer behavior because consumers are more likely to avoid spending more money on taxes. If high gasoline taxes encourage people to drive less, then they can be used to correct environmental externalities associated with automobile use, including pollution, congestion, and accidents (Parry, Walls, and Harrington [73]). Meanwhile, previous studies have suggested that the optimal gasoline tax in the US is more than twice the current level, given the fact that gasoline has an inelastic demand. Parry and Small [72] suggest the US optimal gasoline tax of 1.01 dollars per gallon by deriving the second-best optimal gasoline tax using external cost components and Ramsey tax.[1]

Although most US states have raised their gasoline taxes in recent years, gasoline taxes are still much lower than the optimal taxes that the literature estimates.[2] Moreover, raising

[1]Using Parry and Small model and considering the oil dependence cost, Lin and Prince [57] present the optimal California state gasoline tax of 1.37 dollars per gallon. Wood [88] also examines the optimal gasoline taxes for Ontario and Toronto in Canada and his findings show the second-best optimal gasoline tax of 40.57 cents per litre in 2006 Canadian dollars.

[2]State gasoline taxes range from 14 to 61 cents per gallon as of July 2019. They are still less than a dollar per gallon even including the federal tax of 18.4 cents.

the gasoline taxes is a politically sensitive issue. Thus, the states have increased by less than 10 cents per gallon only. In this paper, I focus on a 23-cent-per-gallon increase in New Jersey's state gasoline tax on November 1, 2016, and explore the impact of this relatively large increase. The extent to which this gasoline tax increase reduces local air pollution can be an important empirical question as climate change and air quality issues become one of the greatest threats.

Understanding the relationship between gasoline tax increase and local air pollution is useful for measuring potential environmental effects of tax increases. A large body of literature estimates consumer responses to gasoline prices and/or taxes and addresses the implications for environment (Lin and Zeng [58]; Austin and Dinan [6]; Sipes and Mendelsohn [86]). However, direct empirical evidence on the impacts of gasoline taxes on air pollution is scarce. In this study, I attempt to take a shortcut by focusing on variations in mobile source air pollution caused by policy changes in gasoline tax. This could also explain the relationship between gasoline taxes and consumption or driving behavior. This paper's research question is closely related to Davis and Kilian [25]. However, Davis and Kilian exploit the historical variation in the US federal and state gasoline taxes to examine the effect of gasoline taxes and carbon emissions.[3]

To evaluate a causal effect of the policy intervention, I use two methods−a difference-in-differences (DID) method and synthetic control method (SCM). First, I apply the DID method to compare pollution levels in treated and non-treated states before and after the gasoline tax increase at the air monitor-day level. I define New Jersey's air monitors and those of other 25 states as treatment and control groups in the DID model, respectively. I seek to estimate the average treatment effect on the treated based on two standard time periods and groups set up. Despite the expansion of gasoline tax increases in recent years, there is little empirical work with a quasi-experimental research design for ex-post policy evaluation of gasoline taxes.[4]

[3]Davis and Kilian [25] show that a 10-cent-per-gallon gasoline tax increase would reduce carbon emissions from the transportation sector in the US by 1.5 percent.

[4]One of the challenges has been a lack of variation in gasoline taxes. The US federal gasoline tax has remained at 18.4 cents per gallon since 1993 and it is only recently that most states have raised or reformed their state gasoline taxes.

Second, I also employ an SCM developed by Abadie, Diamond, and Hainmueller [1] to estimate the aggregate-level effect of a policy intervention. The identifying assumption of the DID approach is that only air monitors in New Jersey are exposed to a treatment in the post-policy period. However, this assumption may be problematic in that the number of air monitors in New Jersey is quite small.[5] With a small number of treated groups, a DID estimation may not perform well due to non-normal distributed data or serially correlated errors (Conley and Taber [21]; Ferman and Pinto [30]). To address this problem, I use a SCM as an alternative approach that weakens the assumptions of the DID method. A policy intervention took place at an aggregate level, which is the state level in this study. Thus, this method would be a reasonable tool using aggregate data and state-level weights to create a synthetic control state. The analysis results show that pollutant concentrations between the actual and synthetic New Jersey after the gasoline tax increase have no difference. Therefore, this supplementary estimation is consistent with the DID estimation.

I focus primarily on variation in the ambient concentrations of mobile source pollutants for 2014−2018. Mobile source pollutants include carbon monoxide (CO), nitrogen oxides (NOx), which consists of nitrogen oxide (NO) and nitrogen dioxide (NO_2), and hydrocarbons (HC).[6] Automobile emissions also contribute to the ambient concentrations of particulate matters ($PM_{2.5}$ and PM_{10}) and sulfur dioxide (SO_2), but the contribution of these pollutants is very small. Moreover, $PM_{2.5}$ and PM_{10} come mostly from diesel engine emissions rather than petrol engine emissions. According to the EPA's 2014 National Emissions Inventory, on-road gasoline vehicles contribute approximately 20.9 million tons/year of CO and 2.4 million tons/year of NO_2 emissions.[7]

The main results suggest that New Jersey's gasoline tax increase negatively affects local air quality, but the results are statistically insignificant. Point estimates show that New Jersey's CO and NOx levels are 3.6 and 1 percent, respectively, lower than those in other states with the gasoline tax increase. However, none of these differences are significant at meaningful levels. The results should be interpreted carefully because estimates are not

[5] In New Jersey, air quality monitors for CO and NOx are 6 and 10, respectively.

[6] Data on variation in hydrocarbons (HC) is not available.

[7] Total emissions of CO and NO_2 are approximately 70.1 and 14.5 million tons/year, respectively. On-road gasoline vehicles lead to 0.2 million tons/year of PM_{10}, 0.06 million tons/year of $PM_{2.5}$, and 0.02 million tons/year of SO_2 emissions.

statistically significant but have a wide confidence interval that may include potentially important impacts. I also find that gasoline tax increase reduces $PM_{2.5}$ and SO_2 levels, which are co-pollutants, but not statistically significant. The results remain consistent across various robustness checks and sensitivity analyses.

Evidence from additional sources shows that gasoline tax increase did not successfully reduce the amount of automobile travel and gasoline sales volume. New Jersey's gasoline tax increase raised vehicle miles traveled (VMT) by 2.4 percent and decreased gasoline sales volume by 0.3 percent. However, these results are not statistically significant. The gasoline tax increase was weak to significantly impact driving behavior in the short run. Switching to more fuel-efficient cars or adjusting drivers' residence may take time. Real impacts of gasoline tax increases may occur in the long run. Additionally, if sales volume has been constant even after the gasoline tax increase, then higher gasoline taxes can be a stable source of local government revenue. Policymakers who are interested in raising transportation funds could pursue policies associated with gasoline tax increases.

Drivers may respond to gasoline tax or price changes because of persistence and salience (Li, Linn, and Muehlegger [56]).[8] Adjusting driving behavior or buying a fuel-efficient car causes drivers to respond more to persistent than transitory change in gasoline taxes or prices. It seems to be related to long-term response rather than short-term response. However, salience could be large for the tax change even in the short run because the media makes more efforts to inform the public of the policy change, especially with considerable changes. Based on the intuition and rationality used in prior research, this paper focuses on the short-term effect of the gasoline tax changes.[9] Although my analysis examines the short-term impacts with a relatively recent policy change, estimation of long-term impacts is left for future work.

This paper proceeds as follows. Section 2 provides the background on gasoline taxes. Section 3 describes the data used for the analysis and Section 4 introduces the empirical approach. Section 5 reports and discusses the results. Section 6 concludes.

[8] This is in line with literature on tax. For example, Chetty, Looney, and Kroft [18] note that consumers' knowledge about sales taxes affects consumer behavior.

[9] Assuming that similar trends in gasoline prices emerge across states. I can focus on the effect of gasoline tax change by ruling out a potential problem of price endogeneity.

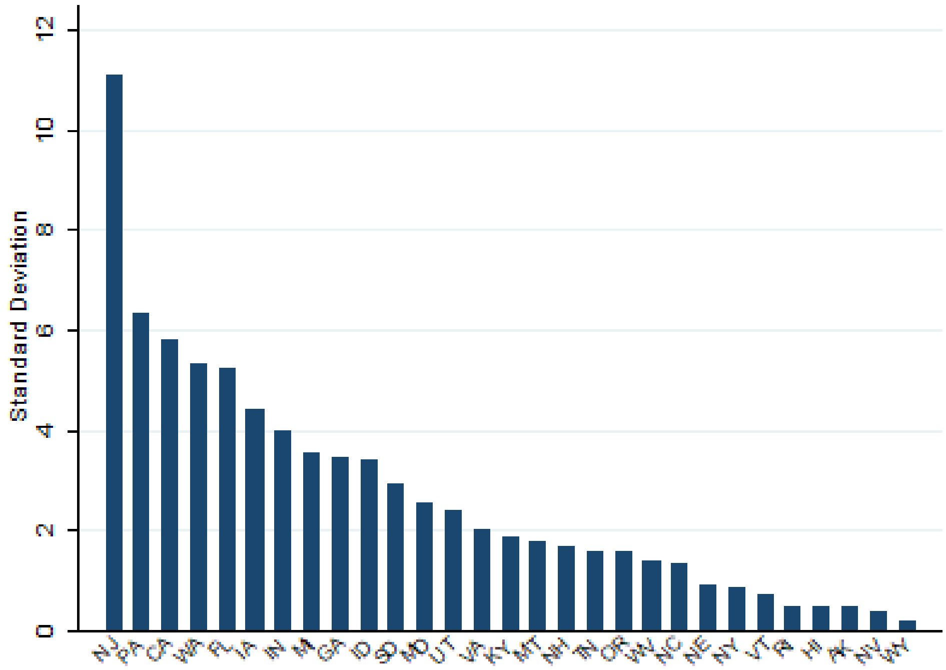


Figure 2.1: Variation in Gasoline Tax by State

Notes: This figure shows the variation in gasoline tax by state for the last five years, 2014−2018. 29 states have changed their state gasoline taxes during the period. Historical data on tax rates comes from the Urban-Brookings Tax Policy Center.

2.2 Background and Motivation

The US government imposed the first federal gasoline tax of 1 cent per gallon in 1932. This tax had been increasing until 1993, and drivers now pay 18.4 cents per gallon on gasoline. According to the US Energy Information Administration (EIA), the federal gasoline tax includes an excise tax of 18.3 cents per gallon and a leaking underground storage tank fee of 0.1 cent per gallon.[10] At the state level, Oregon was the first to levy a gasoline tax of 1 cent per gallon in 1919, and after only over a decade, every state had a gasoline tax. State taxes include excise taxes, environmental taxes, special taxes, and inspection fees. As of August 2018, the national average of total state gasoline taxes is 28.62 cents per gallon (EIA).

[10] As of August 2018, the federal diesel tax is 24.4 cents per gallon including an excise tax of 24.3 cents per gallon and a leaking underground storage tank fee of 0.1 cents per gallon.

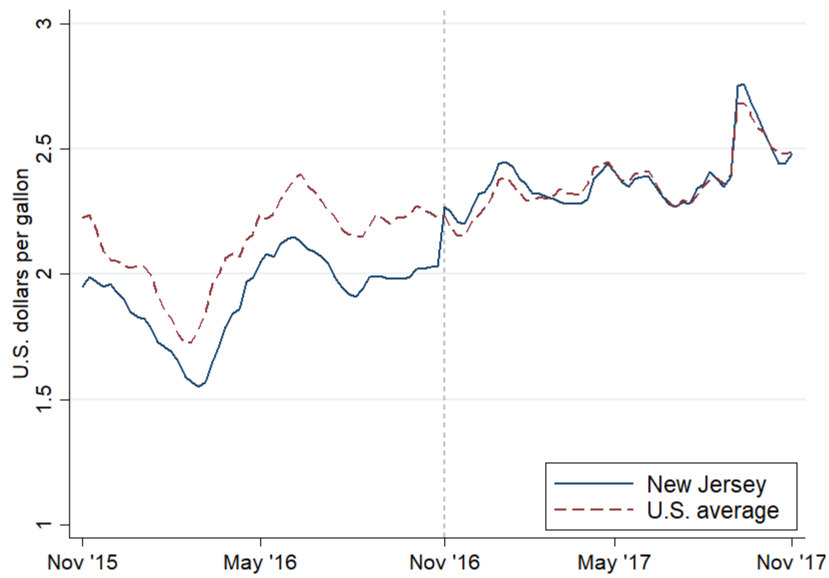


Figure 2.2: Average Retail Prices for Regular Gasoline

Notes: This figure shows the weekly retail prices for regular gasoline (U.S. dollars per gallon, including taxes) both in New Jersey and in the U.S. for the period, 2015−2017. Solid line indicates New Jersey while dashed line indicates all states in the U.S. Vertical dashed bar indicates the policy change date, November, 2016. Gasoline price data comes from the U.S. EIA and GasBuddy.com.

State gasoline taxes have been increasing recently, though the federal gasoline tax remains at 18.4 cents per gallon for the last 25 years. According to a 2018 report from the Federal Highway Administration (FHWA), 32 states, including the District of Columbia, have raised their state gasoline taxes since 2011. Among these states, New Jersey had the largest tax hike. Most states raised their state gasoline taxes by less than 10 cents per gallon, whereas New Jersey increased its state gasoline tax from 14.5 to 37.5 cents per gallon in 2016. This gasoline tax increase was the first in New Jersey since 1988. Before the tax increase, New Jersey drivers paid the second lowest state gasoline tax in the nation, at 14.5 cents per gallon. However, as of January 2017, the state gasoline tax was the eighth highest in 50 states. Figure 2.1 describes the variation in gasoline tax by state and over time, showing that New Jersey has a remarkably large variation.

Crude oil prices, distribution and marketing, and taxes determine retail gasoline costs (EIA, 2018). Figure 2.2 demonstrates that a gasoline tax is one of the key factors that determine gasoline prices. When gasoline tax increased on November 1, 2016, the retail gasoline prices in New Jersey increased by 23 cents per gallon, which is equal to the amount of the gasoline tax increase. However, gasoline is a relatively inelastic product, that is, its price changes have less effect on automobile travel demand. Literature suggests that the optimal gasoline tax should be more than one dollar per gallon (Parry and Small [72]; Parry et al. [71]; Lin and Prince [57]). Hence, one may be concerned that a gasoline tax of 55.9 cents per gallon in New Jersey is still too low to affect driving behavior.[11] Therefore, this study could in part shed light on the previous empirical literature on optimal gasoline taxes, focusing on the short-term impacts of a state gasoline tax.

2.3 Data

Detailed data on air pollutant concentrations come from the Air Quality System (AQS) database of the US EPA from November 2014 to September 2018.[12] This database combines ambient pollutant concentrations measured at more than 4,000 monitoring stations owned and operated mainly by state environmental agencies. Furthermore, information on the exact locations of monitoring stations is gathered. Data include daily (or weekly for some pollutants) and hourly measurements of pollutant concentrations. For the pollutants of interest, the daily measurements used are the 8-hour maximum of CO, 1-hour maximum of NOx, 24-hour average of $PM_{2.5}$, and 1-hour maximum of SO_2, which correspond with those of NAAQS.

Specifically, I use data for air monitors in 25 states as well as in New Jersey because the 25 states did not experience any change in gasoline tax policy from November 2014 to September 2018.[13] Each monitor tracks different types of air pollutants. For example, some monitors observe CO only, whereas others observe all criteria pollutants. Because of

[11] As of January 2018, New Jersey has the gasoline tax of 55.9 cents per gallon including the federal tax of 18.4 and state tax of 37.5 cents per gallon.

[12] I dropped data for October 2018 because New Jersey experienced another policy change in gasoline tax on October 1, 2018

[13] The twenty five states include Alabama, Arizona, Arkansas, Colorado, Connecticut, Delaware, District of Columbia, Illinois, Kansas, Louisiana, Maine, Massachusetts, Minnesota, Mississippi, Missouri, Nevada, New Hampshire, New Mexico, North Dakota, Ohio, Oklahoma, Texas, Vermont, Wisconsin, and Wyoming.

this limitation, I separately identify the causal effects of gasoline tax increase on multiple pollutants. SO_2 is the most frequently monitored pollutant, as well as NOx, CO, and $PM_{2.5}$.

Weather conditions play an important role in air pollution levels but are independent of gasoline taxes, so I include controls for primary meteorological variables in the analysis. Daily weather data also come from EPA's AQS database, including daily average temperature and relative humidity. Although monitors do not provide precipitation and snowfall information, the relative humidity that presents water vapor in the air can be used. Moreover, not all air monitors observe meteorological variables. After merging the weather and air quality data by monitor and date, I found some monitors where the weather data was missing. To address this issue, I identified the closest weather monitor for each air monitor using geographical information system software.

I also employ gasoline sales volumes, vehicle miles traveled (VMT), and public transportation use as different outcome variables. These three variables are available only at a monthly state-level, whereas air pollutant and weather variables are based on daily monitor level data. Data on gasoline volumes and VMT come from monthly motor fuel reports and traffic volume trends, respectively, provided by the FHWA. For gasoline sales volumes data, each state reports to the FHWA the amount of on-highway fuel use on a monthly basis. Based on hourly traffic count data reported by each state, the FHWA adjusts the data and then estimates VMT. Public transportation use data come from the National Transit Database that reports the number of monthly passenger trips by transit mode. All transit properties that receive urbanized area formula grant from the Federal Transit Administration are required to report passenger trips, whereas small transit properties are exempted from the monthly report.

Table 2.1 summarizes the descriptive statistics for ambient air pollutants and additional outcome variables from November 2014 to September 2018. Columns 1−2 and 4−5 show mean values for treatment and control groups during pre- and post-treatment periods. Columns 3 and 6 present the results from a t-test based on the comparison between variables in pre- and post-treatment periods. Table 2.1 reveals the most notable feature, that is, the New Jersey's air pollutant concentrations, except SO_2, are mostly higher than those in other control states. In contrast, this table shows that air pollution levels in New Jersey and

Table 2.1: Descriptive Statistics

	New Jersey			*Control States*		
	Pre-period (1)	*Post-period* (2)	(1)−(2) t-stat (3)	*Pre-period* (4)	*Post-period* (5)	(4)−(5) t-stat (6)
Mobile source pollutants						
CO (ppm)	.47	.43	−5.90	.40	.38	−10.23
	[.29]	[.27]	(.000)	[.30]	[.28]	(.000)
NO_x (ppb)	60.6	58.0	−2.37	30.3	28.4	−11.77
	[57.1]	[55.1]	(.018)	[41.0]	[37.3]	(.000)
Co-pollutants						
$PM_{2.5}$ ($\mu g/m^3$)	8.2	7.7	−6.39	7.9	7.8	−0.31
	[4.7]	[4.0]	(.000)	[4.7]	[4.5]	(.754)
SO_2 (ppb)	.91	.63	−6.92	3.4	3.3	−1.28
	[1.5]	[2.7]	(.000)	[10.7]	[10.9]	(.200)
Additional outcomes						
Gasoline volume	348	272	−3.21	198	165	−2.38
(million gallons/month)	[18]	[116]	(.003)	[227]	[215]	(.018)
Vehicle miles traveled	6304	6344	0.26	4300	4405	0.40
(million vehicle miles/month)	[531]	[498]	(.798)	[4351]	[4606]	(.692)
Public transportation	24.2	23.2	−2.85	1.56	1.46	−20.92
(million riderships/month)	[1.3]	[1.1]	(.007)	[.8]	[.8]	(.000)

Notes: This table shows means for mobile source pollutants, co-pollutants, and additional outcome variables in both New Jersey and control states. Pre-treatment period in columns 1 and 4 is based on data from November 2014 to October 2016. Post-treatment period in columns 2 and 5 is based on data from November 2016 to September 2018. The entries in square brackets show the standard deviation of each variable. Columns 3 and 6 indicate the t-statistic and the p-value in brackets for the hypothesis test that the mean value does not differ between pre- and post-treatment periods.

control states are significantly lower in the post-treatment period than in the pre-treatment period. CO concentrations have fallen in both states, but have fallen further in New Jersey. This finding shows the importance of causal analysis to evaluating this policy.

2.4 Empirical Strategy

2.4.1 Difference-in-Differences

The empirical analysis aims to identify the extent to which New Jersey's gasoline tax increase affects local air quality. I use a difference-in-differences (DID) method based on New Jersey's ambient air concentrations compared with other control states' concentrations. Before using the DID method, I should check the common trends assumption to validate a sample of observations, that is, any trend in the concentrations of air pollutants are the same across the treatment and control groups. My DID framework relies on the assumption that states that have not increased gasoline taxes are appropriate counterfactuals for New Jersey that has increased gasoline taxes. Figure 2.3 shows that air pollution levels in the two groups have similar trends in pre-treatment periods, implying that using air monitors in 25 states as a control group is appropriate.

Throughout my empirical discussion, I define air monitors in New Jersey as a treatment group, against which air monitors in other 25 states are compared. I apply the DID method using observations of ambient air pollution at the daily monitor level. The most straightforward approach is given as follows:

$$ln(y_{it}) = \beta_0 + \beta_1 Post + \beta_2 Treat + \beta_3 Post \times Treat + \delta_i + \epsilon_{it}, \tag{2.1}$$

where y_{it} denotes one of the measures of air pollutants as recorded at monitor i on date t. *Post* is a dummy variable that is 1 for all dates after the gasoline tax increase in New Jersey and 0 otherwise. *Treat* is another dummy variable that is one for air monitors located in New Jersey and 0 otherwise. Thus, β_3 is the DID estimate that is of primary interest in this analysis.

Equation (2.1) contains monitor fixed effects, δ_i, controlling for unobserved monitor attributes that affect air quality. These fixed effects help the DID estimates to avoid an upward bias by the trend that air monitors in New Jersey generally have higher air pollution levels than those in the control states before and after treatment. I also allow the error term, ϵ_{it}, to be correlated across all observations within the same monitor and date. The underlying assumption is that monitor-specific unobserved factors that affect air pollutant concentrations are constant over time. In other words, the identification of β_3 requires that $E[(Post \times Treat) \cdot \epsilon_{it} | \delta_i] = 0$.

To improve the precision of estimates and control for time-varying factors that affect treated and control monitors differently, I add additional variables to Equation (2.1). I

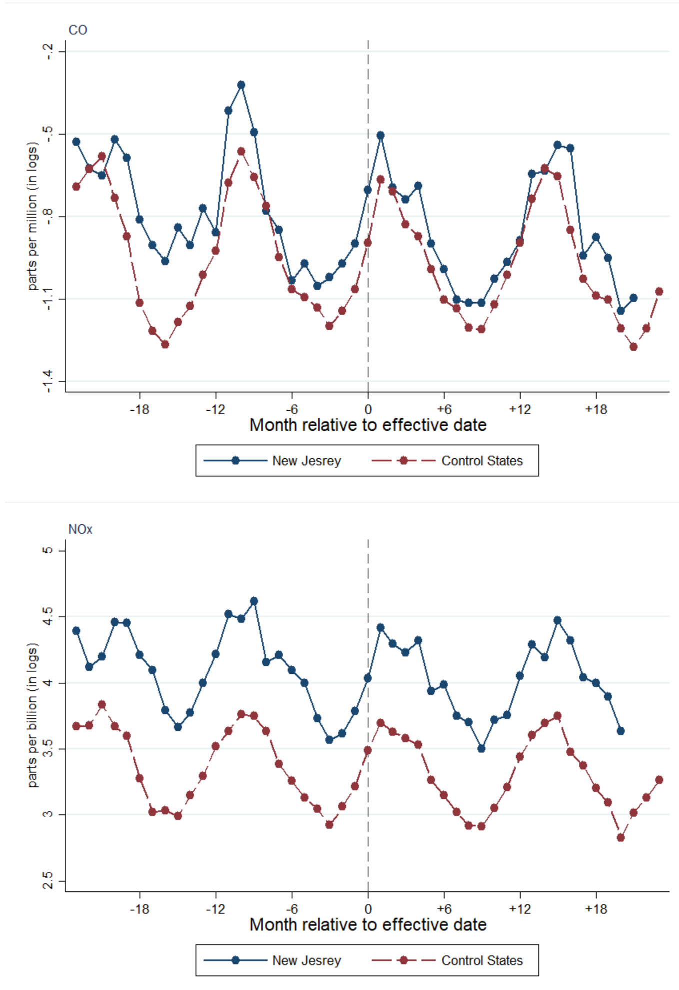


Figure 2.3: Mean Monthly Pollution Levels in both New Jersey and Control States

Notes: This figure shows the mean monthly concentrations of mobile source pollutants, CO (upper panel) and NOx (lower panel), from 2014 to 2018. Solid lines indicate New Jersey while dashed lines indicate control states. Vertical dashed bar represents the gasoline tax change date, November, 2016.
Sources: Author's calculation from the U.S. EPA's air quality system database.

estimate the preferred specification for air pollutant concentrations as follows:

$$\begin{aligned} ln(y_{it}) = \beta_0 + \beta_1 Post + \beta_2 Treat + \beta_3 Post \times Treat + \lambda \boldsymbol{W_{it}} \\ + \mu \boldsymbol{D_t} + \delta_i + \epsilon_{it}, \end{aligned} \tag{2.2}$$

where $\boldsymbol{W_{it}}$ denotes a vector of daily weather variables for monitor i on date t including average temperature and relative humidity. If I omit weather variables which heavily affect the ambient air concentrations, then the regression models may suffer from omitted variable bias. $\boldsymbol{D_t}$ is a vector including dummy variables for year, month, and day-of-week to control for the times that are common across all areas. The identification assumption of Equation (2.2) is more relaxed than that of Equation (2.1), indicating that the interaction term of *Post* and *Treat* does not correlate with unobserved factors conditional on the covariates. Now the identification of β_3 requires that $E[(Post \times Treat) \cdot \epsilon_{it}|\boldsymbol{W_{it}}, \boldsymbol{D_t}, \delta_i] = 0$. Additionally, I estimate all regression models with the cluster-robust standard errors to allow for common shocks over time within a monitor.

2.4.2 Synthetic Control Method

Using a DID strategy may be challenging with a small number of air monitors in New Jersey. A limited number of treated units results in the treated unit observations to be unlikely normally distributed, leading to biased estimates. To address this potential issue, I conduct an additional analysis with a synthetic control method (SCM) developed by Ababie, Diamond, and Hainmueller [1], which does not rely on large sample theory and is valid in small samples.[14]

Following a simple model of Ababie, Diamond, and Hainmueller [1], I let I_{st} be an indicator for treatment for state s at time t. The outcome variable which is one of the air pollutant concentrations, Y_{st}, is the sum of a time-varying treatment effect, $\alpha_{st}D_{st}$, and the outcome variable for states s at time t in the absence of a gasoline tax policy change, Y_{st}^N.

$$\begin{aligned} Y_{st} &= \alpha_{st}I_{st} + Y_{st}^N \\ &= \alpha_{st}I_{st} + (\phi_t + \rho_t\boldsymbol{Z_s} + \gamma_s\tau_t + e_{st}), \end{aligned} \tag{2.3}$$

where ϕ_t is an unknown time factor, $\boldsymbol{Z_s}$ is a $(r \times 1)$ vector of observed covariates that are unaffected by the treatment, ρ_t is a $(1 \times r)$ vector of unknown parameters, γ_s is a $(F \times 1)$ vector of unknown factor loadings, τ_t is a $(1 \times F)$ vector of unknown factors, and the error terms e_{st} are unobserved transitory shocks at state level with a zero mean. If I let the first state be the treated state, then the treatment effect is estimated by approximating the

[14] I implement a SCM analysis using the Stata packages *synth* and *synth_runner*.

unknown Y_{1t}^N with a weighted average of control states.

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{s \in S} w_s Y_{st}, \tag{2.4}$$

where w_s denotes the weight on the sth state which is unaffected by the policy intervention.

Ababie, Diamond, and Hainmueller [1] focus on minimizing the distance between the treatment and control groups in terms of outcomes in the pre-treatment period and predicted outcomes in the post-treatment period. As noted by them, I suppose that $\boldsymbol{X}$ and $\boldsymbol{X_s^0}$ are the $(K \times 1)$ vectors of predictors for New Jersey and for each sth state in the control group, and $\boldsymbol{V}$ is a $(K \times K)$ diagonal matrix with non-negative entries. Then, the optimal vector of weights, $\boldsymbol{W^*(V)}$, should minimize the squared distance conditional on $\boldsymbol{V}$ as follows:

$$(\boldsymbol{X} - \sum_{s \in S} w_s \boldsymbol{X_s^0})' \boldsymbol{V} (\boldsymbol{X} - \sum_{s \in S} w_s \boldsymbol{X_s^0}), \tag{2.5}$$

where $w_s \geq 0$ and $\sum_s w_s = 1$, for all s. Finally, I find the optimal $\boldsymbol{W}^*(\boldsymbol{V})$ to minimize the mean squared prediction error.

2.5 Empirical Results

2.5.1 Main Results

Table 2.2 presents the DID estimates of the short-term effects of New Jersey's gasoline tax increase on air quality.[15] My DID model is based on a four-year time window−two years before and two years after the effective date of the gasoline tax increase. Davis [24] notes that researchers should consider at least two-year air pollution data to control for seasonal variation in air quality. This table reports the results from different model specifications adding or removing controls. Column 1 displays the point estimates from fitting Equation (2.1), which includes monitor fixed effects only. Columns 2 and 3 progressively add controls to the specification based on Equation (2.1). The panels report the results for the two dependent variables associated with mobile source pollutants, namely, CO and NOx.

[15] Table A2.1 in Appendix reports the full point estimates of the DID regression.

Table 2.2: Effect of Gasoline Tax Increase on Air Quality

	(1)	(2)	(3)
Panel A. Log(CO)			
Post×Treat	−.022	−.040	−.036
	(.072)	(.074)	(.075)
N	150,803	150,803	150,803
R^2 (within-monitor)	.001	.121	.159
Panel B. Log(NO_x)			
Post×Treat	.032	−.003	−.010
	(.031)	(.028)	(.026)
N	242,844	242,844	242,844
R^2 (within-monitor)	.000	.123	.167
Monitor FE	✓	✓	✓
Time dummies		✓	✓
Weather			✓

Notes: This table shows the coefficients of DID regression of mobile source pollutants on the Post × Treat interaction. Panel A indicates the estimated effect on CO and Panel B indicates the estimated effect on NOx. Each column shows the results with three different model specifications adding or removing controls which are time dummies and weather variables. Standard errors presented in parentheses are clustered at monitor level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

The DID estimates suggest that the effects are negative but statistically insignificant. The point estimates in column 1 indicate that the gasoline tax increase results in 2.2 percent decrease in CO but 3.2 percent increase in NOx. These findings are not statistically significant at any confidence level. Column 3 shows that the gasoline tax increase reduces CO and NOx concentrations by 3.6 and 1 percent, respectively, which are not statistically significant. Relatively large decreases in CO concentration cannot be ruled out because the estimates are imprecise.[16] Weather variables and/or time dummies in the regression have less effect on the point estimate magnitude, though the regression's R^2 increases from .001 (.000) to .159 (.167). Increase in R^2 demonstrates that the weather variables and time dummies in the regression explain part of the estimated effects.

[16]The null hypothesis of a 5 percent decrease cannot be rejected at any significance level. Although the results for CO would be interpreted as 'statistically insignificant', the size of the effect is important. The confidence interval may include potentially important effects.

Table 2.3: Announcement Effect on Air Quality

	(1)	(2)	(3)
Panel A. Log(CO)			
Post×Treat	.007	.007	.036
	(.039)	(.039)	(.039)
N	6,644	6,644	6,644
R^2 (within-monitor)	.026	.038	.081
Panel B. Log(NO_x)			
Post×Treat	−.039	−.041	−.037
	(.025)	(.025)	(.025)
N	10,697	10,697	10,697
R^2 (within-monitor)	.029	.077	.130
Monitor FE	✓	✓	✓
Time dummies		✓	✓
Weather			✓

Notes: This table shows the coefficients of DID regression of mobile source pollutants on the Post × Treat interaction. Panel A indicates the estimated effect on CO and Panel B indicates the estimated effect on NOx. Each column shows the results with three different model specifications adding or removing controls which are time dummies and weather variables. Standard errors presented in parentheses are clustered at monitor level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

Although New York and Pennsylvania, which are neighboring states of New Jersey, are not controlled in my DID framework, New Jersey's gasoline tax increase may affect its consumer behavior and that of the neighboring states. Drivers in the two neighboring states may no longer have incentive to go to gas stations near New Jersey's border. Therefore, if a behavioral change or change in cross-border shopping exists, the estimates used in this study may be affected, which may lead to biased estimates. Retail gasoline prices of the three states are examined before and after the effective date of New Jersey's gasoline tax increase to identify potential problems. Figure A2.1 shows that the retail prices of regular gasoline are still lower in New Jersey than in the two neighboring states after the gasoline tax increase. This finding implies that drivers in New Jersey and in the two states are less likely to change their behavior.

As mentioned earlier, the DID method for policy evaluation typically sets up a post-treatment period as the time period after a policy implementation. However, if consumers respond to the announcement rather than the actual policy implementation, then policy evaluation estimates may be biased or inaccurate. Announcement of tax rate changes could have important implications for current earnings because current consumption may be affected by future tax rate changes (Auerbach and Kotlikoff [5]). Economists assume that individuals are a forward-looking group who respond environmental changes in economic decision-making because expectations are crucial to economic analysis (Blundell, Francesconi, and van der Klaauw [9]). Coglianese et al. [19] demonstrate that forward-looking drivers will accelerate gas tank refilling in the days leading up to gasoline tax increases.

I examine changes in air pollutants associated with automobile use between the announcement date and effective date of a gasoline tax increase. On September 30, 2016, the state government of New Jersey announced that a 23-cent increase in state gasoline tax would take effect on November 1, 2016. Based on my DID framework, I explore the impact of the policy announcement on air quality by using a narrow time window pre- and post-treatment periods, that is, one month before and one month after the announcement date. Table 2.3 presents the DID regression results with three different model specifications shown in Table 2.2. The announcement of gasoline tax increase does not significantly impact New Jersey's local air pollution, though column 3 shows 3.6 percent increase in CO, which is not statistically significant. This finding implies that no sudden change in gasoline purchases responds to expected gasoline price changes.

Table 2.4 presents the DID estimates of the effect of gasoline tax increase on co-pollutants, which are less associated with mobile source pollutants. As discussed previously, automobile emissions also contribute to air pollution levels of $PM_{2.5}$, PM_{10}, and SO_2, though the contribution is very small compared with main mobile source pollutants. Therefore, I would expect no significant effect of the gasoline tax increase on these co-pollutant concentrations. I obtain the results by estimating the same specifications presented in Table 2.2 but using two different dependent variables. The point estimates indicate that gasoline tax increase and $PM_{2.5}$ and SO_2 have no statistically significant relationship, though column 1 shows 3.1

Table 2.4: Effect on Co-pollutants

	(1)	(2)	(3)
Panel A. Log($PM_{2.5}$)			
Post×Treat	−.031**	−.019	−.016
	(.013)	(.013)	(.014)
N	142,812	142,812	142,812
R^2 (within-monitor)	.000	.047	.072
Panel B. Log(SO_2)			
Post×Treat	−.029	−.034	−.053
	(.068)	(.065)	(.070)
N	248,494	248,494	248,494
R^2 (within-monitor)	.008	.026	.062
Monitor FE	✓	✓	✓
Time dummies		✓	✓
Weather			✓

Notes: This table shows the coefficients of DID regression of co-pollutants on the Post × Treat interaction. Panel A indicates the estimated effect on $PM_{2.5}$ and Panel B indicates the estimated effect on SO_2. Each column shows the results with three different model specifications adding or removing controls which are time dummies and weather variables. Standard errors presented in parentheses are clustered at monitor level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

percent decrease in $PM_{2.5}$, which is a biased estimate without controls. The regression with PM_{10} is not estimated due to the lack of available data.

2.5.2 Robustness Checks

I conduct various robustness checks to address potential issues and provide alternative explanations for my results. First, I control for linear time trends that are specific to treated and control states by including interactions of state indicators with the linear time index. They allow air pollutants in each state to adjust at a different linear rate over time. Controlling for group-specific linear time trends may relax the parallel trends assumption in a DID model (Mora and Reggio [62]). Therefore, if the treatment effect in a DID regression is not sensitive to including group-specific linear time trends, then researchers may consider the baseline result more credible. Column 1 of Table 2.5 reports estimation results of my

DID model with state-specific linear time trends. These results suggest that the inclusion of time trends has less effect on point estimates. The linear time trends would be captured indirectly by time dummies in the main specification, which reflect unobserved time shocks.

Previous studies using high-frequency data have added lagged pollution measures, which are lagged dependent variables to the regression model to examine the persistence of air pollution effects on their results (Graff Zivin and Neidell [38]; Chen and Whalley [17]; Schwartz [84]). Following a consistent approach in the literature addressing this potential concern, I include one- and two-day lags of the air pollution outcome to examine whether the shorter-term lags affect my main findings. Column 2 reports these estimate results.[17] For CO and NOx, point estimates magnitudes are somewhat smaller than those in Table 2.2, and the point estimates are not statistically significant. Adding lagged pollution measures increases the regression's R^2 because they are correlated with an outcome variable.

One may be concerned that the effects of a gasoline tax increase vary across areas (e.g., urban and rural areas). This difference may be due to better public transportation systems in urban areas, which would be alternatives to automobile use. Additionally, a small number of monitors in rural areas may be outliers that are likely to change the results of data analysis because air monitors are mainly distributed in densely populated areas in this study. Most treated monitors in New Jersey are also located in urban areas. In column 3, I use air pollution data only for core-based statistical areas (CBSA), eliminating rural air monitor data.[18] The estimated CO magnitude is consistent with the main result, but the NOx estimate is positive. Thus, a relatively low pollutant level of numerous rural areas in a control group may be more likely to lower the control group's average level, resulting in the DID estimate.

States have unique geography, climate, and life styles. Thus, another potential concern with my estimates is whether air monitors in 25 control states used in this study represent a good control group of the DID model. To address this potential concern, I use an alternative

[17]Table A2.2 in Appendix reports the full point estimates of lagged dependent variables, suggesting that the point estimate remains unchanged and the estimated coefficients of one- and two-day lags are positive and statistically significant.

[18]The office of management and budget defines CBSA as US geographic areas that include one or more counties anchored by urban areas of at least 10,000 population. The EPA's air quality data contain information on CBSA.

Table 2.5: Robustness Checks

	Time trend (1)	*Lagged dep.* (2)	*Urban only* (3)	*Eastern only* (4)
Panel A. Log(CO)				
Post×Treat	−.007	−.010	−.030	−.034
	(.102)	(.029)	(.075)	(.077)
N	150,803	147,695	138,660	35,579
R^2 (within-monitor)	.174	.445	.168	.195
Panel B. Log(NO_x)				
Post×Treat	−.028	−.007	.021	.033
	(.070)	(.017)	(.024)	(.031)
N	242,844	239,716	175,428	39,184
R^2 (within-monitor)	.169	.275	.218	.293

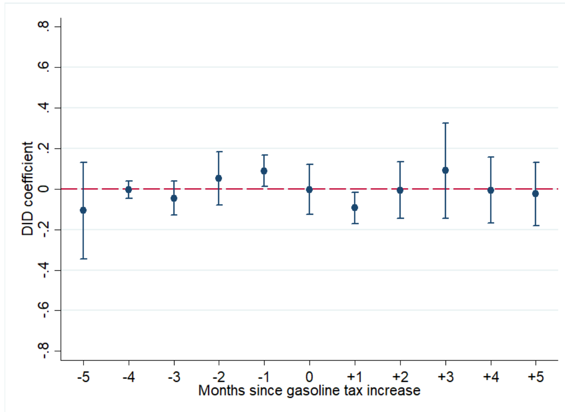
Notes: This table shows results of separate DID regressions testing the robustness of the model. Column 1 shows regression results including state-specific linear time trends. Column 2 shows regression results including lagged dependent variables as control variables. Column 3 shows regression results using data for monitors in urban area. Column 4 shows regression results using data for monitors in northeastern states. All regression models include weather variables, year, month, day-of-week time dummies, and monitor fixed effects. Standard errors presented in parentheses are clustered at monitor level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

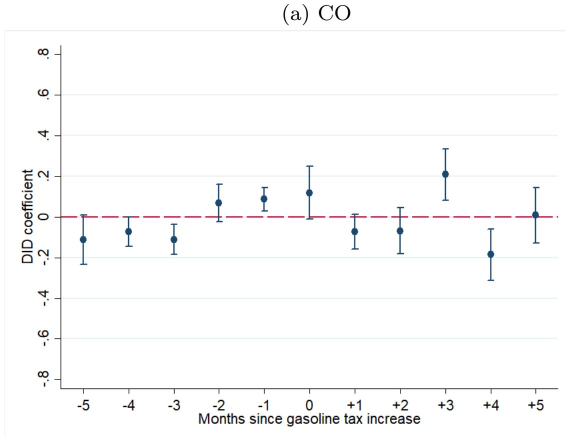
control group as monitors located in 6 nearby states and Washington, D.C. in the Northeast region to which New Jersey belongs, as shown in column 4.[19] Column 3 shows that the estimated CO magnitude is the same as the main result, but the NOx estimate is positive. Overall, the estimated effects of gasoline tax increase on air pollution are not statistically significant at any level across all four columns.

In many DID studies, researchers often implement another test for the common trend assumption by including treatment leads and lags. Plotting the test results provides a simple graphical representation. In general, if the effect of policy intervention is significant in the DID model, then pre-trending does not occur before the policy change. However, after the policy change, a sharp break in the trend is observed, leading to the long-term effect of policy dissipates or larger long-term effects of the policy. However, as my main results suggest no

[19]Six nearby states in the Northeast region are Connecticut, Delaware, Maine, Massachusetts, New Hampshire, and Vermont. These states did not have any change in gasoline tax policy in 2014−2018.



(a) CO



(b) NOx

Figure 2.4: Estimates of Gasoline Tax Increase with Leads and Lags

Notes: This figure plots the estimated coefficients obtained from a specification including treatment leads and lags. Coefficients are defined as months relative to the month the gasoline tax increase begins in New Jersey.

significant effect of a gasoline tax increase, I would expect no anticipation effects and phase-in effects in this test.

Specifically, I estimate the following model using month-level monitor data:[20]

$$ln(y_{it}) = \alpha + \sum_{j=-m}^{q} \beta_j DID_{it+j} + \lambda \boldsymbol{W_{it}} + \mu \boldsymbol{D_t} + \delta_i + \epsilon_{it}, \tag{2.6}$$

[20]Since leads and lags of treatment indicator with high-frequency data may not be feasible in this test, I use monthly data which is an aggregated time level, not daily data.

where DID_{it} is an indicator of the treatment, which is the same as $Post \times Treat$ in the main specification. I have included q leads and m lags of the treatment effect, rather than a single treatment effect. In other words, β_j is the coefficient on the jth lead or lag. If the coefficients on all leads of the treatment are zero, then the common trend assumption is valid.

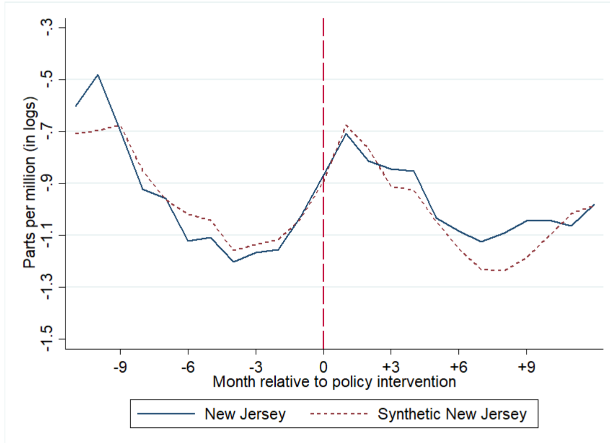
To explore these dynamics, Figure 2.4 shows a plot of the coefficients of the DID model augmented with treatment leads and lags. Specifically, I add indicator variables for $1-5$ months before New Jersey's gasoline tax increase and $0-5$ months after the tax increase.[21] For CO and NOx, the coefficients on the leads are close to zero, suggesting weak evidence of an anticipatory effect. This finding is consistent with the results in Table 2.3. As expected, no sharp break exists even when the tax increase begins, and the coefficients are close to zero in the first few months after the tax increase. Figure 2.4 provides further evidence of the validity of the identification strategy used in this study, which support the main findings.

2.5.3 Synthetic Control Method

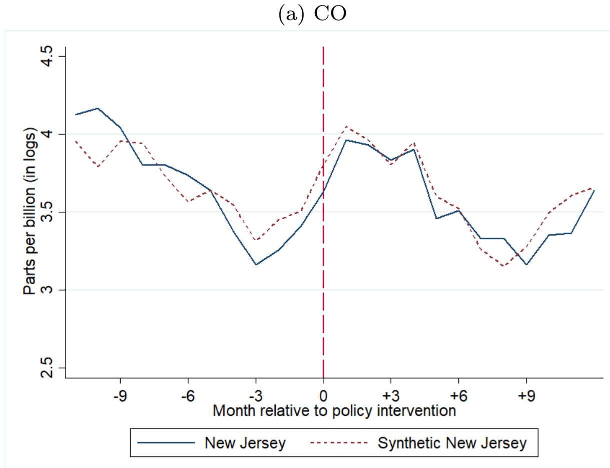
SCM provides pre- and post-treatment trends for New Jersey's air pollutant concentrations compared with the concentrations of synthetic control state. In particular, pre-treatment trends should follow each other closely before New Jersey's gasoline tax increase in this study. This analysis aims to examine New Jersey's air pollution changes after November 2016 in the absence of the tax increase. A synthetic New Jersey, which is a combined state in donor pool, most closely resembles actual New Jersey in terms of pre-treatment values of air pollutant concentrations predictors. Table A2.4 in the Appendix presents a comparison between the pre-treatment factors of the actual New Jersey and those of the synthetic New Jersey for each analysis with two air pollution outcomes. The predictors closely match the actual and synthetic New Jersey in the months that provide the close pre-treatment fit in the SCM estimation. The selected control states in the donor pool and their weights are presented in Table A2.5 in the Appendix.

Figure 2.5 shows the trends of monthly CO and NOx concentrations for the actual and synthetic New Jersey using a two-year window. The estimated impact of gasoline tax increase on air quality indicates the differences in pollutant concentrations between the actual and

[21] Table A2.3 in the Appendix reports estimates of leads and lags of the treatment effect.



(a) CO

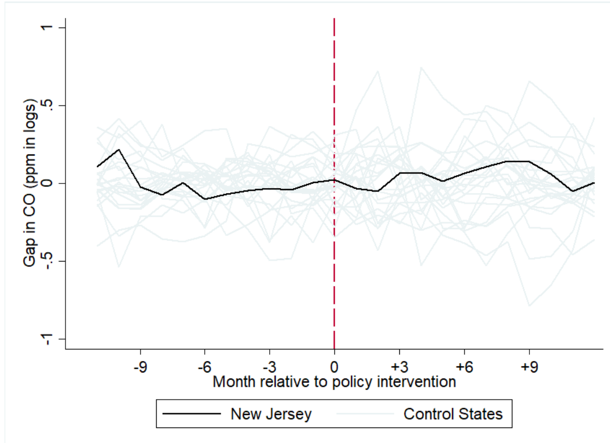


(b) NOx

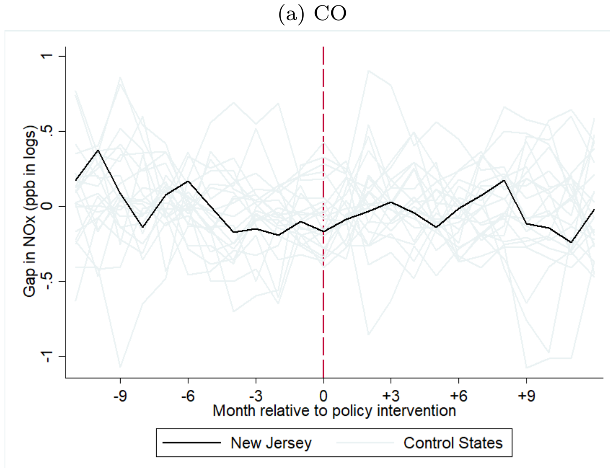
Figure 2.5: Air Pollution Trends: New Jersey vs Synthetic New Jersey

Notes: This figure shows the trends of monthly mobile source pollutant concentrations, CO and NOx. Solid lines indicate New Jersey while dashed lines indicate synthetic New Jersey. Vertical dashed bar represents the policy effective date, November 2016.

synthetic New Jersey in the post-treatment periods. The synthetic New Jersey's trend closely follows the actual New Jersey's trend before the policy intervention. This finding implies that the synthetic New Jersey is a good counterpart of the actual New Jersey. The two trends have no differences even after the policy intervention. Therefore, New Jersey's gasoline tax increase is less associated with air quality improvement. Although I examine the effect of



(a) CO



(b) NOx

Figure 2.6: Pollutant Concentrations Gaps in New Jersey and Placebo Gaps in All Control States

Notes: This figure shows the gaps in concentrations of each pollutant in New Jersey and placebo gaps in all control states in donor pool. Black lines denote the gaps estimated for New Jersey while gray lines denote the gaps associated with each of all control states of the test. The gray lines indicate the difference in concentrations between each state in donor pool and its synthetic state.

a gasoline tax increase at the aggregate level using monthly state-level data, this result is consistent with the earlier DID estimation observed at the disaggregated level.

To evaluate the significance of the estimates, I conduct additional SCM analyses assuming that the remaining states in donor pool receive treatment. Although researchers only use the SCM model where the synthetic unit closely matches the treated unit, they could fail to hold the same condition for the remaining control units in the donor pool (Ferman and Pinto [29]). Following Ababie, Diamond, and Hainmueller [1], I re-estimate assuming that one of the remaining states in the donor pool, other than New Jersey, is treated. A graphical presentation of how the trends of the actual New Jersey differ from those of the remaining control states is a useful information. Figure 2.6 shows the differences between the synthetic and actual New Jersey and the remaining states in the donor pool. The gray lines indicate the difference in pollutant concentrations between each state in the donor pool and its synthetic state. The black lines represent the gap estimated for New Jersey. I reaffirm the insignificant impacts of gasoline tax increase because New Jersey and other states show similar trends in the study period.

2.5.4 Additional Evidence

In this section, I seek to find additional explanations for the insignificant change in air pollution similar to Davis' [24] additional evidence for examining behavioral responses. Davis notes that understanding behavioral responses is important to policy changes in automobile use to explain air quality effects. His work investigates a driving restriction in Mexico City in 1989 that was based on the last digit of the vehicle's license plate. He finds that the program did not improve air quality and a decrease in gasoline sales or increase in public transportation has no evidence. Based on Davis' approach with plausible reasons, I explore gasoline consumption changes, VMT, and public transportation use in this study. These factors are important in determining New Jersey's air quality.

Table 2.6 reports the point estimates for the effect of a gasoline tax increase on additional outcome variables. A model specification of the DID estimation slightly differs from the baseline specification because these outcomes are based on monthly state-level data. It should be noted that clustering at the state level in this analysis, I may fail to control for within-cluster error correlation with small standard errors. Cluster-robust standard errors will be biased downward when clusters are few (Cameron and Miller [13]). Researchers generally apply bootstrap approach to clustering to address this issue. Therefore, I implement a cluster bootstrap by re-sampling with replacement 1,000 times from the original sample of clusters.

Table 2.6: Additional Outcomes

Dependent variable:	Gasoline volume	VMT	Public transportation
	(1)	(2)	(3)
Post×Treat	−.003	.024	−.006
	(.036)	(.041)	(.022)
N	1,097	1,170	1,170
R^2 (within-state)	.815	.591	.879

Notes: This table shows results of separate DID regressions using different outcome variables. Three outcome variables are the natural logs of of gasoline sales volume, VMT, and public transportation ridership using state-month level data. Cluster-robust standard errors presented in parentheses are calculated by using 1,000 bootstrap samples at state level.

*** Statistical significance at the 1 percent level.

** Statistical significance at the 5 percent level.

* Statistical significance at the 10 percent level.

As shown previously in Table 2.1, gasoline consumption has further decreased in New Jersey compared with the control states because the gasoline tax has increased. The simple comparison could be valuable for examining the impact of gasoline tax increase on driving behavior. However, controlling for time and weather and using fixed effects in a regression analysis, I find no strong evidence that New Jersey's gasoline tax increase significantly reduced local gasoline sales volume. The point estimate in column 1 shows that the gasoline tax increase has decreased New Jersey's gasoline sales volume by 0.3 percent relative to the volume changes in the control states. However, this point estimate is not statistically significant. Due to the lack of evidence of the reduction in gasoline sales volume, driving behavior change is not expected.

A large literature has employed traffic volume trends and VMT to explore driving behavior changes. Similar to gasoline sales volume, air quality improvement could not be supported without evidence on reduction in VMT as it directly relates to automobile use and its emission. Column 2 shows that New Jersey's VMT has increased by 2.4 percent after the gasoline tax increase compared with that of the control states. The relationship between gasoline tax increase and VMT is not statistically significant. Additionally, many public transportation options, such as subway and bus systems, may correlate with automobile travel behavior (Shaheen and Cohen [85]). Table 2.6 also reports that New Jersey's public transportation ridership has decreased by 0.6 percent in post-treatment periods. This finding indicates that public transportation does not change with a gasoline tax increase. Therefore, the results from these additional analyses help explain that New Jersey's gasoline tax has no significant effects on local air quality.

2.6 Conclusion

Transport is a major air pollution source. Hence, a gasoline tax is an important policy instrument for reducing automobile emissions. Previous studies have shown that drivers respond to gasoline tax changes differently from price changes. Despite the importance of gasoline taxes and driving behaviors, studies investigating a direct relationship between gasoline tax policy and automobile pollution are limited. Thus, this study aims to find causal evidence of the impact of gasoline tax increase on air quality by focusing on New Jersey's state policy change.

My analysis relies on variation differences in air quality across New Jersey and the control states, before and after New Jersey's gasoline tax increase in 2016. Automobile pollution estimates are negative but not statistically significant. The DID results are consistent with the SCM results. The study results are also robust to alternative model specifications. Furthermore, New Jersey's gasoline tax increase was not associated with any significant change in gasoline sales volume and VMT, which represent the measure of automobile use.

I do not calculate an optimal level of gasoline tax in this study but find that the current tax increase is still too low to have a meaningful improvement in air quality, at least in the short term. My results correspond to the findings of previous studies that the optimal gasoline tax should be more than one dollar per gallon. Although I provide new evidence on air quality effects of a recent gasoline tax reform, exploring long-term effects is a reasonable next step for understanding driving behavior changes. The short-term effects need to be examined in light of the recent growing interest in gasoline tax increase in federal and state governments. Nevertheless, more precise and reliable effects of gasoline tax increase may be observed under a longer period.

Several potential directions for future research are available. First, applying a DID model with multiple treatments to examine air quality effects of gasoline tax increase would be interesting. New Jersey had another tax increase of 4.3 cents per gallon in October 2018. Thus, the effects of its gasoline tax increase on air quality could be measured using a DID model with multiple treatments in multiple time periods. Second, the relationship between gasoline taxes and health outcomes may be assessed. Although a large literature has explored the health benefits of regulations and policies associated with automobiles, empirical works on gasoline tax reform's effects on health outcomes are scarce.

Chapter 3

Effect of Temperature on Crime in California's Urban and Rural Areas

3.1 Introduction

Crime generates substantial economic and social costs for individuals, communities, and nations. According to the Bureau of Justice Statistics (BJS), the cost of the criminal justice system was approximately $280 billion in fiscal year 2012 (BJS, 2016).[1] The total cost should be much greater when considering non-financial costs, such as the cost of medical care and property damage losses. Thus, it is important to design effective policies and strategies that remove the causes and prevent the occurrence of crimes. Most existing research on factors influencing crime has focused on income (inequality), educational attainment, and public expenditure on police (e.g., Buonanno [11]; Di Tella and Schargrodsky [27]; Oreopoulos [68]; Draca, Machin, and Witt [28]; Machin, Marie, and Vujić [59]). Furthermore, recent literature has begun to consider air pollution or weather conditions as determinants of crime and suggested that solving environmental problems can play an important part in reducing crime costs (Herrnstadt and Muehlegger [43]; Goin, Rudolph, and Ahern [36]; Bondy, Roth, and Sager [10]).

[1] The cost includes federal, state, and local governments' spending on police protection, corrections, court, and incarceration.

Crime is a bigger social problem in large cities than in small cities.[2] Most previous research on urban and rural differences in crime has addressed economic and social conditions (Glaeser, Sacerdote, and Scheinkman [35]). Glaeser and Sacerdote [34] describe the correlation between crime and city size by using three main channels−higher returns to crime, lower probability of recognition, and residents' characteristics in urban areas.[3] Greater access to the wealthy and higher victim density in a large city or an urban area causes greater returns to crime. Criminal activity in large cities might be less likely to be discovered due to having many suspects in areas of high population density.

Meanwhile, existing research in psychology and criminology has proposed a hypothesis that weather is strongly associated with criminal activities (Anderson [2]; Cohn [20]; Field [31]; Rotton and Cohn [79]). Temperature has a positive effect on violent crime and other aggressive behavior because high temperatures induce feelings of anger and hostility. The relationship between weather and crime also comes from a more convincing basis, namely, rational choice theory in economics. A canonical model of crime proposed by Becker [7] explains that the decision to commit a crime is based on the expected cost and benefit. Criminal activities are no different from non-criminal activities in that an individual would commit a crime if his or her expected utility from committing the crime outweighs the utility from non-criminal activities. In this model, weather conditions could be a potential variable influencing the decision-making about committing criminal acts.

Previous empirical studies in economics have provided evidence that criminal activities are more likely to occur in higher temperatures. Jacob, Lefgren, and Moretti [47] investigated the short-run dynamics of criminal behavior with weather and found that hot weather leads to an increase in both violent and property crime rates, whereas high rainfall decreases violent crime rates. Using a 30-year panel of monthly crime and weather data for U.S. counties, Ranson [77] examined the effects of weather on monthly criminal activities and

[2]There is a general consensus that crime rates or volumes are much higher in large cities than in small cities (e.g., Laub [51]; Ladbrook [50]; Glaeser and Sacerdote [34]; Nolan III [66]; Chang, Kim, and Jeon [15]). Previous studies argue that differences in crimes between large and small cities can be attributed to differences in their population compositions and characteristics such as its size, density, and heterogeneity.

[3]Gleaser and Sacerdote [34] show that a one-percent increase in population size lead to approximately 0.1 percent increase in serious crimes per capita, using the Uniform Crime Reports (UCR) data.

demonstrated a strong relationship between temperature and crime.[4] Across a variety of offenses, warmer temperatures are correlated with higher violent crime rates.

I question whether the effects of weather impact urban-rural crime differences. To the best of my knowledge, there is no empirical study on the effects of weather on urban-rural crime differences, though weather patterns are known to have a significant impact on criminal activity. I address the gap in the literature and provide the first estimates of the effects of temperature on crime across urban and rural areas. My analysis draws on a panel dataset of monthly crime and weather in California cities for ten years, 2006−2015. Crime data comes from the U.S. Federal Bureau of Investigation's Uniform Crime Reporting (UCR) program, and I focus more on the number of monthly arrests for violent and property crimes from 380 local law enforcement agencies throughout California. I use arrest data as a proxy for crime levels as is common in this literature (e.g., Sah [81]; Glaeser, Sacerdote, and Scheinkman [35]; Glaeser and Sacerdote [34]).

This study makes an important contribution using monthly crime data. Relevant studies have used aggregate annual crime data because collecting monthly crime data has been challenging. Using annual crime and weather data would result in less precise estimates because annual weather cannot reflect seasonality or month-to-month variation over a year. That is, the actual effects might be attenuated using aggregate weather data. Although arrest counts or rates used in this study are not the same as crime counts or rates, they help indicate how crime patterns have changed and represent the lower bound of the incidence of crime.[5] Thus, disaggregated data helps reveal crime patterns that may be concealed by aggregate data.

I examine urban-rural crime differences by defining a metropolitan area with a population of 50,000 or more as an urban area and areas with smaller populations as rural. This definition of metro and non-metro areas is used as a common classification in the literature to compare urban and rural areas (e.g., Porter et al. [76]; Plummer and Headd [74]; Lee and Xu [54]). I use 155 cities as urban areas and 225 cities as rural in 380 California cities in this study. My empirical strategy is based on a fixed-effects Poisson regression model which

[4] I use city-level data, while Ranson used county-level data. I attempted to have more (or at least the same number of) years of data than Ranson, but monthly historical weather data by city was limited.

[5] It is likely that the actual numbers of crimes are understated because not all crimes result in arrests.

is appropriate for a count outcome variable. I explore the marginal effects of temperature for metro and non-metro areas to see more clearly the relationship between temperature and urban-rural crime differences.

The main results suggest that higher temperatures increase the number of violent crimes, which is consistent with existing research that demonstrates the relationship between hot weather and aggressive behavior. For a one-degree increase in temperature, the monthly number of violent crimes in California increases by 0.2 percent. However, results show that the number of property crimes is not correlated with temperature. On average, the number of violent crimes is 8.7 percent higher in urban areas than in rural areas, holding temperature constant. The number of violent crimes in both areas tends to increase in proportion to temperature, but the marginal effects of temperature are smaller in urban areas than in rural areas. This implies that higher temperatures have more influence on violent criminal activities in small cities.

These results have important implications for crime policy. Understanding the relationship between temperature and crime can help local law enforcement agencies allocate limited resources more efficiently. If higher temperatures are correlated with more crimes, local agencies may put more police officers on duty in warmer areas and during warmer days to reduce the economic and social costs of crimes. Policymakers may also try to increase the size of large cities' police forces or introduce more punitive sentencing policies to reduce violent crime levels in urban areas. Moreover, policymakers should develop policies that make small cities and towns safer, particularly in hot weather.

I conduct several sensitivity analyses and robustness checks, including alternative specifications and combinations of controls. Baseline results were confirmed using alternative temperature bins and population subgroups. Adding or removing control variables does not affect the estimated coefficients presented in my preferred specification. Results are unchanged when I use mean or minimum temperature rather than maximum temperature as a predictor variable. To address concerns about count data, I also apply a simple fixed-effects regression model after calculating arrest rates and find consistent estimates.

3.2 Data

3.2.1 Data Sources

My analysis relies on a combination of two primary data sources: criminal and weather records. For crime information, I used monthly crime data at the city level in California for the ten-year period (2006−2015) provided by the FBI's UCR program.[6] Each month, law enforcement agencies across the United States collect and provide detailed crime records to the FBI, which then creates combined reports. The UCR has information not only on the total monthly number of crimes by city but also on the type of crimes. The FBI classifies criminal offenses into two groups according to their seriousness: Part I offenses and Part II offenses. To narrow the focus to major crimes, data for the Part I offenses was collected, which was divided into two categories: violent and property crimes. Violent crimes include murder or manslaughter, forcible rape, robbery, and aggravated assault. Property crimes include burglary, larceny-theft, motor vehicle theft, and arson.

I focused on the number of arrests for various offenses from 380 local law enforcement agencies, including city police departments in California. The Inter-university Consortium for Political and Social Research (ICPSR) provides monthly arrest count data rather than crime incident data. As mentioned earlier, using disaggregated data helps reveal crime patterns that may be concealed by aggregate data. Also, I restricted the sample of this study to crime data reported by city law enforcement agencies. It is difficult to control for city-to-city variation in crimes because there is no city information from county and state agencies. Thus, only data from city police departments was used.[7]

The second main component of my dataset is monthly weather data, particularly temperature, which might be predictive of variation in crime levels. Temperature data comes from the US National Climatic Data Center's Global Summary of the Month. There are 314 weather monitoring stations throughout California that have readings during the sample period of 2006 to 2015. Some weather stations are missing observations, so I restricted the

[6] I obtained the UCR crime data from the Inter-university Consortium for Political and Social Research (ICPSR) which restructures the original data to a rectangular format.

[7] Arrest data from county and state law enforcement agencies accounts for about 7 percent of all samples in my dataset.

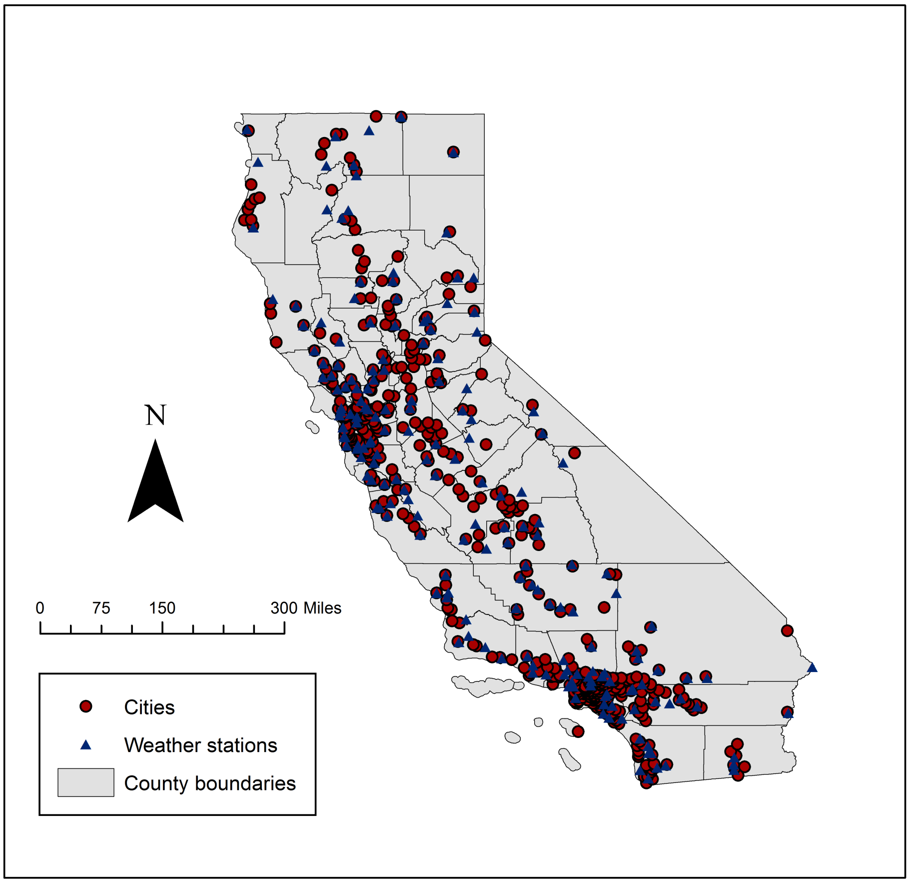


Figure 3.1: Cities and Weather Stations in California

Notes: The points in this figure represent locations of cities and weather monitoring stations throughout California.

sample to consistently observed stations during that period. Then, I assigned temperature to cities by linking temperature to the weather monitoring stations closest to the city centroid. For each city, I identified the nearest weather station using ArcGIS software and ended up using data from 195 stations located near California cities (see Figure 3.1).

The dataset also includes socioeconomic characteristics associated with crimes including population, sex, race, age, poverty status, median household income, and unemployment rate. These control variables rely on data at the county-year level. For that reason, I assume that socioeconomic characteristics of counties represent those of cities that belong to the county. Detailed data on socioeconomic characteristics come from the Census Bureau's

American Community Survey program. In a sensitivity analysis, I considered city-specific time trends to eliminate the effect of exogenous factors on crimes.

To identify urban-rural crime differences, I divided California into two areas based on population size−metro and non-metro areas. Among other information about local law enforcement agencies, UCR data collected from ICPSR indicates whether agencies are located in a metro area. The definition of a metro area in the data follows the U.S. Office of Management and Budget and that is used by the Census Bureau.[8] A metropolitan statistical area has at least one urbanized area with a population of 50,000 or more, plus adjacent communities having a high degree of social and economic integration. As mentioned above, this definition of metro and non-metro areas is often used in literature to classify urban and rural areas in the U.S.

3.2.2 Summary Statistics

Table 3.1 shows the summary statistics for urban and rural areas across key variables, based on the city-month level. In particular, the numbers of arrests for violent and property crimes are much higher in metro than in non-metro areas. The relationship between city size and crime is well established. Crime statistics in California are consistent with crime trends at the national level. According to the BJS, the annual violent crime rates of U.S. are 22.2 per 1,000 persons in urban, 19.3 in suburban, and 18.3 in rural areas, respectively. The annual property crime rates are 148.8 per 1,000 households in urban, 101.7 in suburban, and 103.2 in rural areas, respectively (BJS, 2014).

The annual average temperature of a large city tends to be 1 to 3 degrees Celsius (1.8 to 5.4 degrees Fahrenheit) warmer than its surroundings due to the urban heat island effect (EPA, 2018). In Table 3.1, metro areas are 0.7 to 0.8 degrees warmer than non-metro areas, but this is not a significant difference in temperature. These temperature differences may result from the unique geographical features of California. The relatively low temperatures in coastal areas of the largest cities (such as Los Angeles and San Francisco) may lower the average temperature of the metro areas. Moreover, non-metro areas include small cities in

[8] The Census Bureau also defines Urbanized Areas as places of 50,000 or more people and Urban Clusters of at 2,500 and less than 50,000 people (US Census Bureau, 2010).

Table 3.1: Summary Statistics

	All areas	Metro	Non-metro
Number of arrests			
Violent	13.43 [15.88]	25.99 [19.48]	6.00 [5.34]
Murder	.48 [1.75]	1.13 [2.67]	.09 [.49]
Forcible rape	.59 [1.59]	1.29 [2.32]	.18 [.62]
Robbery	3.22 [5.48]	7.02 [7.25]	.98 [1.80]
Aggravated assault	9.14 [8.67]	16.56 [9.30]	4.75 [4.08]
Property	21.18 [22.46]	41.26 [23.95]	9.30 [9.14]
Burglary	7.57 [7.99]	14.54 [8.50]	3.45 [3.64]
Larceny-theft	10.42 [10.89]	20.10 [10.77]	4.69 [5.65]
Motor vehicle theft	2.80 [4.89]	5.81 [6.68]	1.02 [1.75]
Arson	.39 [1.16]	.81 [1.67]	.14 [.58]
Temperature ($^\circ C$)			
Average temperature	16.7 [5.7]	17.2 [5.2]	16.4 [6.0]
Maximum temperature	23.3 [6.9]	23.8 [6.3]	23.1 [7.2]
Control variables			
Population (1000s)	88.3 [338.9]	204.2 [536.2]	19.8 [12.7]
Male (%)	50.1 [1.1]	49.9 [0.8]	50.2 [1.3]
Age	34.8 [3.5]	34.7 [3.0]	34.9 [3.8]
White (%)	65.6 [11.5]	62.5 [10.4]	67.4 [11.8]
Poverty (%)	14.0 [4.7]	13.3 [4.0]	14.5 [5.0]
Income (1000s of dollars)	59.5 [13.1]	62.0 [11.4]	58.1 [13.8]
Unemployment (%)	9.4 [3.4]	9.1 [3.3]	9.6 [3.4]
Number of cities	380	155	225

Notes: This table shows means and standard deviations for both violent and property crime arrests, temperature, and population at the city-month level. Data for socioeconomic characteristics is based on the county-year level. I use data for the sample period of years 2006−2015.

the desert region in California. These small cities have relatively high average temperatures, leading to increasing the average temperature of non-metro areas.

As mentioned in the introduction, previous studies have explained urban-rural crime differences by linking social and economic characteristics. Table 3.1 also presents summary statistics for the socioeconomic factors, which may influence criminal activities. It is no surprise that population size is much larger in metro areas than in non-metro areas. Metro areas have a lower proportion of white people than non-metro areas because a large city tends to be more ethnically diverse compared to a small city. The ratios of sex and average age do not differ between metro and non-metro areas. Existing studies also address significant impacts of both income and unemployment on crimes (e.g., Gould, Weinberg, and Mustard [37]). Table 3.1 shows that economic and labor market conditions are worse in non-metro areas than in metro areas in California. Criminal activities in rural areas may be driven by economic factors.

3.3 Methodology

The summary statistics in the previous section report that metro areas have a high temperature and many crime arrests. This section introduces an econometric model to identify a causal relationship between temperature and crime. First, the effects of temperature on crime were estimated as in previous empirical studies. Then, I evaluated whether higher temperatures aggravate urban and rural crime differences. If temperature is a principal factor influencing criminal behavior and urban areas have higher temperatures than rural areas, it is reasonable to hypothesize that higher temperatures widen the gap between urban and rural crimes.

Based on the panel structure of count data, I employ a fixed-effects Poisson regression model introduced in detail by Hausman, Hall, and Griliches [42]. I assume that $Crime_{imt}$, the number of arrests for a given crime in month m of year t in city i, has a Poisson distribution with a probability density function given by

$$f(Crime_{imt} \mid \boldsymbol{X}_{imt}) = \frac{exp(-\mu(\boldsymbol{X}_{imt}))\mu(\boldsymbol{X}_{imt})^{Crime_{imt}}}{Crime_{imt}!} \tag{3.1}$$

where $\boldsymbol{X}_{imt}$ denotes a vector of all observed covariates and $\mu(\boldsymbol{X}_{imt})$ is a parametric form for the conditional mean of $Crime_{imt}$ given $\boldsymbol{X}_{imt}$, $E(Crime_{imt} \mid \boldsymbol{X}_{imt})$. Following the standard assumption, $\boldsymbol{X}_{imt}$ takes an exponential form:

$$\boldsymbol{X}_{imt} = exp(\beta Temp_{imt} + \delta_i + \lambda_m + \rho_t). \tag{3.2}$$

A variable $Temp_{imt}$ is the average monthly maximum temperature observed in that month. δ_i is city fixed effects to control for unobservable factors within a city influencing crimes. λ_m and ρ_t are month and year fixed effects, respectively, which capture seasonal and temporal patterns of criminal activity. All regression models are estimated with robust standard errors clustered at the city level.

Poisson regression is a workhorse procedure in count data analysis because a Poisson distribution expresses the probability of any discrete number of random events (i.e., 0, 1, 2, ...) (Osgood [69]). There are several advantages to using a Poisson regression model in crime analysis. Ranson [77] uses a Poisson regression approach based on three considerations. First, there may be several zero values for arrest counts in crime data, especially in a small city or area. A simple log-linear regression approach is inappropriate because it considers these zero values as missing values. Second, although the Poisson distribution is not an exact fit for the distribution of crime data, the Poisson regression by maximum likelihood estimation yields unbiased estimates. Moreover, the Poisson regression can eliminate a potential incidental parameters problem due to multiplicative separability from the linear form of the exponential link function (Cameron and Trivedi [14]).

I believe that Equation (3.1) yields credible estimates of a reasonable amount of variation within each city over time. Nevertheless, I am not certain that the effects of temperature are homogeneous across all cities because a fixed-effects model merely shows the average within-city effect. How the effects vary across different city groups is of interest since relevant literature generally assumes that the relationship between weather and crime is uniform across regions. One might estimate separate regressions for each group and compare the effect of temperature in different subgroups, but this strategy may be quite unwieldy and inefficient when the effect of temperature depends on the city.[9] A better and simpler solution is to estimate a single fixed-effects model by incorporating a city group dummy and an interaction term for group with temperature. This would also be appropriate to examine accurately marginal effect of temperature for each group.

Adding a metro dummy and an interaction term between temperature and metro, I estimate the effect of temperature on crime across metro and non-metro areas. More formally,

[9] When researchers want to compare the coefficients and see their differences in separate regressions, they should calculate a p-value for the difference. Alternatively, simply including an interaction term in a regression model, they can get the p-value for the interaction term which provides a significant test for the difference in the coefficient.

I modify the Equation (3.2) as follows:

$$\boldsymbol{X}_{imt} = exp(\gamma_1 Temp_{imt} + \gamma_2 Metro_i + \gamma_3 Temp_{imt} \times Metro_i + \delta_i + \lambda_m + \rho_t), \tag{3.3}$$

where the variable $Metro_i$ is a dummy variable equal to 1 if city i is in a metro area. Thus, I am primarily interested in the coefficient γ_3, which tells how much the effect of temperature differs between metro and non-metro areas. The coefficient γ_1 is now the unique effect of temperature on crime only for non-metro areas. Furthermore, four additional specifications include alternatives to the variable $Temp_{imt}$ in sensitivity analyses. I attempt to categorize temperature into five groups based on 5-degree bins instead of using a continuous variable. I also use alternative variables, which are average monthly mean temperature, average monthly minimum temperature, and the number of days with the maximum temperature above 21.1 degrees Celsius (70 degrees Fahrenheit) rather than average monthly maximum temperature.

3.4 Results

3.4.1 Main results

Table 3.2 shows fixed-effects Poisson regression estimates of the effect of temperature on crime. Columns 1 to 3 in Table 3.2 display the results from regressions using the number of violent crimes as an outcome variable, and columns 4 to 6 display the results using the number of property crimes as an outcome variable. Columns 1 and 4 present the results from fitting Equation (3.2), and columns 3 and 6 show the results from fitting Equation (3.3). For violent crimes, the number of arrests is 1.002 ($= e^{.002}$) to 1.003 ($= e^{.003}$) times greater for an additional unit in temperature, that is, a 0.2 to 0.3 percent increase in the number of arrests for every one-degree increase in temperature. However, there is no relationship between temperature and property crimes. These results are consistent with the findings of relevant previous research suggesting a positive relationship between temperature and violent crimes. Higher temperatures are likely to produce aggressive behavior and feelings of anger and hostility (Anderson [2]).

Table 3.2: Effect of Temperature on Crime

	Violent crimes			Property crimes		
	(1)	(2)	(3)	(4)	(5)	(6)
Temp	.002***	.002***	.003***	−.000	−.000	.000
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Metro		.087**	.141***		.000	.014
		(.039)	(.046)		(.032)	(.039)
Temp×Metro			−.002**			−.001
			(.001)			(.001)
N	16,600	16,600	16,600	16,600	16,600	16,600
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year and month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows results of separate regressions using temperature, metro, and interaction term. Two dependent variables are the number of violent crime arrests in columns 1 to 3 and the number of property crime arrests in columns 4 to 6. All regression models include city, year, month fixed effects, weather, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.
*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

Column 2 reports that the number of violent crimes is 8.7 percent higher in metro areas than non-metro areas, but there is no property crime difference between metro and non-metro areas, as seen in column 5. This case assumes that the effect of temperature is uniform across metro and non-metro areas, but the intercept changes in the regression model.[10] In columns 3 and 6, I add an interaction term to the regression model between temperature and metro areas to examine not only the main effect of temperature but also how the temperature effect varies between metro and non-metro areas. I find a significant interaction effect, as shown in column 3, suggesting that for every one-degree increase in temperature, the number of violent crimes is, on average, 0.2 percent less in metro areas than in non-metro areas. That is, the main effect of temperature in metro areas is a 0.1 percent increase in the number of violent crimes, while the main effect in non-metro areas is a 0.3 percent increase in the number of violent crimes. This suggests that the effect of temperature on violent crimes is larger in non-metro areas than in metro areas.

[10] That is, the coefficient for *Metro* shows how the intercept changes for metro areas.

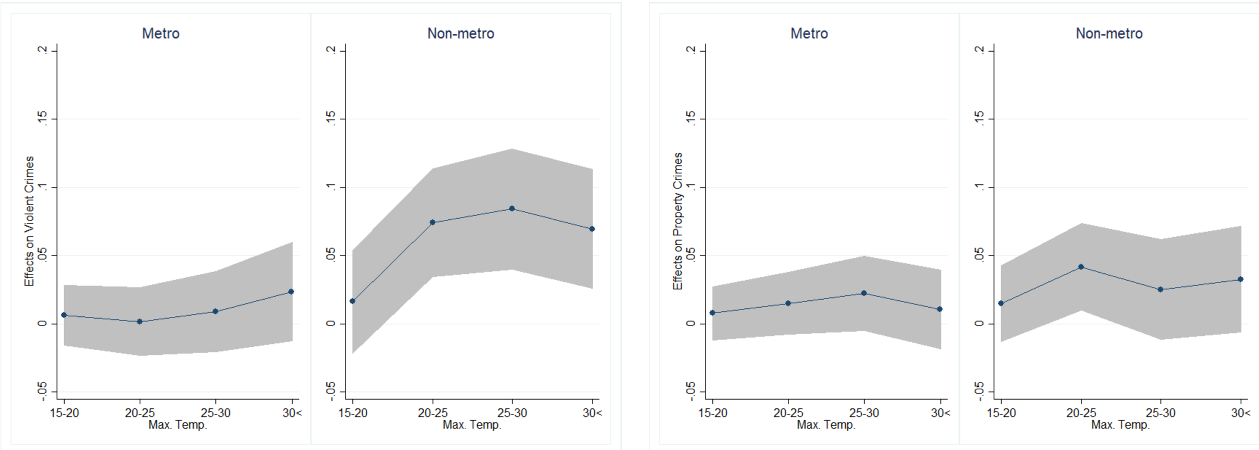
Table 3.3: Temperature Dummies

	Violent crimes			Property crimes		
	(1)	(2)	(3)	(4)	(5)	(6)
$Temp_{15}$	.015	.014	.016	.011	.011	.015
	(.011)	(.011)	(.019)	(.008)	(.008)	(.014)
$Temp_{20}$	.028**	.028**	.074***	.023**	.023**	.042**
	(.012)	(.012)	(.020)	(.010)	(.010)	(.016)
$Temp_{25}$	.037**	.037**	.084***	.024**	.024**	.025
	(.015)	(.015)	(.023)	(.012)	(.012)	(.019)
$Temp_{30}$	.044**	.044**	.070***	.017	.017	.033*
	(.018)	(.018)	(.022)	(.014)	(.014)	(.020)
Metro		.083**	.128***		.003	.017
		(.040)	(.043)		(.030)	(.033)
$Temp_{15} \times Metro$			−.010			−.007
			(.021)			(.017)
$Temp_{20} \times Metro$			−.072***			−.027
			(.021)			(.018)
$Temp_{25} \times Metro$			−.075***			−.003
			(.023)			(.020)
$Temp_{30} \times Metro$			−.046**			−.022
			(.022)			(.020)
N	16,600	16,600	16,600	16,600	16,600	16,600
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year and month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows results of separate regressions using dummies for temperature and metro, and interaction term. Two dependent variables are the number of violent crime arrests in columns 1 to 3 and the number of property crime arrests in columns 4 to 6. All regression models include city, year, month fixed effects, weather, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

Table 3.3 shows the results of the extension of the baseline model. In this extension, I use temperature indicators, a discontinuous variable, defined by five average monthly maximum temperature bins (below 15, 15−20, 20−25, 25−30, and over 30 degrees). The temperature bins then interact with a metro dummy variable. This approach contributes to a flexible statistical model to allow for arbitrary non-linearities in the relationship between temperature and crime. Simultaneously regressing crimes on these indicators recovers their respective



(a) Violent crimes (b) Property crimes

Figure 3.2: Marginal Effects of Temperature
Notes: The plots show how the marginal effects of temperature on crimes for metro and non-metro areas. After estimating a Poisson regression model, I calculate the effect of a discrete change of temperature bins, holding all other covariates at their means. Shaded areas correspond to the 95% confidence interval of the margin.

coefficients, which describes the nonlinear response of crimes. As the coefficients on the set of temperature indicators are identified up to a common constant, the regression omits one temperature bin, as seen in Table 3.3, which I choose to be the lower temperature bin (below 15 degrees bin). Therefore, all temperature indicators are measured relative to this indicator for below 15 degrees.

The results of column 3 in Table 3.3 are consistent with the results of column 3 in Table 3.2, suggesting a positive relationship between temperature and violent crimes. In particular, non-metro areas have a higher number of violent crimes with higher temperatures. At temperatures ranging from 25 to 30 degrees, the number of violent crimes in non-metro areas is 8.4 percent higher than the baseline when the temperature is below 15 degrees. Although metro areas also have a positive relationship between temperature and violent crimes, the temperature effects in metro areas are not as large as the effects in non-metro areas. Perhaps the incidence of crime in large cities is generally high enough as compared with small cities, so the marginal effects of temperature are relatively small.

Figure 3.2 depicts the results in Table 3.3 and focuses on the marginal effect of temperature. With a temperature indicator variable, the marginal effect shows how the number of crimes changes as the temperature indicator changes. Margins are calculated from predictions of a previously Poisson model at fixed values of temperature bins and metro dummy, holding all other covariates at their means. Figure 3.2 demonstrates graphically that the marginal effects of temperature on violent crimes are larger in non-metro areas, as presented in column 3 in Table 3.3. However, the marginal effects on violent crimes in non-metro areas show a slight decline when the temperature is above 30 degrees. For property crimes, there is no considerable heterogeneous effect of temperature across metro and non-metro areas, as seen in Figure 3.2 (b).

One might be concerned that if the effects of temperature vary across different types of crime, using the number of aggregate violent or property crimes as an outcome variable could result in misleading conclusions in this analysis. For example, the effects of temperature on crime may depend on whether each offense is more likely to occur outdoors or indoors. To address this possible concern, I re-estimate Poisson regression models using the number of each offense as an outcome variable. As mentioned previously, violent crimes are a category of offenses that include murder and non-negligent manslaughter, forcible rape, robbery, and aggravated assault. Property crimes include burglary, larceny-theft, motor vehicle theft, and arson. Based on Equation (3.3), Table 3.4 reports the regression results for each offense.

For violent crimes, higher temperatures cause more forcible rape, robbery, and aggravated assault.[11] The number of these three offenses is much greater in metro areas than in non-metro areas. The number of robberies and aggravated assaults increases by 0.7 and 0.3 percent, respectively, for every one-degree increase in temperature. Metro areas have 37.7 percent more robberies than non-metro areas. However, the temperature effects are smaller in metro areas than in non-metro areas. Given the significant effects on robbery and aggravated assault, the results for violent crimes in Table 3.2 may be mainly driven by these two offenses.

[11] A one-degree increase in temperature leads to 1.7 percent fewer murder and non-negligent manslaughter. In Figure 3.2, I find that this reducing effect is driven by non-metro areas. The number of murder and non-negligent manslaughter in non-metro areas is large with very low temperatures but small with very high temperatures. Meanwhile, there is no temperature effect in metro areas.

Table 3.4: Effect of Temperature on Each Offense

	Violent crimes			
	Murder (1)	Forcible rape (2)	Robbery (3)	Agg. assault (4)
Temp	−.017*	.005	.007**	.003***
	(.009)	(.006)	(.003)	(.001)
Metro	−.404	.644*	.377***	.073*
	(.329)	(.351)	(.085)	(.041)
Temp×*Metro*	.017*	−.003	−.007**	−.001*
	(.009)	(.006)	(.003)	(.001)
N	12,709	14,742	16,095	16,600
	Property crimes			
	Burglary (5)	Larceny-theft (6)	Vehicle theft (7)	Arson (8)
Temp	.001	−.002*	.003	.014*
	(.002)	(.001)	(.002)	(.008)
Metro	.057	−.037	.199	−.822***
	(.057)	(.046)	(.146)	(.282)
Temp×*Metro*	−.002	.000	−.002	.013*
	(.001)	(.001)	(.002)	(.007)
N	16,563	16,453	16,384	14,319
City FEs	Yes	Yes	Yes	Yes
Year and month FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table shows results of separate regressions for each Part I offense of the UCR program. All regression models have interaction terms between temperature and metro. All regression models include city, year, month fixed effects, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

For property crimes, higher temperatures produce more arson but fewer larceny-theft, while there are no significant relationships between temperature and burglary and vehicle theft.

3.4.2 Sensitivity Analyses

In this section, I conduct a variety of sensitivity analyses to illustrate how the main results can be strongly robust to alternative model specifications. First, I follow a common exercise

Table 3.5: Sensitivity Analysis

	(1)	(2)	(3)	(4)
Panel A. Violent crimes				
Temp	$-.006^{***}$	$.003^{***}$	$.003^{***}$	$.003^{***}$
	(.001)	(.001)	(.001)	(.001)
Metro	$.126^{***}$	$.139^{***}$	$.141^{***}$	$.125^{***}$
	(.044)	(.045)	(.046)	(.061)
Temp×*Metro*	$-.002^{**}$	$-.002^{**}$	$-.002^{**}$	$-.002^{***}$
	(.001)	(.001)	(.001)	(.001)
N	16,600	16,600	16,600	16,600
Panel B. Property crimes				
Temp	−.000	.000	.000	.−.000
	(.001)	(.001)	(.001)	(.001)
Metro	.016	.017	.014	.064
	(.037)	(.038)	(.039)	(.069)
Temp×*Metro*	−.001	−.001	−.001	−.001
	(.001)	(.001)	(.001)	(.001)
N	16,600	16,600	16,600	16,600
City FEs	Yes	Yes	Yes	Yes
Year and month FEs	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes
City-specific time trends	No	No	No	Yes

Notes: This table shows results of separate model specifications by including or excluding year and month fixed effects, weather controls, socioeconomic factors, and city-specific linear time trends. All regression models have interaction terms between temperature and metro. Two dependent variables are the number of violent crime arrests in Panel A and the number of property crime arrests in Panel B. Standard errors presented in parentheses are clustered at city level.

*** Statistical significance at the 1 percent level.

** Statistical significance at the 5 percent level.

* Statistical significance at the 10 percent level.

by adding or removing control variables in my baseline specification. Table 3.5 presents the regression results of four different model specifications, and column 3 replicates the results from my preferred specification to facilitate comparison. Overall, the estimated coefficients do not change excessively in different specifications, implying that my empirical strategy is a good approach to control for confounding variables. Column 2 shows that it is important to eliminate potential bias from unobserved factors that change over time. I also include city-specific linear time trends in column 4 because different cities might change the regulatory

Table 3.6: Different Population Groups

	Violent crimes			Property crimes		
	(1)	(2)	(3)	(4)	(5)	(6)
Temp	.002***	.002***	.005**	−.000	−.000	−.001
	(.001)	(.001)	(.002)	(.001)	(.001)	(.003)
$City_{10}$		−.058	−.014		.122**	.034
		(.083)	(.104)		(.055)	(.093)
$City_{25}$		.048	.100		.144	.209*
		(.090)	(.108)		(.105)	(.125)
$City_{50}$		.136	.183		.144	.170
		(.099)	(.113)		(.109)	(.128)
$City_{100}$		.249**	.392***		.110	.119
		(.113)	(.125)		(.111)	(.129)
$Temp \times City_{10}$			−.002			.004
			(.002)			(.003)
$Temp \times City_{25}$			−.002			−.002
			(.003)			(.003)
$Temp \times City_{50}$			−.002			−.000
			(.002)			(.003)
$Temp \times City_{100}$			−.005**			.001
			(.002)			(.003)
N	16,600	16,600	16,600	16,600	16,600	16,600
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year and month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows results of separate regressions using temperature, city dummies, and interaction term. Two dependent variables are the number of violent crime arrests in columns 1 to 3 and the number of property crime arrests in columns 4 to 6. All regression models include city, year, month fixed effects, weather, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

regime at different times (Angrist and Pischke [3]). The results remain unchanged with time trends.

In the main specification, I have simply defined two city groups as metro and non-metro based on a population of 50,000 or more. Now I apply an alternative definition of city groups as another robustness check. Following the city group information in the UCR data provided by ICPSR, I group cities into five population size-based bins in place of metro/non-metro−below 10,000, 10,000−25,000, 25,000−50,000, 50,000−100,000, and above 100,000.[12] The population bins are also interacted with a temperature variable to see how the effects of temperature vary across these city groups.[13] In Table 3.6, the regression drops one population bin, which I chose to be the below-10,000 bin. The estimates are similar to the main results in Table 3.2. Columns 2 and 3 show that a city group with a population of 100,000 or more has a higher number of violent crimes, therefore several violent crimes in metro areas may be determined by this city group.

I have focused on average monthly maximum temperature rather than average monthly mean temperature because maximum temperature is a significant measure for predicting criminal behavior associated with hot weather. Maximum temperature is also likely to be highly correlated with other relevant temperature measures (Graff Zivin and Neidell [39]). However, mean temperature can also be used as an alternative predictor variable in the robustness check if the mean was precisely calculated from a large number of temperature samples.[14] Column 1 in Table 3.7 reports the results from a regression (where *Temp* is mean temperature) showing that mean temperature can be a good alternative temperature measure in this analysis.

Anecdotal evidence suggests that crimes occur most often at night. To address this issue, Ranson [77] checks whether crime is affected by maximum or minimum temperature in his sensitivity analyses. He shows that the effects of maximum temperature and minimum temperature are similar and concludes that the two temperature measures affect criminal activities through similar mechanisms. Following Ranson's approach, average monthly minimum temperature instead of average monthly maximum temperature is used, as shown in column 2 in Table 3.7. I get the same results as using the maximum temperature, which does not change the baseline results.

[12] Numbers of cities in each group are 60 (below 10,000), 93 (10,000−25,000), 75 (25,000−50,000), 92 (50,000−100,000), and 60 (above 100,000).

[13] $City_{10}$ is a dummy variable for population of 10,000−25,000 and $City_{100}$ is a dummy variable for population of above 100,000 in Table 3.6.

[14] Mean temperature cannot be often determined precisely when only a limited number of samples are taken.

Table 3.7: Alternative Predictor Variables

	Mean temperature (1)	Min. temperature (2)	Days above 21.1°C (3)
Panel A. Violent crimes			
Temp	.004***	.004***	.003***
	(.001)	(.002)	(.001)
Metro	.123***	.109***	.123***
	(.044)	(.041)	(.041)
Temp×Metro	−.002**	−.002**	−.002***
	(.001)	(.001)	(.001)
N	16,600	16,600	16,600
Panel B. Property crimes			
Temp	.000	.000	.001**
	(.001)	(.002)	(.001)
Metro	.021	.017	.001
	(.036)	(.034)	(.034)
Temp×Metro	−.001	−.001	−.001
	(.001)	(.001)	(.001)
N	16,600	16,600	16,600
City FEs	Yes	Yes	Yes
Year and month FEs	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: This table shows results of separate regressions using mean temperature as *Temp* in column 1, minimum temperature in column 2, and number of day with maximum temperature above 21.1 degrees Celsius (70 degrees Fahrenheit) as *Temp* in column 3. All regression models have interaction terms between *Temp* and *Metro*. Two dependent variables are the number of violent crime arrests in Panel A and the number of property crime arrests in Panel B. All regression models include city, year, month fixed effects, weather, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

An additional concern is that using average monthly maximum temperature might be problematic because it disregards the distribution of weather variables around the average (Schlenker and Roberts [82]; Dell, Jones, and Olken [26]). To overcome this problem, previous research uses the concept of degree days, which handles the nonlinear effect of weather variables. Column 3 in Table 3.7 shows the results from the estimation using the number of days maximum temperature above 21.1 degrees Celsius (70 degrees Fahrenheit) instead of average maximum temperature.[15] The results are statistically significant and similar in magnitude to the main results.

3.4.3 National Incident-Based Reporting System Offenses

Existing literature has focused on Part I offenses of the UCR program because they are serious felonies plaguing society. Although Part II offenses consist of less serious crimes, they cannot be negligible, considering many negative consequences of crime.[16] Furthermore, the National Incident-Based Reporting System (NIBRS) classifies crime categories into two groups: 23 Group A offenses and 10 Group B offenses, while the UCR program includes 8 Part I offenses and 21 Part II offenses.[17] The NIBRS defines more broadly serious crimes in Group A offenses, which include all Part I offenses and some Part II offenses of the UCR program. Following the NIBRS definition, I now examine the effects of temperature on other serious crimes and how the effects vary across metro and non-metro areas.[18]

Table 3.8 reports the regression results using each of eight other serious offenses as an outcome variable. Incidences of all these offenses are greater in metro areas than in non-metro areas, though there are statistically significant effects only on fraud and sex offenses. A one-degree increase in temperature is associated with a 0.6 percent increase in the number of sex offenses, while it leads to a 0.6 percent decrease in forgery and a 0.4 percent decrease in weapons-carry. In particular, the result that higher temperatures are accompanied by a greater incidence of sex offenses is in line with the findings of previous studies (Field [31]; Rotton [78]; McLean [61]). Warmer weather is not claimed to directly promote sex

[15] When I use the number of days maximum temperature above 32.2 degrees Celsius (90 degrees Fahrenheit), the estimated effect of temperature is not statistically significant, with smaller magnitudes. This is in the same line with literature suggesting that when temperatures are too high, people are less likely to be outdoors, leading to the low risk of committing crime outdoors.

[16] According to the UCR program, Part II offenses include simple assaults, forgery and counterfeiting, fraud, embezzlement, stolen property, vandalism, weapons, prostitution and commercialized vice, sex offenses, drug abuse violations, gambling, offenses against the family and children, driving under the influence, liquor laws, drunkenness, disorderly conduct, vagrancy, all other offenses, suspicion, curfew and loitering laws, and runaways.

[17] According to the Federal Bureau of Investigation (FBI), the NIBRS is a part of the UCR program. In general, collecting and reporting data in the NIBRS are the same as those in the UCR. However, the NIBRS includes much greater detail information and is known as an improved version of the UCR.

[18] Since all law enforcement agencies report the Part II offenses information as well as the Part I offenses, I am able to obtain data for other offenses in the Group A offenses of the NIBRS.

Table 3.8: Group A Offenses of NIBRS

	Simple assault (1)	Forgery (2)	Fraud (3)	Embezzlement (4)
Temp	.001	−.006*	−.001	.011
	(.001)	(.003)	(.003)	(.007)
Metro	.064	.167	.445***	.011
	(.048)	(.125)	(.111)	(.166)
Temp×*Metro*	−.001	.001	−.003	−.001
	(.001)	(.003)	(.003)	(.006)
N	16,600	15,919	15,966	13,741
	Weapons (5)	Prostitution (6)	Sex offenses (7)	Drug (8)
Temp	−.004**	−.002	.006**	−.001
	(.002)	(.007)	(.003)	(.001)
Metro	.035	.197	.211**	.012
	(.059)	(.383)	(.090)	(.036)
Temp×*Metro*	.003*	.009	−.004	.000
	(.002)	(.007)	(.003)	(.001)
N	16,532	10,746	16,329	16,600
City FEs	Yes	Yes	Yes	Yes
Year and month FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table shows results of separate regressions for other Group A offenses of the NIBRS in place of Part I offenses of the UCR program. All regression models have interaction terms between temperature and metro. All regression models include city, year, month fixed effects, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

offenses, but rather it is likely that more people are outside, and in turn, offenders can have more potential targets (McLean [61]). Table 3.8 provides no evidence that the effects of temperature on each offense vary between metro and non-metro areas.

3.4.4 Simple Fixed-Effects Model

Since crime outcomes in my analysis are based on discrete events, I have employed a Poisson regression model which is the most common approach for count data. Crime statistics are typically available in two main formats−crime counts and crime rates. Crime counts are

Table 3.9: Simple Fixed Effects Estimation

	Violent crimes			Property crimes		
	(1)	(2)	(3)	(4)	(5)	(6)
Temp	.256**	.256**	.283***	.657	.657	.604
	(.112)	(.112)	(.106)	(.519)	(.519)	(.465)
Metro		1.114	3.615*		−2.978	−7.842
		(1.349)	(1.895)		(2.758)	(6.620)
Temp×Metro			−.106*			.206
			(.064)			(.223)
N	16,600	16,600	16,600	16,600	16,600	16,600
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year and month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows results of separate FE regressions using temperature, metro, and interaction term. Two dependent variables are arrest rates for violent crimes in columns 1 to 3 and arrest rates for property crimes in columns 4 to 6. All regression models include city, year, month fixed effects, weather, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.

*** Statistical significance at the 1 percent level.

** Statistical significance at the 5 percent level.

* Statistical significance at the 10 percent level.

simply the number of crimes that occurred over a specified period in a city or county. Crime rates are generally defined as the number of crimes per 100,000 residents in the population. Thus, I can also explore the effects of temperature on crime by using crime rates which can be treated as continuous. Given the continuous outcome variable, it is more straightforward to estimate a simple fixed-effects model. Specifically, I fit the following model:

$$Crime_{imt} = \alpha_0 + \alpha_1 Temp_{imt} + \alpha_2 Metro_i + \alpha_3 Temp_{imt} \times Metro_i + \boldsymbol{X}_{it} + \delta_i + \lambda_m + \rho_t + \epsilon_{imt}, \quad (3.4)$$

where the outcome variable $Crime_{imt}$ is an arrest rate for a given crime in month m of year t in city i. Other variables are the same as those used in a fixed-effects Poisson model specification.

Table 3.9 displays the results from simple fixed-effects regression estimates based on Equation (3.4). Column 3 suggests that a one-degree increase in temperature increases arrest rates for violent crimes by 0.283 per 100,000 population, and metro areas have 3.615

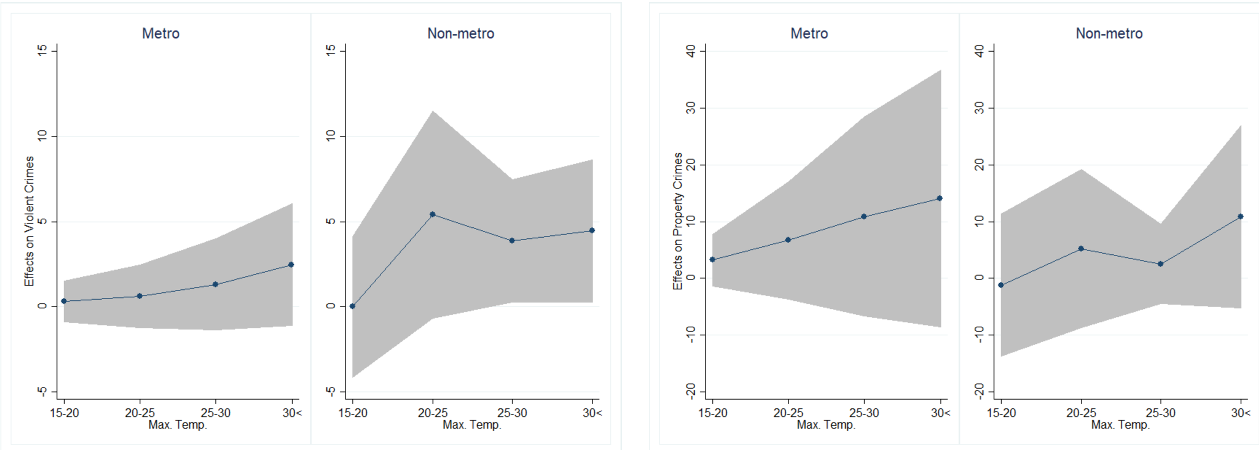
Table 3.10: Temperature Dummies (Simple FE Model)

	Violent crimes			Property crimes		
	(1)	(2)	(3)	(4)	(5)	(6)
$Temp_{15}$	.311	.311	−.001	−.025	−.024	−1.196
	(1.649)	(1.649)	(2.108)	(4.333)	(4.333)	(6.409)
$Temp_{20}$	3.796*	3.794*	5.414*	5.164	5.170	5.239
	(2.247)	(2.247)	(3.102)	(5.138)	(5.138)	(7.092)
$Temp_{25}$	3.175**	3.174**	3.861**	5.348	5.350	2.534
	(1.464)	(1.464)	(1.835)	(4.640)	(4.641)	(3.565)
$Temp_{30}$	3.955**	3.953**	4.452**	11.549	11.557	10.879
	(1.828)	(1.828)	(2.131)	(8.699)	(8.700)	(8.198)
Metro		.883	2.896		−3.126	−6.578
		(1.407)	(2.067)		(2.760)	(7.438)
$Temp_{15} \times Metro$			.325			4.434
			(1.756)			(7.644)
$Temp_{20} \times Metro$			−4.802*			1.436
			(2.727)			(7.581)
$Temp_{25} \times Metro$			−2.550			8.331
			(1.998)			(7.632)
$Temp_{30} \times Metro$			−1.984			3.150
			(1.836)			(6.114)
N	16,600	16,600	16,600	16,600	16,600	16,600
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year and month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows results of separate FE regressions using dummies for temperature and metro, and interaction term. Two dependent variables are the number of violent crime arrests in columns 1 to 3 and the number of property crime arrests in columns 4 to 6. All regression models include city, year, month fixed effects, weather, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

more arrest rates than non-metro areas. The results are consistent with the results in a fixed-effect Poisson regression analysis. For property crimes, the estimated coefficients show relatively large effects, though they are not statistically significant. This might result from the difference in the means of violent crime rates and property crime rates. The monthly



(a) Violent crimes

(b) Property crimes

Figure 3.3: Marginal Effects of Temperature (Simple FE Model)

Notes: The plots show how the marginal effects of temperature on crimes for metro and non-metro areas. After estimating a simple fixed-effect regression model, I calculate the effect of a discrete change of temperature bins, holding all other covariates at their means. Shaded areas correspond to the 95% confidence interval of the margin.

mean values of arrest rates for violent and property crimes are 30.2 and 57.6 per 100,000 population, respectively.[19]

Table 3.10 reports the results of the extension of Equation (3.4) using separate indicators for every five-degree temperature increment, as shown in Table 3.3. The results are similar to the findings in Table 3.3, showing the higher arrest rates for both violent and property crimes at a certain temperature ranges relative to below-15 degrees. However, the effects on property crimes are not statistically significant. Figure 3.3 plots the marginal effect of temperature to show how the arrest rates change if temperature levels are changed. Higher temperatures are likely to increase arrest rates both in metro areas and non-metro areas. Non-metro areas show a relatively large increase in arrest rates for violent crimes at 20−25 degrees which is consistent with Figure 3.2. Results with less precise estimates in Table 3.10 make the confidence interval wider in Figure 3.3.

[19] In California, average arrest rates for violent crimes in metro and non-metro areas are 21.4 and 35.4 per 100,000 population, respectively. Average arrest rates for property crimes in metro and non-metro areas are 36.1 and 70.2 per 100,000 population, respectively. The arrest rates in rural areas have outpaced the arrest rates in urban areas as of the early 2000s because California's rural counties have experienced significantly jail admission rates (Neusteter et al. [65]).

3.5 Conclusion

This paper provides a causal relationship between temperature and crime, though the estimated effects are not large. Using a Poisson regression approach for monthly data on the number of crimes, I build on previous research suggesting that temperature plays an important role in criminal activities, that is, higher temperatures are associated with more violent crimes. As hypothesized, higher temperatures increase the likelihood of aggressive behavior that, in turn, causes an increase in interpersonal crimes. The main results also show that crime levels are much higher in urban areas than in rural areas, which is consistent with previous studies. I also find that higher temperatures have a greater impact on rural crimes.

There are two limitations to this study. First, California tends to be warmer than most other states and has no large temperature variation from month to month. The small variation in temperature might be less likely to cause unusual changes in human behaviors and routines. Given this potential issue, this study may be underestimating the effects of temperature on criminal activities. Second, using crime or arrest data is likely to cause a potential endogeneity problem, such as different employment levels of police in different law enforcement agencies. To address the endogeneity of police levels, Levitt [55] uses instrumental variables that predict changes in the size of the police force. Therefore, future works may extend this study by gathering more data from different regions, overcoming endogeneity issues, and including more years to examine the long-term effects.

Bibliography

[1] Abadie, A., Diamond, A., and 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. 33, 42, 43, 54

[2] Anderson, C. A. (1989). Temperature and aggression: Ubiquitous effects of heat on occurrence of human violence. *Psychological Bulletin*, 106(1):74. 59, 68

[3] Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press. 75

[4] Association, A. L. et al. (2014). State of the air report 2014. *Chicago, IL: American Lung Association*. 2

[5] Auerbach, A. J. and Kotlikoff, L. J. (1983). National savings, economic welfare, and the structure of taxation. In *Behavioral simulation methods in tax policy analysis*, pages 459–498. University of Chicago Press. 46

[6] Austin, D. and Dinan, T. (2005). Clearing the air: The costs and consequences of higher cafe standards and increased gasoline taxes. *Journal of Environmental Economics and Management*, 50(3):562–582. 32

[7] Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pages 13–68. Springer. 59

[8] Becker, S. O., Egger, P. H., and Von Ehrlich, M. (2013). Absorptive capacity and the growth and investment effects of regional transfers: A regression discontinuity design with heterogeneous treatment effects. *American Economic Journal: Economic Policy*, 5(4):29–77. 19

[9] Blundell, R. W., Francesconi, M., and Van der Klaauw, W. (2011). Anatomy of welfare reform evaluation: Announcement and implementation effects. 46

[10] Bondy, M., Roth, S., and Sager, L. (2020). Crime is in the air: The contemporaneous relationship between air pollution and crime. *Journal of the Association of Environmental and Resource Economists*, 7(3):555–585. 58

[11] Buonanno, P. (2003). The socioeconomic determinants of crime. a review of the literature. 58

[12] Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, 101(3):442–451. 13

[13] Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372. 54

[14] Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press. 67

[15] Chang, Y. S., Kim, H. E., and Jeon, S. (2019). Do larger cities experience lower crime rates? a scaling analysis of 758 cities in the us. *Sustainability*, 11(11):3111. 59

[16] Chay, K. Y. and Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *The Quarterly Journal of Economics*, 118(3):1121–1167. 4

[17] Chen, Y. and Whalley, A. (2012). Green infrastructure: The effects of urban rail transit on air quality. *American Economic Journal: Economic Policy*, 4(1):58–97. 48

[18] Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77. 34

[19] Coglianese, J., Davis, L. W., Kilian, L., and Stock, J. H. (2017). Anticipation, tax avoidance, and the price elasticity of gasoline demand. *Journal of Applied Econometrics*, 32(1):1–15. 46

[20] Cohn, E. G. (1990). Weather and crime. *The British Journal of Criminology*, 30(1):51–64. 59

[21] Conley, T. G. and Taber, C. R. (2011). Inference with "difference in differences" with a small number of policy changes. *The Review of Economics and Statistics*, 93(1):113–125. 33

[22] Currie, J. and Neidell, M. (2005). Air pollution and infant health: What can we learn from california's recent experience? *The Quarterly Journal of Economics*, 120(3):1003–1030. 4

[23] Currie, J., Zivin, J. G., Mullins, J., and Neidell, M. (2014). What do we know about short-and long-term effects of early-life exposure to pollution? *Annu. Rev. Resour. Econ.*, 6(1):217–247. 4

[24] Davis, L. W. (2008). The effect of driving restrictions on air quality in mexico city. *Journal of Political Economy*, 116(1):38–81. 43, 54

[25] Davis, L. W. and Kilian, L. (2011). Estimating the effect of a gasoline tax on carbon emissions. *Journal of Applied Econometrics*, 26(7):1187–1214. 32

[26] Dell, M., Jones, B. F., and Olken, B. A. (2014). What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature*, 52(3):740–98. 77

[27] Di Tella, R. and Schargrodsky, E. (2004). Do police reduce crime? estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, 94(1):115–133. 58

[28] Draca, M., Machin, S., and Witt, R. (2011). Panic on the streets of london: Police, crime, and the july 2005 terror attacks. *American Economic Review*, 101(5):2157–81. 58

[29] Ferman, B. and Pinto, C. (2016). Revisiting the synthetic control estimator. 54

[30] Ferman, B. and Pinto, C. (2019). Inference in differences-in-differences with few treated groups and heteroskedasticity. *Review of Economics and Statistics*, 101(3):452–467. 33

[31] Field, S. (1992). The effect of temperature on crime. *The British Journal of Criminology*, 32(3):340–351. 59, 78

[32] Gelman, A. and Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456. 12, 16, 19

[33] Gelman, A. and Zelizer, A. (2015). Evidence on the deleterious impact of sustained use of polynomial regression on causal inference. *Research & Politics*, 2(1):2053168015569830. 12

[34] Glaeser, E. L. and Sacerdote, B. (1999). Why is there more crime in cities? *Journal of Political Economy*, 107(S6):S225–S258. 59, 60

[35] Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and social interactions. *The Quarterly Journal of Economics*, 111(2):507–548. 59, 60

[36] Goin, D. E., Rudolph, K. E., and Ahern, J. (2017). Impact of drought on crime in california: A synthetic control approach. *PLoS one*, 12(10). 58

[37] Gould, E. D., Weinberg, B. A., and Mustard, D. B. (2002). Crime rates and local labor market opportunities in the united states: 1979–1997. *Review of Economics and Statistics*, 84(1):45–61. 66

[38] Graff Zivin, J. and Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review*, 102(7):3652–73. 48

[39] Graff Zivin, J. and Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1):1–26. 76

[40] Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209. 12

[41] Hamilton, T. L. and Wichman, C. J. (2018). Bicycle infrastructure and traffic congestion: Evidence from dc's capital bikeshare. *Journal of Environmental Economics and Management*, 87:72–93. 6

[42] Hausman, J., Hall, B. H., Griliches, Z., et al. (1984). Econometric models for count data with an application to the patents-r&d relationship. *Econometrica*, 52(4):909–938. 66

[43] Herrnstadt, E. and Muehlegger, E. (2015). Air pollution and criminal activity: Evidence from chicago microdata. Technical report, National Bureau of Economic Research. 58

[44] Holladay, J. S., Price, M. K., and Wanamaker, M. (2015). The perverse impact of calling for energy conservation. *Journal of Economic Behavior & Organization*, 110:1–18. 15

[45] Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959. 17

[46] Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635. 11, 26

[47] Jacob, B., Lefgren, L., and Moretti, E. (2007). The dynamics of criminal behavior evidence from weather shocks. *Journal of Human Resources*, 42(3):489–527. 59

[48] Jacob, R., Zhu, P., Somers, M.-A., and Bloom, H. (2012). A practical guide to regression discontinuity. *MDRC*. 11

[49] Janke, K. (2014). Air pollution, avoidance behaviour and children's respiratory health: Evidence from england. *Journal of Health Economics*, 38:23–42. 4

[50] Ladbrook, D. A. (1988). Why are crime rates higher in urban than in rural areas?—evidence from japan. *Australian & New Zealand Journal of Criminology*, 21(2):81–103. 59

[51] Laub, J. H. (1983). Urbanism, race, and crime. *Journal of Research in Crime and Delinquency*, 20(2):183–198. 59

[52] Lee, D. S. and Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2):655–674. 12, 17

[53] Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355. 11, 12, 13, 17, 25, 27

[54] Lee, J. and Xu, J. (2020). Why do businesses grow faster in urban areas than in rural areas? *Regional Science and Urban Economics*, 81:103521. 60

[55] Levitt, S. D. (2002). Using electoral cycles in police hiring to estimate the effects of police on crime: Reply. *American Economic Review*, 92(4):1244–1250. 83

[56] Li, S., Linn, J., and Muehlegger, E. (2014). Gasoline taxes and consumer behavior. *American Economic Journal: Economic Policy*, 6(4):302–42. 34

[57] Lin, C.-Y. C. and Prince, L. (2009). The optimal gas tax for california. *Energy Policy*, 37(12):5173–5183. 31, 37

[58] Lin, C.-Y. C. and Prince, L. (2013). Gasoline price volatility and the elasticity of demand for gasoline. *Energy Economics*, 38:111–117. 32

[59] Machin, S., Marie, O., and Vujić, S. (2011). The crime reducing effect of education. *The Economic Journal*, 121(552):463–484. 58

[60] McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714. 13

[61] McLean, I. (2007). Climatic effects on incidence of sexual assault. *Journal of Forensic and Legal Medicine*, 14(1):16–19. 78, 79

[62] Mora, R. and Reggio, I. (2012). Treatment effect identification using alternative parallel assumptions. 47

[63] Neidell, M. (2009). Information, avoidance behavior, and health the effect of ozone on asthma hospitalizations. *Journal of Human Resources*, 44(2):450–478. 4

[64] Neidell, M. (2010). Air quality warnings and outdoor activities: Evidence from southern california using a regression discontinuity design. *Journal of Epidemiology & Community Health*, 64(10):921–926. 3

[65] Neusteter, S. R., Subramanian, R., Trone, J., Khogali, M., and Reed, C. (2019). *Gatekeepers: The Role of Police in Ending Mass Incarceration*. Vera Institute of Justice. 82

[66] Nolan III, J. J. (2004). Establishing the statistical relationship between population size and ucr crime rate: Its impact and implications. *Journal of Criminal Justice*, 32(6):547–555. 59

[67] Noonan, D. S. (2014). Smoggy with a chance of altruism: The effects of ozone alerts on outdoor recreation and driving in a tlanta. *Policy Studies Journal*, 42(1):122–145. 3

[68] Oreopoulos, P. (2007). Do dropouts drop out too soon? wealth, health and happiness from compulsory schooling. *Journal of Public Economics*, 91(11-12):2213–2229. 58

[69] Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology*, 16(1):21–43. 67

[70] O'brien, O., Cheshire, J., and Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, 34:262–273. 19

[71] Parry, I. W., Heine, M. D., Lis, E., and Li, S. (2014). *Getting energy prices right: From principle to practice*. International Monetary Fund. 37

[72] Parry, I. W. and Small, K. A. (2005). Does britain or the united states have the right gasoline tax? *American Economic Review*, 95(4):1276–1289. 31, 37

[73] Parry, I. W., Walls, M., and Harrington, W. (2007). Automobile externalities and policies. *Journal of Economic Literature*, 45(2):373–399. 31

[74] Plummer, L. A., Headd, B., et al. (2008). Rural and urban establishment births and deaths using the us census bureau's business information tracking series. *An Office of Advocacy Working Paper*. 60

[75] Porter, J. (2003). Estimation in the regression discontinuity model. *Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison*, 2003:5–19. 12

[76] Porter, M. E., Ketels, C. H., Miller, K., and Bryden, R. (2004). Competitiveness in rural us regions: Learning and research agenda. *Washington, DC: US Economic Development Administration (EDA)*. 60

[77] Ranson, M. (2014). Crime, weather, and climate change. *Journal of Environmental Economics and Management*, 67(3):274–302. 59, 67, 76

[78] Rotton, J. (1993). Atmospheric and temporal correlates of sex crimes: Endogenous factors do not explain seasonal differences in rape. *Environment and Behavior*, 25(4):625–642. 78

[79] Rotton, J. and Cohn, E. G. (2004). Outdoor temperature, climate control, and criminal assault: The spatial and temporal ecology of violence. *Environment and Behavior*, 36(2):276–306. 59

[80] Saberian, S., Heyes, A., and Rivers, N. (2017). Alerts work! air quality warnings and cycling. *Resource and Energy Economics*, 49:165–185. 3, 21

[81] Sah, R. K. (1991). Social osmosis and patterns of crime. *Journal of Political Economy*, 99(6):1272–1295. 60

[82] Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106(37):15594–15598. 77

[83] Schlenker, W. and Walker, W. R. (2016). Airports, air pollution, and contemporaneous health. *The Review of Economic Studies*, 83(2):768–809. 4

[84] Schwartz, J. (2000). The distributed lag between air pollution and daily deaths. *Epidemiology*, 11(3):320–326. 48

[85] Shaheen, S. and Cohen, A. (2018). Is it time for a public transit renaissance?: Navigating travel behavior, technology, and business model shifts in a brave new world. *Journal of Public Transportation*, 21(1):8. 55

[86] Sipes, K. N. and Mendelsohn, R. (2001). The effectiveness of gasoline taxation to manage air pollution. *Ecological Economics*, 36(2):299–309. 32

[87] Vogel, P. (2016). Service network design of bike sharing systems. In *Service Network Design of Bike Sharing Systems*, pages 113–135. Springer. 22

[88] Wood, J. (2015). Is it time to raise the gas tax? optimal gasoline taxes for ontario and toronto. *Canadian Public Policy*, 41(3):179–190. 31

[89] Xie, X.-F. and Wang, Z. J. (2018). Examining travel patterns and characteristics in a bikesharing network and implications for data-driven decision supports: Case study in the washington dc area. *Journal of Transport Geography*, 71:84–102. 13

[90] Zivin, J. G. and Neidell, M. (2009). Days of haze: Environmental information disclosure and intertemporal avoidance behavior. *Journal of Environmental Economics and Management*, 58(2):119–128. 3, 27, 29

Appendices

Table A1.1: Baseline Results for 2013−2018

	Full sample	2-Bandwidth	5-Bandwidth	10-Bandwidth
	(1)	(2)	(3)	(4)
Panel A. Log(trips)				
Alert	−.079***	−.056***	−.056***	−.035***
	(.005)	(.013)	(.010)	(.008)
Panel B. Log(duration)				
Alert	−.124***	−.068***	−.115***	−.083***
	(.008)	(.024)	(.018)	(.014)
N	331,010	12,386	21,893	30,493

Notes: This table shows the results of separate RDD regressions using different bandwidths as well as the full sample. After dropping data for the first two years, I estimate a baseline RDD model. All regressions include weather and pollution variables, day of the week, holidays, year, month, and station fixed effects. Standard errors presented in parentheses are clustered at station level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

Table A1.2: Sensitivity to Bandwidth Choices

	6-Bandwidth (1)	7-Bandwidth (2)	8-Bandwidth (3)	9-Bandwidth (4)
Panel A. Log(trips)				
Alert	−.022**	−.022**	−.021***	−.021***
	(.009)	(.009)	(.007)	(.007)
Panel B. Log(duration)				
Alert	−.043***	−.043***	−.051***	−.051***
	(.016)	(.016)	(.012)	(.012)
N	27,693	27,693	35,591	35,985

Notes: This table shows the results of separate RDD regressions using different bandwidth choices (6, 7, 8, and 9). All regressions include weather and pollution variables, day of the week, holidays, year, month, and station fixed effects. Standard errors presented in parentheses are clustered at station level.

*** Statistical significance at the 1 percent level.

** Statistical significance at the 5 percent level.

* Statistical significance at the 10 percent level.

Table A1.3: Subgroup Analyses by Ward and Time of Day

	Weekday				Weekend
	Morning peak	Afternoon peak	Midday off-peak	Night off-peak	
	(1)	(2)	(3)	(4)	(5)
Ward 1	.015	−.009	.017	.012	−.013
	(.025)	(.023)	(.017)	(.015)	(.030)
Ward 2	.006	.007	−.057***	−.011	−.031
	(.017)	(.014)	(.011)	(.010)	(.020)
Ward 3	.040	.007	.004	−.017	.041
	(.040)	(.051)	(.024)	(.023)	(.054)
Ward 4	−.008	.079	−.042*	−.062**	−.048
	(.038)	(.061)	(.023)	(.027)	(.102)
Ward 5	−.044	.001	.005	.018	−.020
	(.038)	(.033)	(.014)	(.024)	(.047)
Ward 6	.003	.022	−.021	−.011	−.039
	(.018)	(.017)	(.013)	(.012)	(.028)
Ward 7	.155***	.085	.033	.046	.085
	(.039)	(.103)	(.043)	(.059)	(.113)
Ward 8	.052	.193***	−.004	.241***	−.307
	(.097)	(.055)	(.054)	(.071)	(.202)

Notes: This table shows the results of separate local RDD regressions by ward and time of day using 5-bandwidth. Outcome variable is the natural log of bike share trip duration. All regressions include weather and pollution variables, day of the week, year, month, hour, and station fixed effects.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

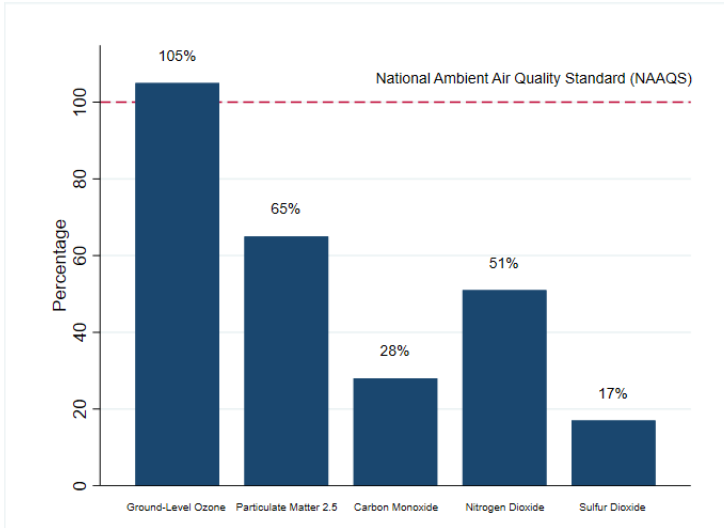


Figure A1.1: Nonattainment of O_3 Standards
Source: District of Columbia's ambient air quality trends report (DDOE, 2014)

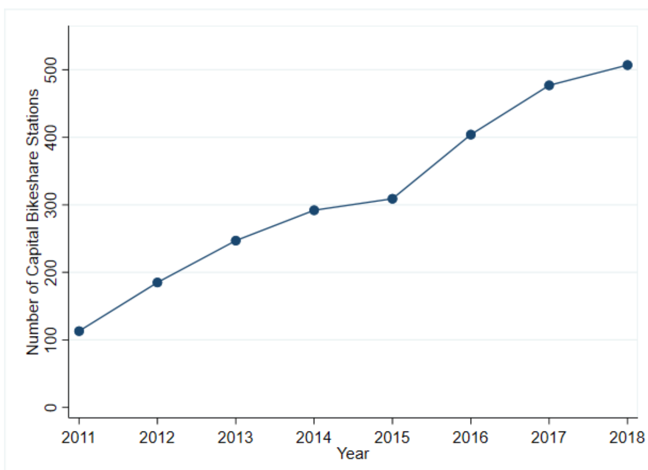


Figure A1.2: Number of Capital Bikeshare Stations, 2011−2018

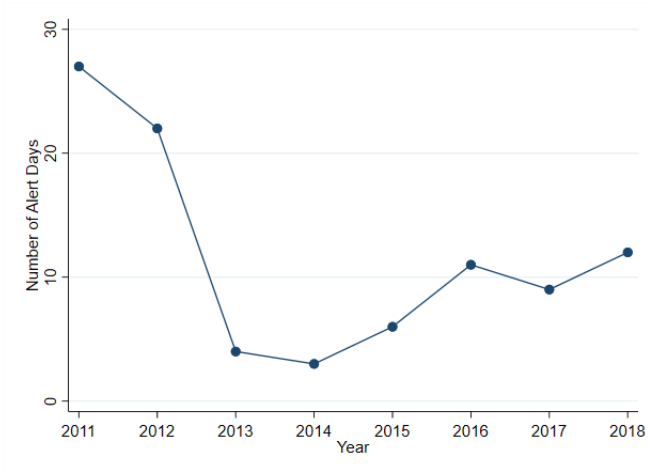
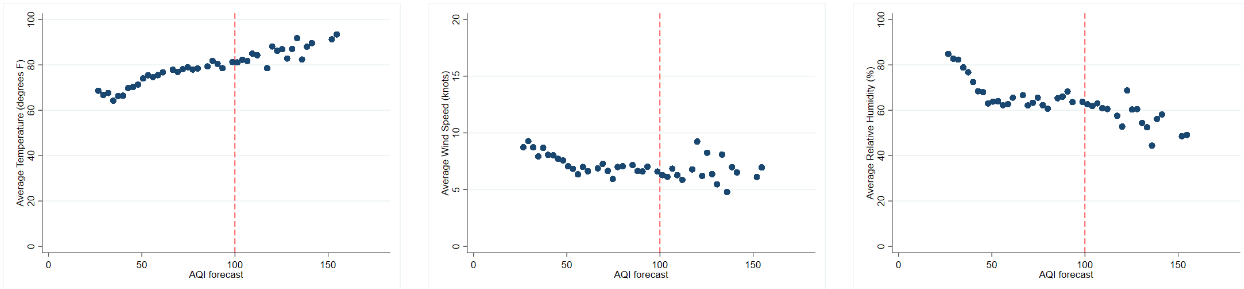
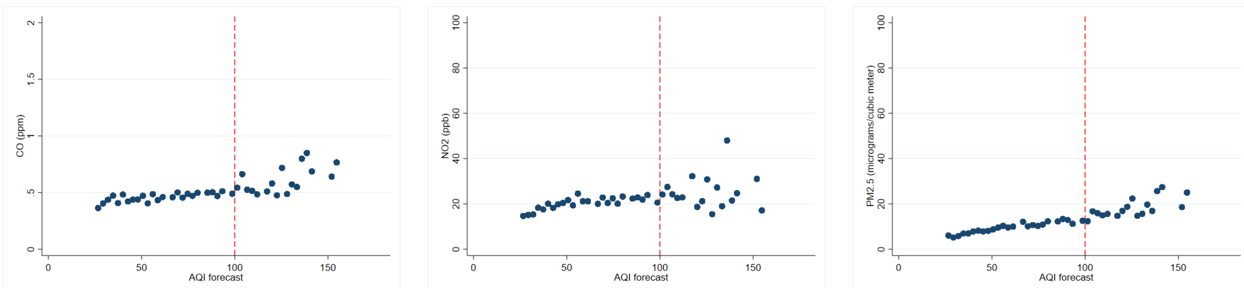


Figure A1.3: Number of Alert Days, 2011−2018



(a) Weather



(b) Pollution

Figure A1.4: Continuity of Weather and Pollution

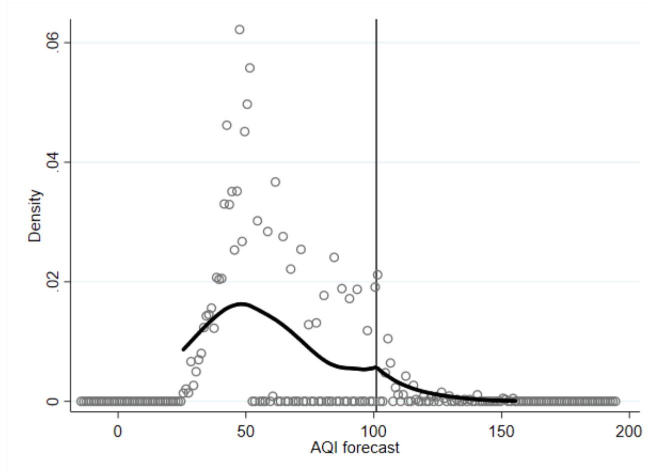
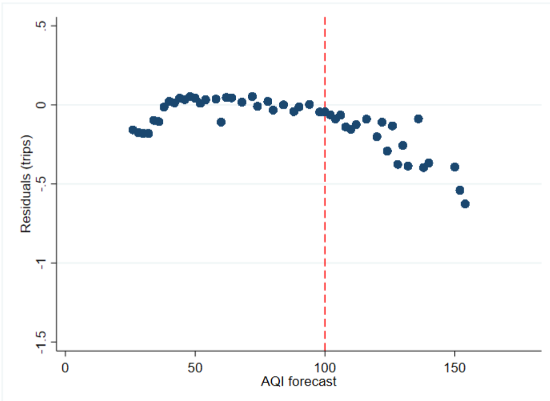
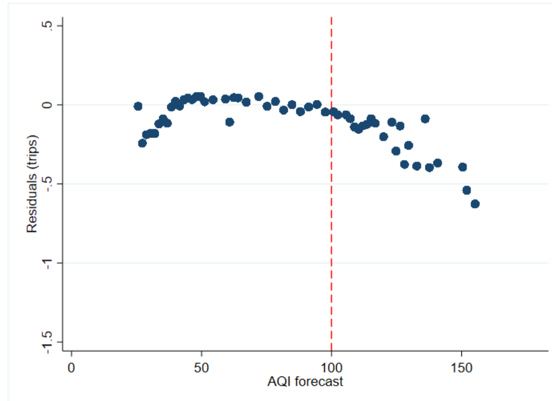


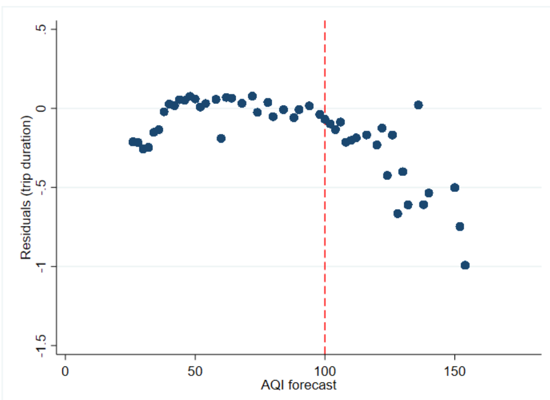
Figure A1.5: Continuity along Assignment Variable



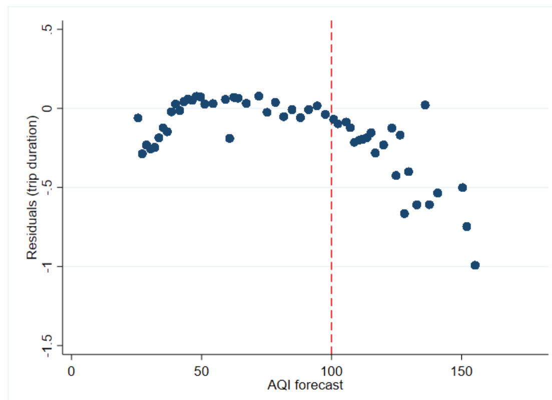
(a) Trips (80 bins)



(b) Trips (100 bins)



(c) Trip duration (80 bins)



(d) Trip duration (100 bins)

Figure A1.6: Residual Plots

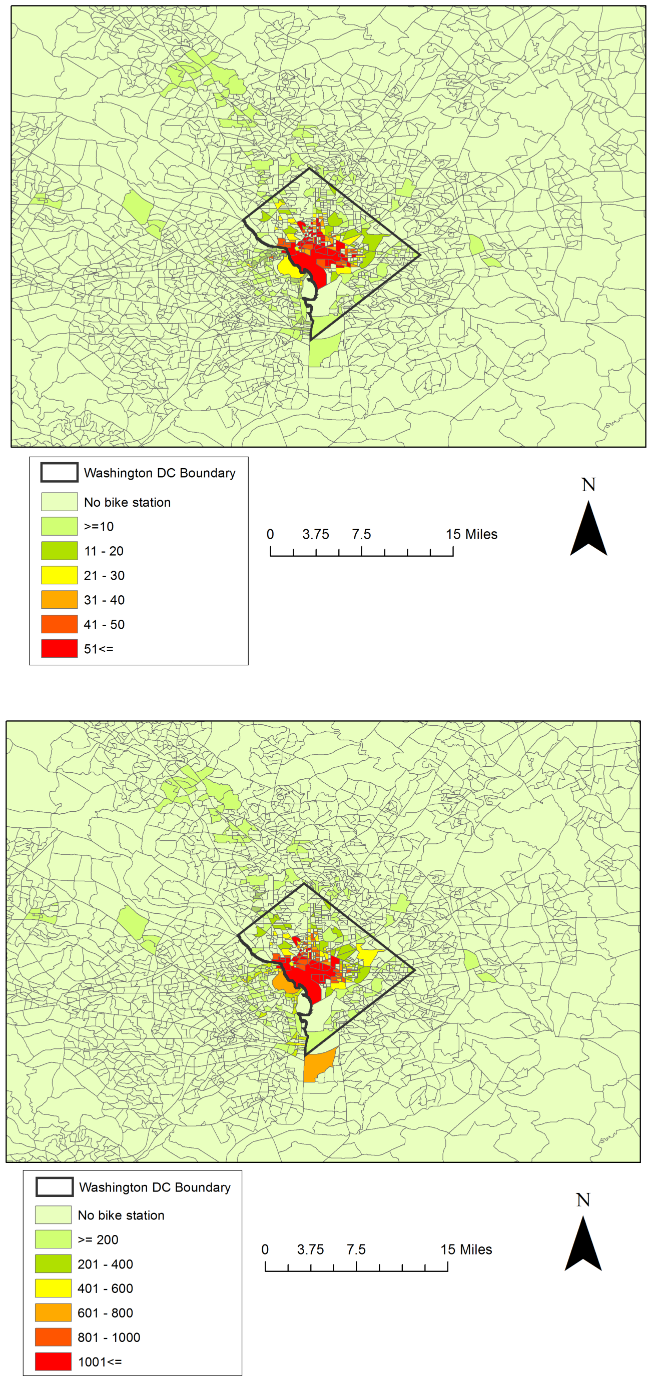


Figure A1.7: Capital Bikeshare Usage

Notes: This figure shows levels of Capital Bikeshare usage by census block group for 2011−2018. An upper panel represents the average daily number of trips and lower panel represents average daily trip duration of all bike stations by each census block group. Red color indicates high values of the number of trips and trip duration while light green color indicates the low values. A thick solid black line is Washington, DC boundary. The bike-share usage is mostly high in the central DC area.

Table A2.1: Effect of Gasoline Tax Increase on Air Quality

	(1)	(2)	(3)
Panel A. Log(CO)			
Post	−.025	−.015	−.011
	(.018)	(.019)	(.020)
Post×Treat	−.022	−.040	−.036
	(.072)	(.074)	(.075)
y2015		.012	.001
		(.020)	(.020)
y2016		−.007	−.032
		(.026)	(.028)
y2017		−.012	−.040
		(.033)	(.035)
y2018		−.032	−.061*
		(.036)	(.037)
temp			−.017***
			(.001)
rhum			.001**
			(.000)
L.temp			−.013***
			(.001)
L.rhum			−.002***
			(.000)
N	150,803	150,803	150,803
R^2 (within-monitor)	.001	.121	.159
Panel B. Log(NO_x)			
Post	−.027*	−.003	−.005
	(.019)	(.074)	(.017)
Post×Treat	.032	−.003	−.010
	(.031)	(.028)	(.026)
y2015		.022	.021
		(.018)	(.019)
y2016		−.041**	−.054**
		(.020)	(.021)
y2017		−.059**	−.069**
		(.028)	(.028)
y2018		−.042	−.064*
		(.034)	(.034)
temp			.016***
			(.001)
rhum			−.003***
			(.001)
L.temp			−.020***
			(.001)
L.rhum			−.006***
			(.001)
N	242,844	242,844	242,844
R^2 (within-monitor)	.000	.123	.167
Monitor FE	✓	✓	✓
Time dummies		✓	✓
Weather			✓

Notes: This table shows the point estimates from DID regression of mobile source pollutants. Panel A indicates the estimated effect on CO and Panel B indicates the estimated effect on NOx. Point estimates of month and day-of-week dummies are not reported in this table. Standard errors presented in parentheses are clustered at monitor level.

*** Statistical significance at the 1 percent level.

** Statistical significance at the 5 percent level.

* Statistical significance at the 10 percent level.

Table A2.2: Lagged Pollution Measure Controls

Dependent variable:	Log(CO)	Log(NO_x)
	(1)	(2)
Post×Treat	−.010	−.007
	(.029)	(.017)
First lag	.538***	.344***
	(.010)	(.008)
Second lag	.065***	.024***
	(.009)	(.007)
N	147,695	239,716
R^2 (within-monitor)	.445	.275

Notes: This table shows results of DID regression including one and two day lags of the pollution outcome. All regression models include weather variables, year, month, day-of-week time dummies, and monitor fixed effects. Standard errors presented in parentheses are clustered at monitor level.

*** Statistical significance at the 1 percent level.

** Statistical significance at the 5 percent level.

* Statistical significance at the 10 percent level.

Table A2.3: Estimates of Gasoline Tax Increase with Leads and Lags

Dependent variable:	Log(CO)	Log(NO_x)
	(1)	(2)
DID_{t-5}	−.107	−.113*
	(.120)	(.062)
DID_{t-4}	−.004	−.074**
	(.021)	(.037)
DID_{t-3}	−.045	−.111***
	(.042)	(.038)
DID_{t-2}	.052	.068
	(.067)	(.046)
DID_{t-1}	.090**	.087***
	(.038)	(.029)
DID_{t0}	−.004	.119*
	(.062)	(.065)
DID_{t+1}	−.094**	−.074*
	(.039)	(.043)
DID_{t+2}	−.006	−.068
	(.071)	(.057)
DID_{t+3}	.090	.208***
	(.119)	(.064)
DID_{t+4}	−.006	−.186***
	(.082)	(.064)
DID_{t+5}	−.025	.008
	(.079)	(.069)
N	3,699	5,953
R^2 (within-monitor)	.282	.276

Notes: This table shows results of DID regression augmented with treatment leads and lags. All regression models include weather variables, year, month time dummies, and monitor fixed effects. Standard errors presented in parentheses are clustered at monitor level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

Table A2.4: Predictor Means of Actual and Synthetic New Jersey

	CO		NOx	
Variables	Actual	Synthetic	Actual	Synthetic
Concentrations	−.942	−.941	3.680	3.684
Temperature	56.729	56.652	56.130	56.198
Humidity	59.502	59.445	60.629	60.683

Notes: This table shows the predictor means of actual and synthetic New Jersey for the pre-treatment months, November 2015 to October 2016. Concentrations of CO and NOx are expressed in part per million in logs and in parts per billion in logs, respectively. Temperature is expressed in degrees Fahrenheit, and relative humidity is expressed in percentage terms.

Table A2.5: State Weights in the Synthetic New Jersey

State	Weight CO	Weight NOx	State	Weight CO	Weight NOx
Alabama	.031	.037	Mississippi	.015	–
Arizona	.022	.004	Missouri	.028	.005
Arkansas	.017	.005	Nevada	.022	.229
Colorado	.209	.003	New Hampshire	.010	.001
Connecticut	.027	–	New Mexico	.013	.004
Delaware	–	.009	North Dakota	.013	.001
District of Columbia	.291	.015	Ohio	.026	.009
Illinois	.017	.643	Oklahoma	.018	.003
Kansas	.019	.002	Texas	.019	.004
Louisiana	.025	.004	Vermont	.012	.005
Maine	.018	.002	Wisconsin	.026	.007
Massachusetts	.022	.004	Wyoming	.018	.001
Minnesota	.081	.004	*Sum*	1	1

Notes: This table shows the weights of selected control states in the donor pool to produce the synthetic New Jersey. Due to the lack of available data, 23 control states are used in the SCM estimation for CO pollution and 22 control states are used in the estimation for NOx pollution.

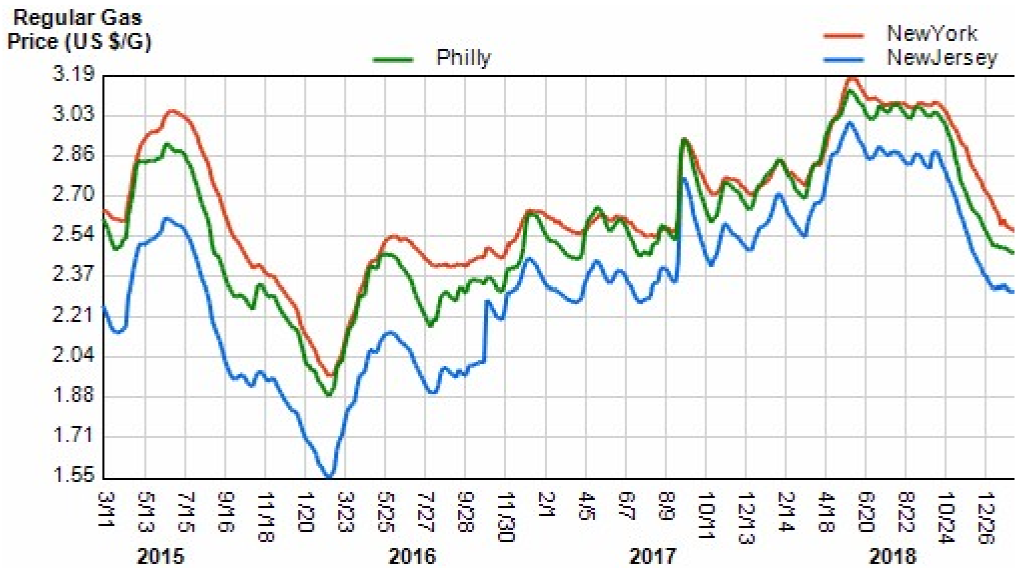


Figure A2.1: Average Retail Prices for Regular Gasoline in Three States

Notes: This figure shows the weekly retail prices for regular gasoline (U.S. dollars per gallon, including taxes) in New Jersey, New York City, and Philadelphia for the period, 2015−2018.
Sources: http://www.GasBuddy.com/

Table A3.1: Full Point Estimates of Baseline Model

	Violent crimes			Property crimes		
	(1)	(2)	(3)	(4)	(5)	(6)
Temp	.002***	.002***	.003***	−.000	−.000	−.000
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Metro		.087**	.141***		.000	.014
		(.039)	(.046)		(.032)	(.039)
Temp×Metro			−.002**			−.001
				(.001)		(.001)
Precipitation	−.000	−.000	−.000	−.000	−.000	−.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Snow	.000	.000	.000**	−.000**	−.000**	−.000**
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Population	.000	.000	.000	.000	.000	.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Male	.015	.015	.015	−.002	−.002	−.002
	(.021)	(.021)	(.021)	(.023)	(.023)	(.023)
Age	−.009	−.009	−.009	.005	.005	.005
	(.011)	(.011)	(.011)	(.010)	(.010)	(.010)
White	.001	.001	.001	−.000	−.000	−.000
	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
Poverty	−.004	−.004	−.004	−.001	−.001	−.001
	(.003)	(.003)	(.003)	(.005)	(.005)	(.005)
Income	−.000	−.000	−.000	.000	.000	.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Unemployment	−.001	−.001	−.000	.001	.001	.001
	(.005)	(.005)	(.005)	(.006)	(.006)	(.006)
yr2007	.029**	.029**	.029**	.002	.002	.002
	(.013)	(.013)	(.013)	(.012)	(.012)	(.012)
yr2008	.030	.027	.027	.003	.003	.003
	(.019)	(.019)	(.001)	(.017)	(.018)	(.018)
yr2009	.019	.012	.012	−.019	−.019	−.019
	(.028)	(.028)	(.028)	(.032)	(.033)	(.033)
yr2010	−.003	−.010	−.010	−.039	−.039	−.039
	(.038)	(.038)	(.038)	(.041)	(.042)	(.042)
yr2011	−.036	−.043	−.044	−.052	−.052	−.053
	(.045)	(.045)	(.045)	(.053)	(.053)	(.053)
yr2012	.000	−.006	−.007	−.063	−.063	−.063
	(.040)	(.040)	(.040)	(.049)	(.049)	(.049)
yr2013	−.037	−.041	−.041	−.114**	−.114**	−.114**
	(.045)	(.040)	(.040)	(.051)	(.052)	(.052)
yr2014	−.011	−.014	−.014	−.088	−.088	−.088
	(.040)	(.040)	(.040)	(.055)	(.055)	(.055)
yr2015	.020	.019	.019	−.169**	−.169**	−.170**
	(.046)	(.046)	(.046)	(.078)	(.078)	(.078)
mo2	−.039***	−.039***	−.039***	−.054***	−.054***	−.054***
	(.011)	(.011)	(.011)	(.008)	(.008)	(.008)
mo3	.036***	.036***	.037***	.010	.010	.011
	(.011)	(.011)	(.011)	(.009)	(.009)	(.009)
mo4	.010	.010	.011	−.016*	−.016*	−.016*
	(.011)	(.011)	(.011)	(.009)	(.009)	(.009)
mo5	.065***	.065***	.066***	.016	.016	.016
	(.013)	(.013)	(.013)	(.011)	(.011)	(.011)
mo6	.012	.012	.014	−.021*	−.021*	−.020*
	(.013)	(.013)	(.013)	(.011)	(.011)	(.011)
mo7	.016	.016	.019	−.019	−.019	−.018
	(.015)	(.015)	(.015)	(.014)	(.014)	(.014)
mo8	.012	.012	.015	−.030**	−.030**	−.030**
	(.016)	(.016)	(.016)	(.013)	(.013)	(.013)
mo9	.025*	.025*	.028*	−.054***	−.054***	−.053***
	(.015)	(.015)	(.015)	(.013)	(.013)	(.013)
mo10	.014	.015	.017	−.033***	−.033***	−.032***
	(.012)	(.012)	(.012)	(.011)	(.011)	(.011)
mo11	−.051***	−.051***	−.050***	−.043***	−.043***	−.043***
	(.010)	(.010)	(.010)	(.009)	(.009)	(.009)
mo12	−.042***	−.042***	−.042***	.010	.010	.010
	(.011)	(.011)	(.011)	(.009)	(.009)	(.009)
N	16,600	16,600	16,600	16,600	16,600	16,600

Notes: This table shows full point estimates of baseline specification. All regression models include city, year, month fixed effects, weather, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.
*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

Table A3.2: Effect of Temperature on Each Offense (Simple FE Model)

	Violent crimes			
	Murder (1)	Forcible rape (2)	Robbery (3)	Agg. assault (4)
Temp	−.011	−.007	−.004	.305***
	(.007)	(.012)	(.052)	(.098)
Metro	−.310	.444	1.823***	1.657
	(.213)	(.301)	(.674)	(1.581)
Temp×Metro	.008	.002	−.045	−.071
	(.005)	(.006)	(.035)	(.049)
N	16,600	16,600	16,600	16,600
	Property crimes			
	Burglary (5)	Larceny theft (6)	Vehicle theft (7)	Arson (8)
Temp	.109	.406	.046	.044
	(.160)	(.473)	(.129)	(.039)
Metro	−1.102	−6.542	.415	−.613
	(2.248)	(4.961)	(1.492)	(.390)
Temp×Metro	.031	.181	−.005	−.001
	(.080)	(.181)	(.052)	(.012)
N	16,600	16,600	16,600	16,600
City FEs	Yes	Yes	Yes	Yes
Year and month FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table shows results of separate simple FE regressions for each Part I offense of the UCR program. All regression models have interaction terms between temperature and metro. All regression models include city, year, month fixed effects, and socioeconomic controls. Standard errors presented in parentheses are clustered at city level.

*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

Vita

Sangwon Lee was born in Seoul, South Korea on June 11, 1982. He earned a Bachelor of Arts degree in Applied Statistics from Konkuk University in February 2010 and a Master of Science degree in Statistics from the University of Texas at Austin in May 2013. He received his Doctor of Philosophy degree in Economics from the University of Tennessee in August 2020.