UNDERSTANDING THE HETEROGENEITY IN DECENTRALIZED ENVIRONMENTAL POLICIES

by

WENBO MENG

B.A., Nankai University, 2015

M.A., University of Colorado Boulder, 2019

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Economics

2023

Committee Members:

Jonathan Hughes

Daniel Kaffine

Tania Barham

Shuang Zhang

Steve Miller

Meng, Wenbo (Ph.D., Economics [Department of Economics])

Understanding the Heterogeneity in Decentralized Environmental Policies

Thesis directed by Associate Professor Jonathan E. Hughes

This dissertation studies the heterogeneity in decentralized air pollution, water pollution, and transportation policies, and identifies implications for optimal policy design. The first chapter investigates whether states exhibit beggar-thy-neighbor and free-riding behaviors when implementing nonpoint-source (NPS) water pollution policies. I find that rivers within 30 km of state borders are less likely to be treated by more decentralized policies. Each behavior leads to a large deadweight loss. The second chapter conducts a large-scale study of the effectiveness of a variety of driving restriction policies in a variety of locations in China. I show that policy details and pollution concentration are the major factors that affect the actual and estimated effects of driving restriction policies, and also a potential explanation for earlier studies that showed driving restriction policies had little effect. In the third chapter, my co-author and I estimate time-varying and location-specific congestion costs across California's highways. We find that toll incidence increases with income. Black and Asian households bear more toll costs per day in location-based congestion tolls compared to white households.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my advisor, Professor Jonathan E. Hughes, for his unwavering support, guidance, and encouragement throughout my Ph.D. journey. His invaluable insights and constructive feedback have been instrumental in shaping my research and academic growth. I am also deeply grateful to Professor Daniel Kaffine, Professor Tania Barham, and Professor Shuang Zhang for their invaluable help and insightful comments that helped refine my ideas and sharpen my analytical skills. I would like to thank Professor Steve Miller for attending my Final Defense and providing valuable feedback. I am grateful to all the faculty members in the Department of Economics at the University of Colorado Boulder for their support and encouragement throughout my Ph.D. program.

CONTENTS

TABLES

Table

FIGURES

Figure

Chapter 1. Beggar-thy-Neighbor or Free-riding? Transboundary Behaviors in Decentralized Water Pollution Policies

Wenbo Meng

1 Introduction

Many public policies are implemented jointly by the central and local governments. When local governments dominate, they optimize policy implementation within their jurisdictions to maximize local welfare, thereby exhibiting different policy behaviors at jurisdictional boundaries. Examples include insufficient regulations near boundaries and transboundary policy spillovers. Since waterways connect jurisdictions, the transboundary analysis of water pollution policies is particularly important (Keiser et al. (2022)). However, the economic literature on water pollution policies is sparse relative to air pollution policies due to limitations in data and difficulties in hydro-spatial computation and causal identification (Keiser and Shapiro (2019)). Some studies of point-source pollution have found potential jurisdiction and enforcement issues for water policies in the U.S., China, India, Brazil, and other countries around the world (Sigman (2005), Chen et al. (2018), Greenstone and Hanna (2014), Lipscomb and Mobarak (2016), Sigman (2002)). However, the literature has been ambiguous in classifying and identifying the policy behaviors. In addition, because nonpoint sources are diffuse, they provide more opportunities for local governments to prioritize treatment geographically.

This paper studies whether state governments have different behaviors when implementing decentralized nonpoint-source (NPS) water pollution policies on near-boundary and transboundary rivers compared with intrastate rivers. I compile a unique and comprehensive dataset of NPS pollution policies and hydrological information to estimate the policy implementation priority and intensity across rivers' distances to state borders. I find states treat near-boundary and transboundary rivers less if the policy is highly decentralized. These policy behaviors are inefficient and lead to large deadweight losses.

The state governments have two types of transboundary policy behaviors. The first is beggar-thy-neighbor behavior, where states that possess the upstream portions of interstate rivers conduct insufficient treatment and let pollution flow into the downstream states. The second is free-riding behavior, where downstream states conduct insufficient treatment because they expect the upstream states to control the source of pollution. I identify every near-boundary and transboundary river in the contiguous United States based on its hydrographic and spatial features. I match every flow segment with its NPS policy treatment status using ArcGIS to study the transboundary policy behaviors.

The major NPS pollution policies in the Clean Water Act (CWA) are Section 305(b) Water Quality Assessment, Section 303(d) Total Maximum Daily Load (TMDL) Program, and Section 319(h) Nonpoint Source Management Program Grants. S305(b) is the first step of NPS pollution policies. S303(d) can only be conducted based on S305(b) assessment outcomes. S319(h) funding is not a follow-up to the other two policies, because many watersheds were approved for S319(h) grants without a TMDL. I differentiate the degree of decentralization of the three NPS policies by their levels of state authority and state-borne costs. S303(d) is the most decentralized policy because states have higher authority and costs in implementing S303(d) than in S305(b) and S319(h). States bear low costs and have low authority in S319(h) since grants are subject to final approval by the federal government. Therefore, S319(h) is the least decentralized among the three policies.

I adopt different empirical strategies to study states' behaviors based on the decision-making process of each NPS policy. S305(b) water quality assessment has binary outcomes, a catchment is either assessed or not depending on various factors that affect decision-making. Therefore, I use a discrete choice model (Probit) to study the effect of these factors on the probability of assessment. The development of TMDLs under S303(d) is based on the water quality assessment results. Only impaired water bodies need treatment. Based on the priority ranking, different water bodies receive TMDLs at different times. Thus, I use a duration model with selection: the first step is a discrete choice model (Probit) that selects the impaired water bodies, and the second step is a proportional hazard model that estimates the probability a catchment receives a TMDL. The amount of S319(h) NPS pollution management grants is based on the watershed restoration plan submitted by states and the Environmental Protection Agency (EPA)'s approval. Different watersheds receive different grant amounts. I conduct a Heckman two-step analysis to select the watersheds treated by S319(h) programs using a Probit model then estimate the effect of boundary factors on grant amount using an OLS model.

The policy's degree of decentralization affects the states' policy behavior at the near-boundary and transboundary rivers. My results show that the border catchments are 45.81% less likely to be assessed under S305(b) and 8.03% less likely to receive a TMDL under S303(d). Catchments within 1-30 km of state borders are between 6.26% to 12.68% less likely to be assessed and between 6.83% to 21.87% less likely to receive a TMDL compared to intrastate catchments. The S319(h) grant amounts fluctuate substantially across distances and portions of rivers. Border subwatersheds receive \$20,185.5 more grants, and a subwatershed within 1-30 km of the state borders receives \$9,714.2 to \$23,957.7 more grants than intrastate subwatersheds. The more decentralized a policy is, the stronger the boundary behaviors. The relatively less decentralized S319(h) funding is granted more for near-boundary and transboundary rivers, indicating these rivers need more pollution management. However, the two more decentralized policies, S305(b) and S303(d), are less likely to be implemented for these rivers, which implies that the state-level policy decisions do not maximize the national welfare.

I construct a theoretical model for the federal and state government's policy decisions. Some interstate watersheds will be treated by the federal government to maximize the national welfare but will not be treated by the state government due to the beggar-thy-neighbor and free-riding behaviors. Therefore, a policy decided by the state government will result in deadweight losses. The empirical estimation results show that S319(h) does not reflect these behaviors. S305(b) reflects the beggar-thy-neighbor behavior but the results fluctuate. The upstream portions of interstate rivers are 9.66-11.15% less likely to be assessed within 2.5-10 km of state borders and are 6.3% less likely to be assessed within 20-25 km of state borders than intrastate rivers. States exhibit both behaviors in S303(d), and the behaviors diminish with the distance to state borders. The upstream portions of interstate rivers within 5 km of the state borders have a 10.28% to 18.9% lower probability of TMDL development, which is consistent with the beggar-thy-neighbor behavior. The downstream portions of interstate rivers within 10 km of the state borders have a 25.71% to 55.81% lower probability of TMDL development, which is consistent with the free-riding behavior.

Using the willingness to pay for clean water estimated in the literature (Hite, Hudson, and Intarapapong (2002), Jordan and Elnagheeb (1993), Chatterjee et al. (2017)), I conduct a back-of-the-envelope calculation and infer that each behavior incurs a large deadweight loss. The downstream free-riding behavior is stronger than the upstream beggar-thy-neighbor behavior in TMDL development but does not necessarily incur a larger deadweight loss, because the interstate rivers have more upstream tributaries than downstream tributaries. The free-riding behavior is affected by the upstream states' environmental and political-economic background. An upstream state that is environmentally friendly or smaller in population or GDP would intensify the free-riding behavior of the downstream state.

The two transboundary policy behaviors have not been studied directly and separately in the literature. Transboundary analyses of other environmental problems like air pollution (Fowlie, Petersen, and Reguant (2021)), endangered species (List, Bulte, and Shogren

(2002)), etc. are not able to differentiate between the two policy behaviors because, unlike water flows, these other settings do not have a determined direction when crossing borders. While the literature on transboundary water pollution has been ambiguous regarding the two types of behaviors. One reason is that the literature uses pollution levels to indicate policy implementation, instead of studying the policy implementation directly. This makes it more difficult to differentiate the two types of policy behaviors and separate the results of policy behaviors from polluting activities. Helland and Whitford (2003)) finds high emissions in air and water from facilities in counties near state borders. Sigman (2014) finds higher interjurisdictional variation in water pollution levels. Sigman (2002) and Lipscomb and Mobarak (2016) find high pollution upstream near the jurisdictional borders. These results can be attributed to both less pollution control and more polluting activities near the boundaries. Sigman (2005) finds high pollution downstream, which can result from both the beggar-thy-neighbor behavior and free-riding behavior in policy implementation. This paper studies the transboundary policy behaviors directly and separates the two behaviors using the policies' implementation status across different portions of interstate rivers.

This paper is also the first transboundary analysis of NPS water pollution policies. Since being established in 1972, the CWA has invested more than $1 trillion in pollution treatment. However, despite these investments, 50% of assessed rivers remain impaired (Kelderman et al. (2022)). The major threat is NPS pollution, which has become the most widespread contributor to water pollution in past decades (Birkeland (2001), Keiser and Shapiro (2019)). However, NPS pollution and regulations have received less attention in the literature. This is because NPS diffuses, which makes it difficult to monitor pollution levels and establish a causal relationship between NPS pollution and NPS policies. I avoid these difficulties by studying the NPS policy implementation status directly. The diffuse nature of NPS leads to more jurisdictional issues and creates more flexibility for the jurisdictional authorities to prioritize policy implementation than point sources, leading to a higher risk of transboundary policy behaviors.

The two transboundary policy behaviors studied in this paper are also applicable to other environmental policies. Because these behaviors result from local authorities' goal of maximizing local welfare, which is a common feature of many public policies. In addition, this paper decomposes the decision-making process of the three NPS policies and models each policy into its characteristics to better understand the policy incentives. Thus, the results in this paper improve our understanding of decentralized environmental policies and specify more flexibility in decision-making that is easier to generalize to other contexts. For example, S305(b) water quality assessment can be compared with other low local cost policies for providing information, such as air pollution monitoring and traffic monitoring. S303(d) TMDL development can be compared with other second-step policies with high-level local government authority, such as driving restrictions, construction of infrastructure and welfare facilities, protection of locally endangered species, etc. Results for S319(h) NPS pollution management granting program can be generalized to other decentralized grants, subsidies, and rebate programs subject to central approval. The heterogeneous near-boundary and transboundary effects across these policies provide valuable insights into the implementation of other environmental policies at jurisdictional boundaries.

This paper also provides an argument for the ongoing debate about environmental federalism. I study the policy behaviors on different portions of rivers and discuss the outcomes of environmental federalism using three different policies. Prior literature has found that the impact of decentralized environmental policies on environmental quality may be insignificant (List and Gerking (2000)), or positive (Millimet (2003), Sigman (2003)), or both, depending on jurisdiction homogeneity (Oates and Schwab (1988)) or decision-making process (Silva and Caplan (1997)). The debate over environmental federalism for water pollution control has intensified in recent years. This is because the Clean Water Act has been unclear in the jurisdiction definition of "waters of the United States" (WOTUS). The 2020 Navigable Waters Protection Rule (NWPR) adopts a narrow federal jurisdiction that excludes the "interstate waters". However, some studies point out that unclear jurisdictional responsibility and treating waters as locally public goods would increase interstate water pollution (Greenstone and Hanna (2014), Keiser et al. (2021), Keiser et al. (2022)). Keiser et al. (2022) proposes future research to connect studies of economic behavior with the nation's hydrological network to improve ex-ante projections of the impacts of future water rules and regulations. This paper fills an important gap in this literature.

2 Policy Background

The series of nonpoint-source (NPS) pollution policies in the CWA are Section 305(b) Water Quality Assessment, Section 303(d) the Total Maximum Daily Load (TMDL) Program, and Section 319(h) Nonpoint-source Management Program Grants. These NPS pollution policies are water-quality based instead of effluent-based as are most point-source pollution policies. This is because nonpoint-source (NPS) water pollution cannot be regulated at the end of the pipe like point sources. Table 1 briefly summarizes the degrees of decentralization of the three NPS policies evaluated by their levels of state authority and state-borne costs.

Table 1: Degrees of Decentralization

NPS Policy	State Authority	State Costs (per year)	Decentralization
S305(b)	Medium	Low ($1.3 m. first, $0.2 m. subsq.)	Medium
S303(d)	High	High ($63-69 m.)	High
S319(h)	Low	Low	Low

2.1 Section 305(b): Water Quality Assessment

Section 305(b) Water Quality Assessment is the prerequisite of Section 303(d). S305(b) requires states to assess the quality of the state's surface and ground waters and submit an integrated report to the U.S. Environmental Protection Agency (EPA) every two years. Not all water bodies are assessed at once. From the establishment of the Clean Water Act in 1972 until the first integrated reporting cycle in 2002, many bodies of water in the nation were still listed as unassessed. As shown in Table 2, states categorize the assessment units (indexed catchments) into five integrated reporting categories (IRC). States set water quality standards for each pollutant by the water's designated uses and decide the water's attainment status. IRC 4A is also referred to as the "S303(d) list," which is the list of impaired water bodies that enter S303(d) for TMDL development. The cost of S305(b) mainly consists of monitoring and administrative costs, while monitoring is conducted jointly by federal, state, and local agencies. The costs borne by the state government are about \$1 million per reporting cycle before 2002, \$1.3 million in 2002, and \$200,000 for subsequent listings (EPA (2001)).[1]

2.2 Section 303(d): Total Maximum Daily Load Program

Section 303(d) is the major federal policy to control NPS water pollution. After states have developed a list of impaired water bodies under S305(b), S303(d) requires states to

[1] S305(b) costs each state government about \$1 million per reporting cycle before the July 2000 TMDL Program Regulation Revisions. "The July 2000 TMDL rule requires states to improve their methodologies for setting priorities, establish schedules for developing TMDLs, increase public participation, provide their lists in a consistent format, and convey the essential information supporting the listing." The states pay about \$1.3million one-time transition costs for the first listing in 2002, and \$200,000 for subsequent listings (EPA (2001)).

Table 2: EPA Integrated Reporting Category

1	All Uses have been assessed and all are supporting Water Quality Standards
2	All assessed Uses are supporting Water Quality Standards but may have one or more Uses that were Not Assessed
3	Insufficient Information to make an assessment decision, or Not Assessed
4A	Impaired by a pollutant but already has a TMDL
4B	Impaired by a pollutant but doesn't need a TMDL since other pollution control measures are in place
4C	Impaired by something that is not a pollutant
5	Impaired by a pollutant and still needs a TMDL

develop TMDLs for the impaired water bodies by priority ranking. "A TMDL is the calculation of the maximum amount of a pollutant allowed to enter a waterbody so that the waterbody will meet and continue to meet water quality standards for that particular pollutant."[2] Thus, a TMDL is a tool to set the pollutant reduction target. To meet the TMDL, the state allocates load reductions among the sources of the pollutant. According to an EPA report in 2011, about 76% of TMDLs were driven by nonpoint sources (EPA (2011)).

S303(d) was established in the Clean Water Act in 1972, but it was given limited attention at the time. Early water quality programs focused on point sources like industrial dischargers and sewage treatment facilities, primarily the National Pollutant Discharge Elimination System (NPDES) program (Craig and Roberts (2015), Keiser and Shapiro (2019)). However, nonpoint sources have become a widespread contributor to water pollution in the U.S. in past decades. As a result, the point source control programs are not able to achieve the desired level of environmental quality in the whole watershed. S303(d) was brought to the table again in the 1990s when citizen organizations began legal actions to facilitate the development of TMDLs (Craig and Roberts (2015)), the Clean Water Action Plan in 1998 highlighted nonpoint-source pollution, funding for states to deal with nonpoint-source pollution almost doubled from $105 million in 1998 to $200 million in 1999 (GAO (2000)).

[2] EPA, https://www.epa.gov/tmdl/overview-identifying-and-restoring-impaired-waters-under-section-303d-cwa, https://www.epa.gov/tmdl/overview-total-maximum-daily-loads-tmdls.

To maintain the past progress in point-source pollution regulation, identify nonattainment water bodies polluted by multiple sources, and implement cost-effective restoration activities, EPA published final revisions to the TMDL regulations, effective after October 30, 2001 (EPA (2001)). These events represent a shift in the focus of water quality management from effluent-based point-source regulation to ambient-based water quality standards for designated uses (NRC (2001)). S303(d) assesses the integrated health of the watersheds and allows states to allocate abatement efforts among point sources and nonpoint sources, meeting the water quality standards at the lowest cost.

States have high authority in TMDL development. EPA reviews, modifies, and approves the TMDLs. EPA may promulgate a state S303(d) list if the state fails to do so, but overall, EPA's role in S303(d) is minimal.[3] The average annual cost for each state to develop TMDLs is estimated to be $63-69 million per year after 2001 (EPA (2001)).

The decentralized implementation of S303(d) leads to welfare maximization at the state level instead of the national level. Different water quality standards across states also increase the difficulty of developing TMDLs at state borders. Therefore, states may have different behaviors when developing TMDLs for near-boundary and transboundary waters.

2.3 Section 319(h): Nonpoint Source Management Program Grants

Section 319(h) provides funding for nonpoint source management programs.[4] States submit work plans and grant applications to the EPA regional offices. Each EPA regional office reviews the state's work plans and provides comments. The EPA region office will

[3]For example, the EPA had prompted the states around the Chesapeake Bay to cooperate in the abatement of nutrient pollution for twenty-five years without sufficient progress, so the EPA had to promulgate the Chesapeake Bay TMDL in 2010 (Craig and Roberts (2015)).

[4]S319(h) grants supports a wide range of nonpoint-source pollution management activities including technical, financial, or incentive programs. Some projects also use the grants to assess water quality or develop TMDLs.

work with the state to ensure that the work plan is consistent with the EPA guidelines.[5] Only the EPA has the final authority to award the grants (EPA (2011)). S319(h) funding is not randomly granted, but it does not have the procedural order as do S305(b) and S303(d). S319 is an important mechanism for implementing TMDLs, but some S319(h) funding was granted in the absence of TMDLs.

Several features of S319(h) make it more centralized than S305(b) and S303(d). First, though the grant application is proposed by the state government, the EPA holds the final authority for grant approval. States prioritize the restoration of impaired waters, but they must account for national welfare in the work plans to get funding approved by the EPA. One of the nine key elements of an effective state program required by the EPA is that "The State strengthens its working partnerships and linkages to appropriate state, interstate, tribal, regional, and local entities (including conservation districts), private sector groups, citizens groups, and Federal agencies."(EPA (2011)). This also provides an incentive for the policy to disproportionately target near-boundary and transboundary waters more. In addition, states can include S319(h) grants in Performance Partnership Grants (PPGs). The PPGs is designed to strengthen the partnerships between the EPA and states and interstate agencies through joint planning and priority setting. Thus, S319(h) work plans in PPGs address the national welfare more.[6] Given these settings, we expect that states exhibit less near-boundary and transboundary behaviors when implementing S319(h) than S305(b) and S303(d).

[5] An effective state program should contain nine key elements.

[6] "Performance Partnership Grants enable States and interstate agencies to combine funds from more than one environmental program grant into a single grant with a single budget." "The Performance Partnership Grant program is designed to: (1) Strengthen partnerships between EPA and State and interstate agencies through joint planning and priority setting and better deployment of resources; (2) Provide State and interstate agencies with flexibility to direct resources where they are most needed to address environmental and public health priorities; (3) Link program activities more effectively with environmental and public health goals and program outcomes; (4) Foster development and implementation of innovative approaches such as pollution prevention, ecosystem management, and community-based environmental protection strategies; and (5) Provide savings by streamlining administrative requirements.", Title 40 - Protection of Environment Chapter I - ENVIRONMENTAL PROTECTION AGENCY Subchapter B - GRANTS AND OTHER FEDERAL ASSISTANCE Part 35 - STATE AND LOCAL ASSISTANCE.

3 Theoretical Model

The federal government makes policy decisions to maximize the national welfare. The state-level decisions may conflict with the federal decisions on watersheds shared with other states. There are two potential transboundary policy behaviors for states: (1) Beggar-thy-neighbor. The upstream states do not get the full benefits of cleaning the interstate rivers so they conduct insufficient controls and allow pollution to flow down into the downstream states. (2) Free-riding. The downstream states conduct insufficient control of interstate rivers because they expect upstream states to control pollution and they can enjoy the benefits without spending on pollution control.

Assume there are $N = \{1, 2, ..., n, ..., N\}$ impaired watersheds within a state, and $T = \{1, 2, ..., t, ..., T\}$ of the N impaired watersheds are interstate watersheds. The benefits and costs of treating the N watersheds within the state's territory are $\mathbb{B} = \{B_1, B_2, ..., B_n, ..., B_N\}$ and $\mathbb{C} = \{C_1, C_2, ..., C_n, ..., C_N\}$.[7] The federal government ranks the net benefits of treating each watershed from high to low and prioritizes the treatment:

$$\max_{F} \quad \textstyle\sum_{n=1}^{F}(B_n - C_n) \tag{1}$$

The federal government will treat the F watersheds that have positive net benefits, among which F_a are intrastate watersheds and F_t are interstate watersheds.

3.1 The Beggar-thy-Neighbor Behavior

Assume the T interstate watersheds lie in the upstream portions of interstate rivers. The upstream state u faces the same problem as the federal government in Equation 1 when

[7] B_n and C_n are continuously distributed nonnegative-valued random variables. $B_n \sim Rayleigh(\sigma)$ and $C_n \sim Rayleigh(\sigma)$.

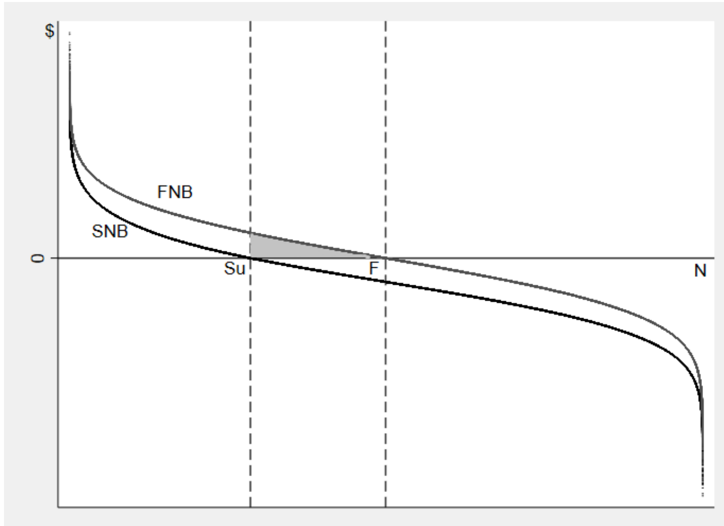
deciding whether or not to treat the intrastate watersheds. However, the state bears the full costs of treating the T upstream watersheds but only enjoys a percentage α_t ($\alpha_t \in (0,1)$) of the benefits. This is because when the upstream portion of an interstate watershed is cleaned, the downstream portion also has less pollution, but the environmental benefits in the downstream are acquired by the downstream states. Thus, some interstate watersheds with positive federal-level net benefits may have negative state-level net benefits. The state government ranks the state-level net benefits of treating each watershed from high to low and prioritizes the treatment. Compared with the federal-level rankings, the state government ranks higher the intrastate watersheds but lower the interstate watersheds. The problem faced by the upstream state u when deciding whether or not to treat the F_t upstream watersheds with positive national-level net benefits is:

$$\begin{aligned} \max_{x} \quad & \textstyle\sum_{t=1}^{F_t - x} (\alpha_t B_t - C_t) \\ \text{s.t.} \quad & \alpha_t B_t - C_t \geq 0 \end{aligned} \tag{2}$$

An interstate watershed t would be chosen to be treated by the federal government but not treated by the state government if $B_t - C_t > 0$ but $\alpha_t B_t - C_t < 0$, which means $\alpha_t < \frac{C_t}{B_t} < 1$. Assume there are x such watersheds, then the upstream state would only choose to treat $F_t - x$ upstream watersheds. In total, the upstream state will treat $F - x = S_u$ watersheds with positive state-level net benefits.[8] Figure 1 shows the federal and state governments' decisions. FNB denotes the federal-level net benefits, SNB denotes the state-level net benefits. The grey shaded area is the deadweight loss due to the state's inefficient policy decision.

[8]The subscript u denotes the policy is implemented in the upstream state u.

Figure 1: Beggar-thy-Neighbor



3.2 The Free-riding Behavior

Assume the T interstate watersheds lie in the downstream portions of interstate rivers. Unlike the situation with the upstream watersheds in Section 3.1, the benefit of treating a downstream watershed, B_t, is the same at the national and state level. This is because the treated downstream portions of interstate rivers do not flow into other states, so the federal and the downstream state d enjoy the same environmental benefits when treating an interstate river from its downstream portions. The state's problem with the intrastate watersheds is the same as the federal government. However, the downstream state d has an incentive to free ride if they expect the upstream state to treat the interstate watersheds and they can enjoy the benefits without spending any costs. The policy implementation status of the upstream state on the upstream portion of the interstate rivers is binary:

$$y = \begin{cases} 1, & \text{if treat} \\ 0, & \text{if do not treat} \end{cases} \tag{3}$$

The downstream state d's expectation for the upstream state u's decision is $Pr(y =$

$1) = \gamma_{du}$, $\gamma_{du} \in (0,1)$. γ is built on the downstream and upstream states' environmental and political-economic backgrounds. Theoretically, the federal government should treat interstate watersheds from their upstream portions to acquire larger environmental benefits. However, in reality, due to different administrative costs and the actual situation of the watersheds, the federal government may choose to treat the downstream portion of an interstate watershed instead of its upstream portion. In the case where the federal government can only treat interstate watersheds at their downstream portions, or if the downstream state government does not have any free-riding policy behaviors, the F_t interstate watersheds with positive federal-level net benefits should be treated. However, if the downstream state has an incentive to free ride, the downstream state government's problem for the F_t interstate watersheds is:

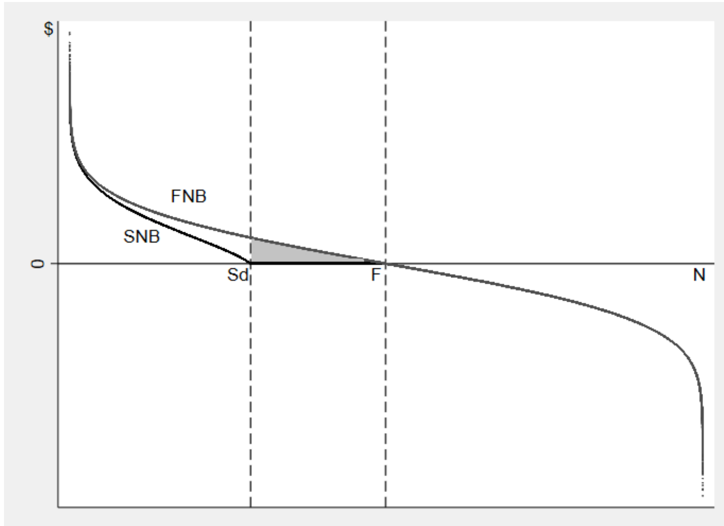
$$\begin{aligned} \max_{z} \quad & \textstyle\sum_{t=z+1}^{F_t}(B_t - C_t) + \sum_{t=1}^{z}(\gamma_{du} B_t) \\ \text{s.t.} \quad & B_t - C_t \geq 0 \\ & \gamma_{du} B_t \geq 0 \end{aligned} \tag{4}$$

A downstream watershed t will not be treated by the state government due to the free-riding behavior if $B_t - C_t > 0$ and $\gamma_{du} B_t > B_t - C_t$, i.e. $1 - \gamma_{du} < \frac{C_t}{B_t} < 1$. Assuming there are z such watersheds, then the downstream state would only choose to treat $F - z = S_d$ watersheds in total.[9] Figure 2 shows the federal and state government's decisions.[10] The grey shaded area is the deadweight loss incurred if the socially optimal solution is to treat these downstream watersheds. In reality, the deadweight loss could be smaller if it is sub-optimal to treat some interstate rivers from their upstream portions or if the downstream state's expectation for the upstream state's pollution control activity is accurate.

[9]The subscript d denotes the policy is implemented in a downstream state d.

[10]The state-level net benefit of not treating one of the z watersheds is $\gamma_{du} B_t$, which is non-negative. However, since the state does not treat these watersheds, the net benefits of these z watersheds do not appear in the state's treating decision. Thus, the net benefits of the z watersheds in Figure 2 are plotted as zeros.

Figure 2: Free Riding



There are two parameters in the transboundary behavior models, α_t and γ_{du}. α_t is the upstream state's share of benefits in treating an interstate watershed, while γ_{du} is the downstream state's expectation for the upstream state's pollution management of interstate watersheds. All else equal, if $z > x$, we will have:

$$\begin{gathered} 1-(1-\gamma_{du}) > 1-\alpha_t \\ \alpha_t+\gamma_{du} > 1 \end{gathered} \tag{5}$$

When α_t or γ_{du} is large enough to meet the above inequality, that is to say, when the upstream state could recover the majority of the benefits in treating the interstate watersheds, or when the downstream state has a high expectation that the upstream state will treat the interstate watersheds, the free-riding behavior leads to a larger percentage of under-managed interstate watersheds than the beggar-thy-neighbor behavior. If one or both of α_t and γ_{du} are small such that $\alpha_t + \gamma_{du} < 1$, the beggar-thy-neighbor behavior would be stronger than the free-riding behavior.

4 Empirical Strategies

To examine if states' have different behaviors when implementing the three NPS pollution policies on near-boundary and transboundary rivers, I study the probability of S305(b) water quality assessment, the priority of S303(d) TMDL development, and the funding level of S319(h) nonpoint-source pollution management programs for different water bodies.

4.1 Probit Model for S305(b) Water Quality Assessment

The assessment status of a catchment is binary with:

$$Assess = \begin{cases} 1, & \text{if assessed} \\ 0, & \text{if not assessed} \end{cases} \tag{6}$$

I use a Probit model to estimate the probability that a state assesses a catchment when controlling for the hydrologic and social-economic factors of the catchment:

$$p_{cs} = Pr[\ Assess_{cs} = 1 | \mathbb{X}_{cs}] = \Phi(\alpha_0 + Distbin_c\beta_1 + \beta_2 CoastCat_c + Str_c\alpha_1 + X_c\alpha_2 + D_s) \tag{7}$$

The dependent variable is the probability that a catchment c in state s is assessed, i.e., it is not in IRC 3 in Table 2. I model the policy behaviors near state borders by dividing the distance of all the catchments to state borders into 9 bins from less than 1 km to more than 30 km and denote them as *Distbin*. This flexible modeling of the distance allows me to observe the changes in the probability of assessment with the catchments' distances to the state borders. β_1 for each distance bin represents the effect of being a near-boundary catchment within that distance bin on the probability that the catchment is assessed. The state government may treat the catchments in the coastal areas differently, so I control for a coastal catchment dummy variable $CoastCat$. I also control for several other stream characteristics Str in the catchment: First, since the main stream may be more likely to be assessed than its tributaries, I control for an index of stream size based on a hierarchy of tributaries. Second, I construct a variable $DistArea = \frac{4Distance}{\sqrt{Area}}$ to represent the percentage of the catchment's distance to state borders to the state's area. This is because if a large state conducts less pollution control within a certain distance of its borders, it can still properly manage pollution over the broad intrastate area. However, if a small state has fewer pollution controls within the same distance of its borders, a large part of the state would be under-managed. Therefore, the distance from the state borders at which the policy behavior occurs is related to the state area. I also control for the average social-economic factors X in each catchment, including county population and annual personal income, because when the states make an assessment decision, they may consider health benefits for nearby residents. In addition, I control for state-fixed effects D_s to account for the state-level factors that affect the assessment decisions. The standard errors are robust.

Equation 7 estimates the policy behaviors on near-boundary catchments. Next, I explore if the states have different policy behaviors on transboundary catchments near state borders in Equation 8. I interact the distance bins with a dummy variable $InterCat$ that equals 1 if the catchment contains segments of interstate rivers and 0 if the catchment only contains intrastate rivers. I also interact the coastal catchments with their distance to the shorelines to check the change in the probability of assessment for coastal catchments across distance bins.

$$\begin{aligned} p_{cs} = Pr[\ Assess_{cs} = 1 | \mathbb{X}_{cs}] \ &= \Phi(\alpha_0 + Distbin_c\beta_1 + Distbin_c \times InterCat_c\beta_2 \\ &+ Distbin_c \times CoastCat_c\beta_3 + Str_c\alpha_1 + X_c\alpha_2 + D_s) \end{aligned} \tag{8}$$

To explore the beggar-thy-neighbor behavior and free-riding behavior separately, in Equation 9, I separate the interstate river into its upstream, downstream, and up/downstream portions and interact them with the distance bins.

$$
\begin{aligned}
p_{cs} = Pr[\ Assess_{cs} = 1|\mathbb{X}_{cs}] \ &= \Phi(\alpha_0 + Distbin_c\beta_1 + Distbin_c \times UpCat_c\beta_2 \\
&+ Distbin_c \times DnCat_c\beta_3 + Distbin_c \times UpDnCat_c\beta_4 \\
&+ Distbin_c \times CoastCat_c\beta_5 + Str_c\alpha_1 + X_c\alpha_2 + D_s)
\end{aligned} \tag{9}
$$

In Equation 9, β_1 for the 0-1 km distance bin is the coefficient for catchments of border flowlines. β_1 for the distance bins between 1 to 30 km are coefficients for catchments that are near-boundary but not transboundary. β_2, β_3, and β_4 are coefficients for different portions of interstate catchments across distance bins within 30 km of state borders. β_5 is the coefficient for coastal catchments across distance bins within 30 km of the shorelines. The omitted category includes catchments further than 30 km from state borders and shorelines, which are the intrastate catchments defined in Table 3.

Table 3: Portions of Rivers

Name	Description	Cat. No.
Interstate river		132,215
Border river	River segments within 1 km of the state borders	14,804
Upstream	The upstream portion of an interstate river within 30 km of the state borders	107,026
Downstream	The downstream portion of an interstate river within 30 km of the state borders	4,534
Up/Downstream	The upstream and downstream portion of an interstate river within 30 km of the state borders	5,851
Coastal river	River segments that flow into the ocean or the Great Lakes and within 30 km of the shoreline	53,844
Non-trans. river	River segments that are near-boundary but not transboundary	45,136
Intrastate river	River segments further than 30 km from the state borders	191,110

4.2 Duration Model with Selection for S303(d) TMDL Development

The state government decides the priority to develop TMDLs for S303(d) listed impaired waters. Therefore, it takes different lengths of time for different catchments to receive TMDLs. I conduct a survival analysis to estimate the hazard rate that the state develops TMDLs for catchments across different locations. A simple duration model will estimate the hazard rate that an impaired catchment receives TMDLs. However, the catchments may not be impaired randomly. For example, catchments close to industrial areas, cultivated areas, livestock areas, or catchments prone to soil erosion may be more likely to be impaired. In this case, the estimation results from a simple duration model would be subject to selection bias. Therefore, I use a duration model with selection proposed by Boehmke, Morey, and Shannon (2006) to eliminate the selection bias. I select the sample of impaired catchments using a Probit model and estimate the hazard rate that state develops TMDLs for a catchment using a proportional hazard model. The Boehmke, Morey, and Shannon (2006) method uses a bivariate exponential distribution to bind together the selection and duration equations.[11]

A catchment's impairment status is binary:

$$Impair = \begin{cases} 1, & \text{if impaired} \\ 0, & \text{if not impaired} \end{cases} \tag{10}$$

The TMDL development status is only observed if $Impair = 1$. The selection equation is:

$$p_{cs} = Pr[\ Impair_{cs} = 1|\mathbb{X}_{cs}] = \Phi(\gamma_0 + Distbin_c\gamma_1 + Str_c\gamma_2 + Str1_c\gamma_3 + X_c\gamma_4) \tag{11}$$

[11] I also run an independent duration model without sample selection in Section A in the Appendix.

In Equation 11, the dependent variable is the probability that a catchment c in state s is listed as impaired. Since past studies find higher pollution near jurisdictional boundaries (Sigman (2005), Sigman (2005), Lipscomb and Mobarak (2016)), I include $Distbin$ to control for the effect of a catchment's distance to state borders on its impairment status. I control for the stream characteristics Str in each catchment as in Equation 7, including the stream size index and the distance/area percentage, because pollution concentration may differ with stream size, and the intensity of polluting activities may differ with the distance from borders relative to the state area. Since the attainment status depends on the state's water quality standards by designated uses and is evaluated at the waterbody/pollutant level, I include another set of stream characteristics $Str1$, including the natural flows[12] and flow speed that affect pollution loads, and the water's designated uses that affect the attainment standards. Variables in $Str1$ affect the catchment's impairment status but are less relevant to the priority of TMDL development. Human activities affect pollution levels and types, so I also include social-economic factors X at the catchment level, including county population and annual personal income.

The duration equation is:

$$h(t)_{cs} = h_0(t)_{cs} exp(\alpha_0 + Distbin_c\beta_1 + \beta_2 CoastCat_c + Str_c\alpha_1 + X_c\alpha_2 + D_s) \tag{12}$$

Equation 12 is a proportional hazards model. $h(t)$ is the hazard rate that a catchment receives a TMDL after t years of impairment. The larger the hazard, the sooner the catchment receives a TMDL. The baseline hazard $h_0(t) = pt^{p-1}$ follows a Weibull distribution. I use $Distbin$ to explore the state's policy behavior on near-boundary catchments. I control for the coastal catchments $CostCat$ because the states may have different behaviors when developing TMDLs for coastal catchments. I control for the same variables that affect

[12]Flow from runoff.

policy decisions as in Equation 7, which include stream characteristics Str, social-economic variables X, and state-fixed effects D_s. The standard errors are robust.

To see whether states develop TMDLs for transboundary rivers with different hazard rates from intrastate rivers, I interact the distance bins with the interstate river dummy and coastal river dummy in Equation 13:

$$\begin{aligned} h(t)_{cs} = h_0(t)_{cs}exp(\alpha_0 + Distbin_c\beta_1 + Distbin_c \times InterCat_c\beta_2 + Distbin_c \times CoastCat_c\beta_3 \\ + Str_c\alpha_1 + X_c\alpha_2 + D_s) \end{aligned} \tag{13}$$

I then divide the interstate river into different portions in Equation 14 to explore the hazard rate of TMDL development for the upstream and downstream portions of interstate rivers:

$$\begin{aligned} h(t)_{cs} = h_0(t)_{cs}exp(\alpha_0 + Distbin_c\beta_1 + Distbin_c \times UpCat_c\beta_2 + Distbin_c \times DnCat_c\beta_3 \\ + Distbin_c \times UpDnCat_c\beta_4 + Distbin_c \times CoastCat_c\beta_5 + Str_c\alpha_1 + X_c\alpha_2 + D_s) \end{aligned} \tag{14}$$

4.3 Heckman Selection Model for S319(h) Nonpoint Source Pollution Management Grants

S319(h) funding is granted at the subwatershed (HUC12) level.[13] Different subwatersheds receive different amounts of grants. Subwatersheds in S319(h) programs are subject to the state government's planning and the federal government's approval, so they are not randomly

[13]Watersheds at the 12-digit hydrologic unit level.

selected. Therefore, I adopt a Heckman two-step analysis to estimate if the approved grant amounts are different for near-boundary and transboundary subwatersheds.

A subwatershed is either managed by the S319(h) program or not:

$$Manage = \begin{cases} 1, & \text{if managed by S319(h)} \\ 0, & \text{if not managed by S319(h)} \end{cases} \tag{15}$$

The grant amount is only observed if $Manage = 1$. So I use a discrete choice model (Probit) for the first-step sample selection:

$$p_{ws} = Pr[\ Manage_{cs} = 1|\mathbb{X}_{ws}] = \Phi(\gamma_0 + Distbin_w\gamma_1 + Str_w\gamma_2 + Str1_w\gamma_3 + X_w\gamma_4 + D_s + D_h) \tag{16}$$

In Equation 16, the dependent variable is the probability that a subwatershed is managed by S319(h) program. Since we expect states to exhibit near-boundary and transboudnary behaviors in pollution management, I control for the subwatershed's distance to state borders, $Distbin$. I include the stream characteristics Str, including the stream size index, because larger streams are more likely to be selected into restoration programs, and the distance/area percentage, because the state's application for watershed restoration grants may be different across the watershed's distance to state borders and the state area. I also include the second set of stream characteristics $Str1$, including the natural flows and flow speed that affect pollution load thus the treatment needs, and the water's designated uses that affect the political authority's incentive to treat. Variables in $Str1$ affect the state's incentive to apply for S319(h) restoration grants but have little impact on the specific amount of grants received by each subwatershed. The political authority's pollution management incentive is also affected by social-economic factors X of the subwatershed. More populous and richer places may receive more government attention for pollution management. So I control for county population and annual personal income. Since the S319(h) programs are proposed by state governments, I include state-fixed effects D_s to control for state-level factors that affect the decision to apply S319(h) grants for a subwatershed. I also control for the watershed-fixed effects D_h at the HUC4 level,[14] because the S319(h) work plans are usually proposed at the watershed level, so HUC12 subwatersheds within the same HUC4 watershed have correlated probabilities to be managed by S319(h) programs. However, since the actual work plans may cover a larger or smaller scale of watersheds than HUC4, and the management practices vary across subwatersheds, the yearly grant amount for each subwatershed in a HUC4 watershed can be very different.[15]

The second-step OLS regression estimates the near-boundary factors' effect on grant amount:

$$Grant_{wsy} = \alpha_0 + Distbin_w\beta_1 + \beta_2 CoastWs_w + Str_w\alpha_1 + X_w\alpha_2 + D_s + D_y + \theta\hat{\lambda_{ws}} + \epsilon_{wsy} \tag{17}$$

In Equation 17, $Grant_{wsy}$ is the total S319(h) grants awarded to subwatershed w in state s in year y. I use $Distbin$ to examine the near-boundary granting trend. I control for coastal subwatersheds $CoastWs$ because the coastal subwatersheds may be subject to different S319(h) granting behaviors and other sources of grants like the EPA's Coastal Zone Act Reauthorization Amendments (CZARA). I control for stream features Str, including the stream size index since larger streams may receive larger amounts of grants, and the distance/area percentage that may affect treatment incentives. I control for the average

[14] A larger scale of watersheds than the HUC12 subwatersheds.

[15] For subwatersheds that have entered the S319(h) programs, the mean and standard deviation of the grant amount for the whole sample is \$230,402 and \$294,522.6, respectively. The average standard deviation of the grant amount for subwatersheds in each HUC4 watershed is \$200,762.1, which is large relative to the whole sample statistics.

social-economic factor X at the subwatershed, including county population and annual personal income, because the political authority may consider benefits for nearby residents when deciding the grant amount. I include the state-fixed effects D_s to account for state-level factors that affect the grant amount. I also include the year-fixed effects D_y to control for the yearly granting trend common for all subwatersheds. $\hat{\lambda_{ws}}$ is an estimate of the inverse Mills ratio derived from the estimation of the selection equation 16. The standard errors are robust.

I then interact the distance bins with interstate and coastal subwatershed dummies and split the interstate dummy into different portions to explore the granting behaviors on transboundary rivers in detail:

$$\begin{aligned} Grant_{wsy} = {} & \alpha_0 + Distbin_w\beta_1 + Distbin_w \times InterWs_w\beta_2 + Distbin_w \times CoastWs_w\beta_3 + Str_w\alpha_1 \\ & + X_w\alpha_2 + D_s + D_y + \theta\hat{\lambda_{ws}} + \epsilon_{wsy} \end{aligned} \tag{18}$$

$$\begin{aligned} Grant_{wsy} = {} & \alpha_0 + Distbin_w\beta_1 + Distbin_w \times UpWs_w\beta_2 + Distbin_w \times DnWs_w\beta_3 \\ & + Distbin_w \times UpDnWs_w\beta_4 + Distbin_w \times CoastWs_w\beta_5 + Str_w\alpha_1 + X_w\alpha_2 \\ & + D_s + D_y + \theta\hat{\lambda_{ws}} + \epsilon_{wsy} \end{aligned} \tag{19}$$

5 Data

5.1 Hydrology and Spatial Data

I acquire the features of all the stream flowlines in the U.S. from the National Hydrography Dataset Plus, Version 2 (NHDPlus V2). The NHDPlus V2 is a comprehensive geospatial dataset that draws stream networks and records stream characteristics by segments. The line features are applicable to upstream and downstream navigation and also contain stream attributes such as natural flows, speed, and index of stream size based on a hierarchy of tributaries. NHDPlus allows interactive modeling with other geospatial data layers. Thus, I match the hydrographic feature with other spatial, temporal, and political data using ArcGIS to construct a unique and comprehensive dataset for modeling the water pollution policy decisions on hydrological variables.

Since the further away a flow segment is from the state borders, the less likely that it is subject to the near-boundary and transboundary policy behaviors, this paper takes the catchments within 30 km of state borders as near-boundary catchments. Table 3 describes the different portions of rivers and the number of catchments in the sample for each portion. There are four types of interstate catchments: A border catchment contains river segments that are within 1 km of the state borders; An upstream catchment contains only the upstream portion of interstate rivers; A downstream catchment contains only the downstream portion of interstate rivers; An up/downstream catchment contains interstate river segments that are the upstream of one state border but the downstream of another state border. A coastal catchment contains river segments that are within 30 km of a shoreline and will flow into the ocean or the Great Lakes. Catchments that are within 30 km of state borders but do not contain river segments that flow to other states, the ocean, or the Great Lakes are near-boundary but non-transboundary. The catchments that are further than 30 km away from the state borders are intrastate catchments.

Figure 3 and Figure 4 show river flowlines in two representative watersheds.[16] The blue flowlines are intrastate rivers. Watershed 07 in Figure 3 mainly covers areas of Minnesota, Wisconsin, Iowa, and Illinois. The border flow segments (within 1 km of state borders) are marked in yellow, the upstream portions of interstate rivers are green, the downstream

[16]The watersheds are at the 2-digit hydraulic unit level.

portions of interstate rivers are red, and the up/downstream portions are pink. In watershed 07, a segment of the Mississippi river forms the state borders, and most of its tributaries are the upstream portions. This may explain why the number of upstream catchments in the sample is much larger than the number of downstream catchments. Watershed 18 in Figure 4 mainly covers California. This watershed has many coastal flowlines.

Figure 3: Watershed 07 Upper Mississippi Region

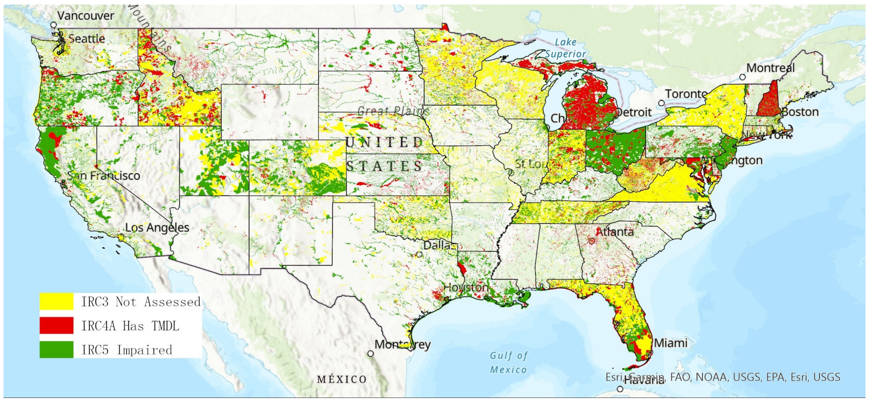
5.2 Policy Data

I obtain the catchment-level S305(b) and S303(d) information from the Assessment, Total Maximum Daily Load Tracking and Implementation System (ATTAINS). The ATTAINS records all the catchments' Integrated Reporting Categories (IRC) from 2002 to 2020, and TMDL development information from 1987 to 2020. Figure 5 presents a map of catchments

Figure 4: Watershed 18 California Region

listed in three categories in the contiguous United States in 2020. I obtain the Section 319(h) projects and grants information at the HUC12 subwatershed level from the Grants Reporting and Tracking System (GRTS).[17]

Figure 5: Distribution of Catchments in Different IRC Categories in 2020



The ATTAINS and GRTS data record the change in policy status in time and across locations, which allows me to match them with the hydrology and spatial data. I also match these data with state boundary, watershed boundary,[18] and shoreline data from the United States Geological Survey (USGS), and the annual county population and annual personal income data from the Census Bureau.[19]

Figure 6 shows the total approved S319(h) grants across subwatersheds in the contiguous United States. The darker area means a larger amount of grants. A comparison of Figure 6 and Figure 39 river borders in the Appendix shows that the subwatersheds of rivers that form the state borders seem to receive larger amounts of S319(h) grants. There are 10 states with river borders. Section 7.2 compares the S319(h) funding behavior for river-border states

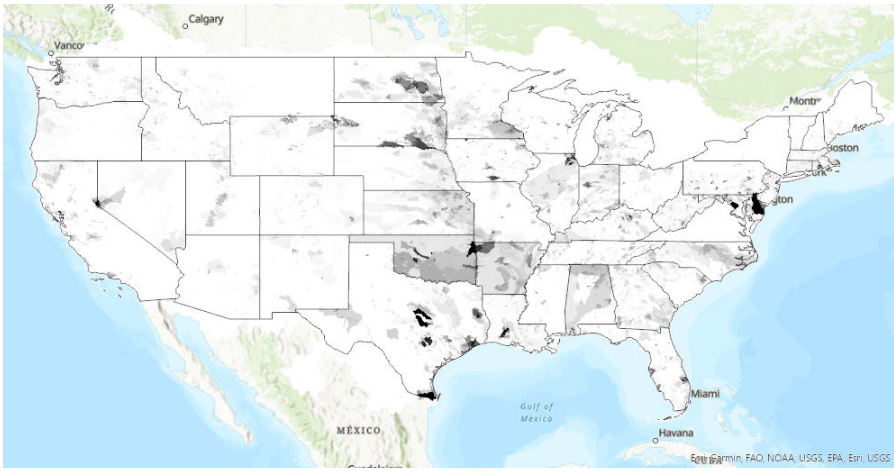
[17]Figure 38 in the Appendix shows an example of a subwatershed, the assessment units (catchments), and the flowlines in the subwatershed. The green boundary denotes the area of a huc12-level subwatershed; The purple and red areas are two assessment units; The blue lines are stream flowlines. The huc12-level subwatershed is a larger area than the catchments. A catchment is an area of land from which the stream drains the water.

[18]See Figure 40 in the Appendix for watershed boundaries at HUC4 level.

[19]Data are aggregated to catchment/subwatershed level. County population and annual personal income are averaged; The stream size index is aggregated to its maximum; The maximum, minimum, and average natural flow and speed are all used; Water uses are accumulated.

and the other states.

Figure 6: Total S319(h) Grants Across Subwatersheds by 2020



6 Results

This section presents the estimation results of near-boundary behavior and transboundary behavior for the three NPS pollution policies. By cutting the effects of boundary factors on policy implementation into upstream and downstream portions and distance bins, I can observe the changing pattern of the policy behaviors. However, this modeling also reduces the power of results in specific distance bins. Therefore, when interpreting the estimation results, it is more valuable and reliable to focus on the overall changing pattern of the policy behaviors rather than the specific estimation results in each distance bin.

6.1 Policy Behaviors on Near-Boundary Rivers

The estimation results for policy behaviors on near-boundary rivers are presented in Table 4. The omitted category contains rivers further than 30 km from state borders (intrastate rivers). The coefficients of interest with 95% confidence intervals are plotted in Figure 7.

States are less likely to implement both S305(b) water quality assessment and S303(d) TMDL development on near-boundary catchments. However, more S319(h) grants are approved for near-boundary subwatersheds.

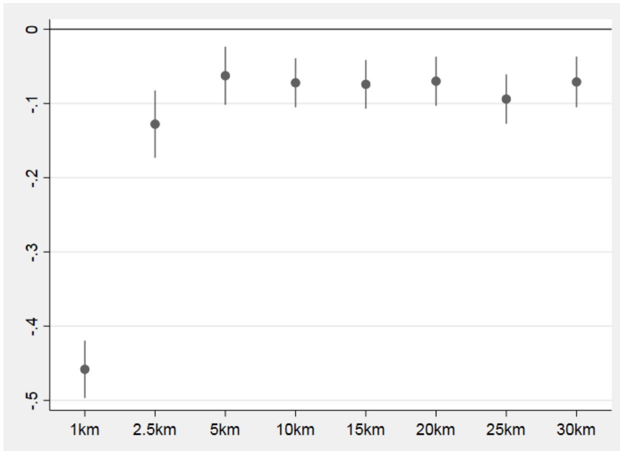
Table 4: Estimation Results: Distance to State Borders

	(1) Assessment	(2) TMDL	(3) Grants
	Distance to state borders		
0-1km (Border)	-0.4581***	-0.0803***	20.1855***
	(0.0197)	(0.0302)	(5.2784)
1-2.5km	-0.1278***	-0.1368***	23.9577***
	(0.0231)	(0.0340)	(7.2959)
2.5-5km	-0.0626***	-0.0150	9.1819
	(0.0200)	(0.0285)	(5.6209)
5-10km	-0.0721***	-0.1103***	17.9849***
	(0.0168)	(0.0254)	(4.6124)
10-15km	-0.0741***	-0.1311***	9.7142**
	(0.0168)	(0.0250)	(4.4656)
15-20km	-0.0700***	-0.0683***	-0.5492
	(0.0169)	(0.0248)	(4.6281)
20-25km	-0.0940***	-0.1722***	12.1329***
	(0.0171)	(0.0264)	(4.5124)
25-30km	-0.0710***	-0.2187***	12.0541***
	(0.0174)	(0.0275)	(4.5337)
CoastCat.	0.1408***	0.1075***	-12.6375***
	(0.0126)	(0.0171)	(4.6956)
Pop. (k)	0.0001**	0.0001***	-0.0064
	(0.0000)	(0.0000)	(0.0070)
Inc. (m)	0.0000**	0.0000***	-0.0000
	(0.0000)	(0.0000)	(0.0001)
Constant	0.8310***	-36.3063***	-157.1660***
	-0.0365	(0.1521)	(57.5881)
p		8.9747***	
		(0.0000)	
rho		-0.2500***	-0.2383**
		(0.0000)	(0.0267)
Str	Yes	Yes	Yes
Year FE	No	No	Yes
State FE	Yes	Yes	Yes
Obs.	194,470	359,045	89,502

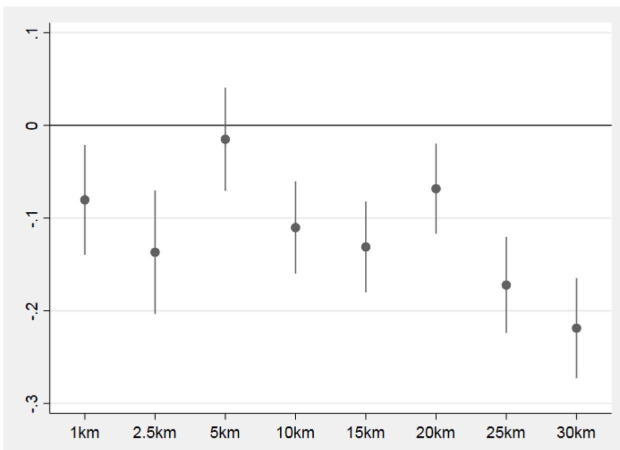
Note: Numbers in parentheses are robust standard errors.

Figure 7: Policy Behaviors on Near-boundary Rivers

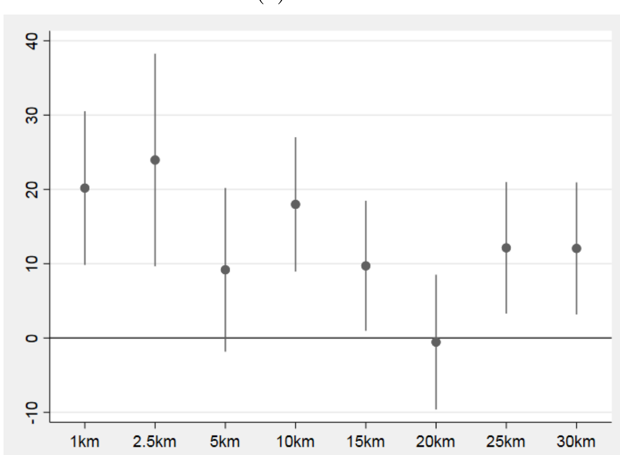
(a) Prob. of Assessment



(b) Hazard of TMDL Development



(c) Grants



Column (1) in Table 4 shows the effects of the catchment's distance to state borders on the probability of S305(b) water quality assessment. The border catchments are 45.81% less likely to be assessed than intrastate catchments. Catchments within 1-30 km of the state borders are 6.26-12.68% less likely to be assessed. Panel (a) of Figure 7 presents these coefficients.

Column (2) of Table 4 presents the estimation results for S303(d) TMDL development. The border catchments have an 8.03% lower hazard to receive TMDLs than intrastate catchments. Being a catchment within 1-30 km of the state borders lowers the hazards of TMDL development by 6.83% to 21.87%, which is greater than the decreased probability of S305(b) water quality assessment. Panel (b) in Figure 7 plots the coefficient estimates.

The estimation results for the approved S319(h) grants are presented in column (3) of Table 4. The border subwatersheds receive \$20,185.5 more grants per year than intrastate subwatersheds. A subwatershed within 1-30 km of the state borders receives \$9,714.2 to \$23,957.7 more grants per year. Panel (c) in Figure 7 shows that the coefficients fluctuate across the distance bins, but coefficients closer to state borders are more positive and statistically significant, indicating higher funding approved for subwatersheds near state borders. Since S319(h) is relatively more centralized, the heterogeneous results from the three policies indicate that the two decentralized policies, S305(b) and S303(d), are not implemented sufficiently near the state borders.

6.2 Policy Behaviors on Transboundary Rivers

States' behaviors on interstate rivers are presented in Table 5 and Figure 8. Column (1) of Table 5 for S305(b) water quality assessment shows that most estimation results for interstate catchments near the state borders are negative, but the standard errors are large. Panel (a) in Figure 8 shows that the coefficients fluctuate substantially.

Table 5: Estimation Results: Transboundary Rivers

	(1) Assessment	(2) TMDL	(3) Grants
0-1km (Border)	-0.4339***	-0.0381	18.4630***
	(0.0196)	(0.0297)	(5.3217)
	Near/Non-trans ×Distance		
1-2.5km	-0.0760**	-0.0329	23.8154*
	(0.0327)	(0.0409)	(12.9397)
2.5-5km	-0.0422	0.0792**	-7.6150
	(0.0270)	(0.0337)	(9.7492)
5-10km	-0.0161	-0.1037***	-2.6812
	(0.0225)	(0.0321)	(7.7115)
10-15km	-0.0624***	-0.1117***	-5.5869
	(0.0236)	(0.0326)	(8.5320)
15-20km	-0.0488**	-0.1062***	-37.3951***
	(0.0243)	(0.0325)	(7.8563)
20-25km	-0.0442*	-0.2784***	-12.2012
	(0.0251)	(0.0356)	(7.9115)
25-30km	-0.0356	-0.2679***	-12.8288*
	(0.0260)	(0.0367)	(6.9954)
	Interstate ×Distance		
1-2.5km	-0.0541	-0.1273**	-1.0035
	(0.0398)	(0.0570)	(14.5626)
2.5-5km	0.0025	-0.1065**	19.9959*
	(0.0323)	(0.0437)	(10.7347)

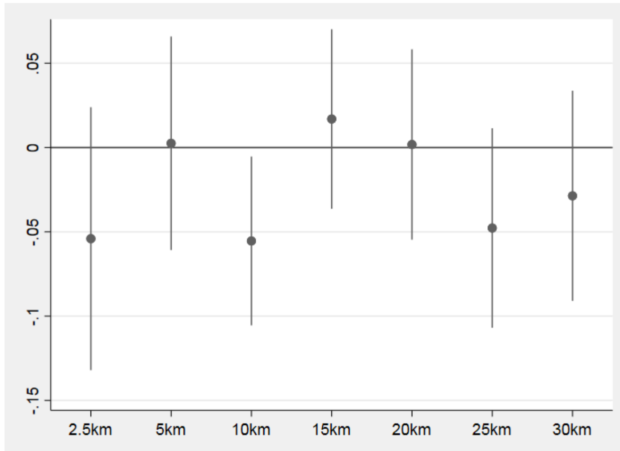
5-10km	-0.0554**	0.0535	27.7800***
	(0.0256)	(0.0374)	(8.2708)
10-15km	0.0169	0.0012	19.1527**
	(0.0272)	(0.0391)	(8.9924)
15-20km	0.0018	0.0844**	48.4971***
	(0.0288)	(0.0409)	(8.8228)
20-25km	-0.0477	0.2171***	29.8973***
	(0.0302)	(0.0463)	(9.0739)
25-30km	-0.0287	0.1097**	32.1216***
	(0.0318)	(0.0501)	(8.7727)
	Coastal ×Distance		
1-2.5km	0.0175	-0.1901***	-6.3073
	(0.0408)	(0.0542)	(17.3371)
2.5-5km	-0.0682**	-0.0692	8.4239
	(0.0345)	(0.0511)	(13.2414)
5-10km	-0.0612**	-0.0221	-15.5261
	(0.0279)	(0.0402)	(9.5890)
10-15km	-0.0328	0.1049***	15.0894
	(0.0292)	(0.0405)	(10.6083)
15-20km	-0.0631**	0.2275***	58.7162**
	(0.0313)	(0.0408)	(24.4602)
20-25km	-0.0359	0.0093	4.9850
	(0.0318)	(0.0449)	(9.3596)
25-30km	-0.0454	0.1115**	52.1040**
	(0.0330)	(0.0443)	(22.1852)
Pop. (k)	0.0001***	0.0001***	-0.0063

	(0.0000)	(0.0000)	(0.0071)
Inc. (m)	0.0000*	0.0000***	-0.0000
	(0.0000)	(0.0000)	(0.0001)
Constant	0.8351***	-36.3235***	-157.6376***
	(0.0365)	(0.1523)	(57.8772)
p		8.9777***	
		(0.0000)	
rho		-0.2500***	-0.2376**
		(0.0000)	(0.0261)
Str	Yes	Yes	Yes
Year FE	No	No	Yes
State FE	Yes	Yes	Yes
Obs.	194,470	359,045	89,502

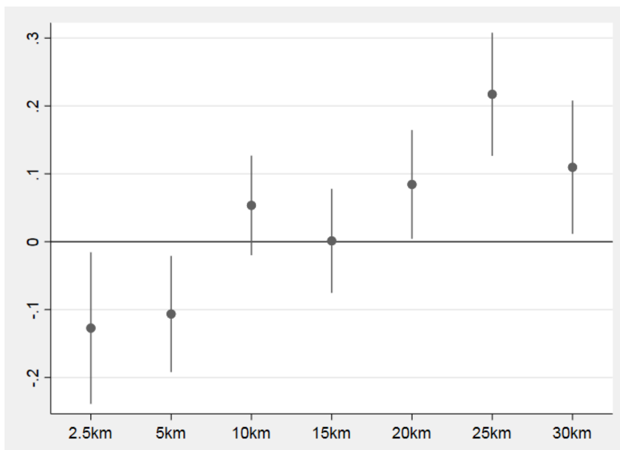
Note: Numbers in parentheses are robust standard errors.

Figure 8: Policy Behaviors on Interstate Rivers

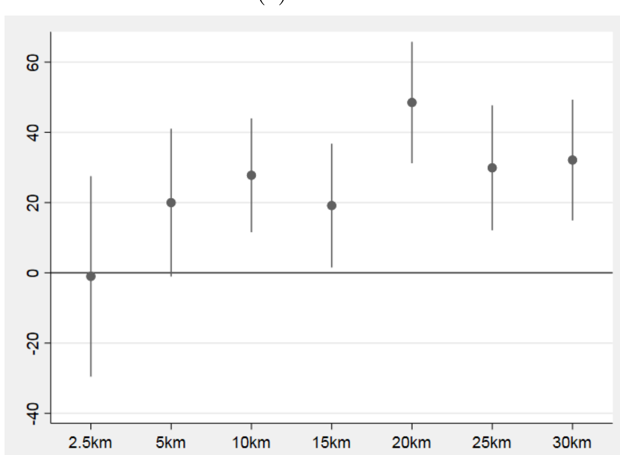
(a) Prob. of Assessment



(b) Hazard of TMDL Development



(c) Grants



Column (2) in Table 5 shows the states' behaviors in S303(d) TMDL development. Interstate catchments within 1-5 km of the state borders have a lower hazard to receive TMDLs than the intrastate catchments, with -12.73% between 1-2.5 km and -10.65% between 2.5-5 km. Panel (b) in Figure 8 shows that the hazards of TMDL development increase as the distance bins become further away from the state borders.

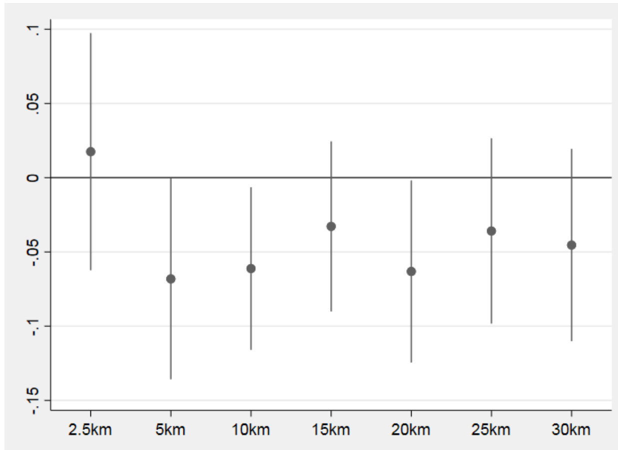
Column (3) in Table 5 shows the results for S319(h) grants. Most estimation results for interstate subwatersheds are positive, indicating more funding granted than intrastate subwatersheds. Panel (c) in Figure 8 shows that the estimation results are insignificant in the 1-2.5 km distance bin, then become larger and statistically significant in bins further from state borders.

The above results show that the states have a significantly lower hazard to develop TMDLs on interstate rivers near state borders. On the contrary, more S319(h) grants are approved for interstate rivers. These results are consistent with our expectation of different behaviors across the policies by their different levels of decentralization.

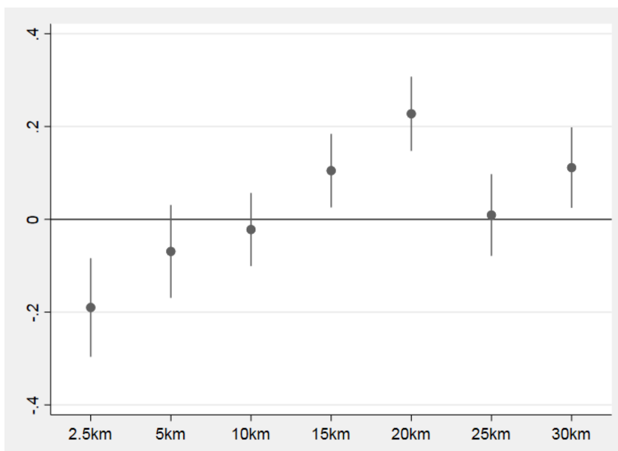
Table 5 and Figure 9 present the estimation results for coastal rivers. The coastal catchments in most of the distance bins within 30 km of the state borders have a lower probability of assessment than intrastate catchments. The states have significantly lower hazards to develop TMDLs for the coastal catchments within 1-2.5 km of the shoreline. The hazard increases as the coastal catchment's distance from the shoreline becomes further. S319(h) grants approved for coastal subwatersheds fluctuate across the distance bins, with insignificant estimates near the shoreline and more positive estimates as the subwatersheds become further away from the shoreline.

Figure 9: Policy Behaviors on Coastal Rivers

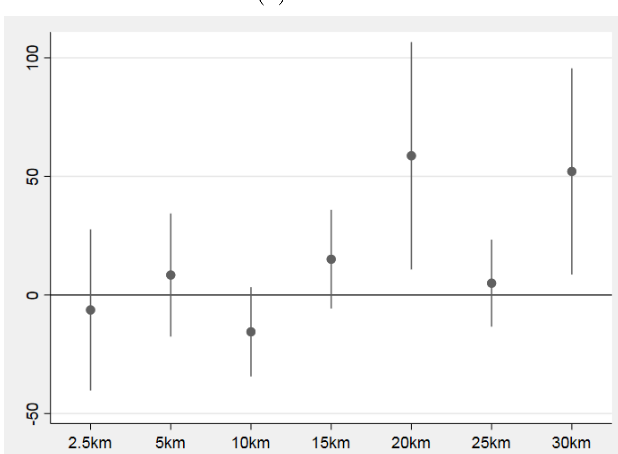
(a) Prob. of Assessment



(b) Hazard of TMDL Development



(c) Grants



6.3 The Beggar-thy-Neighbor and Free-riding Behaviors

I divide interstate rivers into upstream and downstream portions to study the states' beggar-thy-neighbor and free-riding behaviors. The estimation results are presented in Table 6, Figure 10, and Figure 11. The estimation results for the downstream portions of interstate rivers have larger standard errors than the upstream portions. This is because there are much fewer downstream portions of interstate rivers in the sample than the upstream portions, as is shown in Table 3 and Figure 3. The three policies exhibit different degrees of policy behaviors.

Table 6: Estimation Results: Different Portions of Interstate Rivers

	(1) Assessment	(2) TMDL	(3) Grants
0-1km (Border)	-0.4270***	-0.0386	19.3400***
	(0.0196)	(0.0297)	(5.2929)
	Near/Non-trans ×Distance		
1-2.5km	-0.0750***	-0.0254	25.4982***
	(0.0275)	(0.0366)	(7.8155)
2.5-5km	-0.0049	0.0759**	5.0651
	(0.0251)	(0.0322)	(6.5122)
5-10km	-0.0021	-0.0985***	12.2901**
	(0.0219)	(0.0312)	(5.6181)
10-15km	-0.0413*	-0.1056***	10.8650*
	(0.0234)	(0.0323)	(6.5345)
15-20km	-0.0343	-0.1012***	-24.5083***

	(0.0242)	(0.0323)	(6.7593)
20-25km	-0.0372	-0.2765***	-8.8900
	(0.0250)	(0.0355)	(6.7492)
25-30km	-0.0318	-0.2647***	-5.0710
	(0.0259)	(0.0366)	(6.7966)
	Upstream ×Distance		
1-2.5km	-0.0698	-0.1890**	-27.2125
	(0.0435)	(0.0798)	(16.7792)
2.5-5km	-0.0966***	-0.1028**	17.6531
	(0.0337)	(0.0505)	(10.9128)
5-10km	-0.1115***	0.0472	4.6954
	(0.0261)	(0.0396)	(7.2291)
10-15km	-0.0276	-0.0303	2.0996
	(0.0275)	(0.0409)	(8.1460)
15-20km	-0.0314	0.0765*	43.6373***
	(0.0291)	(0.0420)	(8.5279)
20-25km	-0.0630**	0.2141***	26.3316***
	(0.0305)	(0.0472)	(8.3015)
25-30km	-0.0432	0.1209**	18.6261**
	(0.0321)	(0.0512)	(8.1982)
	Downstream ×Distance		
1-2.5km	-0.4069***	-0.5581**	0.6925
	(0.1188)	(0.2348)	(48.8711)
2.5-5km	0.0482	0.1628	-18.8042
	(0.0905)	(0.1353)	(20.5534)
5-10km	0.0909	-0.2571**	40.2383***

	(0.0753)	(0.1220)	(15.6058)
10-15km	0.1206	-0.1066	-3.4005
	(0.0865)	(0.1168)	(12.5784)
15-20km	0.1097	-0.2114	24.5957
	(0.1045)	(0.1431)	(15.8137)
20-25km	0.0657	0.2971*	34.5460**
	(0.1067)	(0.1609)	(16.9658)
25-30km	0.1910	-0.2853	22.7175
	(0.1338)	(0.1946)	(17.1189)
	Up/Downstream ×Distance		
1-2.5km	-0.0120	-0.4505***	-2.2313
	(0.1053)	(0.1372)	(21.1495)
2.5-5km	0.0559	-0.3318***	-4.8435
	(0.0825)	(0.0917)	(15.0475)
5-10km	0.1659**	0.1690**	19.5644
	(0.0735)	(0.0708)	(17.7545)
10-15km	0.1046	0.2376***	-30.6352***
	(0.0962)	(0.0818)	(11.2364)
15-20km	0.1939	0.3076***	-14.2043
	(0.1221)	(0.1185)	(19.8327)
20-25km	-0.0523	0.1321	39.2648
	(0.1322)	(0.1561)	(26.1891)
25-30km	0.1417	0.0678	55.5285*
	(0.1496)	(0.1747)	(30.8769)
	Coastal ×Distance		
1-2.5km	0.0056	-0.1921***	-12.6686

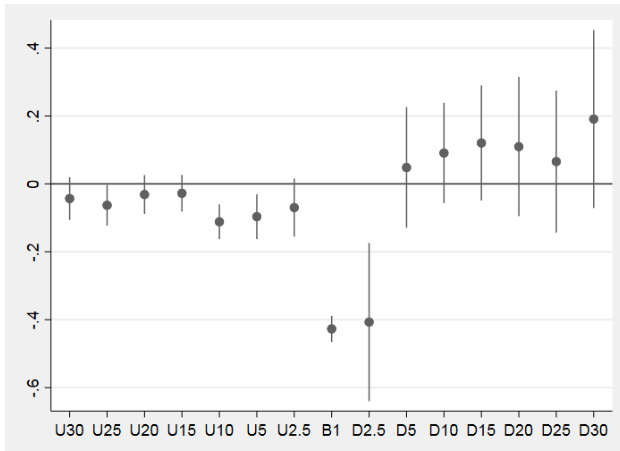
	(0.0405)	(0.0539)	(17.2225)
2.5-5km	-0.0875**	-0.0708	0.2333
	(0.0344)	(0.0509)	(12.8879)
5-10km	-0.0756***	-0.0257	-21.6251**
	(0.0278)	(0.0402)	(9.3146)
10-15km	-0.0476	0.1015**	7.3941
	(0.0292)	(0.0404)	(10.3165)
15-20km	-0.0740**	0.2248***	53.2533**
	(0.0313)	(0.0408)	(24.1073)
20-25km	-0.0439	0.0076	1.0212
	(0.0318)	(0.0449)	(9.2705)
25-30km	-0.0510	0.1104**	48.9589**
	(0.0330)	(0.0442)	(22.0969)
Pop. (k)	0.0001***	0.0001***	-0.0046
	(0.0000)	(0.0000)	(0.0071)
Inc. (m)	0.0000**	0.0000***	-0.0000
	(0.0000)	(0.0000)	(0.0001)
Constant	0.8356***	-36.3287***	-156.9401***
	(0.0365)	(0.1523)	(57.6267)
p		8.9797***	
		(0.0000)	
rho		-0.2500***	-0.2374**
		(0.0000)	(0.0261)
Str	Yes	Yes	Yes
Year FE	No	No	Yes
State FE	Yes	Yes	Yes

Obs.	194,470	359,045	89,502

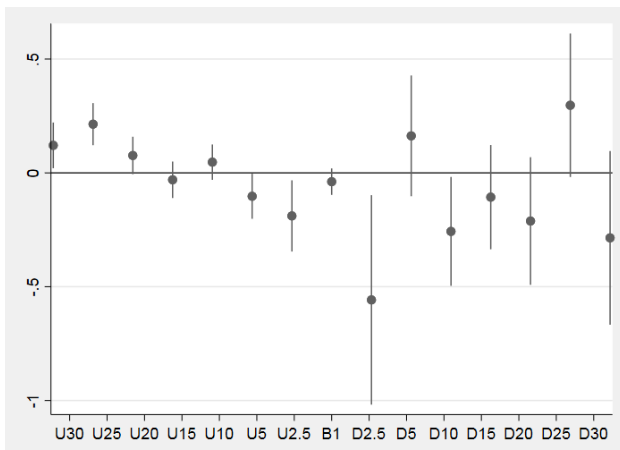
Note: Numbers in parentheses are robust standard errors.

Figure 10: The Beggar-thy-neighbor Behavior Upstream and Free-riding Behavior Downstream of Interstate Rivers

(a) Prob. of Assessment



(b) Hazard of TMDL Development



(c) Grants

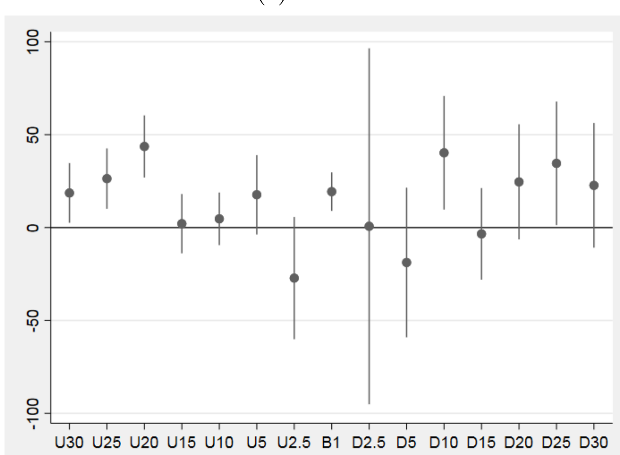
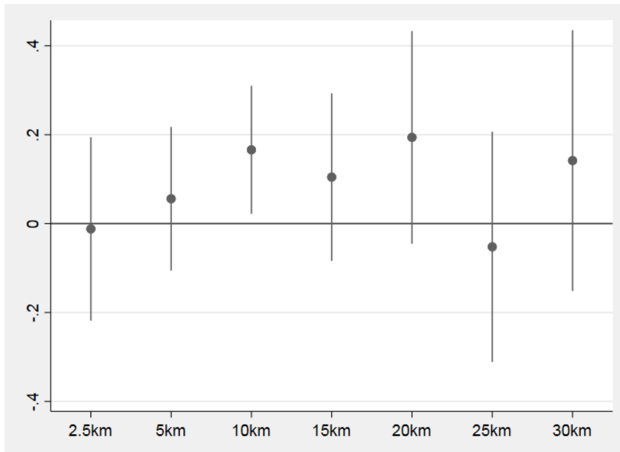
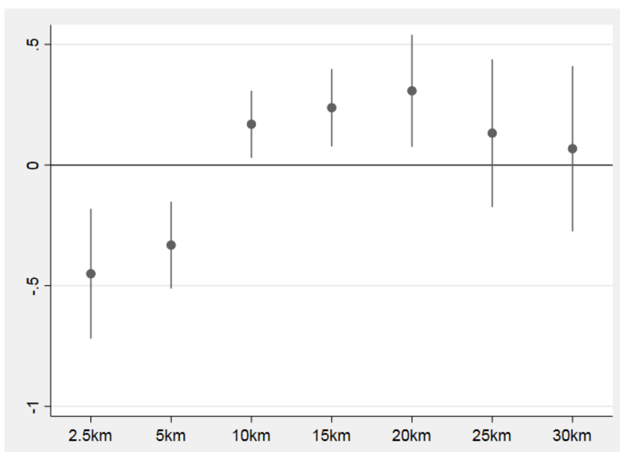


Figure 11: The Beggar-thy-neighbor Behavior and Free-riding Behavior in the Up/Downstream of Interstate Rivers

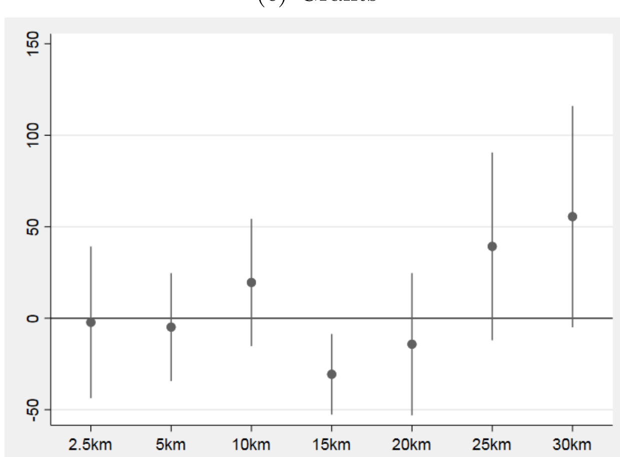
(a) Prob. of Assessment



(b) Hazard of TMDL Development



(c) Grants



Column (1) of Table 6 shows the estimation results for the probability of S305(b) water quality assessment. The coefficients for the upstream catchments fluctuate, but all the coefficients are negative, indicating a lower probability of assessment. The statistically significant results indicate that the upstream catchments are 9.66-11.15% less likely to be assessed than intrastate catchments within 2.5-10 km of state borders and are 6.3% less likely to be assessed within 20-25 km of state borders. The coefficients for downstream catchments also fluctuate but are mostly insignificant, indicating a similar probability of assessment with intrastate catchments, except for catchments within 2.5 km downstream of state borders, which are 40.69% significantly less likely to be assessed. Catchments that are both upstream and downstream mostly have insignificant estimation results. The coefficients are plotted in Panel (a) of Figure 10 and Figure 11.

When implementing S305(b), the state governments exhibit beggar-thy-neighbor behaviors across a long distance from state borders, but the free-riding behavior is not clear. The lower probability of water quality assessment in the 1-2.5 km bin downstream could be a spurious estimate. This is because the estimation result in the nearby 2.5-5 km bin is close to zero and insignificant, as are all the other estimates in the downstream portions. Possible explanations for the single-side policy behavior are that the upstream state governments are less motivated to assess the quality of water bodies that are going to flow away from their territory, so they exhibit beggar-thy-neighbor behavior in some distance bins. But since the assessment cost is low, the behavior is also statistically insignificant in many distance bins. For the downstream portions of interstate rivers, because they continue to flow in the downstream states and the assessment costs are low, the downstream states assess the downstream portions of interstate rivers with a similar probability to their intrastate rivers. Most importantly, even if the upstream state assesses the upstream water quality, it will not provide much free-rider benefit to the downstream state. The downstream state is not able to make accurate inferences about the downstream water quality based on the upstream assessment results, especially when the catchment is far away from state borders. Therefore, the downstream states have a low incentive to free ride in S305(b) water quality assessment.

Column (2) of Table 6 presents the estimation results for S303(d) TMDL development. Panel (b) in Figure 10 shows that in most distance bins, the closer the upstream catchment to state borders the lower the hazards of TMDL development. States have an 18.9% lower hazard to develop TMDLs for the upstream catchments within 1-2.5 km of the state borders and a 10.28% lower hazard for catchments within 2.5-5 km of the state borders. This is the beggar-thy-neighbor behavior. Compared with the upstream catchments, the estimation results for the downstream catchments are larger and mostly negative, indicating lower hazards of TMDL development than intrastate catchments. Though the standard errors are larger due to the small sample size of the downstream portions of interstate rivers, the few statistically significant estimation results are more negative as the catchment becomes closer to state borders, indicating stronger free-riding behavior. Specifically, the downstream catchments have a 55.81% significantly lower hazard of TMDL development in the 1-2.5 km distance bin and a 25.71% significantly lower hazard in the 5-10 km distance bin. Catchments that are both upstream and downstream also have a lower hazard of TMDL development within 5 km of the state borders, with -45.05% in the 1-2.5 km bin and -33.18% in the 2.5-5 km bin, as shown in panel (b) of Figure 11. This is the mixed result of beggar-thy-neighbor and free-riding behaviors. The hazard of TMDL development has increasing trends in the upstream, downstream, and up/downstream portions of interstate rivers as the catchments' distance becomes further away from state borders. This implies that the transboundary behaviors diminish with the rivers' distance from state borders.

The state governments exhibit both the beggar-thy-neighbor behavior and free-riding behavior in S303(d). The magnitudes of the coefficients indicate that the downstream free-riding behavior is stronger than the upstream beggar-thy-neighbor behavior in the U.S. This is one of the situations discussed in Section 3, indicating that states could recover the majority of the benefits of developing TMDLs for interstate watersheds, or the downstream state has a high expectation that the upstream state will develop TMDLs for the interstate watersheds. However, the back-of-the-envelope calculation in Section 6.4 shows that the downstream free-riding behavior does not incur larger losses than the upstream beggar-thy-neighbor behavior. This is because the number of upstream tributaries of interstate rivers substantially exceeds the downstream tributaries.

Column (3) of Table 6 and panel (c) of Figure 10 and Figure 11 present the estimation results for S319(h) grants. The estimated coefficients for the upstream, downstream, and up/downstream portions of interstate rivers fluctuate substantially, and most estimation results are statistically insignificant. Thus, different portions of interstate rivers in most distance bins do not receive significantly different amounts of grants than intrastate rivers. Interstate subwatersheds in some distance bins receive a larger amount of grants than intrastate subwatersheds.

The S319(h) grant amounts do not reflect the state governments' beggar-thy-neighbor or free-riding behaviors. This is because the federal government has high authority in S319(h). Section 6.1 and Section 6.2 show that larger amounts of grants are approved for near-boundary rivers and some portions of interstate rivers and coastal rivers. These results imply that the near-boundary rivers and some portions of interstate rivers need more NPS pollution controls to maximize the national welfare. However, the probabilities of S305(b) assessment, especially the hazards of S303(d) TMDL development, are the opposite, which indicates that the state-level policy decision conflicts with the federal-level decision, leading to insufficient pollution controls for near-boundary and transboundary rivers.

6.4 The Costs of Beggar-thy-Neighbor and Free-riding Behaviors

Section 6.3 shows that states exhibit beggar-thy-neighbor behavior and free-riding behavior in TMDL development. Using the parameter estimates in Section 6.3, I conduct back-of-the-envelope calculations for the deadweight losses incurred in S303(d) TMDL development by each behavior. I use people's willingness to pay (WTP) for clean water in the literature to infer the unrecovered net benefits of pollution control. For example, Hite, Hudson, and Intarapapong (2002) conducted a survey in Mississippi, 1999, and estimates that people's WTP for agricultural NPS pollution abatement is \$46.97 for 10% abatement and \$49.94 for 20% abatement. Jordan and Elnagheeb (1993)'s 1991 survey in Georgia estimates that people are willing to pay \$5.49 more in their monthly water bill for treat water. Chatterjee et al. (2017)'s 2016 survey in Florida estimates a WTP of \$6.22 in the monthly water bill. Using these WTPs and the total population of counties that have interstate rivers within each distance bin that are undertreated, I infer the magnitude of the deadweight loss caused by each behavior as is indicated in Figure 1 and Figure 2.

Table 7 presents the estimated deadweight loss. The states' beggar-thy-neighbor (BTN) behavior in TMDL development has resulted in a \$786 million loss if the TMDL development could have led to 10% agricultural NPS pollution abatement, or a \$836 million loss with a 20% abatement, or a \$73 to \$113 million loss in the drinking water quality improvement each month. The free-riding (FR) behavior has led to a \$766 or \$814 million loss if the TMDL development could have led to a 10% or 20% agricultural NPS pollution abatement, or a \$71 to \$110 million loss in the monthly drinking water quality improvement. Catchments in both the upstream and downstream of interstate rivers are affected by both behaviors. The deadweight loss incurred in these catchments is \$965 or \$1026 million for 10% or 20% agricultural NPS pollution abatement or \$89 to \$138 million in the monthly water bill. This is a raw calculation because the true benefits of unconducted pollution management vary across time and locations. In addition, Keiser, Kling, and Shapiro (2019) and Keiser and

Shapiro (2019) state that many studies on the benefit of water pollution abatement have ignored the nouse values and health benefits, so the unrecovered benefits could be even larger.

Table 7: Deadweight Loss due to Insufficient TMDL Development

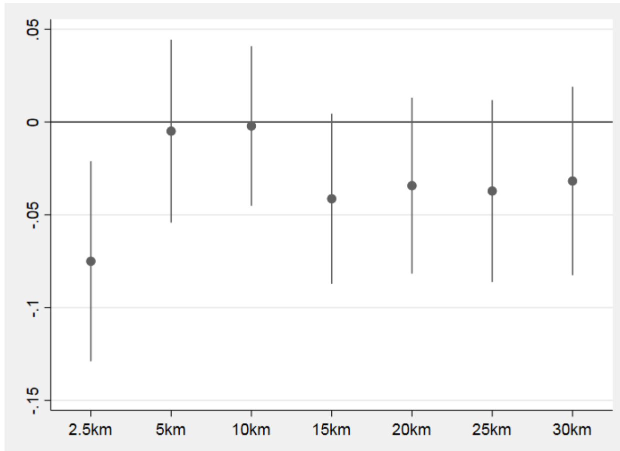
				deadweight Loss (mil. $ in 2020)			
	Dist.	Hazard	Pop. (m)	H. (2002) 10%	H. (2002) 20%	C. (2017)	J. (1993)
	1-2.5km	0.1890	34.5557	475.4811	505.5466	43.8728	68.1251
BTN	2.5-5km	0.1028	41.5063	310.6417	330.2841	28.6630	44.5075
	Total			786.1228	835.8308	72.5358	112.6326
	1-2.5km	0.5581	11.5362	468.7330	498.3719	43.2501	67.1582
FR	5-10km	0.2571	15.8833	297.3009	316.0997	27.4320	42.5961
	Total			766.0339	814.4716	70.6822	109.7543
	1-2.5km	0.4505	12.6005	413.2709	439.4028	38.1326	59.2118
Both	2.5-5km	0.3318	22.8543	552.0731	586.9817	50.9399	79.0989
	Total			965.3440	1026.3840	89.0726	138.3107

6.5 Policy Behaviors on Near-boundary but Non-transboundary Rivers

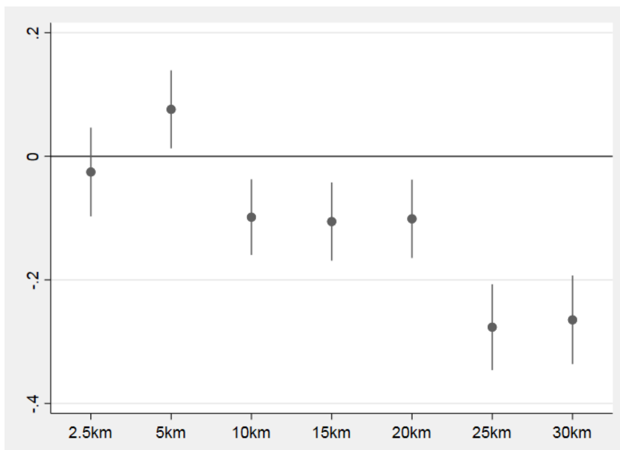
Table 5 and Table 6 also present the coefficients for catchments and subwatersheds that are near-boundary but not transboundary. Though the estimation results fluctuate and are not the same in the two regressions, they share similar trends. Figure 12 plots the coefficients in Table 6. The probabilities of S305(b) water quality assessment and S303(d) TMDL development are mostly lower for the non-transboundary catchments than intrastate catchments. The S319(h) for non-transboundary subwatersheds has positive and significant results at least within 2.5 km of state borders. The estimation results for non-transboundary rivers have similar variation patterns to the results for all near-boundary rivers in Figure 7, which are different from the variation patterns of interstate rivers. These results imply that the state governments exhibit different policy behaviors on interstate rivers than near-boundary rivers that do not cross borders.

Figure 12: Policy Behaviors on Near-boundary but Non-transboundary Rivers

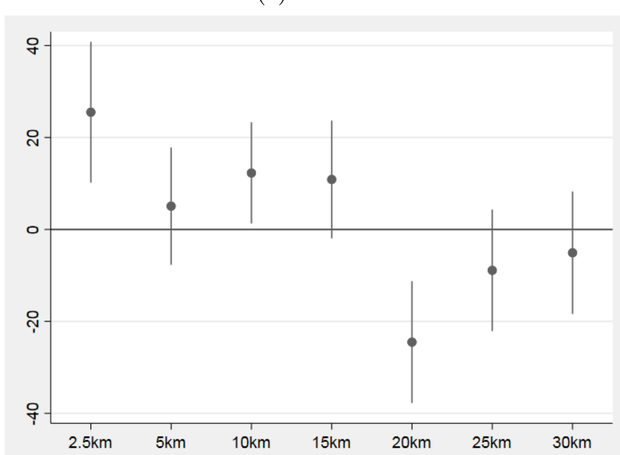
(a) Prob. of Assessment



(b) Hazard of TMDL Development



(c) Grants

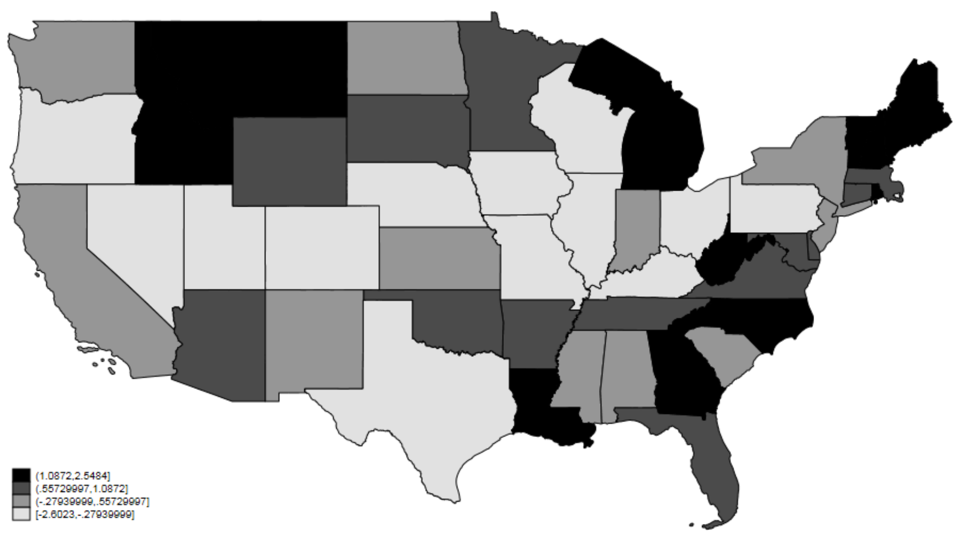


6.6 Other Parameters of Interest

The estimation results for county population and personal income in Table 4, Table 5, and Table 6 are all statistically significant, positive, and small for S305(b) water quality assessment and S303(d) TMDL development, but insignificant for grants approval. This implies that states take the socio-economic background of the catchment into consideration when making decisions on S305(b) and S303(d). More populated and richer catchments have slightly higher probabilities of assessment and TMDL development. But S319(h) grant approval is not significantly affected by these factors.

Since states have the highest level of authority in S303(d) among the three policies, I plot the coefficients of the state-fixed effect for TMDL development in Table 5. Figure 13 shows that the northern and southeastern states have higher hazards to develop TMDLs.

Figure 13: State-fixed Effects on Hazards of TMDL Development



The shape parameter p of the duration model for S303(d) TMDL development is higher than 8.97 in all three regressions (Equation 12, Equation 13, and Equation 14), which implies that the hazard that state develops a TMDL increases dramatically with time. For example, after 48 years of the establishment of the CWA, in 2020, a catchment is 42.34833 times more likely to receive TMDLs per year than in 2002, i.e. $(\frac{48}{30})^{8.97-1}$.

The error correlation ρ between the first and second-step regressions in the duration model with selection for S303(d) TMDL development is -0.25 in all three regressions. This has reached the limit restriction set by Boehmke, Morey, and Shannon (2006), which indicates strong correlations between the first and second-step regressions and that ignoring sample selection leads to biased estimates.[20] The negative error correlation suggests that if the probability that a catchment is listed as impaired is greater than average, the state's expected probability to develop TMDLs for the catchment is less than average. One possible explanation is: A watershed can be listed as impaired due to many factors, and the lack of proper pollution management is one of them. Since S303(d) TMDL development is also an NPS pollution management policy, the state may not develop TMDL timely for the watershed either. The error correlation ρ in the Heckman selection model for S319(h) grants is around -0.24 in all three regressions. The negative error correlation suggests that subwatersheds that are more likely to enter S319(h) programs do not necessarily receive a larger amount of grants.

7 Extensions

7.1 Political Influence on Policy Behavior

The above analysis shows that states exhibit near-boundary and transboundary behaviors in S303(d) TMDL development. A state's policy behaviors are affected by the state's political ideology. Therefore, in this section, I explore the potential effects of political ideology on states' behavior in TMDL development. The first variable I use is the state governor's party. I add two control variables for the political party of the state governor in the year that the

[20] Section A in the Appendix investigates the selection bias by running an independent duration model.

TMDL is developed and the year before in Equation 14. The political party dummy is 0 if the governor is a Democrat, and 1 if the governor is a Republican.[21] The coefficients of the political party dummy are presented in Table 8. A state with a Republican governor in the current year has a 46.54% lower hazard to develop TMDLs for a catchment than a state with a Democratic governor. A state with a Republican governor in the previous year has a 94.02% lower hazard.

Table 8: The Effect of the State Governor's Political Party on the Hazard of TMDL Development

	Proportional Hazard
Republican Governer in Current Year	-0.4654***
	(0.1019)
Republican Governer in Previous Year	-0.9402***
	(0.1062)

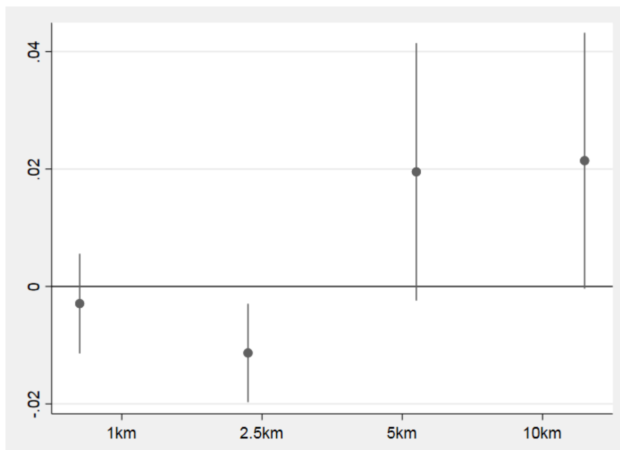
The second factor I check is the League of Conservation Voters (LCV) score. The LCV score records the members of Congress' voting on environmental issues. The average lifetime LCV score is 70.8 and 76.1 for the Democratic senate and house, respectively, and 22.5 and 22 for the Republican senate and house. This is consistent with the results above that the Democratic are more environmentally friendly than the Republicans, so a state with a democratic governor has a higher probability to develop TMDLs. I check if a more environmentally-friendly state would have milder policy behavior, including insufficient policy implementation at near-boundary rivers, and the beggar-thy-neighbor and free-riding behaviors at interstate rivers. I run Equation 14 for each state to get state-specific coefficient estimates. Then I regress the coefficient estimates on the state Congress's average LCV score in the year of TMDL development. Based on the results from previous sections,

[21]The situation that the current governor is independent in the year of or the year before the TMDL development is rare so the independent party is dropped.

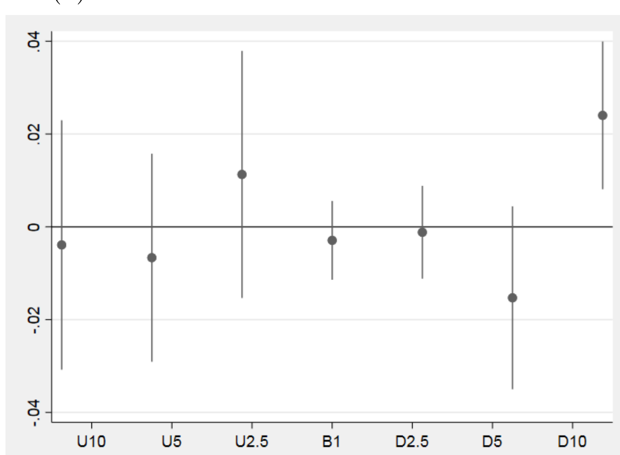
the beggar-thy-neighbor and free-riding behaviors mostly occur within 10 km of the state borders, so I only present the influence of LCV scores on the policy behaviors within 10 km of the state borders. The results are plotted in Figure 14. The magnitudes of the estimated LCV influences are very small and most results are statistically insignificant at a 95% confidence level, indicating that a state's LCV score has little effect on its near-boundary and transboundary policy behaviors.

Figure 14: Influence of State LCV Score on Policy Behaviors

(a) Distance Bins



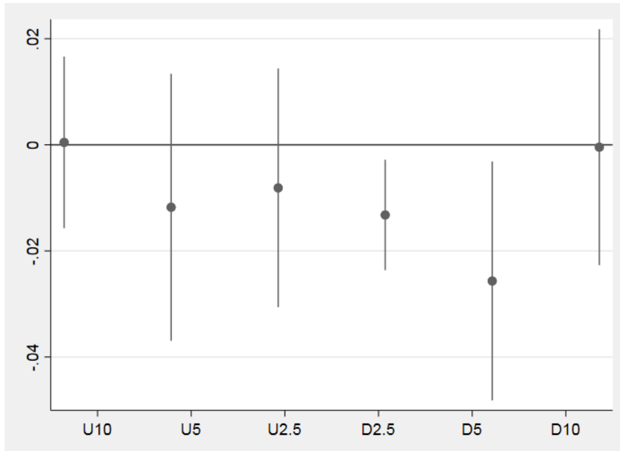
(b) Different Portions of the Interstate Rivers



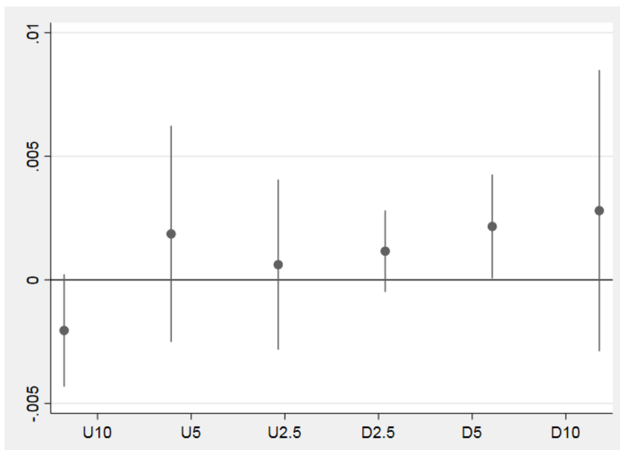
The interstate rivers connect the state with its neighboring states. Therefore, in addition to the state's own influence, the neighboring state's political-economic features may also affect the state's behavior in the pollution management of interstate rivers. First, I run Equation 14 for each state separately. Then I regress each state's estimated coefficients on the features of its neighboring states. The results are presented in Table 9 and Figure 15. Panel (a) in Figure 15 shows that the downstream state's LCV score does not affect the upstream state's beggar-thy-neighbor behavior. However, the higher the upstream state's LCV score, the stronger the downstream state's free-riding behavior. 1 unit increase in the upstream state's LCV score will lower the proportional change of the downstream state's TMDL development hazard in the downstream catchments by 1.32% in the 1-2.5 km bin and by 2.57% in the 2.5-5 km bin. This is consistent with the theoretical model in Section 3.2: The degree of free-riding behavior in the downstream state depends on their expectation γ_{du} of the upstream state's pollution management. The higher the LCV score in the upstream state, the larger the expectation γ_{du} that the downstream state has for the upstream state's pollution management on interstate rivers, thus the stronger the free-riding behavior in the downstream state.

Figure 15: Influence on Policy Behaviors from the Neighboring State

(a) LCV



(b) GDP



(c) Population

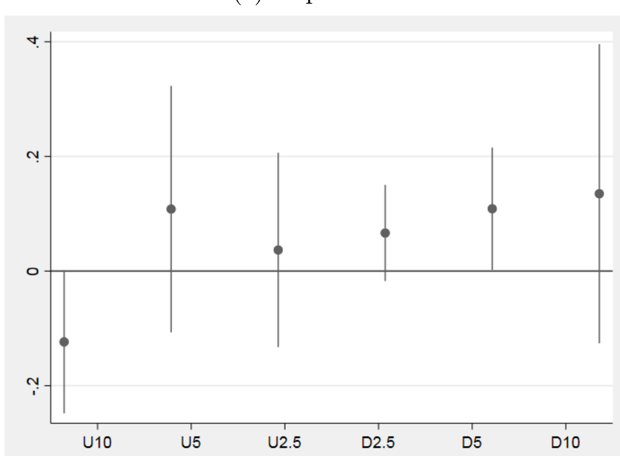


Table 9: Influence on Policy Behaviors from the Upstream/Downstream State

	Distance to State Borders		
	1-2.5km	2.5-5km	5-10km
Dn. State	BTN Upstream		
LCV	-0.0081	-0.0118	0.0005
	(0.0114)	(0.0127)	(0.0082)
GDP (b.)	0.0006	0.0019	-0.0020*
	(0.0017)	(0.0022)	(0.0012)
Pop. (m.)	0.0368	0.1081	-0.1236*
	(0.0859)	(0.1089)	(0.0631)
Up. State	FR Downstream		
LCV	-0.0132**	-0.0257**	-0.0005
	(0.0053)	(0.0114)	(0.0113)
GDP (b.)	0.0012	0.0022**	0.0028
	(0.0008)	(0.0011)	(0.0029)
Pop. (m.)	0.0664	0.1088**	0.1349
	(0.0426)	(0.0540)	(0.1322)

Panel (b) and panel (c) show the effect of a neighboring state's GDP and population on the state's policy behavior. The downstream state's effect on the upstream state's behavior is only negative and statistically significant within the 5-10 km distance bin to state borders, but the other two distance bins that are closer to state borders both have positive and insignificant results. Therefore, it is hard to infer a causal relationship between the downstream state's characteristics and the upstream state's behavior. For the downstream state, a 1 billion increase in the upstream state GDP or a 1 million increase in the upstream state population will increase the proportional change of the downstream state's TMDL development hazard for downstream catchments within the 2.5-5 km bin by 0.22% and 10.88%, respectively. The estimation results in the other two distance bins are insignificant but also positive, indicating that the upstream state's GDP and population slightly affect the downstream state's policy behavior. These results imply that the upstream state's beggar-thy-neighbor behavior is not affected by the downstream state's political-economic scale. A political-economically “large” downstream state does not have higher bargaining power over the water pollution management of the upstream state. While a “large” upstream state with high GDP and a huge population slightly incentivizes the downstream state to manage pollution more and free-riding less. This could result from higher bargaining power of the upstream state at the transboundary rivers. Or, a larger upstream state signals more pollution flows from the upstream, so the downstream state needs to take more actions to treat the downstream watersheds.

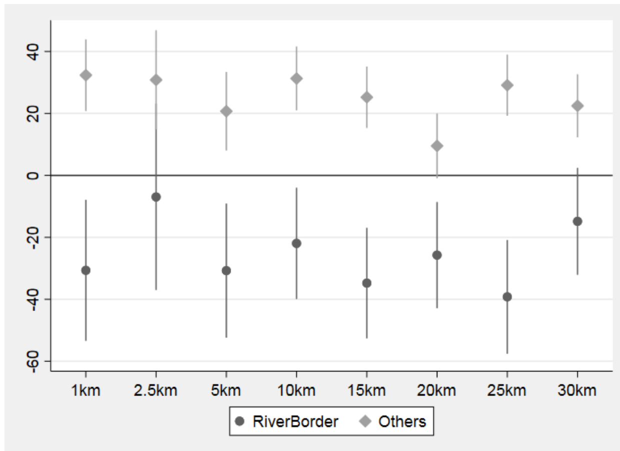
The above results show that the upstream state’s beggar-thy-neighbor behavior is relatively independent. The downstream state’s free-riding behavior depends on the downstream state’s expectation for the upstream state’s pollution management, which is built on the upstream state’s political-economic features. A more environmentally friendly or political-economically “smaller” upstream state will intensify the downstream state’s free-riding behavior.

7.2 Grants Approval in States with River Borders

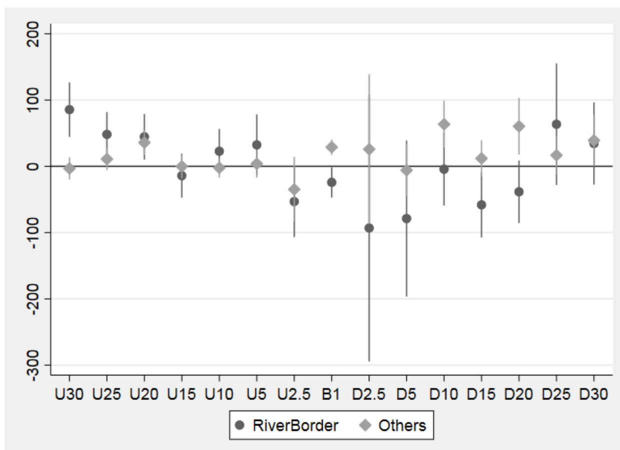
Table 4 shows that near-boundary subwatersheds receive a larger amount of grants than intrastate subwatersheds. Figure 6 and Figure 39 suggest that more grants may have been approved for rivers that form the state borders. Figure 3 also shows that the distribution of the upstream and the downstream portions of interstate rivers are very different for rivers that form the state borders and rivers that only cross the state borders. Most of the tributaries of the border rivers are upstream of the river. Therefore, I run the Heckman selection model of approved grants for the 10 states that have river borders and the rest states separately to compare their coefficients. The results are presented in Figure 16.

Figure 16: Comparison of Grants Approval for States with River Borders and the Other States

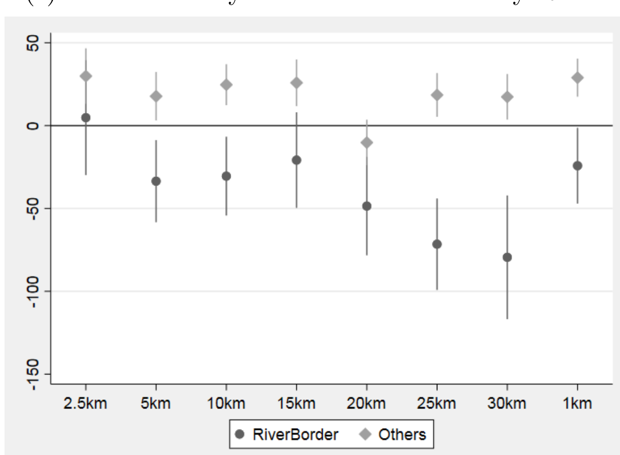
(a) Distance Bins



(b) Different Portions of Interstate Rivers



(c) Near-boundary but Non-transboundary Rivers



Panel (a) of Figure 16 shows that in the states with river borders, the near-boundary rivers receive fewer grants than intrastate rivers. While in the states that do not have river borders, the near-boundary rivers receive more grants. Panel (b) for different portions of interstate rivers show that in the river-border states, the upstream portions of interstate rivers mostly receive larger amounts of grants than intrastate rivers. The downstream portions of interstate rivers mostly receive fewer grants than intrastate rivers. The upstream and downstream portions of interstate rivers in the non-river-border states do not receive significantly different grants from the intrastate rivers in most distance bins, with the few distance bins in the downstream portions receiving larger amounts of grants. Panel (c) presents the coefficients for non-transboundary rivers. Non-transboundary rivers receive fewer amounts of grants than intrastate rivers in river-border states but receive larger amounts of grants in non-river-border states. The variation patterns of the coefficients for non-transboundary rivers are similar to those in Panel (a) for all near-boundary rivers.

In conclusion, in states that do not have river borders, rivers that lie close to the state borders but do not cross the state borders receive relatively more grants; In states with river borders, the upstream portions of interstate rivers receive relatively more grants. These are the two driving factors of the larger amounts of grants for near-boundary and transboundary rivers estimated in Section 13.

7.3 The Probability of Impairment

I conduct a Heckman two-step analysis to estimate the probability that a catchment is listed as impaired. The impairment is one outcome of water quality assessment, it is not a decision made by the government like the three policies discussed above. However, being listed as impaired requires that the catchment is assessed. So I run a first-step selection of the assessed catchment then run the second-step Probit model for the probability of impairment.[22]

The estimated coefficients for near-boundary, interstate, and non-transboundary catchments are plotted in Figure 17. The near-boundary catchments are more likely to be listed as impaired. The interstate catchments, however, have a lower probability to be impaired when they are close to the state borders. The high probability of impairment for near-boundary catchments is mainly driven by the near-boundary but non-transboundary catchments. These results are consistent with the results in Section 13, which show that the near-boundary catchments receive a larger amount of S319(h) pollution management grants, while different portions of interstate rivers in most distance bins do not receive significantly different amount of grants than intrastate rivers.

[22] First step:

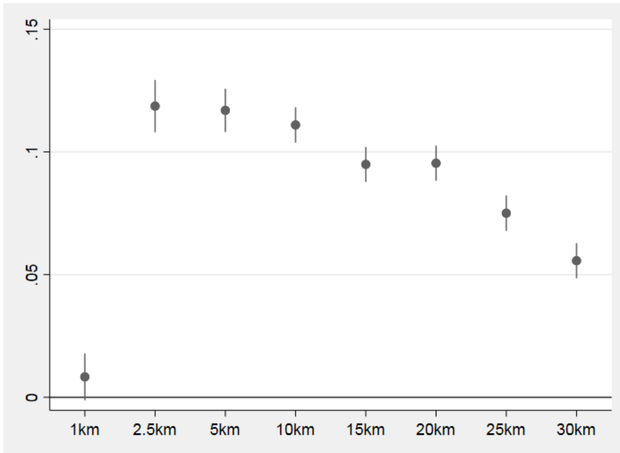
$$p_{cs} = Pr[\ Assess_{cs} = 1|\mathbb{X}_{cs}] = \Phi(\alpha_0 + Distbin_c\beta_1 + Str_c\alpha_1 + X_c\alpha_2 + D_s) \tag{20}$$

Second step:

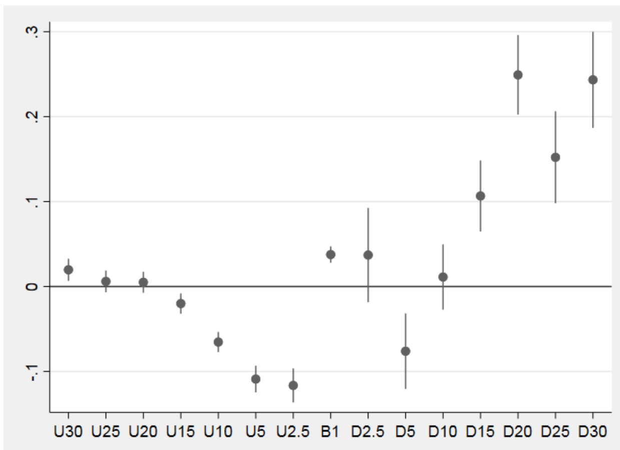
$$p_{cs} = Pr[\ Impair_{cs} = 1|\mathbb{X}_{cs}] = \Phi(\alpha_0 + Distbin_c\beta_1 + \beta_2 CoastCat_c + Str_c\alpha_1 + Str1_c\alpha_2 + X_c\alpha_3) \tag{21}$$

Figure 17: The Probability of Being Listed as Impaired

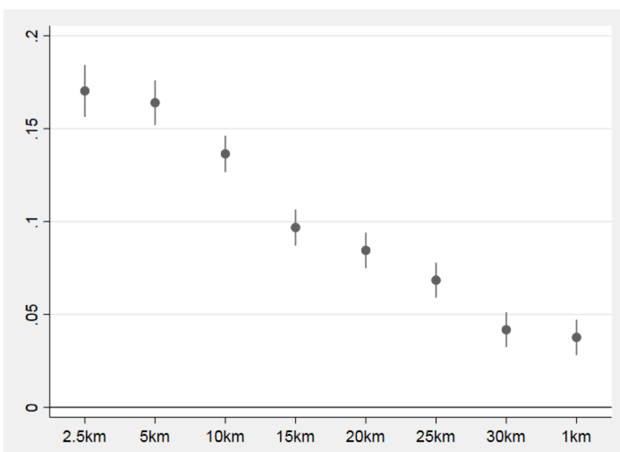
(a) Distance Bins



(b) Different Portions of the Interstate Rivers



(c) Near-boundary but Non-transboundary Rivers



8 Conclusions

This paper studies states' behaviors in the implementation of three nonpoint-source water pollution policies on near-boundary and transboundary rivers. Among S305(b) water quality assessment, S303(d) TMDL development, and S319(h) nonpoint source pollution management grant. S303(d) is the most decentralized policy among the three policies, and S319(h) is the list decentralized. Using different models depending on each policy's characteristics, this paper finds that states have a lower probability to assess the water quality and develop TMDLs for near-boundary rivers. However, more S319(h) grants are approved for near-boundary rivers.

The state governments have two transboundary policy behaviors: the beggar-thy-neighbor behavior that the upstream states conduct less pollution control to let pollution flow to the downstream states, and the free-riding behavior that the downstream state conduct insufficient pollution control. The beggar-thy-neighbor behavior is not affected by the downstream states' characteristics while the free-riding behavior depends on the upstream state's environmental and political-economic background. States exhibit both behaviors in S303(d) TMDL development, which results in large deadweight losses.

Chapter 2. Explaining the Heterogeneity in the Effect of Driving Restriction Policies on Air Quality: Evidence from Chinese Cities

Wenbo Meng

9 Introduction

Due to rapid growth in the use of vehicles around the world, vehicle emissions have become one of the largest sources of air pollution. This trend has led to serious health effects including increased deaths from stroke, lung cancer, and heart disease, and has brought about climate change driven by carbon dioxide emissions from the combustion of transportation fuel (Chay and Greenstone (2003), Cohen et al. (2005), Knittel, Miller, and Sanders (2016), Schlenker and Walker (2016)). To reduce vehicle emissions, cities around the world including Athens, Bogotá, Mexico City, San José, Paris, Milan, as well as many Chinese cities have adopted driving restriction policies. However, the effectiveness of these policies in reducing air pollution is still an open empirical question.

The effects of driving restriction policies shown in past studies range from significant reductions in air pollution (Viard and Fu (2015), Liu, Yan, and Dong (2016)), to no reduction in air pollution (Davis (2008), Sun, Zheng, and Wang (2014), Ye (2017), Zhang, Lawell, and Umanskaya (2017)). Most studies on this topic focus on a specific driving restriction policy in a specific city and use a small number of pollutants as a proxy for air quality, making it impossible to compare across studies. Moreover, earlier studies have ignored details of implementation rules that may lead to endogenous treatment. To address these challenges, I exploit detailed and comprehensive panel data on air quality across a large number of Chinese cities with and without driving restrictions. I analyze multiple air pollutants and use policy implementation criteria to isolate exogenous policy variation. I find that driving restriction policies are effective in reducing vehicle emissions. The heterogeneity in policy effect mainly results from the policy's implementation criteria and the air pollution concentration during the restriction period.

I collect the universe of available Chinese air quality monitor data to construct the daily maximum of eight measures of air quality: AQI, CO, NO_2, SO_2, O_3, 8-hour O_3, $PM2.5$, $PM10$, from 2015 to 2019. The responses of these measures to driving restriction policies are different and locally dependent. I classify the details of 357 driving restriction policies in 54 treated cities, including implementation criteria, intensity, length, and area. I find that the policy implementation criteria and pollution concentration influence the magnitude of the policy effect, and the estimation results are biased positively if the policy is endogenously implemented.

I exploit two different identification strategies. First, I estimate the short-run treatment effect for each air quality monitor and each driving restriction policy using a regression discontinuity in time (RDiT) approach. Second, I estimate the average policy effect using a panel fixed-effect approach. The RDiT model is useful for examining spatial and policy heterogeneity. The panel fixed-effect model exploits cross-sectional variation, it is better at capturing the average policy effect in the long run. In most treated cities, driving restriction policies were put in place and removed multiple times during the sample period. In addition, different cities enacted driving restrictions at different times. Combined, this provides variation to identify the policy effect.

Selection bias, reverse causality, and endogeneity are the main challenges in performing empirical work on pollution control policies. Considering that cities with severe air pollution or more vehicles are more likely to select into treatment, I explore a large number of Chinese cities and driving restriction policies and compare their estimation results in the RDiT approach; I use propensity score matching to obtain a sample of 90 control cities that are similar in observables to the 41 treated cities in the panel fixed-effect approach. Reverse causality and endogeneity arise because sometimes governments start driving restriction policies on days when air quality is poor, or on days that more drivers than usual take to the roads. Thus, regulated days may not be randomly selected but may instead be endogenous to a variety of economic, political, and environmental factors. However, much of the existing literature treats driving restriction policies as exogenous shocks to air quality. To tackle this problem, I classify the driving restriction policies in the sample into four categories by their implementation criteria: air pollution alert triggered policies, traffic-triggered policies, seasonal and regular policies, and events-triggered policies. I compare the estimation results under these different policies.

The RDiT estimation results show that the effects of driving restriction policies triggered by air pollution alerts are biased positively. Policies implemented to reduce traffic or improve the long-run ambient air quality are less subject to the endogeneity problem, and more of their estimates show reductions in air pollution. Further heterogeneity analysis shows that driving restriction policies are less effective with high pollution concentration; restricting more vehicles does not necessarily reduce more air pollution. The city's social-economic background and weather conditions also influence the policy effect. When I conduct a placebo test using false starting dates for the same driving restriction policies, the density distributions of the estimation results become symmetric about zero, indicating the RDiT estimation is robust.

The average treatment effect estimated by the panel fixed-effect model shows that AQI, $PM2.5$, $PM10$ increase by about 20% under air pollution alert triggered policies. Traffic-triggered policies, seasonal and regular policies reduce AQI, $PM2.5$, $PM10$, CO, and NO_2 significantly from 7.26% to 14.61%. Back-of-the-envelope calculations show that the reduction in CO and NO_2 are reasonable for driving restriction policies that effectively limit vehicle use. O_3 and 8-hour O_3 do not respond to driving restriction policies instantly as the other pollutants do. SO_2 concentrations do not decline much with driving restriction policies but are affected by industrial pollution reduction policies. These results are consistent with the RDiT estimation results.

This paper contributes to the literature in several areas. First, it conducts a large-scale study of driving restriction policies across many cities. Compared to earlier studies that focused on a handful of cities, the large number of cities and policies studied here greatly improve our understanding of the effectiveness of such policies. Many of the Chinese cities studied here have driving restriction policies similar to other cities in the world, making the results easier to generalize to other contexts. Second, I use the variety of cities studied to explore heterogeneity across policy types, length, implementation criteria, weather, and social-economic characteristics of the cities, and show how specific characteristics of driving restriction policies impact their effectiveness in reducing air pollution. Third, I exploit a rather complete set of air pollutants and show how driving restriction policies are effective in reducing specific pollutants. Finally, the heterogeneous effects studied using the RDiT model provide valuable insights into the design of future driving restriction policies.

Using a large number of cities and driving restriction policies, this paper identifies and explains the heterogeneity in the effect of driving restriction policies on air quality. Davis (2008) is the first empirical study on this topic. Using an RDiT approach, Davis studies a driving restriction policy, Hoy No Circula (HNC), in Mexico City and gets insignificant estimation results. HNC is a regular policy. This paper shows that though this kind of policy generally correlates to more negative RDiT estimation results than air pollution alert triggered policies, it is not completely free from the reverse causality problem at implementation. Many RDiT estimation results are insignificant and close to zero. Possible short-run explanations are low compliance rate (Davis (2008)) and heavy congestion (Sun, Zheng, and Wang (2014).

Viard and Fu (2015) studies two driving restriction policies in Beijing. The first policy is a traffic-triggered odd-even (OE) policy.[23] They find that the Air Pollution Index (API) decreases by about 18% and $PM10$ decreases by about 30%.[24] These results lie near the modes of the density distributions of the RDiT estimation results for AQI and $PM10$ under traffic-triggered policies in this paper. For the one-day-per-week (ODW) regular policy implemented later on,[25] Viard and Fu (2015) finds about 21% reduction in API and 27% reduction in $PM10$. This policy is similar to HNC in the implementation criterion but receives a much better effect. This is most likely attributable to the high compliance rate in Beijing. In their study, the OE policy does not have a proportionally larger effect than the ODW policy; in fact, the estimation results are very similar. This nonlinear relationship between vehicle numbers and air pollution level is consistent with the discussion in this paper and the conclusions in Sun, Zheng, and Wang (2014). The upgrading of driving restriction policies should be careful. Continuing to strengthen the policy incurs high costs but may not receive the expected benefits.

The estimation bias caused by the endogeneity at policy implementation can be carefully controlled in some situations. Mullins and Bharadwaj (2015) studies a string of air pollution alert triggered policies. They use a difference-in-difference approach and the counterfactual is days that "should" have had the alert but did not. They find approximately 20% reduction in $PM10$. Fu and Gu (2017) uses a similar way to construct the counterfactual. They compare API at the same toll stations and during the same holidays in two years with and without a highway toll. They find that eliminating highway toll increases API by 20%. In the empirical study on air pollution policies, it is not easy to find an ideal counterfactual. This paper seeks to identify the policy effect by decomposing the details of a large number of policies and cities and finds that the endogeneity at policy implementation biases the

[23] This policy is an odd-even (OE) policy set to restrict 50% vehicles on road each day. It was implemented before the 2008 Beijing Olympic Games. The purpose of the policy is to show a good city image, and many tourists were entering Beijing during that period. So this policy can be classified as the traffic-triggered policy defined in this paper.

[24] API depends on $PM10$, NO_2, and SO_2.

[25] This policy is set to restrict 20% vehicles on road each weekday.

estimation results rather than eliminates the true policy effects. The true effect of driving restriction policies is revealed under the traffic-triggered policies and seasonal and regular policies.

Traffic regulations are costly in that they restrict drivers' preferred transportation mode and induce other costly behaviors like purchasing additional vehicles (Davis (2008)). The health benefit is usually proportional to the amount of reduced air pollution (Small and Kazimi (1995)). Therefore, it would be beneficial for policymakers to have good information on the policy effect in a specific city in advance. The effectiveness of policies should not be assessed without adequate analysis of the policy details, a variety of pollutants, and local factors.

The remainder of the paper proceeds as follows. Section 10 explains the mechanisms of driving restriction policies; Section 11 presents the empirical approaches; Section 16 describes data; Section 13 presents the estimation results and explores the heterogeneous policy effects; Section 14 concludes.

10 Driving Restriction Policies

Driving restriction policies limit vehicle use based on the last digits of vehicles' license plate numbers; certain numbers are restricted from roadways at particular times. The policy has the advantage that its target is straightforward, the start and stop are flexible, and the instructions are easy to follow. Therefore, since the 1980s, many cities around the world have implemented driving restriction policies with the main goal to reduce urban air pollution.[26] With such popularity, it is important to evaluate the effectiveness of driving

[26]Some policies restrict two numbers per weekday, such as the "Hoy No Circula" policy in Mexico City, Mexico, which began in 1989, and the driving restriction in San José, Costa Rica, 2005. Some policies are more intense, they restrict odd and even numbers in turns on each day. New Delhi, India has implemented this type of policy several times since 2016. Rome, Italy, also adopted the policy temporarily in 2015.

restriction policies and explore how changes in policy settings would influence the policy effect. However, the effect of driving restriction policy is related to urban traffic conditions, economic development, human behavior, weather conditions, etc. As such, the ex-post policy evaluations tend to be city-specific. For the ex-ante decision of implementation and ex-post adjustment of policy details, it is important to understand the "clean" effect of driving restriction policies then combine the analysis with city characteristics.

The flexibility in the adoption and design of driving restriction policies also creates several empirical challenges. First, selection bias arises since the treated areas are not randomly assigned. Second, the restriction on driving may be based on expectations of increased driving, which makes the implementation decision endogenous. Third, the potential mutual influence between air quality and the assignments of driving restriction policies leads to reverse causality in the estimation results. Mullins and Bharadwaj (2015) states that it is hard to determine if a subsequent drop in air pollution is because of the driving restriction policy.

Studying driving restriction policies across a large number of Chinese cities can tackle these challenges. Of the 340 cities and regions in mainland China by second-order administrative division, 64 have adopted driving restriction policies at some point. Figure 18 presents the geographical distribution of these cities.[27] Beijing was the first Chinese city to adopt a driving restriction policy, beginning on July 20, 2008, to improve air quality during the Olympic Games. Later that year, the city government implemented another driving restriction policy as a regular policy. More Chinese cities adopted driving restriction policies

Some policies have regular application periods, like the "Paris Respire" in Paris, France, beginning 2016, which banned motor vehicles from driving in certain districts on Sundays and public holidays. Some policies have more detailed requirements based on vehicles' characteristics, such as the "Pico y Placa" in Bogotá, Columbia that began in 1998. On each weekday the policy restricts four numbers for private vehicles and two numbers for public vehicles. Other Columbian cities adopted this policy later, as did Quito, Ecuador in 2010.

[27]Source of the basemap: Institute of Geographic Sciences and Natural Resources Research, CAS, Resource and Environment Science and Data Center.

thereafter. City governments have full discretion in the decisions to implement driving restriction policies. Influence from higher-order governments exists but is limited. Therefore, the assignments of driving restriction policies across Chinese cities are largely independent. With the variety of city characteristics and policy details, the driving restriction policies in many Chinese cities are likely to be comparable to the policies in other cities around the world. And the average effect of these policies in China is likely to reveal the policy effect in general.

Figure 18: Distribution of Treated Cities

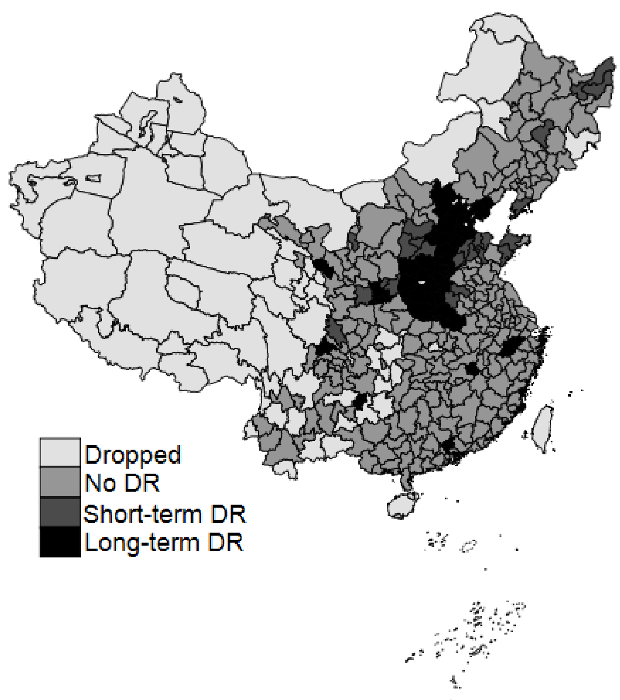


Table 10 summarizes the details of 357 driving restriction policies in 54 Chinese cities.[28] I list four criteria by the potential endogeneity at policy implementation. The first imple-

[28] A driving restriction policy is identified as "one" policy if it is continuously implemented and not divisible. For example, during the application of a regular policy, if the government increased the policy intensity on a few days when there were air pollution alerts, these intense restrictions are not counted as separate policies; if a policy with different implementation criterion is implemented right after a policy, they are counted as two separate policies.

After dropping cities that are only treated before the sample period and cities that are under treatment during the whole sample period, there are 54 cities left in the treatment group.

mentation criterion is the air pollution alert.[29] The driving restriction policies started and ended following the alerts. Some cities also applied industrial pollution reduction policies at the same time. These driving restriction policies are similar to the policies in Italy and Norway, the Environmental Episodes in Santiago, Chile, and the ozone smog alert in southern California. Since these policies were implemented when air quality is poor and removed when air quality improves, there is a potential concern of reverse causality.

The second implementation criterion is an expected traffic increase, which includes city or higher-level sports games, tourism activities, national exams,[30] and big events with road closures. Since these policies regulate traffic when traffic is expected to be higher than usual, they are potentially endogenous.

The third implementation criterion focuses on the long-run concerns over ambient air quality, leading to seasonal and regular policies.[31] For example, high temperatures in the summer accelerate the formation of ozone, frequent thermal inversions in the winter trap pollutants locally, and winter heating in the northern part of China leads to more combustion of fossil fuels. Therefore, some cities applied driving restriction policies in the summer or

[29]The air pollution alerts are based on the Ambient Air Quality Index (AQI) forecasts, and they are affected by weather conditions. Level III alert: $AQI > 200$ and is forecasted to last for more than 48 hours; Level II alert: $AQI > 200$ and is forecasted to last for more than 72 hours; Level I alert: $AQI > 200$ and is forecasted to last for more than 96 hours, $AQI > 300$ and is forecasted to last for more than 48 hours, or $AQI > 500$.

[30]In each city, the National College Entrance Examination (NCEE) takes place on June 7th, 8th each year; the High School Entrance Examination takes place around the end of June each year. Traffic increases near the examination sites.

[31]All the regular policies are long-term policies, but not all the seasonal policies are long-term. Some seasonal policies are applied for less than a month, like 26 to 29 days. Some are even shorter. For example, Xuchang applied a driving restriction in November 2015 for four days, and the reason published on the government website is "to utilize the weather conditions these days to alleviate the winter pollution". Since the starting day of the policy is not associated with an increase in air pollution, this policy is classified as a seasonal policy. Some seasonal policies started before the beginning of the sample period, January 2nd, 2015, but ended in January 2015; some seasonal policies started in December 2019, but the sample period ends on December 31, 2019. The lengths of these policies that lie within the sample period are less than a month.

Table 10: Policy Details

Category	Number of Policies	Number of Cities
	Implementation Criteria	
Air pollution alert	224	38
Expected traffic increase	46	13
Seasonal or regular policy	76	37
Restriction	11	9
	Type	
2 digits each day (ODW)	152	35
5 digits each day (OE)	183	46
drive 4 days rest 4 days	1	1
1 digit each day	15	2
3 digits each day	3	2
4 digits each day	2	2
non-local vehicles	1	1
	Length	
$\leq$ 7 days	221	40
$>$ 7 days & $\leq$ 15 days	39	24
$>$ 15 days & $\leq$ 30 days	33	28
$>$ 30 days & $\leq$ 60 days	16	14
$>$ 60 days & $\leq$ 90 days	0	0
$>$ 90 days & $\leq$ 120 days	12	12
$>$ 120 days	36	25
	Hours per Day	
$\leq$ 6 hours	16	6
$>$ 6 hours & $\leq$ 15 hours	188	41
$>$ 15 hours	167	44

winter. Similar to some Latin American cities, some Chinese cities also applied driving restriction policies all year long. Though these policies were implemented due to air pollution concerns, their starting days do not necessarily correspond to more severe air pollution than usual. The air quality in that season or city is poor in general. Thus, these policies are less subject to reverse causality than the pollution alert triggered policies.

There were several driving restriction policies implemented due to big events that did not cause much increase in traffic. The city governments implemented the policies to show a better city image,[32] or even the image of a nearby city.[33] These policies are plausibly exogenous.

I categorize the driving restriction policies in the sample into seven types by the restriction intensity. The two most frequently adopted types in China and other cities around the world are the odd-even restriction (OE) and one-day-per-week restriction (ODW). Under the OE policy, vehicles with an odd number as the last digit of their license plate can only drive on odd-numbered dates, and even-digit vehicles can only drive on even-numbered dates. This type of driving restriction policy bans half of the vehicles from the road each day. Under the ODW policy, two of the last digits of plate numbers are banned each day. This driving restriction policy bans 20% of vehicles each day.

Table 10 also summarizes the lengths of the driving restriction policies in the sample. Policies longer than 30 days are marked as long-term policies in Figure 18. I also subdivide the hours restricted each day into three categories.[34] Restrictions less than 6 hours each day are mostly peak-hour restrictions; restrictions less than 15 hours cover most of the day time;

[32] Such as the Car-free Day in Luoyang and the 2015 China Langfang International Economic and Trade Fair in Langfang.

[33] Cities of Hebei Province implemented driving restriction policies and industrial pollution reduction policies during the 2015 China Victory Day parade in Beijing.

[34] The total number of different hour lengths exceeds the number of driving restriction policies because some policies adjusted the hour lengths during application; some driving restriction policies started or ended halfway during a day, so the number of hours applied on the starting day or ending day are less than the normal application length.

restrictions longer than 15 hours extend to night time and are mostly 24-hour restrictions.

Past studies have listed some behavioral responses that reduce the effectiveness of driving restriction policies (Eskeland and Feyzioglu (1997), Davis (2008), de Grange and Troncoso (2011), Troncoso, De Grange, and Cifuentes (2012), Ye (2017), Gu, Deakin, and Long (2017), Zhang, Lawell, and Umanskaya (2017), Chen et al. (2020)). Overall, they can be classified into two categories: 1) long-term responses, such as purchasing a second car, and substitution to less fuel-efficient cars; 2) short-term responses, such as inter-temporal substitution (driving more in unrestricted time), inter-spatial substitution (detour the unregulated area), and noncompliance.[35] I explore these behaviors through the heterogeneity checks and the comparison of results from the RDiT approach and the panel fixed-effect approach.

11 Empirical Strategy

The RDiT model is good at estimating the short-run effects of driving restriction policies. It is less prone to bias from unobserved time-varying factors and allows us to more easily see the spatial heterogeneity in the treatment effects. However, RDiT lacks cross-sectional variation and many applications use observations far from the temporal thresholds (Hausman and Rapson (2018)). In contrast, the panel fixed-effect model exploits cross-sectional variation and time-series variation. It shows the average effect of driving restriction policies. Considering the pros and cons of the RDiT model and the panel fixed-effect model, I exploit both identification strategies as complements to each other.

[35] In China, the Road Monitoring System almost eliminates the possibility of escape from violation. Viard and Fu (2015)) finds high compliance rates using data on license-plates entering the garage in Beijing. Liu et al. (2020) uses license-plates records to find that the noncompliance rate is 5.92% for the ODW policy and 11.73% for the OE policy in Langfang. Thus, the noncompliance behavior is unlikely to be a large factor that negates the policy effect in China.

11.1 Regression Discontinuity in Time Approach

A sharp regression discontinuity in time (RDiT) identification is possible because air pollutants dissipate quickly without exogenous shocks (Li et al. (2010), Zhang et al. (2015)). Daily pollution levels depend on vehicle emissions on the given day. Since different cities implemented driving restriction policies at different times, the RD thresholds are city-specific. Moreover, air pollution concentration is sensitive to many factors like road density and building height (Sini, Anquetin, and Mestayer (1996), Huang et al. (2008)). Even within the same city, the air quality and policy effects in different areas can be different. Therefore, I estimate an RDiT model for each air quality monitor i and each driving restriction policy to conduct a heterogeneity analysis.[36] Equation 22 is monitor-specific. The dependent variable $Pollution_t$ is the logarithm of one of the eight measures of air quality at monitor i on day t. Each regression estimates a vector of treatment effects α for the driving restriction policies that monitor i experienced. The policies are represented by a vector of the policy dummies, $Policy$. W_{ct} is a non-linear polynomial of the weather variables in city c.[37] f_t includes indicator variables for the month of the year and a Chebyshev time polynomial.[38] Standard errors are robust and clustered at the month of the sample to account for correlations in pollution shocks across months for each monitor.

$$Pollution_t = \beta_0 + Policy\alpha + W_{ct}\beta_1 + f_t\beta_2 + \epsilon_t \tag{22}$$

[36]See Section C in the Appendix for city-level RDiT.

[37]W_{ct} includes: the daily average of air temperature, 6-hour accumulated precipitation, wind speed, sky coverage, and their square terms; one-day lag of temperature and precipitation, the interaction terms of temperature and its lag, precipitation and its lag, and temperature and precipitation; the daily maximum and minimum of temperature and precipitation.

[38]A Chebyshev polynomial is a good approximation for continuous functions (Mason and Handscomb (2002), Auffhammer and Kellogg (2011)).

11.2 Pooled Panel Fixed-Effect Approach

The treatment group includes cities that have implemented driving restriction policies; the control group includes cities that have never implemented the policy. The driving restriction policies in the sample are staggered. Cities in the treatment group are not treated at the same time and are not necessarily treated once. Identification of the treatment effect is based on the differences in policy implementation status across different cities. Recent literature has expressed concerns on a potential pitfall of the two-way fixed-effect estimators: they may recover a weighted average of treatment effects in which some effects are weighted negatively (Goodman-Bacon (2018), Imai and Kim (2019), Steigerwald, Vazquez-Bare, and Maier (2019), de Chaisemartin and d'Haultfoeuille (2020)). Thus, I test the signs of the weights assigned to the estimators in Section 13.3.

$$\begin{aligned} Pollution_{it} = \beta_0 + \alpha_1 PolicyAP_{ct} + \alpha_2 PolicyT_{ct} + \alpha_3 PolicySR_{ct} + \alpha_4 PolicyE_{ct} \\ + \beta_1 Stag_{ct} + \beta_2 Fest_t + X_{cy}\beta_3 + W_{ct}\beta_4 + W_{ct}\tau\beta_5 + T + D_i + \epsilon_{it} \end{aligned} \tag{23}$$

Equation 23 estimates the policy effects across different implementation criteria. $PolicyAP$ is the dummy variable for air pollution alert triggered policies; $PolicyT$ is the dummy variable for traffic-triggered policies; $PolicySR$ is the dummy variable for seasonal and regular policies; $PolicyE$ is the dummy variable for the several events-triggered policies. The dummy variables equal 1 for monitors in the treatment group cities on days treated by the certain kind of driving restriction policies, and 0 otherwise. $Stag_{ct}$ represents days with air stagnation.[39] Pollutants are poorly dissipated under such meteorological conditions. $Fest_t$ indicates days that are legal festivals. X_{cy} is a vector of the Gross Regional Product (GRP) and population

[39] According to the National Oceanic and Atmospheric Administration (NOAA)'s definition of Air Stagnation Index, I calculate the Air Stagnation dummy as days with wind speed less than 8m/s and no precipitation.

of city c in year y. W_{ct} is the same weather polynomial as in Equation 22, it captures weather shocks to pollution levels. $W_{ct}\tau$ represents interaction terms of the weather polynomial W_{ct} and a time vector τ, which includes the day-of-week and seasons.[40] T is a time vector that includes day-of-week, month-of-year, and year-fixed effects. The day-of-week fixed effects capture air quality and traffic patterns on different days of a week. The month-of-year fixed effects capture seasonality and traveling behaviors in different months. The year-fixed effects control for yearly air quality trends common to all cities. The monitor-fixed effects D_i control for unobserved monitor characteristics. ϵ_{it} is the error term. I cluster the standard errors in two dimensions, city and month of the sample, to account for pollution shocks that are collected at the city, and in the month of the sample.

Selection bias is a potential concern if cities in the treatment group suffer more from vehicle emissions than cities in the control group. To account for observable differences in the control cities that are related to pollution and driving, I calculate the propensity scores ($p(D|X)$) of the sample cities using their city characteristics and balance treated and control cities in the regression using $weight = \frac{p(D|X)}{1-p(D|X)}$.

12 Data

12.1 Air Quality Data

I obtain air quality data from January 2, 2015, to December 31, 2019, from the China National Environmental Monitoring Center. Researchers have expressed concerns about data manipulation from this data source (Chen et al. (2012), Ghanem and Zhang (2014), Ito and Zhang (2016)), but many papers have shown that data manipulation has not occurred since

[40] W_{ct} interacts with the day-of-week variable to allow for variation in weather-affected pollution formation on weekdays and weekends. W_{ct} interacts with seasons to allow weather effects to vary across different seasons.

2013 (Stoerk (2016), Liang et al. (2016)).[41] The data set reports hourly readings of the Ambient Air Quality Index (AQI), CO, NO_2, SO_2, O_3, 8-hour O_3, $PM2.5$, and $PM10$ from 1498 air quality monitors that are active in the sample period. Dropping air quality monitors located in parks, forests, or near waters, there are 253 monitors left in the 54 treated cities. 44 of the monitors lie outside the restricted areas. AQI is calculated from the other seven pollutants.[42] Since the city governments comply with the National Ambient Air Quality Standards by controlling the maximum pollution concentration, the data are aggregated to their daily maximum. Table 11 presents the descriptive statistics of the pollutants in the sample period.

CO is generated from the incomplete combustion of carbon fuels. Vehicle emissions account for about 17.75% to 35.50% of total CO emissions in China. The percentage of NO_2 coming from vehicle emissions ranges widely from 1.85% to 64.79%. Little SO_2 comes

[41] This is confirmed here using US embassy air quality data. The US embassy has one air quality monitor in each of the five Chinese cities: Beijing, Shanghai, Guangzhou, Shenyang, and Chengdu. These air quality monitors report hourly O_3 concentrations in Beijing, and hourly $PM2.5$ concentrations in all five cities. The US embassy monitors are not located in the same places as the Chinese monitors, so the observations should not be the same. The correlation coefficients between the US embassy monitors and all the Chinese monitors in each city are high. The coefficients are 0.8763 for O_3 in Beijing, 0.9062, 0.9268, 0.8566, 0.8677, and 0.8157 for $PM2.5$ in Beijing, Shanghai, Shenyang, Chengdu, and Guangzhou, respectively. One of the political targets for cities in mainland China is to reduce the number of "unhealthy" days each year. The threshold of "unhealthy" $PM2.5$ is an hourly average of $30ug/m^3$. I count the number of days that $PM2.5$ exceeds the "unhealthy" threshold for all the monitors in each city, each year, and present the results in Figure 42 in the Appendix. The number of "unhealthy" days counted by the US embassy monitors does not always exceed the days counted by the Chinese monitors. Since the "unhealthy" days counted by Chinese monitors also vary from one another, it is reasonable for the US embassy monitors to count more or fewer days than some of the Chinese monitors. These results support the conclusions in the literature that the air quality data reported by the China National Environmental Monitors have not been manipulated in the sample period.

[42] As shown in equation 24, AQI is the maximum of n Individual Air Quality Index $IAQI_i$. $IAQI_i$ is an indicator that corresponds to a certain level of pollutant i, which includes hourly CO, NO_2, SO_2 and their 24-hour averages, 24-hour $PM2.5$, 24-hour $PM10$, O_3 and 8-hour O_3 (MEE (2012)):

$$AQI = max\{IAQI_1, IAQI_2, ..., IAQI_n\} \tag{24}$$

Table 11: Descriptive Statistics

Variable	Unit	Obs.	Mean	Std.Dev.	Min	Max
			Pollutants			
AQI	id.	483,035	138.08	90.68	8	500
CO	ug/m^3	483,035	1913.72	1761.71	1	90000
NO_2	ug/m^3	483,035	68.25	34.29	1	766
SO_2	ug/m^3	483,035	53.18	74.93	1	1597
O_3	ug/m^3	483,035	113.45	58.98	1	1200
8h-O_3	ug/m^3	483,035	99.60	52.56	1	300
$PPM2.5$	ug/m^3	483,035	95.33	80.13	1	2008
$PM10$	ug/m^3	483,035	168.78	162.88	1	52996
			Weather Variables			
Temperature	0.1 $°C$	483,035	141.86	110.77	-285.63	361.38
Precipitation	0.1 mm	483,035	9.89	36.50	0	1040
Wind Speed	0.1 m/s	483,035	22.74	14.82	0	140
Sky Coverage	id.(0-19)	483,035	4.72	3.01	0	9

from vehicle emissions. The share of O_3 depends on the level of NO_x and VOC, and it is largely affected by weather conditions. The shares of $PM2.5$ and $PM10$ coming from vehicle emissions vary widely across cities. Table 12 summarizes these pollutants' main sources, the shares from vehicle emissions, and the approximate reductions under driving restriction policies. Reducing vehicle emissions does not reduce a pollutant much if its share from vehicles emissions is small, let alone many driving restriction policies are not applied 24 hours a day. See Section B in the Appendix for detailed back-of-the-envelope calculations.

Table 12: The Shares of Pollutants from Vehicle Emissions

Pollutant	VE Share	Main Source	Min. Effect of ODW Policy	Max. Effect of OE Policy
CO	17.75% - 35.50%	Vehicles or other machinery burning fossil fuels.	3.55%	17.75%
NO_2	0.19% - 32.40%	Vehicles, power plants.	0.038%	16.20%
SO_2	About 1%	Industrial facilities like power plants.		
O_3	Weather sensitive	Vehicles, power plants.		
PM	City specific	Various dusty sources and formed by other pollutants.		

Notes: See Section B in the Appendix for references and the calculation process.

12.2 Weather Data

Because pollution reacts, diffuses, and is transported in the atmosphere, ambient air pollution levels are highly dependent on weather conditions. For example, lower temperatures increase vehicle emissions of CO, and ozone forms with warm temperatures and sunlight. I obtain weather data from the National Climatic Data Center (NCDC), which, every three hours, provides readings of air temperature, wind speed rate, sky total coverage, and six-hour accumulated liquid precipitation depth at more than 410 weather stations across China.

Some cities do not have a weather station. I impute weather data for these cities in the following manner: I calculate the centroids of the air quality monitors in these cities and identify all existing weather stations located within 65 miles in Vincenty distance to these centroids.[43] I calculate the mean observations from these weather stations by inverse weighting their Vincenty distances. I use these means as the weather observations of cities without weather stations.[44] Table 11 presents the summary statistics of the weather variables.

12.3 City Characteristics

64 cities in mainland China have ever implemented driving restriction policies. After dropping cities that are only treated before the sample period and cities that are under

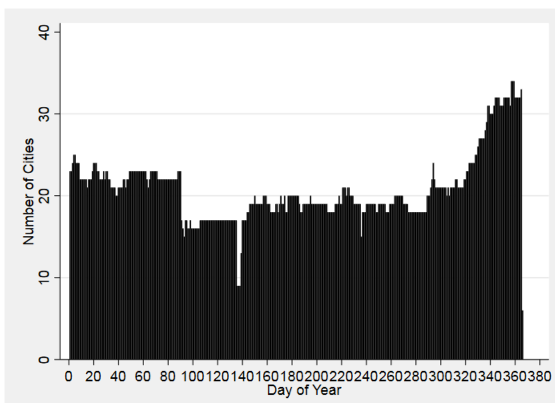
[43]Vincenty distance is the distance between two points on the surface of a spheroid.

[44]To check whether the predicted weather observations are accurate, I predict weather observations for existing weather stations. The correlation coefficient between the actual and predicted observations for each weather station is 0.9699 on average for daily maximum temperature. But they are only 0.5599 for daily accumulated six-hour precipitation, 0.3796 for daily maximum sky coverage, and 0.5629 for daily maximum wind speed on average. I omit these three variables from the RDiT regression and present the density distribution of the estimation results in Figure 47 in the Appendix. Figure 47 looks very similar to Figure 22, indicating that these variables do not affect the results much and the imprecisely predicted observations do little harm to the estimation.

treatment during the whole sample period, there are 54 cities left in the treatment group. These cities applied 357 driving restriction policies in total in the sample period.

Figure 19 demonstrates the temporal variation of the driving restriction policies. Almost every day of the year there are some cities applied the policies. Winter has more driving restriction policies than the other seasons. Figure 18 presents the geographical distribution of the treated cities.[45] All 64 treated cities lie in eastern and middle China. Eastern China is more developed and more populous than the west (See Figure 43 and Figure 44 in the Appendix[46]). Therefore, from all cities that have never applied driving restriction policies, I drop the western cities and keep the rest 214 cities as the control group, so that the economic development and geographical locations are not too different between the treatment group and the control group.

Figure 19: Number of Cities Applied Driving Restriction Policies Across Day-of-year



To check whether the control cities are a good counterfactual for the treated cities, I conduct propensity score matching using 12 demographic and economic variables in 2013 acquired from the 2014 China City Statistical Yearbook, before any cities in the treatment group had implemented driving restriction policies. They include social-economic variables

[45] The "Long-term DR" legend represents cities that have ever applied long-term driving restriction policies; the "Short-term DR" represents cities that have only applied short-term driving restriction policies.

[46] Source: Institute of Geographic Sciences and Natural Resources Research, CAS, Resource and Environment Science and Data Center.

(GRP, population, urban area, wage), transportation variables (the number of buses, the number of taxis, road area), air quality variables (the emissions of industrial SO_2 and industrial dust), climate variables (average summer and winter temperatures), and a location dummy variable indicating Northern or Southern cities. The details of the matching process are presented in Section E in the Appendix. I exploit more city characteristics in the heterogeneity analysis in Section 13.2. Table 34 in the Appendix presents the summary statistics of all these variables. The yearly GRP and population used in the pooled panel fixed-effect regression are from the 2015 to 2019 National Economic and Social Development Statistical Bulletin.

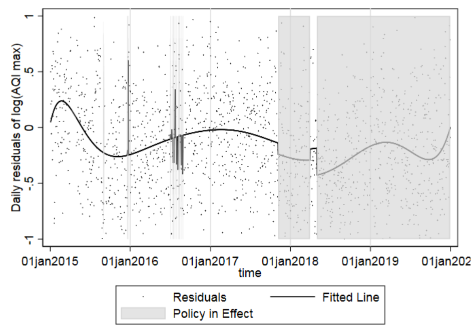
13 Results

13.1 Regression Discontinuity in Time Results

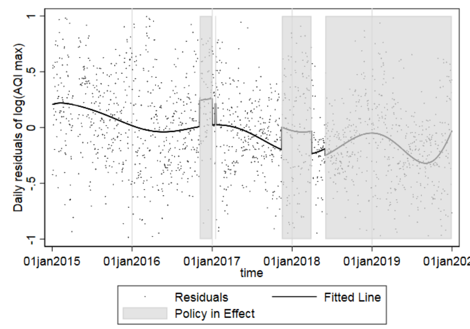
Figure 20 plots the regression results for AQI, CO, NO_2, and $PM10$ at two representative air quality monitors. The results for the other pollutants at the same monitors are presented in Figure 46 in the Appendix. The grey areas are the periods with driving restriction policies. The fitted line is the time series of predicted values of the treatment effect. It is obtained from estimating Equation 22 and then centered so that its mean value is 0. The residuals are noisy. There are sharp changes in pollution levels when the driving restriction policies turn on and off, and the effects of driving restriction policies on these pollutants exhibit heterogeneity across different air quality monitors and driving restriction policies.

Figure 20: Daily Maximum Pollutant Concentration

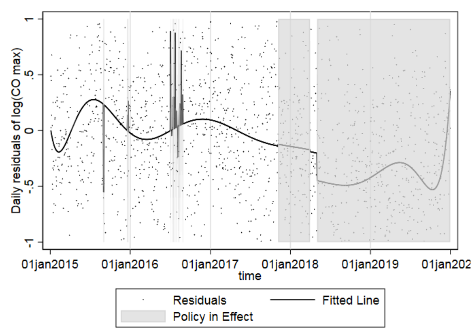
(a) AQI: Monitor A1042, Qinhuangdao



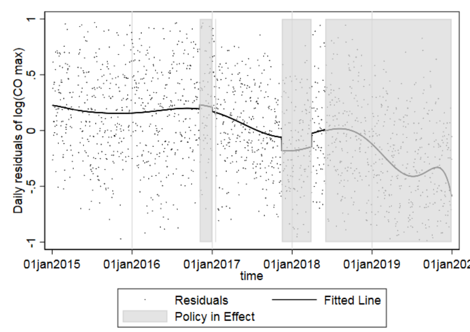
(b) AQI: Monitor A1820, Anyang



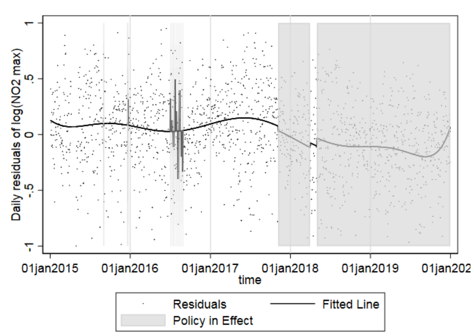
(c) CO: Monitor A1042, Qinhuangdao



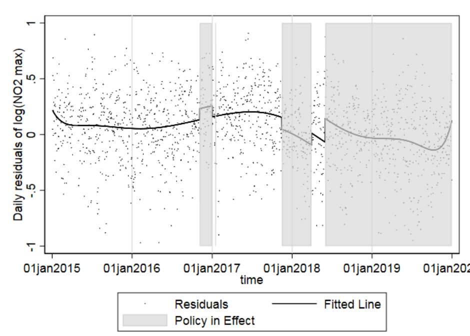
(d) CO: Monitor A1820, Anyang



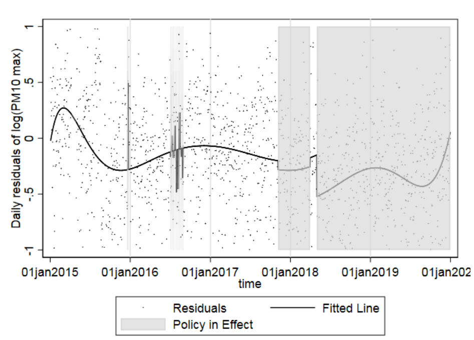
(e) NO_2: Monitor A1042, Qinhuangdao



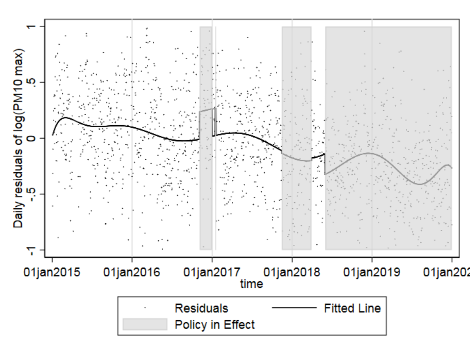
(f) NO_2: Monitor A1820, Anyang



(g) $PM10$: Monitor A1042, Qinhuangdao



(h) $PM10$: Monitor A1820, Anyang



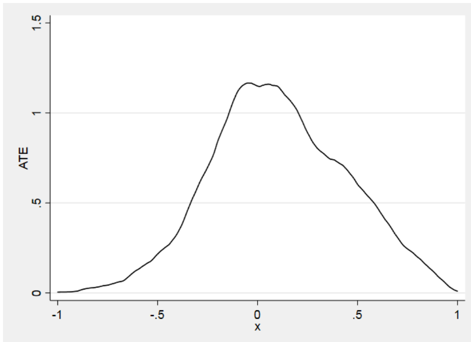
The density distributions of the RDiT estimation results of 357 driving restriction policies at 253 air quality monitors are presented in Figure 21.[47] The estimation results are heterogeneous for all the pollutants. It is natural for past studies to find that driving restriction policies are effective in some cities while ineffective in others. But we can infer the general pattern of the policy effect by controlling the implementation criteria.

---

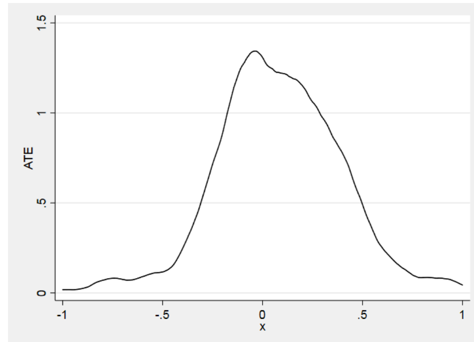
[47] The average AIC and BIC values for CO in Table 36 of the Appendix show that the higher the order of the time polynomial the better the model. Since infinite orders are impossible, I use a tenth-order polynomial in these regressions.

Figure 21: Distribution of the RDiT Estimation Results

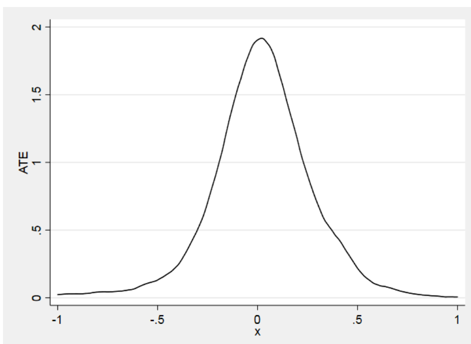
(a) AQI



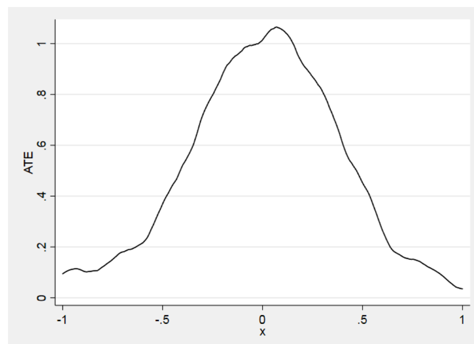
(b) CO



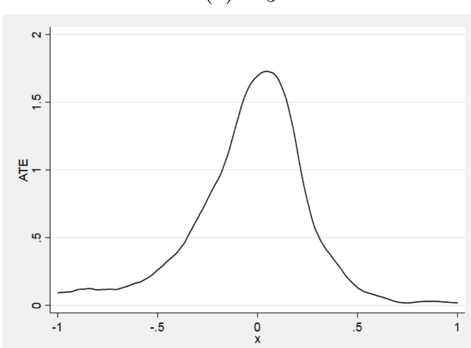
(c) NO_2



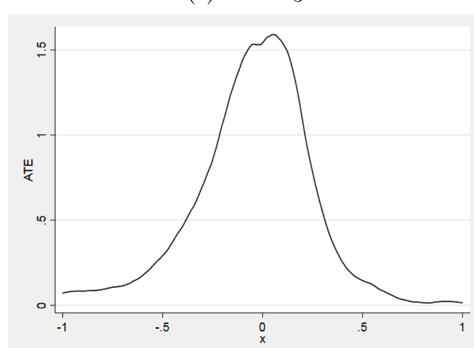
(d) SO_2



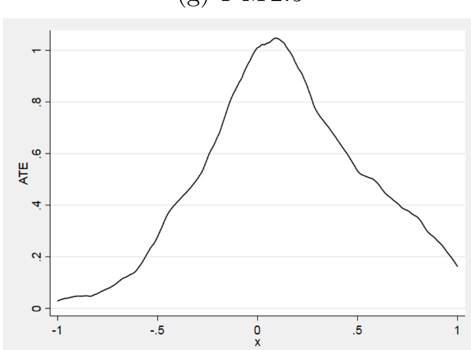
(e) O_3



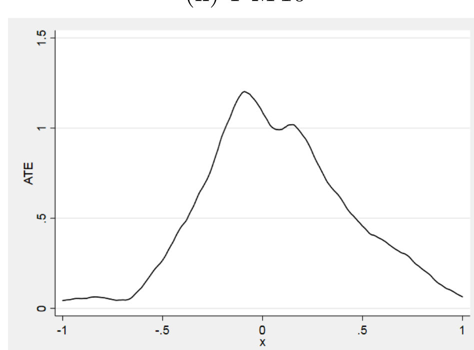
(f) 8-h O_3



(g) $PM2.5$



(h) $PM10$



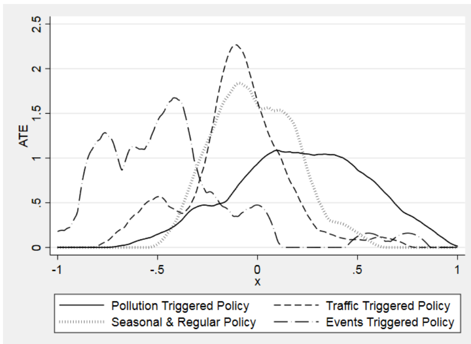
The density distributions of the RDiT results by policy implementation criteria are presented in Figure 22. The estimation results of air pollution alert triggered policies distribute positively for all pollutants. The estimation results of traffic-triggered policies shift negatively for all pollutants except for ozone. The seasonal and regular policies also have more negative results than air pollution alert triggered policies. Though the distributions for plausibly exogenous events-triggered policies are multimodal because the sample size is small, we can still observe negative shifts. These distribution patterns show that the reverse causality at policy implementation due to increased air pollution biases the RDiT estimation results positively. Governments implemented these driving restrictions to reduce air pollution when the air pollution is higher than usual, so it looks like that it is the driving restriction policy that makes the air quality worse. Though not completely free from reverse causality, seasonal and regular policies are less subject to this problem than the alert triggered policies, so they have relatively more negative estimation results. The other endogenously implemented policies, i.e. the traffic-triggered policies, however, do not seem to bias the estimation results much. Governments implemented these policies to reduce the number of vehicles on the road when they expect more vehicles than usual, so an x percentage restriction is not able to reduce vehicles by x percentage, i.e. the policy effect is discounted or even eliminated. However, the actual situation may not be so disappointing. First, the traffic-triggered policies are based on "expected" traffic. Real traffic does not increase as much as the amount reduced by the policies. Tourism activities usually can not attract more than 20% of vehicles of the city, let alone 50%, as set by the ODW and OE policies. Road closures for conferences or exams may cause congestion nearby, but this effect can easily be offset by the driving restrictions throughout the central urban area. Second, air pollution levels may not change with vehicle numbers linearly. When the air pollution alerts are triggered, air pollution concentration is high, reducing vehicles does not reduce the pollution much. When there is an expected increase in traffic, air pollution concentration is at its normal level, reducing vehicles would reduce more air pollution. This inference is further verified in Section 13.2.

The negative tendency of the estimation results for traffic-triggered policies does not show in O_3 and 8-hour O_3. As is discussed in Section 12.1 and Section B in the Appendix, ozone is not emitted by vehicles directly and its formation depends on the levels of other pollutants. When the levels of the other pollutants change because of driving restriction policies, the change in ozone is ambiguous. Thus, though long-term ozone levels can reflect ambient air quality, ozone does not serve well as an instant indicator of vehicle emissions when policies turn on and off frequently. The density distributions for SO_2 are more complicated. As the fuel standards become stricter, SO_2 is not the major pollutant from vehicle emissions in China anymore. However, many driving restriction policies implemented due to air pollution alerts and big events are accompanied by requirements of industrial pollution reduction, which leads to a reduction of SO_2 at the same time. Therefore, there are also negative shifts in the estimation results for SO_2 under traffic and events-triggered policies. Compared to the other pollutants, the estimation results for SO_2 under seasonal and regular policies are more flatly distributed and symmetric about zero, because seasonal and regular policies usually do not require industrial pollution reduction at the same time.

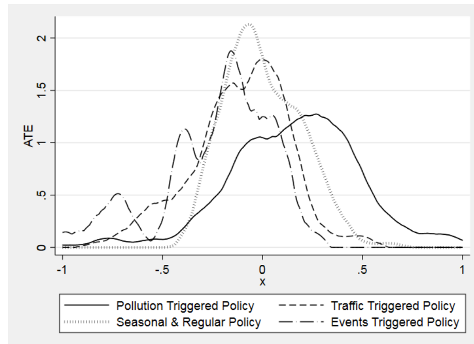
Figure 23 presents the density distributions of the estimation results that are significant at 90% confidence level. Even the statistically significant results have mixed signs. Estimates close to zero are mostly insignificant. Except for ozone, all the other pollutants have more positive and significant results under air pollution alert triggered policies, more negative and significant results under traffic-triggered and events-triggered policies. Under the seasonal and regular policies, AQI and $PM10$ have more negative results, the other pollutants have similar amounts of results in both signs. The significant results further verify the above inference that estimation results for pollution-induced policies are biased positively. The "clean" policy effects are mostly negative.

Figure 22: Distribution of the RDiT Estimation Results by Policy Implementation Criteria

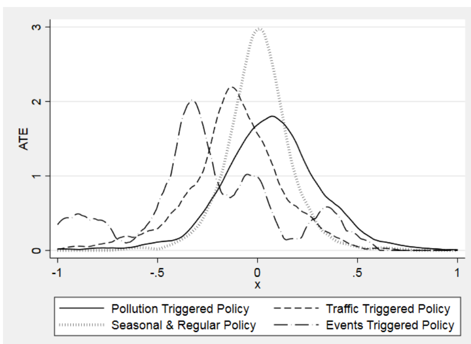
(a) *AQI*



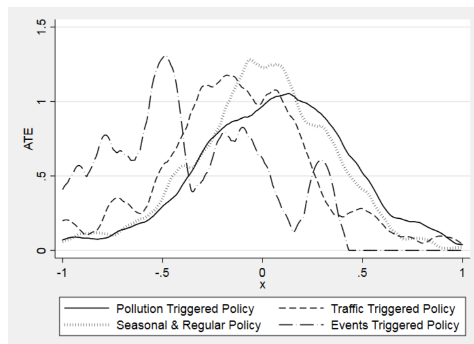
(b) *CO*



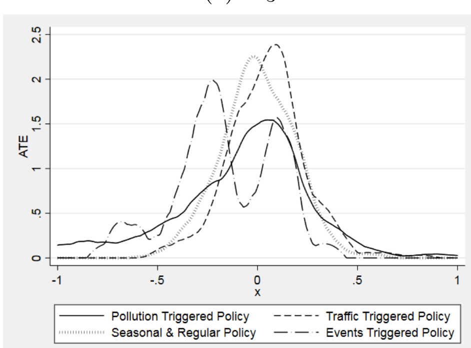
(c) NO_2



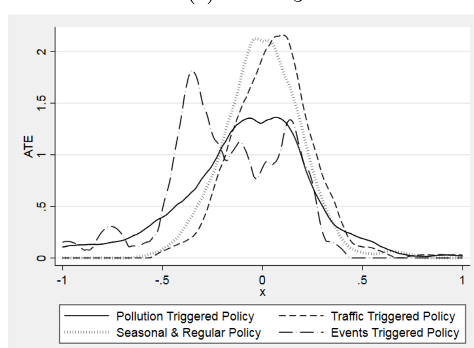
(d) SO_2



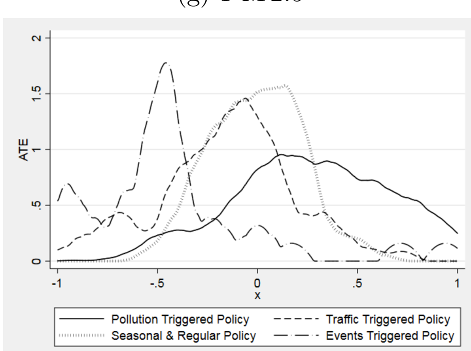
(e) O_3



(f) 8-h O_3



(g) $PM2.5$



(h) $PM10$

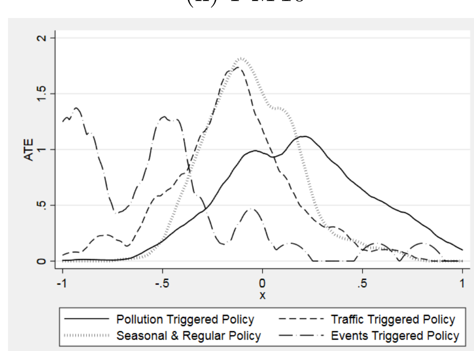
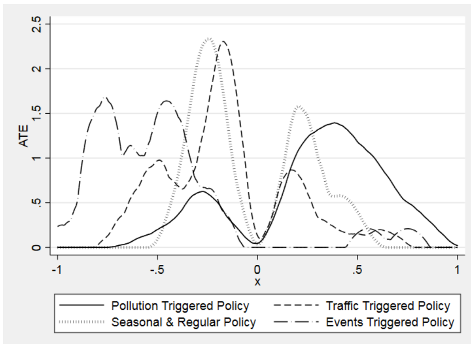
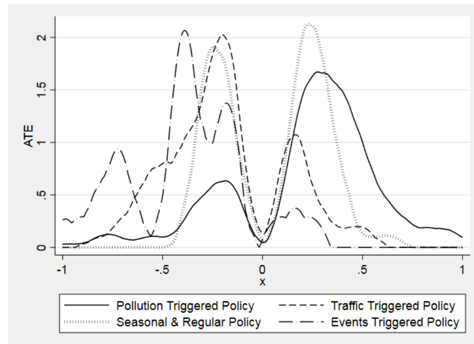


Figure 23: Distribution of Statistically Significant RDiT Estimation Results by Policy Implementation Criteria

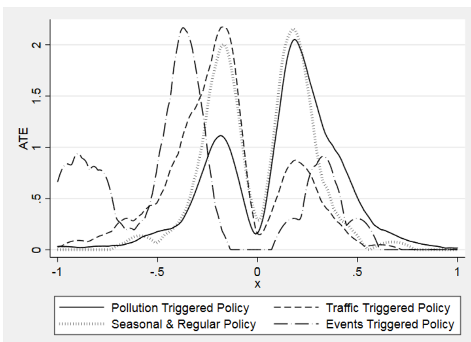
(a) AQI



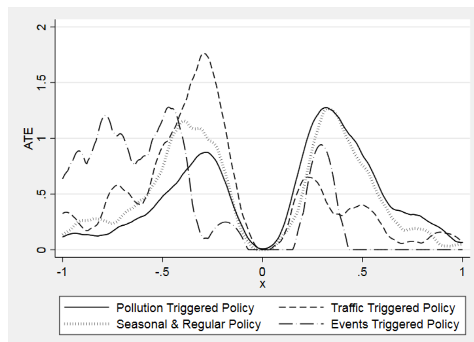
(b) CO



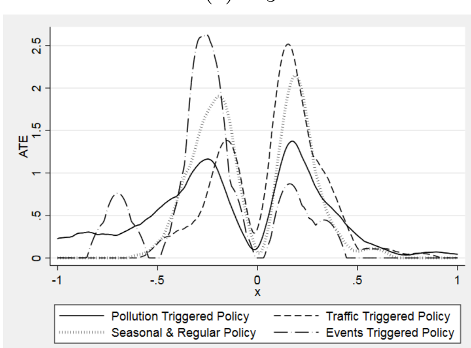
(c) NO_2



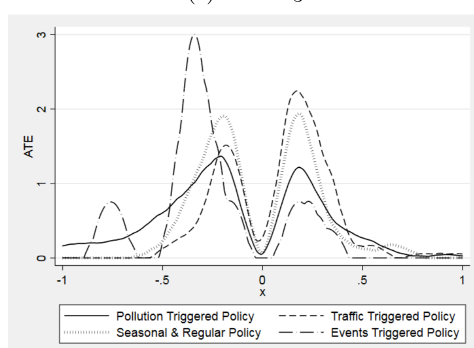
(d) SO_2



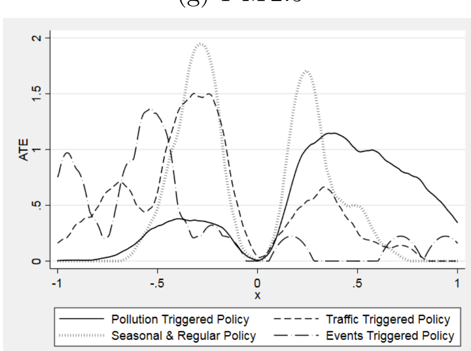
(e) O_3



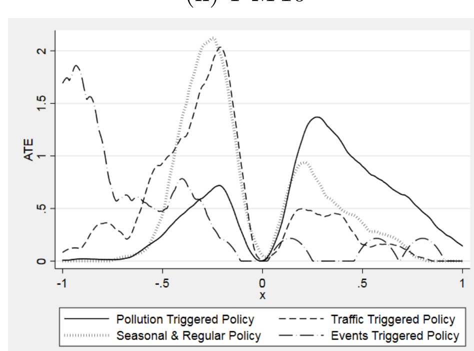
(f) 8-h O_3



(g) $PM2.5$



(h) $PM10$



13.2 Heterogeneity in Policy Effect

To uncover the factors that influence the policy effect, I regress the estimated treatment effect of each driving restriction policy at each air quality monitor on several sets of control variables separately: 1) the average pollution concentration at the air quality monitor during the restriction periods; 2) the details of the driving restriction policies; 3) an indicator of whether the monitor locates in the restricted areas or not; 4) the weather conditions during the restriction period; 5) city characteristics, including economic variables and climate variables listed in Table 34 in the Appendix. Standard errors are robust for all regressions and are clustered at the city level for regressions on 2), 4), and 5).[48]

[48] I block bootstrap the full RDiT analysis in Section D in the Appendix.

Table 13: Influence of Pollution Concentration on the RDiT Estimation Results

	AQI	CO	NO_2	SO_2	$O3$	$8-hO_3$	$PM2.5$	$PM10$
Pollution	0.0024	0.0001	0.0045	0.0010	0.0028	0.0029	0.0027	0.0019
	(0.0001)	(0.0000)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Obs.	1,618	1,618	1,618	1,618	1,618	1,618	1,618	1,618
R-sq.	0.5421	0.2882	0.3157	0.0874	0.2072	0.1794	0.6220	0.5680

Note: Numbers in parentheses are standard errors.

Table 11 shows the magnitudes and ranges of the pollution concentrations in the sample. Together with Table 13, they show that higher pollution concentration is associated with more positive treatment effects. In addition to the positive bias caused by the potential endogeneity at policy implementation, other explanations are: first, when a pollutant's level is high, the share of the pollutant coming from vehicles might be lower than usual. The severe pollution could come from increased driving, but also other sources like industrial production, coal-fired heating, etc. In this situation, reducing vehicles does not reduce pollution as much as usual. Second, pollution may not decrease with the number of vehicles on road linearly. There may exist a concave relationship between pollution level and vehicle numbers. Sun, Zheng, and Wang (2014) states that this probably attributes to alleviated congestion; improved traffic condition worsens the air quality. The R-squares are relatively large for AQI, $PM2.5$, and $PM10$, indicating their concentrations predict the policy effects better than the other pollutants. The R-squares for SO_2 is particularly lower than the other pollutants. This is consistent with the expectation that SO_2 almost does not respond to driving restriction policies and thus the policy effects are less predictable by SO_2 concentration.

The first section of Table 14 shows the correlations between the policy implementation criteria and the corresponding estimated policy effects. Compared to the omitted category, i.e. the air pollution alert triggered policies, all the other kinds of policies correlate to more negative estimation results. Except for NO_2, which does not necessarily have better policy effects under seasonal and regular policies. The second section reports the coefficients for policy types. The omitted type is the ODW policy. Types other than OE are not presented here because the sample size is small. Compared with the ODW policies that restrict 20% of vehicles, OE policies that restrict 50% of vehicles do not necessarily correlate to more negative estimation results. The standard errors are large, compared to the magnitudes of the coefficients. This implies again that the policy effect, i.e. air pollution reduction, may not change with the number of vehicles on road linearly. This can also be seen in the hours restricted each day. Governments are more likely to implement longer-hour driving restrictions when the air pollution is severe. But policies covering longer than peak hours do not correlate to better policy effects for most pollutants, except for ozone. These policies even correlate to more positive estimation results for SO_2 and $PM10$. Section "Length" shows that, compared to policies less than 7 days, longer-term policies generally correlate to more negative policy effects for AQI, $PM2.5$, and $PM10$. These conclusions are further verified in the city-level RDiT estimation in Section C of the Appendix.

Table 14: Influence of Policy Details on the RDiT Estimation Results

	AQI	CO	NO_2	SO_2	$O3$	$8-hO_3$	$PM2.5$	$PM10$
				Implementation Criteria				
Traffic	-0.3321	-0.2864	-0.2086	-0.1470	0.1398	0.1624	-0.4716	-0.3643
	(0.0255)	(0.0249)	(0.0229)	(0.0339)	(0.0193)	(0.0200)	(0.0354)	(0.0300)
Seasonal	-0.1161	-0.0985	-0.0101	-0.0923	0.1096	0.1241	-0.1351	-0.1318
	(0.0247)	(0.0263)	(0.0208)	(0.0463)	(0.0293)	(0.0310)	(0.0305)	(0.0310)
Events	-0.6624	-0.4779	-0.3325	-0.4807	-0.1205	-0.1150	-0.8816	-0.7652
	(0.0618)	(0.0633)	(0.0582)	(0.0790)	(0.0559)	(0.0563)	(0.0875)	(0.0779)
				Type				
OE	-0.0588	-0.0058	0.0157	-0.0622	0.0064	-0.0023	-0.0193	-0.0466
	(0.0211)	(0.0203)	(0.0149)	(0.0240)	(0.0203)	(0.0201)	(0.0278)	(0.0235)
Others	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
				Length				
7-15D	-0.0744	-0.0340	-0.0302	-0.0710	0.0343	0.0422	-0.1194	-0.1251
	(0.0228)	(0.0262)	(0.0206)	(0.0374)	(0.0286)	(0.0297)	(0.0285)	(0.0259)
15-30D	-0.2372	-0.1881	-0.1198	-0.1542	-0.0174	0.0096	-0.2910	-0.2492
	(0.0247)	(0.0235)	(0.0202)	(0.0348)	(0.0343)	(0.0347)	(0.0299)	(0.0290)

30-60D	-0.1430	-0.0495	-0.0256	0.1415	-0.0064	-0.0131	-0.1905	-0.1494
	(0.0320)	(0.0296)	(0.0222)	(0.0517)	(0.0332)	(0.0349)	(0.0359)	(0.0371)
90-120D	-0.1192	-0.0708	0.0016	0.0425	-0.0195	-0.0253	-0.1260	-0.0823
	(0.0323)	(0.0404)	(0.0278)	(0.0659)	(0.0361)	(0.0377)	(0.0395)	(0.0395)
> 120D	-0.2281	-0.0880	-0.0722	0.0172	-0.0472	-0.0528	-0.2981	-0.2280
	(0.0341)	(0.0349)	(0.0259)	(0.0570)	(0.0358)	(0.0382)	(0.0427)	(0.0432)
				Hours				
6-15H	0.0256	0.1359	-0.0144	0.2814	-0.1132	-0.1052	0.0347	0.1197
	(0.0811)	(0.0945)	(0.0690)	(0.1232)	(0.0464)	(0.0395)	(0.1346)	(0.0945)
> 15H	0.0957	0.1885	0.0091	0.3323	-0.2176	-0.1998	0.1141	0.2212
	(0.0799)	(0.0935)	(0.0687)	(0.1228)	(0.0470)	(0.0403)	(0.1331)	(0.0933)
Obs.	1,667	1,667	1,667	1,667	1,667	1,667	1,667	1,667
R-sq.	0.2487	0.1602	0.1058	0.0981	0.0701	0.0732	0.2391	0.2405

Note: Numbers in parentheses are standard errors.

Driving restriction policies may not affect the air quality equally for the whole urban area. There may be an inter-spatial spillover effect on air quality outside of the restricted area in the treated city. Table 15 shows the coefficients of regressing the RDiT estimation results on the dummy variable of monitor location. The standard errors are large. This shows that the spillover effect is negligible. Detour behavior hardly influences the policy effect. The heterogeneous policy effects across weather variables and city characteristics are presented in Table 35 and Table 37 in the Appendix. The correlations between the policy effects and these variables are relatively weak compared to the pollution concentration and policy details. The coefficients are difficult to interpret because these variables depend on each other. Removing or adding some variables from the regression would affect the coefficients of the other variables. All these variables together influence the treatment effects, but it is difficult to infer the role of a single variable.

Table 15: Influence of Being Monitors Outside the Restricted Areas on the RDiT Estimation Results

	AQI	CO	NO_2	SO_2	$O3$	$8-hO_3$	$PM2.5$	$PM10$
SP.Monitor	0.0162	0.0057	-0.0106	0.0288	0.0278	0.0303	0.0116	0.0225
	(0.0225)	(0.0215)	(0.0191)	(0.0279)	(0.0224)	(0.0219)	(0.0285)	(0.0266)
Obs.	1,618	1,618	1,618	1,618	1,618	1,618	1,618	1,618
R-sq.	0.0002	0.0000	0.0002	0.0006	0.009	0.0011	0.0001	0.0004

Note: Numbers in parentheses are standard errors.

In short, policy implementation criteria affect the accuracy of the estimation results; together with the other policy details and pollution concentration, they determine the policy effects. Social-economic variables and weather conditions also cause heterogeneity in the policy effects.

13.3 Pooled Panel Fixed-Effect Results

In this section, I conduct a pooled panel analysis to estimate the average policy effects by implementation criteria. To ensure that the treated cities and the control cities are not too different, I use propensity score matching to trim the sample.[49] There are 41 cities (185 air quality monitors) in the treatment group and 90 cities (402 air quality monitors) in the control group after trimming. Table 16 presents the estimation results.[50]

[49] See Section E in the Appendix.

[50] The coefficients of the policy dummies in equation 23 estimate a weighted sum of ATTs. I conduct a test as in de Chaisemartin and d'Haultfoeuille (2020) and found that no ATT receives negative weight.

Table 16: Panel Fixed-Effect Results

	AQI	CO	NO_2	SO_2	$O3$	$8-hO_3$	$PM2.5$	$PM10$
AP Alert	0.2078***	0.0808	0.0603	0.0320	-0.0835	-0.1043	0.2611***	0.1991***
	(0.0585)	(0.0771)	(0.0403)	(0.1310)	(0.0862)	(0.0976)	(0.0761)	(0.0682)
Traffic	-0.0985*	-0.0643	-0.1461*	-0.1251	-0.0030	0.0007	-0.1170*	-0.0871**
	(0.0552)	(0.0515)	(0.0811)	(0.1043)	(0.0594)	(0.0622)	(0.0669)	(0.0434)
Seasonal	0.0265	-0.0726**	-0.0273	-0.0993	0.0317	0.0314	0.0110	0.0226
	(0.0285)	(0.0333)	(0.0240)	(0.0597)	(0.0295)	(0.0319)	(0.0383)	(0.0306)
Events	-0.3499	-0.2233	-0.0269	-0.6841***	0.0238	0.0578	-0.4040	-0.4308
	(0.2683)	(0.2306)	(0.2339)	(0.2416)	(0.3069)	(0.3043)	(0.4070)	(0.3490)
Stag.	0.0118	-0.0256**	0.0340***	0.0762***	0.0729***	0.0804***	-0.0232	0.0385***
	(0.0134)	(0.0117)	(0.0100)	(0.0142)	(0.0124)	(0.0128)	(0.0193)	(0.0141)
Fest.	-0.0126	-0.0618**	-0.1784***	-0.0492	-0.0047	0.0167	-0.0165	-0.0155
	(0.0422)	(0.0265)	(0.0353)	(0.0423)	(0.0181)	(0.0191)	(0.0559)	(0.0477)
CityChar.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. $\times$ Seas.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. $\times$ DOW	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

DOW FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Monitor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MOS Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PS Wgt.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	973,713	973,713	973,713	973,713	973,713	973,713	973,713	973,713

Note: $***$ Significance at 1 percent level; $**$ Significance at 5 percent level; $*$ Significance at 10 percent level. Numbers in parentheses are standard errors.

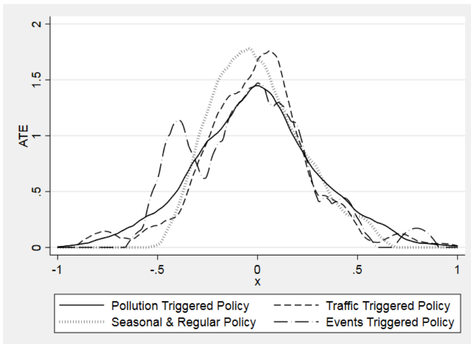
The estimation results are consistent with the different chemical properties of these pollutants. AQI increases by 20.78% under the air pollution alert triggered policies. This is not surprising since the alerts are based on AQI levels. $PM2.5$ and $PM10$ also increase largely by 26.11% and 19.91%, respectively. This implies that $PM2.5$ and $PM10$ may be the driving pollutants of the AQI alerts. During legal festivals, driving behavior is different from usual. CO and NO_2 both decrease significantly. Especially NO_2, which decreases by 17.84%, showing that it is very sensitive to traffic change. This explains its large reduction of 14.61% under the traffic-triggered policies. $PM2.5$ and $PM10$ are also major pollutants in vehicle emissions, so they also decrease significantly under these policies. CO's reduction under the traffic-triggered policies has a large standard error. But it decreases significantly under the seasonal and regular policies. Air stagnation often occurs in winter, which is also the main season for the implementation of seasonal policies. Most pollutants increase in air stagnation days, but CO does not seem to be trapped locally as the other pollutants. The reductions of NO_2 and CO match with the back-of-the-envelope calculations of effective driving restriction policies presented in Table 12. SO_2 only decreases significantly under the events-triggered policies, so we can infer that its reduction is largely driven by the industrial pollution reduction policies applied at the same time. Consistent with the RDiT estimation results, O_3 and 8-hour 0_3 do not have significant changes under any kinds of policies.

13.4 Placebo Test

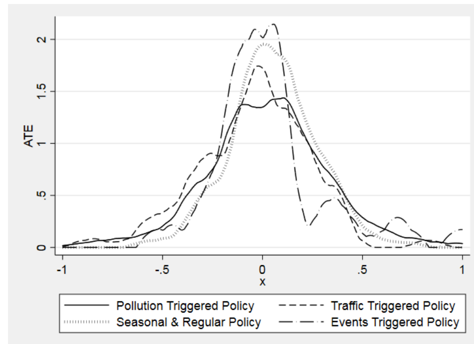
For each driving restriction policy applied at an air quality monitor, I reassign the starting date of that driving restriction policy randomly across the sample period but keep the same policy length, criterion, type, etc. That is to say, each air quality monitor experiences the same driving restriction policies as in the real world but at different times in the sample period. Figure 24 presents the density distributions of the RDiT estimation results. Compared with Figure 22, the differences in estimation results by policy implementation criteria disappear, all the results distribute symmetrically about zero. This shows that the RDiT estimation in Section 13.1 is robust.

Figure 24: Distribution of Placebo RDiT Estimation Results by Policy Implementation Criteria

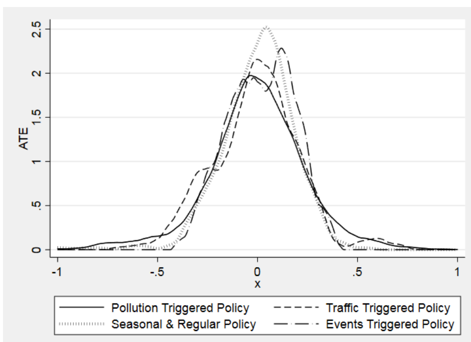
(a) AQI



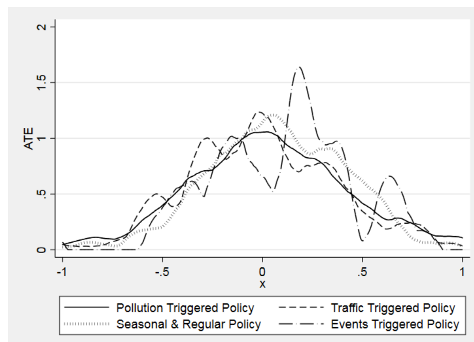
(b) CO



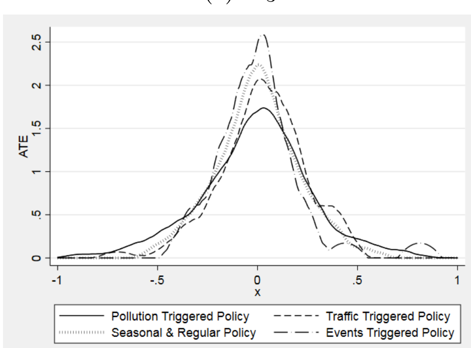
(c) NO_2



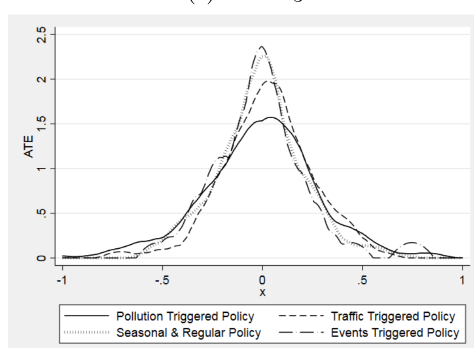
(d) SO_2



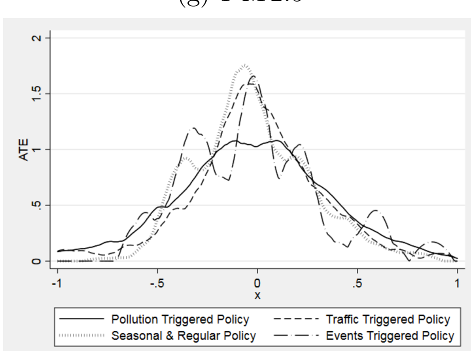
(e) O_3



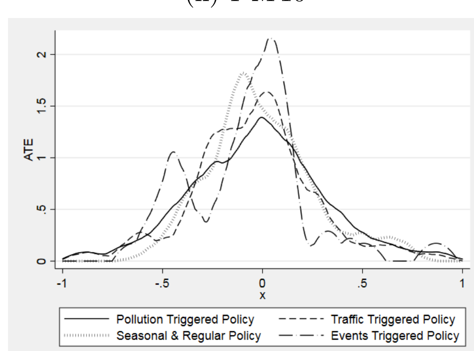
(f) 8-h O_3



(g) $PM2.5$



(h) $PM10$



14 Conclusions

This paper studies the effect of driving restriction policies on air quality in Chinese cities using eight measures of air quality. Results from both the regression discontinuity in time approach and the panel fixed-effect approach show that the effects of driving restriction policies are heterogeneous across numerous factors, but in general, they can significantly reduce CO, NO_2, $PM2.5$, $PM10$, and AQI. The magnitudes of the panel estimation results for CO and NO_2 match with the back-of-the-envelope calculations for effective driving restriction policies. Thus, we can conclude that the driving restriction policies are effective in reducing air pollution.

The estimation results are biased positively if the endogeneity in policy implementation is not well controlled. The estimation results under the traffic-triggered policies and seasonal and regular policies reflect more real policy effects. The heterogeneity analysis shows that pollution concentration and policy details, especially implementation criteria, are the main factors that influence the policy effects. In addition to the influences of the city's social-economic background and the weather conditions during the restriction periods, the policy effect is diminished under severe air pollution, heavy congestion, or low compliance rate; the estimation results also vary with model specifications.

Chapter 3. Who Pays for Congestion? The Incidence of an Optimal Congestion Toll on California Highways

Jonathan Hughes and Wenbo Meng

15 Introduction

External costs due to traffic congestion exceed $120 billion annually in the US. Internationally, traffic congestion is a major issue in both developed and developing countries. Factors such as wasted time, fuel consumption, environmental damage, and negative health impacts affect the broader community beyond just drivers, making it a pressing concern. Economists tend to favor congestion tolls as a solution to alleviate traffic congestion. However, since traffic conditions vary substantially, optimal congestion tolls, *i.e. Pigouvian taxes*, will vary by location and time of day. This raises concerns about equity issues if commuting patterns vary based on demographics such as income or race.

Policy makers and the general public have highlighted the potential for congestion charges to disproportionately impact low-income individuals and minority communities who may have limited transportation alternatives or face higher travel costs (Murphy (2017), Dawid (2019), UCLA (2019), Berger (2008)). However, the existing evidence on the equity of congestion tolling is limited and mixed, with a few small area studies and simulations producing conflicting results, including suggestions that congestion charges could be progressive, regressive, or neutral (Arnott, De Palma, and Lindsey (1994), Santos and Rojey (2004), Schweitzer and Taylor (2008), Krol (2016), Ke and Gkritza (2018)). To address this concern, this paper seeks to conduct a definitive study on congestion tolling in California, a large populous state with high levels of existing congestion.

We investigate the potential distributional effects of congestion charges. We utilize detailed traffic flow data to estimate marginal external congestion costs (MECs) for thousands of locations in California. We estimate time-varying and location-specific congestion costs. The potential endogeneity of driving demand poses a threat to estimating the marginal external costs of congestion. If for instance, drivers are more (less) likely to drive during uncongested (congested) times, then estimates of the travel time-vehicle density relationship would be biased (Yang, Purevjav, and Li (2020)). We account for potential bias in marginal cost estimation by exploiting the exogenous shift in driving demand due to COVID-19 in an instrumental variables framework. We find the two-stage least squares estimates for marginal external congestion costs diverge more from our base estimates as vehicle density increases, suggesting endogeneity may be a major concern when the road is very congested.

We leverage MEC estimates to inform various toll policies. By assigning tolls to individual trips using detailed data from the National Household Travel Survey (NHTS), we estimate the total toll cost for each household under different policies. We analyze the distribution of toll costs with various aspects of household travel behavior and demographics. Our analysis demonstrates that tolls based on location and/or hour are differentially paid among income and racial groups, with significant implications for policymakers and researchers studying sustainable transportation practices.

Our findings show that highway tolls are progressive rather than regressive. Households with higher incomes bear a greater burden of tolls. For instance, households with income above \$200,000 spend \$0.9785 more per day under monitor-based tolls, \$0.3693 more per day under hourly-based tolls, and \$1.2388 more per day under monitor and hourly-based tolls than those with income below \$35,000. Our analysis also suggests that households with higher incomes tend to travel more frequently on congested routes but less during peak hours. Black and Asian households pay a higher toll per day than white households under the monitor-based toll policy, with coefficients of \$0.5145 and \$0.6705, respectively, while Indian households pay a lower toll of -\$0.4226 per day. Further results suggest that different racial groups tend to cluster in areas with varying congestion situations, leading to varying toll costs.

Our study provides policymakers with insights into equity issues related to congestion tolling, highlighting the need to consider equity concerns in designing and implementing congestion pricing policies to promote sustainable and equitable transportation practices. By accounting for differences in travel behavior and congestion costs across locations, times, and demographics, we can identify toll policies that are both efficient and equitable, providing a sustainable solution to alleviate congestion while ensuring that costs are distributed fairly.

The remainder of the paper proceeds as follows. Section 16 describes traffic data, weather data, and household travel data; Section 17 constructs the model for travel time and vehicle density relationships, presents the statistics of the estimates of real-time marginal external congestion costs (MEC), and investigates the distributional effect of several second-best congestion tolls calculated from real-time marginal external congestion costs; Section 18 concludes.

16 Data

16.1 Traffic Data

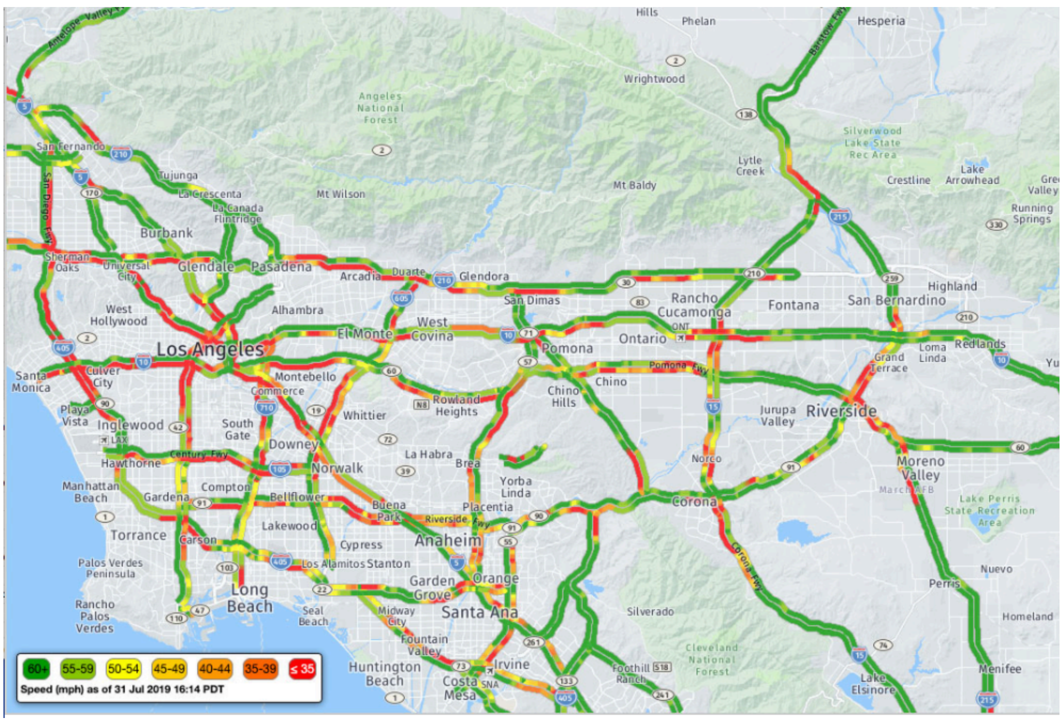
We use hourly data on traffic speed and flow from the California Freeway Performance Measurement (PeMS) system (California Department of Transportation, 2020) from January 1, 2019, through December 31, 2020. We observe traffic conditions from approximately 40,000 detectors at over 9,300 locations on major highways within the state. Data include observations from 42 of California's 58 counties and focus on the most heavily populated regions and well-travel routes. In addition to traffic conditions, we record information on the location of each traffic monitoring monitor, the highway route, direction of travel, number of lanes and lane type, *i.e.* mainline or carpool (HOV) lane, on-ramp, off-ramp or interchange. We focus our analysis on mainline lanes and exclude observations from carpool lanes or ramps. The California Department of Transportation (Caltrans) divides the state into 12 administrative districts, shown in Figure 25. The PeMS network covers highways in Districts 3 through 8 and 10 through 12. In each district, PeMS sensors monitor traffic conditions on the region's major highways. Figure 26 shows the PeMS network in District 7, Los Angeles.

Figure 25: Caltrans Districts in the PeMS Monitoring Network

Table 17 presents summary statistics on our traffic data.[51] There are tens of millions of observations. Average speeds across all hours of the day approach free-flow levels, near 65 miles per hour. For the majority of observations, across locations and hours, traffic is essentially uncongested. In 2019, before COVID, the 25th percentile speed is 62 miles per

[51]See Table 33 in the Appendix for summary statistics in a narrower sample period, the statistics are similar to the wider sample.

Figure 26: District 7 PeMS Monitoring Network



hour, only slightly less than free flow speed. However, for about 25 percent of observations traffic is congested and slowdowns can be severe. Speeds at the 10th, 5th, and 1st percentiles are approximately 55, 52, and 21 miles per hour, respectively. Speeds are lower and traffic flows and densities are higher during 2019. Average densities and flows decreased somewhat during 2020 due to COVID-19, leading to somewhat higher average speeds. This foreshadows our instrumental variable strategy where we propose using COVID-19 as an exogenous change in driving demand.

Table 17: Summary Statistics by Year: 2019.1.1 - 2020.12.31

2019							
	Obs.	Mean	Min.	P25	P50	P75	Max.
Speed (mph)	771,412	62.9	3.0	62.0	65.4	67.8	92.2
Flow (veh./hr.)	771,412	2,841.3	0.0	1,044.0	2,436.0	4365.0	20362.0
Density (veh./lane-mi.)	771,412	13.6	0.0	4.9	11.4	18.8	465.0
Lanes	771,412	3.59	1.0	3.0	4.0	4.0	8.0
North	771,412	0.31	0	0	0	1	1
South	771,412	0.31	0	0	0	1	1
East	771,412	0.19	0	0	0	0	1
West	771,412	0.19	0	0	0	0	1
2020							
	Obs.	Mean	Min.	P25	P50	P75	Max.
Speed (mph)	773,931	64.3	3.0	62.8	65.8	67.9	99.0
Flow (veh./hr.)	773,931	2,450.2	0.0	841.0	1,959.0	3709.0	17354.0
Density (veh./lane-mi.)	773,931	11.0	0.0	4.0	9.0	15.8	275.8
Lanes	773,931	3.6	1.0	3.0	4.0	4.0	8.0
North	773,931	0.3	0	0	0	1	1
South	773,931	0.3	0	0	0	1	1
East	773,931	0.2	0	0	0	0	1
West	773,931	0.2	0	0	0	0	1

16.2 Weather Data

To account for the potential effect of weather on traffic conditions we collect local hourly weather data from the California Irrigation Management Information System (CIMIS). We use air temperature and precipitation records to infer rainy hours and snowy hours. This will allow us to estimate travel time-density relationships under different weather conditions.

To determine the weather conditions at each PeMs monitor, we first calculate the distance between the monitor and each weather station. For weather stations located within a 50-mile radius, as measured by Vincenty distance, we select the 10 closest stations. In cases where no stations are located within 50 miles, we choose the closest available station. To calculate the mean observations from these weather stations, we use inverse weighting based on their Vincenty distances. These means provide us with the weather observations at each PeMs monitor. We also differentiate between precipitation types based on the air temperature. When the temperature is below $32°F$, we identify precipitation as snow, whereas temperatures above the freezing point correspond to rainfall.

16.3 Household Travel Data

We obtain data on household travel behavior from the 2017 National Household Travel Survey (NHTS) and California Add-On. The NHTS is a nationally representative survey administered by the Federal Highway Administration (FHWA) to collect data on daily travel by U.S. residents that includes an inventory of all trips taken within a 24-hour period by all household members aged 5 or older. The California Add-On is an additional survey conducted on a subset of California households to collect state-specific travel data. The NHTS and California Add-On provide detailed information on household home locations at the 5-digit zip code level. Upon approval, we will obtain accurate coordinates for these locations, which will allow us to conduct geospatial analyses of travel patterns and the distributional effect of the marginal external congestion costs.

The NHTS and California Add-On surveys record detailed information on every trip taken by every household member in a day, including trip distance, start and end time. We utilize this information to identify the PeMs monitors located within the trip distance for each trip. We assume regulators adopt congestion tolling that equals the marginal external congestion costs that vary by location and time. We assume that the second-best tolls borne by the person during the trip are the average MEC at these PeMs monitors, which allows us to estimate the preliminary toll incidence by household demographics. Ideally, we will estimate the tolls using exact trip start and end locations and optimal routing with OpenStreetMap upon approval of the data application.

17 Empirical Approach

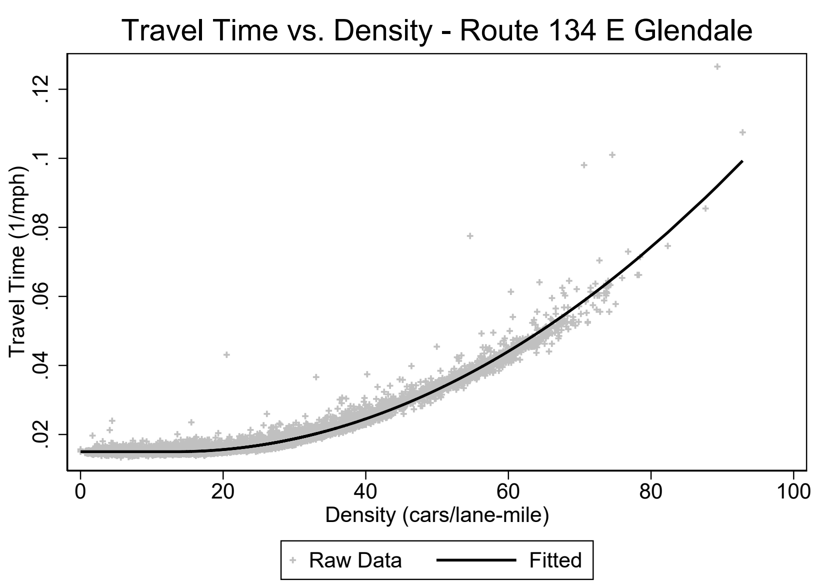
In this section, we will examine how households are impacted by various toll policies. As measuring the actual marginal external congestion costs with significant variability can be costly or impractical, policymakers rely on second-best policies based on average values. We propose a methodology to use readily available information to calculate the real-time MECs. While our ultimate goal is to analyze the toll incidence for different households using the optimal (real-time) congestion toll, we currently assess the tolls that households would face under the second-best tolls calculated by average MECs.

17.1 Estimating Marginal External Congestion Costs

Our goal is to estimate the marginal cost due to an additional highway user (vehicle). To do this we first estimate the relationship between travel time (T) and vehicle density (n). Unlike the relationship between travel time and highway flow, which can be backward bending, the travel time-density relationship is monotonic. The slope of the travel-time density relationship, $\frac{\partial T}{\partial n}$, gives the marginal effect of an additional highway user on travel time and the marginal external (time) cost is the product of the marginal time effect and the number of highway users: $n\frac{\partial T}{\partial n}$. To monetize this cost, we use an average value of time of \$21.51 per hour consistent with Small, Winston, and Yan (2005) and Bento, Roth, and Waxman (2020). An important feature of this approach is that once the travel-time vehicle-density relationships are estimated, the marginal external congestion costs can be calculated from an observed density.

During uncongested periods, changes in highway use have no effect on average travel time. For instance, adding an additional vehicle during the early morning hours will not reduce speeds for other drivers. However, during congested periods, changes in highway use affect speeds and average travel times. Figure 27 illustrates these features for a representative location in District 7 (Los Angeles). For vehicle densities less than approximately 18 vehicles/lane-mile, increasing vehicle density does not affect travel time. Above this level, increasing highway use increases travel time.

Figure 27: Representative TT-density Fit: Route 134



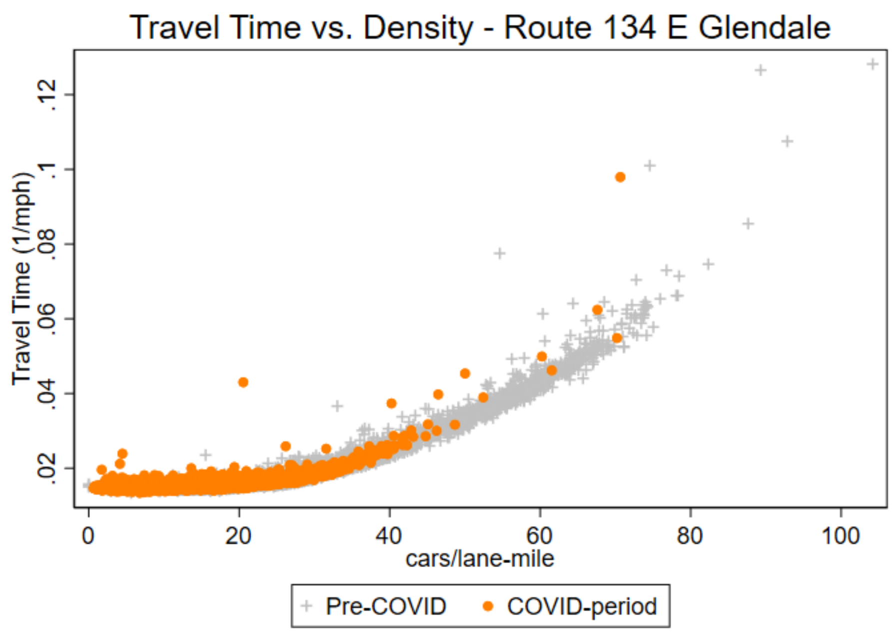
To capture these effects empirically, we estimate models of the form:

$$T_{it} = \beta_0(n_{it} < c) + [\beta_0 + \beta_1(n_{it} - c) + \beta_2(n_{it} - c)^2] * (n_{it} \geq c) + Weather_{it}\alpha + \epsilon_{it} \qquad (25)$$

where T_{it} is average travel time at location i at time t and n_{it} is vehicle density. In the piecewise relationship, c is the critical vehicle density corresponding to the onset of congestion, β_0 is the uncongested travel time, $(n_{it} < c)$ and $(n_{it} \geq c)$ are indicator functions. $Weather_{it}$ is a nonlinear polynomial that includes the first, second, and third order of rain and snow accumulations. Because precipitation may affect driving demand and also the relationships between travel time and vehicle density. For instance, if precipitation reduces highway capacity, *i.e.* traffic speeds are lower for a given number of vehicles, *and* fewer drivers travel on rainy days, our estimated travel-time vehicle density relationships would be biased. For each highway location, we fit Equation 25 to the data using non-linear least squares and estimate the parameters c, β_0, β_1 and β_2. The fitted relationship for the representative highway location is shown in Figure 27. The slopes $\frac{\partial T}{\partial n}$ are easily calculated as: 0 for $n_{it} < c$ and $\beta_1 + 2\beta_2 n_{it}$ for $n_{it} \geq c$.

As noted by Yang, Purevjav, and Li (2020), estimates of Equation 25 could be biased due to endogeneity if, for instance, drivers' decisions to drive respond to traffic conditions. We address this possibility using an instrumental variables strategy. During 2020, travel restrictions and behavioral changes led to large changes in travel demand. For instance, Figure 28 plots travel times and vehicle density in the year before COVID-19 and during the COVID period. Here, as in Hughes and Kaffine (2019), we define the COVID period as the time after California Governor Gavin Newsom signed an executive order implementing public health measures on March 12, 2020. We see a large reduction in travel times and vehicle densities during the COVID period. Interestingly, the travel time-density relationship during the COVID period follows the same pattern as during the period before COVID.

Figure 28: Shift in Driving Demand (Vehicle Density) during the COVID Period



We operationalize the shift in travel demand as an instrumental variable in the following manner. We first aggregate hourly vehicle travel (flow) within each Caltrans district to the week level. This eliminates concerns over intertemporal substitution across days of the week or hours of the day during the COVID period. We then interact the total weekly flow with a set of hour-of-day fixed-effects to allow the effects of reduced travel demand during COVID to vary by time of day. Using these instruments, we implement a two-stage estimation. In the first stage, we regress vehicle density (n_{it}) on our instruments and use the parameter estimates to predict vehicle density $\hat{n}_{it}$. In the second stage, we replace n_{it} with our estimate $\hat{n}_{it}$ in our non-linear travel-time density estimation (Equation 25).

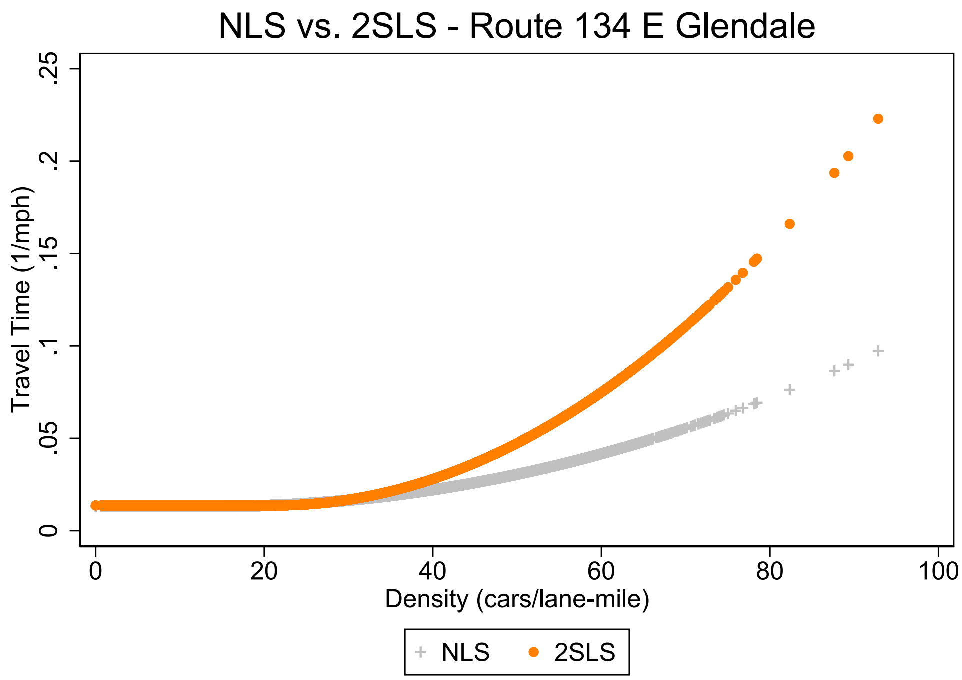
Figure 29 presents two examples of the results of this procedure. The top panel shows travel-time density relationships estimated using nonlinear least squares and our instrumental variables procedure for the representative section of highway shown previously. The bottom panel presents results from a different representative highway. In both cases, the two-stage least squares estimates are steeper (as expected) than the base estimates. However, the differences are relatively small. In the first example (top panel) the two curves lie nearly on top of one-another. The differences are somewhat larger in the second example (lower panel), yet the two curves are quite similar. We explore differences in marginal cost estimates in detail for a large set of highway locations in our discussion of results below.

First, we compare marginal external congestion costs calculated using non-linear least squares (naïve) estimates for the travel-time vehicle-density relationships with those calculated using our two-stage least squares instrumental variables procedure. We construct travel-time vehicle-density relationships for each monitor in all nine Caltrans districts and calculate marginal external congestion costs as described above. Table 18 presents summary statistics on each set of cost estimates. Mean marginal external costs are much higher with the 2SLS approach. The difference in mean values is driven mainly by a longer right tail in the distribution of costs estimates using the 2SLS travel-time density relationships. At the 75th percentile, marginal external costs are zero in both sets of estimates. At the 90th, 95th, and 99th percentiles, values are smaller using the NLS curves, \$0.20, \$0.36, and \$1.17 per vehicle-mile compared with \$0.25, \$0.71, and \$4.26 using the 2SLS estimates.

Table 18: Comparison of NLS and 2SLS MECs

	NLS	2SLS
Observations	1.91E+08	1.91E+08
Mean	\$0.07	\$716.61
Min.	\$0.00	\$0.00
p1	\$0.00	\$0.00
p5	\$0.00	\$0.00
p10	\$0.00	\$0.00
p25	\$0.00	\$0.00
p50	\$0.00	\$0.00
p75	\$0.00	\$0.00
p90	\$0.20	\$0.25
p95	\$0.36	\$0.71
p99	\$1.17	\$4.26

Figure 29: Representative TT-density Estimates: NLS vs. 2SLS



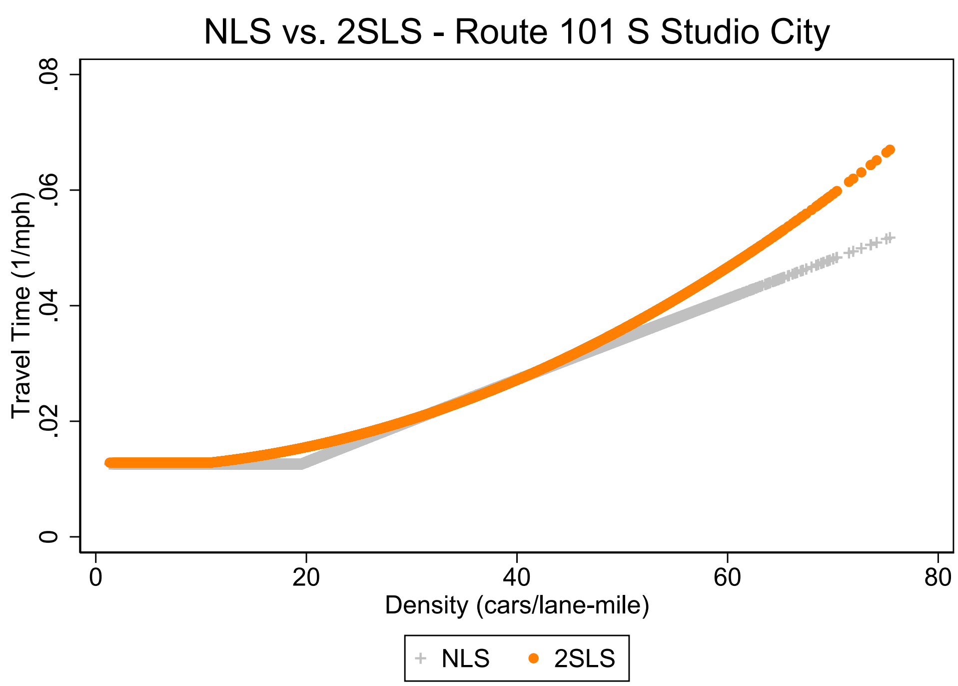


Figure 30 presents graphical evidence, plotting the distributions of positive marginal external cost estimates, from the two procedures for monitors in District 7. Here we see 2SLS yields fewer small positive estimates, but more moderately large estimates. Overall the distributions of estimates are not very similar. Comparisons of cost estimates under NLS and 2SLS for the other Caltrans districts will be instructive, however, this preliminary analysis suggests endogeneity may lead to a relatively large bias in the estimation of marginal external congestion costs using traffic speed and flow data. Therefore, we will use 2SLS estimates in the analysis below.

Figure 30: Comparison of MECs Calculated Using NLS and 2SLS Procedures

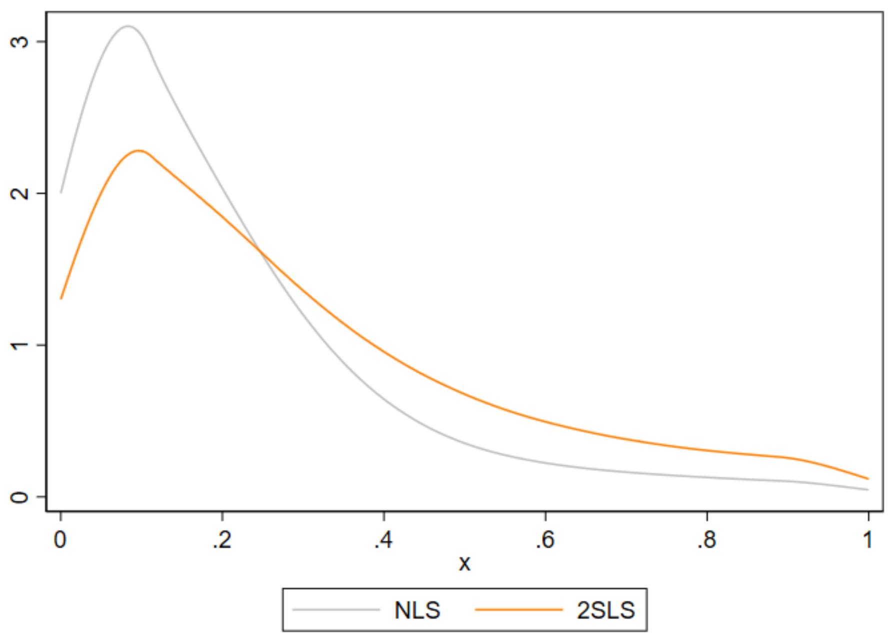


Table 20 summarizes the 2SLS marginal external cost estimates across the Caltrans districts in the PeMS network. Consistent with the summary statistics for highway speed presented in Table 17, the average highway location is, on average, uncongested. The median marginal external congestion cost is zero across every district. However, during congested periods, marginal external costs can be quite large. At the 95th percentile, marginal external congestion costs range from \$0.00 per vehicle mile in District 6 to \$1.54 per vehicle mile in District 7. Figure 31 presents the average MECs calculated using the 2SLS method at different PeMs monitors. The high MECs are clustered in specific areas. The NLS MEC estimates in Table 19 have smaller values, with \$0.07 to \$0.70 in the 95th percentile and \$0.27 to \$1.78 in the 99th percentile. These values are broadly consistent with average congestion cost values used in the literature Parry, Walls, and Harrington (2007).

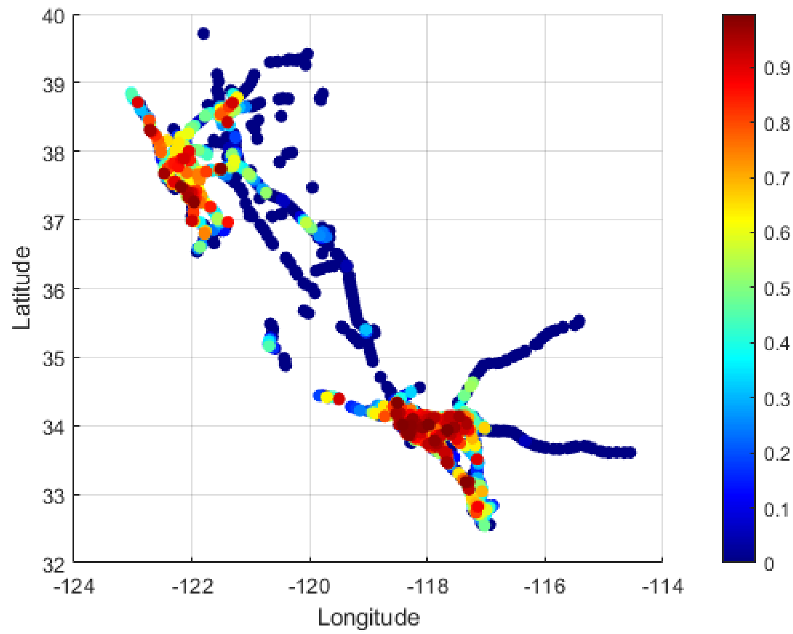
Table 19: NLS MEC Summary Statistics by Caltrans District

	MEC (\$/car-mile)						
District	Obs.	Mean	Min.	P5	P50	P95	P99
D3 - Sacramento	8,686,364	0.06	0.00	0.00	0.00	0.26	0.87
D4 - Bay Area	41,500,000	0.06	0.00	0.00	0.00	0.32	1.06
D5 - Central Coast	4,751,172	0.04	0.00	0.00	0.00	0.21	0.63
D6 - Central Valley	8,801,064	0.01	0.00	0.00	0.00	0.07	0.27
D7 - Los Angeles	32,800,000	0.13	0.00	0.00	0.00	0.70	1.78
D8 - Inland Empire	18,400,000	0.06	0.00	0.00	0.00	0.35	1.06
D10 - Stockton	10,700,000	0.03	0.00	0.00	0.00	0.16	0.45
D11 - San Diego	12,100,000	0.05	0.00	0.00	0.00	0.25	1.09
D12 - Orange County	16,700,000	0.05	0.00	0.00	0.00	0.32	0.69

Table 20: 2SLS MEC Summary Statistics by Caltrans District

	MEC (\$/car-mile)						
District	Obs.	Mean	Min.	P5	P50	P95	P99
D3 - Sacramento	8,703,896	0.22	0.00	0.00	0.00	0.45	3.73
D4 - Bay Area	41,500,000	0.30	0.00	0.00	0.00	0.62	4.45
D5 - Central Coast	4,751,172	0.12	0.00	0.00	0.00	0.20	1.97
D6 - Central Valley	8,818,596	0.04	0.00	0.00	0.00	0.00	0.29
D7 - Los Angeles	32,800,000	116.32	0.00	0.00	0.00	1.54	5.73
D8 - Inland Empire	18,400,000	0.21	0.00	0.00	0.00	0.69	3.79
D10 - Stockton	10,700,000	0.12	0.00	0.00	0.00	0.12	1.04
D11 - San Diego	12,100,000	0.17	0.00	0.00	0.00	0.39	3.76
D12 - Orange County	16,700,000	0.17	0.00	0.00	0.00	0.70	3.10

Figure 31: Average 2SLS MECs for All Districts: 3D Top View



Congestion costs vary substantially across the time dimension. Table 21 shows average 2SLS MECs, across all PeMS districts, by hour of day. Intuitively, marginal external costs are zero during the night and early morning hours when few vehicles are on the highway. Costs peak during the morning and evening commute hours, between 6 am and 9 am and between 4 pm and 5 pm. The 95th percentile of morning peak costs are approximately \$0.82 to \$2.06 per vehicle-mile and evening costs are approximately \$3.00 to \$3.22 per vehicle-mile. Using the NLS method, Table 22 shows similar peak hours with smaller MEC estimates. The 95th percentile of morning peak costs are approximately \$0.35 to \$0.60 per vehicle-mile and evening costs are approximately \$1.04 to \$1.08 per vehicle-mile.

Table 21: 2SLS MEC Summary Statistics by Hour of Day

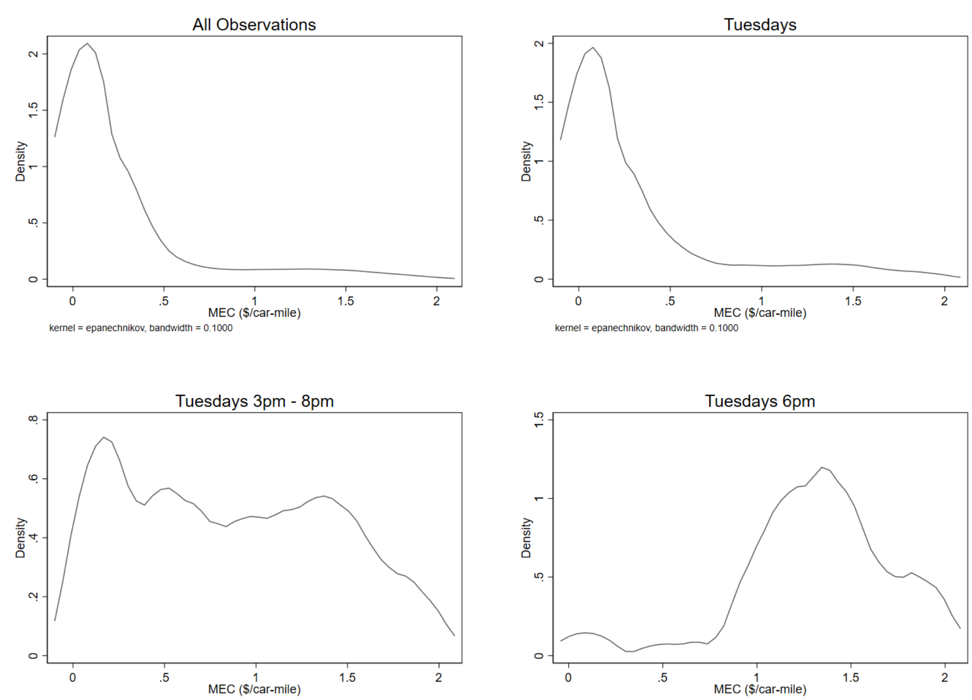
	MEC ($/car-mile)						
Hour	Obs.	Mean	Min.	P5	P50	P95	P99
0	6,444,058	0.00	0.00	0.00	0.00	0.00	0.00
1	6,444,038	0.00	0.00	0.00	0.00	0.00	0.00
2	6,426,420	0.00	0.00	0.00	0.00	0.00	0.00
3	6,444,034	0.00	0.00	0.00	0.00	0.00	0.00
4	6,444,020	0.02	0.00	0.00	0.00	0.00	0.00
5	6,444,132	3.40	0.00	0.00	0.00	0.03	0.63
6	6,444,140	142.39	0.00	0.00	0.00	0.82	4.43
7	6,444,136	152.62	0.00	0.00	0.00	2.06	9.50
8	6,444,130	251.41	0.00	0.00	0.00	1.94	9.45
9	6,444,131	19.50	0.00	0.00	0.00	0.89	5.31
10	6,444,136	2.95	0.00	0.00	0.00	0.59	3.04
11	6,444,132	1.50	0.00	0.00	0.00	0.59	2.58
12	6,444,143	2.91	0.00	0.00	0.00	0.72	2.90
13	6,444,143	3.04	0.00	0.00	0.00	0.91	3.37
14	6,435,183	0.60	0.00	0.00	0.00	1.43	4.90
15	6,435,172	3.65	0.00	0.00	0.00	2.33	7.16
16	6,435,176	4.65	0.00	0.00	0.00	3.00	9.02
17	6,435,181	5.52	0.00	0.00	0.00	3.22	10.01
18	6,435,183	1.26	0.00	0.00	0.00	1.43	5.47
19	6,435,187	0.54	0.00	0.00	0.00	0.31	1.81
20	6,435,298	0.03	0.00	0.00	0.00	0.08	0.48
21	6,435,200	0.06	0.00	0.00	0.00	0.00	0.15
22	6,435,207	0.02	0.00	0.00	0.00	0.00	0.07
23	6,435,213	0.00	0.00	0.00	0.00	0.00	0.00

Table 22: NLS MEC Summary Statistics by Hour of Day

	MEC ($/car-mile)						
Hour	Obs.	Mean	Min.	P5	P50	P95	P99
0	6,443,327	0.00	0.00	0.00	0.00	0.00	0.00
1	6,443,307	0.00	0.00	0.00	0.00	0.00	0.00
2	6,425,691	0.00	0.00	0.00	0.00	0.00	0.00
3	6,443,303	0.00	0.00	0.00	0.00	0.00	0.00
4	6,443,289	0.00	0.00	0.00	0.00	0.00	0.08
5	6,443,401	0.02	0.00	0.00	0.00	0.10	0.33
6	6,443,409	0.06	0.00	0.00	0.00	0.35	0.92
7	6,443,405	0.12	0.00	0.00	0.00	0.59	1.62
8	6,443,399	0.12	0.00	0.00	0.00	0.60	1.69
9	6,443,400	0.08	0.00	0.00	0.00	0.38	1.18
10	6,443,405	0.07	0.00	0.00	0.00	0.32	0.85
11	6,443,401	0.08	0.00	0.00	0.00	0.34	0.80
12	6,443,412	0.09	0.00	0.00	0.00	0.38	0.91
13	6,443,412	0.11	0.00	0.00	0.00	0.44	1.09
14	6,434,453	0.14	0.00	0.00	0.00	0.59	1.40
15	6,434,442	0.18	0.00	0.00	0.00	0.86	1.88
16	6,434,446	0.21	0.00	0.00	0.00	1.04	2.23
17	6,434,451	0.21	0.00	0.00	0.00	1.08	2.34
18	6,434,453	0.12	0.00	0.00	0.00	0.65	1.75
19	6,434,457	0.05	0.00	0.00	0.00	0.25	0.85
20	6,434,568	0.02	0.00	0.00	0.00	0.10	0.35
21	6,434,470	0.01	0.00	0.00	0.00	0.03	0.15
22	6,434,477	0.00	0.00	0.00	0.00	0.09	0.09
23	6,434,483	0.00	0.00	0.00	0.00	0.02	0.02

These hourly averages are consistent with variations in marginal external congestion costs within particular highway locations. Figure 32 illustrates this fact for a representative location in District 7 using NLS MECs. The first panel plots the density of costs across all weekdays and hours of the day. Mean costs are centered around zero. The long right tail extends to costs exceeding \$1.00 per vehicle-mile. The second panel focuses on Tuesdays. Again, there is substantial variation in costs across hours of the day. The third panel focuses on hours around the afternoon peak. This distribution is multimodal roughly corresponding to hourly peaks during the period. The fourth panel focuses on a single hour, 6 pm. Here we see the hourly averages hide substantial variations in costs within an hour. Mean costs are approximately \$1.50 per vehicle-hour but can be as high low as \$1.00 and as high as \$2.00 per vehicle lane-mile. This variability suggests average tolls will likely do a poor job of correcting traffic externalities. We quantify these shortcomings in the section below.

Figure 32: Distribution of MEC Across Different time periods for Representative monitor (Route 134 E Glandale)



17.2 Distributional Effect of Second-Best Tolling

The variability in marginal external congestion costs documented in Figure 32 suggests congestion costs based on average values may substantially overstate or understate costs much, if not most, of the time. Because measuring actual costs may be expensive or impractical in settings with substantial variability, policymakers instead set second-best policies (taxes) based on average values. We imagine several policy scenarios, denoting the level at which second-best congestion tolls based on average congestion costs across various time and spatial dimensions, e.g. tolls based on monitor average MECs that do not vary with time, average hourly tolls that are the same across locations, and tolls vary by each monitor and hour. We evaluate the tolls that households would face under these different policies. Ultimately we will use the optimal (real-time) congestion toll to check the toll incidence for different households, for now, we use the second-best averages.

We compute the average toll for a single trip under each policy and then aggregate the tolls paid by households for all their trips in the sample. We then regress the total tolls in each scenario on household demographic variables, such as income and race, as shown in Equation 26. These variables help us understand the distributional effects of the toll policies within the sample. Currently, we use the average MECs in 2019 as tolls in 2017, because we do not have the precise trip start and end locations and the route information, eventually, we will use the actual MECs. Future iterations will also use NHTS sampling weights to construct population estimates.

$$Toll_h = \beta_0 + \beta_1 Demogr_h + \epsilon_h \tag{26}$$

Table 23 illustrates that households with higher incomes bear a larger burden of congestion tolls under all three types of policies, particularly under the location-specific policy. For instance, households with an income above \$200,000 spend \$0.9785 more per day under monitor-based tolls, \$0.3693 more per day under hourly-based tolls, and \$1.2388 more per day under monitor and hourly-based tolls than households with an income below \$35,000. To examine the toll increment for each income group over the course of a year, we conduct a back-of-the-envelope calculation and divide the annual increment in toll spending by the median income of each income group. The results are presented in Table 24. we can see the percentage expenditure in toll to income is also higher for richer households under location-based tolls, not just the absolute expenditure. Households with an income higher than \$200,000 spend 0.18% more their income in tolls than poor households. Figure 33 presents the increasing pattern of the percentage expenditure more straightforwardly.

Table 23: Distribution of Household Total Tolls per Day by Income Level

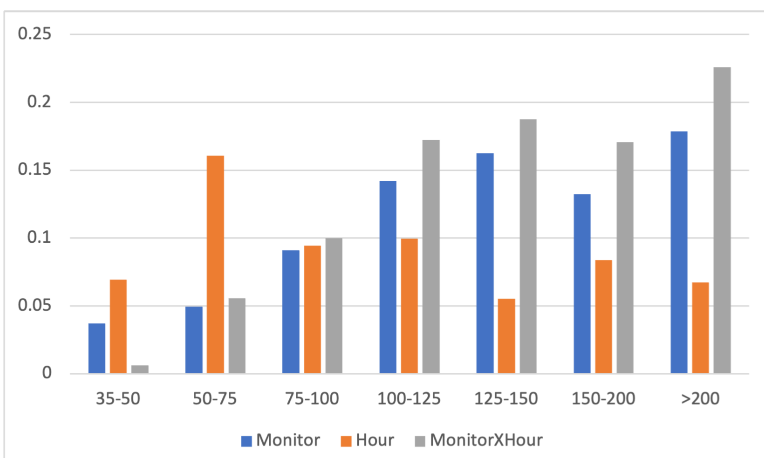
	Second-best Policy		
Income(k)	Monitor	Hour	Monitor×Hour
35-50	0.0436	0.0816	0.0074
	(0.0457)	(0.0958)	(0.0565)
50-75	0.0857**	0.2775***	0.0959*
	(0.0392)	(0.0890)	(0.0513)
75-100	0.2196***	0.2280***	0.2408***
	(0.0407)	(0.0868)	(0.0526)
100-125	0.4397***	0.3088***	0.5336***
	(0.0489)	(0.0952)	(0.0630)
125-150	0.6141***	0.2098*	0.7093***
	(0.0662)	(0.1072)	(0.0792)
150-200	0.6342***	0.4020***	0.8176***
	(0.0578)	(0.1112)	(0.0738)
≥200	0.9785***	0.3693***	1.2388***
	(0.0565)	(0.0982)	(0.0737)
Obs.	15,149	9,130	15,072

These results indicate that the second-best tolls based on average MECs are progressive, possibly because richer households can afford more vehicle trips, and they may drive more

Table 24: Percentage Annual Toll Expenditure Increment by Income Level

	Percentage Toll Expenditure (%)		
Income(k)	Monitor	Hour	Monitor×Hour
35-50	0.04	0.07	0.01
50-75	0.05	0.16	0.06
75-100	0.09	0.09	0.10
100-125	0.14	0.10	0.17
125-150	0.16	0.06	0.19
150-200	0.13	0.08	0.17
≥200	0.18	0.07	0.23

Figure 33: Percentage Annual Toll Expenditure Increment by Income Level



frequently on more popular (congested) routes. We explore the possible reasons by looking at the average toll spending across trips within each household and present the results in Table 25. The monitor-based toll expenditure still increases with household income, but not the hourly-based toll expenditure. This implies that richer households travel more in congested areas, but may be less in congested hours.

Table 25: Distribution of Household Mean Tolls per Day by Income Level

	Second-best Policy		
Income(k)	Monitor	Hour	Monitor×Hour
35-50	-0.0055	-0.0259	-0.0003
	(0.0055)	(0.0364)	(0.0087)
50-75	0.0004	0.0151	0.0090
	(0.0049)	(0.0325)	(0.0074)
75-100	0.0115**	-0.0212	0.0204***
	(0.0049)	(0.0327)	(0.0074)
100-125	0.0235***	-0.0249	0.0378***
	(0.0050)	(0.0347)	(0.0078)
125-150	0.0464***	-0.0918**	0.0543***
	(0.0064)	(0.0383)	(0.0087)
150-200	0.0459***	-0.0738**	0.0701***
	(0.0055)	(0.0372)	(0.0085)
≥200	0.0690***	-0.0890***	0.0932***
	(0.0052)	(0.0341)	(0.0078)
Obs.	15,149	9,130	15,072

Table 26 provides insights into the impact of highway toll policies on different racial groups. The coefficients in the table demonstrate the difference in the congestion tolls paid by each non-white racial group compared to the white racial group. The results reveal that under the monitor-based toll policy, Black and Asian households pay a higher toll than white households by \$0.5145 and \$0.6705, respectively, both of which are statistically significant. In contrast, Indian households pay a lower toll than white households (-\$0.4226). The coefficients for different racial groups under the hour-based policy are not statistically significant, and those for the monitor×hour-based policy are similar to those for the monitor-based policy. Table 27 aggregates the daily toll increment to an annual level for the three racial groups and the two policies with statistically significant results. Compared to white households, Black households spend \$187.79 more under the monitor-based toll, Asian households spend \$244.73 more, and Indian households spend \$154.25 less.

Table 26: Distribution of Household Total Tolls per Day by Race

	Second-best Policy		
Race	Monitor	Hour	Monitor×Hour
Black	0.5145***	-0.0471	0.4779***
	(0.1077)	(0.1341)	(0.1204)
Asian	0.6705***	0.0183	0.8187***
	(0.0481)	(0.0846)	(0.0672)
Indian	-0.4226***	0.1350	-0.4427**
	(0.1134)	(0.4586)	(0.1775)
Islander	-0.0618	-0.1169	-0.1607
	(0.1285)	(0.3249)	(0.1515)
Multiple	0.0737	-0.2069*	0.0848
	(0.0675)	(0.1190)	(0.0805)
Other	0.2419***	-0.0889	0.2999***
	(0.0890)	(0.1502)	(0.1114)
Obs.	15,432	9,281	15,353

Table 27: Annual Toll Expenditure Increment by Race

	Toll Expenditure (\$)	
Race	Monitor	Monitor×Hour
Black	187.79	174.43
Asian	244.73	298.83
Indian	-154.25	-161.59

When examining household average toll spending in Table 28, Black and Asian households also spend about \$0.07 more per day than white households under monitor-based policies, indicating that Black and Asian households travel more in congested areas. In contrast, Indian households spend less, indicating they travel less in congested areas. For the hour-based toll policy, Asian households spend less on average than white households, indicating they travel less during congested hours.

Table 28: Distribution of Household Mean Tolls by Race

	Second-best Policy		
Race	Monitor	Hour	Monitor×Hour
Black	0.0710***	-0.0180	0.0835***
	(0.0081)	(0.0515)	(0.0123)
Asian	0.0759***	-0.1205***	0.0877***
	(0.0044)	(0.0276)	(0.0068)
Indian	-0.0483***	-0.1001	-0.0695***
	(0.0160)	(0.1250)	(0.0249)
Islander	-0.0005	-0.0860	-0.0199
	(0.0150)	(0.1126)	(0.0225)
Multiple	0.0025	-0.0962**	0.0075
	(0.0064)	(0.0438)	(0.0097)
Other	0.0070	-0.1396***	0.0080
	(0.0072)	(0.0469)	(0.0112)
Obs.	15,432	9,281	15,353

18 Conclusions

Using detailed data on traffic flows on California highways, we construct estimates of the travel time-vehicle density relationship and calculate the marginal external costs of traffic congestion for each monitor-hour in the sample. The study finds that the marginal external costs of traffic congestion are generally low, but can be substantial in congested cities during peak travel periods. We use these estimates for marginal external congestion costs as the basis for several potential second-best tolling policies. By utilizing detailed trip data from the

NHTS, we assign tolls to households and investigate the incidence of these tolls. The analysis reveals that households in richer income groups pay higher tolls under both location-based and time-based toll policies. Furthermore, Black and Asian households pay higher tolls under location-based policies. Future research will refine the analysis of the distributional effect of the marginal external congestion costs and examine additional household demographics and travel behaviors. Additionally, we have yet to consider spillover effects on roads other than highways, but this is a topic we plan to investigate in the future.

References

Alvarez, Robert, Martin Weilenmann, and Jean-Yves Favez. 2008. "Evidence of increased mass fraction of NO2 within real-world NOx emissions of modern light vehicles - derived from a reliable online measuring method." *Atmospheric Environment* 42 (19):4699–4707.

Arnott, Richard, André De Palma, and Robin Lindsey. 1994. "The welfare effects of congestion tolls with heterogeneous commuters." *Journal of Transport Economics and Policy* :139–161.

Auffhammer, Maximilian and Ryan Kellogg. 2011. "Clearing the air? The effects of gasoline content regulation on air quality." *American Economic Review* 101 (6):2687–2722.

Bento, Antonio, Kevin Roth, and Andrew R Waxman. 2020. "Avoiding traffic congestion externalities? the value of urgency." Tech. rep., National Bureau of Economic Research.

Berger, Joseph. 2008. "Congestion Pricing: Just Another Regressive Tax?" *The New York Times* URL `http://www.keepnycfree.com/media/files/2008-04-20_Regressive_Tax(NYTimes).pdf`.

Birkeland, Sarah. 2001. "EPA's TMDL program." *Ecology Law Quarterly* :297–325.

Boehmke, Frederick J, Daniel S Morey, and Megan Shannon. 2006. "Selection bias and continuous-time duration models: consequences and a proposed solution." *American Journal of Political Science* 50 (1):192–207.

California Department of Transportation. 2020. "Performance Measurement System." http://pems.dot.ca.gov/, Accessed on April 29, 2020.

Carslaw, David C. 2005. "Evidence of an increasing NO2/NOx emissions ratio from road traffic emissions." *Atmospheric Environment* 39 (26):4793–4802.

Chatterjee, Chiradip, Russell Triplett, Christopher K Johnson, and Parvez Ahmed. 2017. "Willingness to pay for safe drinking water: A contingent valuation study in Jacksonville, FL." *Journal of environmental management* 203:413–421.

Chay, Kenneth Y and Michael Greenstone. 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.

Chen, Sen, Xun Zheng, Haitao Yin, and Ya Liu. 2020. "Did Chinese cities that implemented driving restrictions see reductions in PM10?" *Transportation Research Part D: Transport and Environment* 79:102208.

Chen, Yuyu, Ginger Zhe Jin, Naresh Kumar, and Guang Shi. 2012. "Gaming in air pollution data? Lessons from China." *The BE Journal of Economic Analysis & Policy* 12 (3).

Chen, Zhao, Matthew E Kahn, Yu Liu, and Zhi Wang. 2018. "The consequences of spatially differentiated water pollution regulation in China." *Journal of Environmental Economics and Management* 88:468–485.

Cohen, Aaron J, H Ross Anderson, Bart Ostro, Kiran Dev Pandey, Michal Krzyzanowski, Nino Künzli, Kersten Gutschmidt, Arden Pope, Isabelle Romieu, Jonathan M Samet et al. 2005. "The global burden of disease due to outdoor air pollution." *Journal of Toxicology and Environmental Health, Part A* 68 (13-14):1301–1307.

Craig, Robin Kundis and Anna M Roberts. 2015. "When will governments regulate nonpoint source pollution: a comparative perspective." *BC Envtl. Aff. L. Rev.* 42:1.

Crump, Richard K, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik. 2009. "Dealing with limited overlap in estimation of average treatment effects." *Biometrika* 96 (1):187–199.

Davis, Lucas W. 2008. "The effect of driving restrictions on air quality in Mexico City." *Journal of Political Economy* 116 (1):38–81.

Dawid, Irvin. 2019. "If Road Pricing is Inherently Unfair and Regressive..." *Planetizen* URL `https://www.planetizen.com/news/2019/02/102730-if-road-pricing-inherently-unfair-and-regressive`.

de Chaisemartin, Clément and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110 (9):2964–96.

de Grange, Louis and Rodrigo Troncoso. 2011. "Impacts of vehicle restrictions on urban transport flows: the case of Santiago, Chile." *Transport Policy* 18 (6):862–869.

EPA. 2001. "The National Costs of the Total Maximum Daily Load Program."

———. 2011. "A National Evaluation of the Clean Water Act Section 319 Program." .

Eskeland, Gunnar S and Tarhan Feyzioglu. 1997. "Rationing can backfire: the "day without a car" in Mexico City." *The World Bank Economic Review* 11 (3):383–408.

Fowlie, Meredith, Claire Petersen, and Mar Reguant. 2021. "Border carbon adjustments when carbon intensity varies across producers: Evidence from california." In *AEA Papers and Proceedings*, vol. 111. 401–05.

Fu, Shihe and Yizhen Gu. 2017. "Highway toll and air pollution: Evidence from Chinese cities." *Journal of Environmental Economics and Management* 83:32–49.

GAO, US Government Accountability Office. 2000. "Key EPA and State Decisions Limited by Inconsistent and Incomplete Data."

Ghanem, Dalia and Junjie Zhang. 2014. "¡®Effortless Perfection:¡¯Do Chinese cities manipulate air pollution data?" *Journal of Environmental Economics and Management* 68 (2):203–225.

Goodman-Bacon, Andrew. 2018. "Difference-in-differences with variation in treatment timing." Tech. rep., National Bureau of Economic Research.

Greenstone, Michael and Rema Hanna. 2014. "Environmental regulations, air and water pollution, and infant mortality in India." *American Economic Review* 104 (10):3038–72.

Gu, Yizhen, Elizabeth Deakin, and Ying Long. 2017. "The effects of driving restrictions on travel behavior evidence from Beijing." *Journal of Urban Economics* 102:106–122.

Hausman, Catherine and David S Rapson. 2018. "Regression discontinuity in time: Considerations for empirical applications." *Annual Review of Resource Economics* 10:533–552.

Helland, Eric and Andrew B Whitford. 2003. "Pollution incidence and political jurisdiction: evidence from the TRI." *Journal of Environmental Economics and management* 46 (3):403–424.

Hite, Diane, Darren Hudson, and Walaiporn Intarapapong. 2002. "Willingness to pay for water quality improvements: The case of precision application technology." *Journal of Agricultural and resource Economics* :433–449.

Huang, Hong, Ryozo Ooka, Hong Chen, Shinsuke Kato, Takeo Takahashi, and Takeaki Watanabe. 2008. "CFD analysis on traffic-induced air pollutant dispersion under non-isothermal condition in a complex urban area in winter." *Journal of wind engineering and industrial aerodynamics* 96 (10-11):1774–1788.

Hughes, Jonathan E and Daniel Kaffine. 2019. "When should drivers be encouraged to carpool in hov lanes?" *Economic Inquiry* 57 (1):667–684.

Imai, Kosuke and In Song Kim. 2019. "On the use of two-way fixed effects regression models for causal inference with panel data." Tech. rep., Harvard University IQSS Working Paper). Cambridge, MA. Retrieved from https ¡.

Itano, Yasuyuki, Makiko Yamagami, and Toshimasa Ohara. 2014. "Estimation of primary NO2/NOx emission ratio from road vehicles using ambient monitoring data." *Stud. Atmos. Sci* 1 (2):1–7.

Ito, Koichiro and Shuang Zhang. 2016. "Willingness to pay for clean air: Evidence from air purifier markets in China." .

Jacobsen, Mark R, Christopher R Knittel, James M Sallee, and Arthur A Van Benthem. 2020. "The use of regression statistics to analyze imperfect pricing policies." *Journal of Political Economy* 128 (5):1826–1876.

Jin, Xiaomeng, Arlene M Fiore, Lee T Murray, Lukas C Valin, Lok N Lamsal, Bryan Duncan, K Folkert Boersma, Isabelle De Smedt, Gonzalo Gonzalez Abad, Kelly Chance et al. 2017. "Evaluating a space-based indicator of surface ozone-NOx-VOC sensitivity over midlatitude source regions and application to decadal trends." *Journal of Geophysical Research: Atmospheres* 122 (19):10–439.

Jordan, Jeffrey L and Abdelmoneim H Elnagheeb. 1993. "Willingness to pay for improvements in drinking water quality." *Water resources research* 29 (2):237–245.

Ke, Yue and Konstantina Gkritza. 2018. "Income and spatial distributional effects of a congestion tax: A hypothetical case of Oregon." *Transport Policy* 71:28–35.

Keiser, David A, Catherine L Kling, and Joseph S Shapiro. 2019. "The low but uncertain measured benefits of US water quality policy." *Proceedings of the National Academy of Sciences* 116 (12):5262–5269.

Keiser, David A, Sheila M Olmstead, Kevin J Boyle, Victor B Flatt, Bonnie L Keeler, Catherine L Kling, Daniel J Phaneuf, Joseph S Shapiro, and Jay P Shimshack. 2021. "A water rule that turns a blind eye to transboundary pollution." *Science* 372 (6539):241–243.

Keiser, David A, Sheila M Olmstead, Kevin J Boyle, Victor B Flatt, Bonnie L Keeler, Daniel J Phaneuf, Joseph S Shapiro, and Jay P Shimshack. 2022. "The Evolution of the ¡°Waters of the United States¡± and the Role of Economics." *Review of Environmental Economics and Policy* 16 (1):146–152.

Keiser, David A and Joseph S Shapiro. 2019. "US water pollution regulation over the past half century: burning waters to crystal springs?" *Journal of Economic Perspectives* 33 (4):51–75.

Kelderman, Keene, Ari Phillips, Tom Pelton, Eric Schaeffer, Paul MacGillis-Falcon, and Courtney Bernhardt. 2022. "The Clean Water Act at 50: Promises Half Kept at the Half-Century Mark." *The Environmental Integrity Project* .

Knittel, Christopher R, Douglas L Miller, and Nicholas J Sanders. 2016. "Caution, drivers! Children present: Traffic, pollution, and infant health." *Review of Economics and Statistics* 98 (2):350–366.

Krol, Robert. 2016. "Tolling the freeway: congestion pricing and the economics of managing traffic." *MERCATUS RESEARCH* .

Li, Can, Si-Chee Tsay, Joshua S Fu, Russell R Dickerson, Qiang Ji, Shaun W Bell, Yang Gao, Wu Zhang, Jianping Huang, Zhanqing Li et al. 2010. "Anthropogenic air pollution observed near dust source regions in northwestern China during springtime 2008." *Journal of Geophysical Research: Atmospheres* 115 (D7).

Liang, Xuan, Shuo Li, Shuyi Zhang, Hui Huang, and Song Xi Chen. 2016. "PM2. 5 data reliability, consistency, and air quality assessment in five Chinese cities." *Journal of Geophysical Research: Atmospheres* 121 (17):10–220.

Lipscomb, Molly and Ahmed Mushfiq Mobarak. 2016. "Decentralization and pollution spillovers: evidence from the re-drawing of county borders in Brazil." *The Review of Economic Studies* 84 (1):464–502.

List, John A, Erwin H Bulte, and Jason F Shogren. 2002. "¡°Beggar thy neighbor:¡± Testing for free riding in state-level endangered species expenditures." *Public Choice* 111 (3):303–315.

List, John A and Shelby Gerking. 2000. "Regulatory federalism and environmental protection in the United States." *Journal of Regional Science* 40 (3):453–471.

Liu, Yan, Zhijun Yan, and Chao Dong. 2016. "Health implications of improved air quality from Beijing's driving restriction policy." *Environmental pollution* 219:323–328.

Liu, Zhiyong, Ruimin Li, Xiaokun Cara Wang, and Pan Shang. 2020. "Noncompliance behavior against vehicle restriction policy: A case study of Langfang, China." *Transportation Research Part A: Policy and Practice* 132:1020–1033.

Mason, John C and David C Handscomb. 2002. *Chebyshev polynomials*. CRC press.

McCubbin, Donald R and Mark A Delucchi. 1999. "The health costs of motor-vehicle-related air pollution." *Journal of Transport Economics and Policy* :253–286.

MEE. 2012. "Ministry of Ecology and Environment of the People's Republic of China: Technical Regulation on Ambient Air Quality Index." .

———. 2016a. "Ministry of Ecology and Environment of the People's Republic of China: China ecological environment annual report." .

———. 2016b. "Ministry of Ecology and Environment of the People's Republic of China: China mobile source environmental management annual report." .

———. 2017. "Ministry of Ecology and Environment of the People's Republic of China: China ecological environment annual report." .

———. 2019. "Ministry of Ecology and Environment of the People's Republic of China: China mobile source environmental management annual report." .

Millimet, Daniel L. 2003. "Assessing the empirical impact of environmental federalism." *Journal of Regional Science* 43 (4):711–733.

Mullins, Jamie and Prashant Bharadwaj. 2015. "Effects of short-term measures to curb air pollution: Evidence from Santiago, Chile." *American Journal of Agricultural Economics* 97 (4):1107–1134.

Murphy, Jarrett. 2017. "Debate Fact Check: Is Congestion Pricing Regressive?" *City Limits* URL `https://citylimits.org/2017/09/07/debate-fact-check-is-congestion-pricing-regressive/`.

NRC, National Research Council. 2001. *Assessing the TMDL approach to water quality management*. National Academies Press.

Oates, Wallace E and Robert M Schwab. 1988. "Economic competition among jurisdictions: efficiency enhancing or distortion inducing?" *Journal of public economics* 35 (3):333–354.

Parry, Ian WH, Margaret Walls, and Winston Harrington. 2007. "Automobile externalities and policies." *Journal of economic literature* 45 (2):373–399.

Salvo, Alberto and Yi Wang. 2017. "Ethanol-blended gasoline policy and ozone pollution in Sao Paulo." *Journal of the Association of Environmental and Resource Economists* 4 (3):731–794.

Santos, Georgina and Laurent Rojey. 2004. "Distributional impacts of road pricing: The truth behind the myth." *Transportation* 31:21–42.

Schlenker, Wolfram and W Reed Walker. 2016. "Airports, air pollution, and contemporaneous health." *The Review of Economic Studies* 83 (2):768–809.

Schweitzer, Lisa and Brian D Taylor. 2008. "Just pricing: the distributional effects of congestion pricing and sales taxes." *Transportation* 35:797–812.

Sigman, Hilary. 2002. "International spillovers and water quality in rivers: do countries free ride?" *American Economic Review* 92 (4):1152–1159.

———. 2003. "Letting states do the dirty work: State responsibility for federal environmental regulation." *National Tax Journal* 56 (1):107–122.

———. 2005. "Transboundary spillovers and decentralization of environmental policies." *Journal of environmental economics and management* 50 (1):82–101.

———. 2014. "Decentralization and environmental quality: an international analysis of water pollution levels and variation." *Land Economics* 90 (1):114–130.

Silva, Emilson CD and Arthur J Caplan. 1997. "Transboundary pollution control in federal systems." *Journal of environmental economics and management* 34 (2):173–186.

Sini, Jean-François, Sandrine Anquetin, and Patrice G Mestayer. 1996. "Pollutant dispersion and thermal effects in urban street canyons." *Atmospheric environment* 30 (15):2659–2677.

Small, Kenneth A and Camilla Kazimi. 1995. "On the costs of air pollution from motor vehicles." *Journal of Transport Economics and policy* :7–32.

Small, Kenneth A, Clifford Winston, and Jia Yan. 2005. "Uncovering the distribution of motorists' preferences for travel time and reliability." *Econometrica* 73 (4):1367–1382.

Steigerwald, Douglas G, Gonzalo Vazquez-Bare, and Jason Maier. 2019. "Measuring the Effect of Environmental Policies using Panel Data." .

Stoerk, Thomas. 2016. "Statistical corruption in Beijing's air quality data has likely ended in 2012." *Atmospheric Environment* 127:365–371.

Sun, Cong, Siqi Zheng, and Rui Wang. 2014. "Restricting driving for better traffic and clearer skies: Did it work in Beijing?" *Transport Policy* 32:34–41.

Troncoso, Rodrigo, Louis De Grange, and Luis A Cifuentes. 2012. "Effects of environmental alerts and pre-emergencies on pollutant concentrations in Santiago, Chile." *Atmospheric environment* 61:550–557.

UCLA. 2019. "Congestion Pricing." *UCLA Institute of Transportation Studies* URL `https://www.its.ucla.edu/news/for-the-press/congestion-pricing/`.

Viard, V Brian and Shihe Fu. 2015. "The effect of Beijing's driving restrictions on pollution and economic activity." *Journal of Public Economics* 125:98–115.

Wang, LT, Q Zhang, JM Hao, and KB He. 2005. "Anthropogenic CO emission inventory of Mainland China." *Acta Scientiae Circumstantiae* 25 (12):1580–1585.

Wild, Robert J, William P Dubé, Kenneth C Aikin, Scott J Eilerman, J Andrew Neuman, Jeff Peischl, Thomas B Ryerson, and Steven S Brown. 2017. "On-road measurements of vehicle NO2/NOx emission ratios in Denver, Colorado, USA." *Atmospheric Environment* 148:182–189.

Yang, Jun, Avralt-Od Purevjav, and Shanjun Li. 2020. "The marginal cost of traffic congestion and road pricing: evidence from a natural experiment in Beijing." *American Economic Journal: Economic Policy* 12 (1):418–53.

Ye, Jingjing. 2017. "Better safe than sorry? Evidence from Lanzhou's driving restriction policy." *China Economic Review* 45:1–21.

Zhang, Renyi, Gehui Wang, Song Guo, Misti L Zamora, Qi Ying, Yun Lin, Weigang Wang, Min Hu, and Yuan Wang. 2015. "Formation of urban fine particulate matter." *Chemical Reviews* 115 (10):3803–3855.

Zhang, Wei, C-Y Cynthia Lin Lawell, and Victoria I Umanskaya. 2017. "The effects of license plate-based driving restrictions on air quality: Theory and empirical evidence." *Journal of Environmental Economics and Management* 82:181–220.

Zhao, B and JZ Ma. 2008. "Development of an air pollutant emission inventory for Tianjin." *Acta Scientiae Circumstantiae* 28 (2):368–375.

Appendix

A Potential Selection Bias in the Survival Analysis of TMDL Development

I run the Weibull duration model without the first-step sample selection to explore the selection bias. Table 29 compares the constant terms in the two models for each regression. The independent duration model uses a sample of impaired catchments, the constant term increases from about -36 in the selected model to about -34, which indicates a larger hazard rate of TMDL development.

Table 29: Comparison of the Constant Terms

Constant	Equation 12	Equation 13	Equation 14
With Selection	-36.3063***	-36.3287***	-36.3235***
	(0.1521)	(0.1523)	(0.1523)
No Selection	-34.3832***	-34.4041***	-34.3992***
	(0.1427)	(0.1429)	(0.1429)

The coefficients of the near-boundary and interstate catchment dummies in the two models are plotted in Figure 41. Panel (a) shows that in the independent duration model, the states have much lower hazards to develop TMDLs for the near-boundary catchments. The difference between the coefficients of the two models is the smallest for the border catchments (within 1 km to state borders), the largest for catchments next to the border catchments, then the difference shrinks as the catchments become further away from the state borders. These differences indicate that states have a lower hazard to develop TMDLs for the impaired near-boundary catchments than for the near-boundary catchments in general, especially the catchments that are closer to state borders but do not lie on the borders. Panel (b) shows that the coefficients for different portions of interstate rivers are similar in the two models. The independent duration model has slightly more positive estimates in the downstream, indicating that states develop TMDLs for the impaired downstream catchments a bit sooner than the downstream catchments in general. Apart from the magnitude of the coefficients, the changing pattern of the proportional hazards in the two models are very similar.

B Back-of-the-Envelope Calculations for Pollutants Shares

The mobile sources of CO emissions in mainland China from 2015 to 2018 are 34.61, 34.19, 33.27, 30.89 million tons, respectively (MEE (2019)). The detailed emission inventory list is not accessible, but we can find some evidence in the literature. In 2001, CO emissions from mobile sources were 32.63 million tons in mainland China, which is similar to the amounts in 2017 and 2018, and it accounts for 21.75% total human source CO emissions (Wang et al. (2005)). In 2003, coal-fire sources contributed to 58% CO emissions in Tianjin (Zhao and Ma (2008)), which indicates that mobile sources accounted for at most 42% of CO emissions. The vehicle share of CO is different across cities and time periods, but we can use the statistics in the literature to do inference. Wang et al. (2005)'s statistics should be close to the average value in China, while Zhao and Ma (2008)'s statistics can be used as an upper bound. Thus, mobile sources account for about 20% to 40% of CO emissions in China. The shares of vehicle-emitted CO to the total mobile-source CO from 2015 to 2018 are 86.9%, 87.7%, 87.8%, and 92.6% (MEE (2019)). The average is 88.75%. With a simple multiplication, vehicle emissions contribute about 17.75% to 35.50% CO.

The mobile-source contribution to NO_x was 31.64% on average in mainland China, 2015

(MEE (2016a)). This percentage varies across different cities, ranging from 10% to 50% (MEE (2019)). NO_x mainly consists of NO and NO_2. The NO_2/NO_x ratio has increased during these years. The chemical reaction between NO and O_3 forms NO_2, so the NO_2/NO_x ratio varies with the concentration of O_3. The NO_2/NO_x ratio from vehicle emissions has a large range from 2% to 70% (Carslaw (2005), Itano, Yamagami, and Ohara (2014), Wild et al. (2017), Alvarez, Weilenmann, and Favez (2008)). In addition, not all NO_2 is directly emitted from vehicles, some is formed by NO_x. The shares of vehicle-emitted NO_2 to the total mobile-source NO_2 from 2015 to 2018 were 92.2%, 92.5%, 92.8% and 92.7% (MEE (2019)). The average is 92.55%. Using the mobile-source NO_x multiply the NO_2/NO_x ratio and the vehicle-to-mobile share NO_2, I calculate a vehicle share NO_2 roughly ranges from 0.19% to 32.40%.

In the 2018 China Annual Environmental Report, SO_2 from vehicle emissions accounts for less than 0.3%. The 2010 MIX Asian Emission Inventory shows that 0.82% SO_2 in mainland China comes from the Transportation sector. Figure 45 in the Appendix shows increasingly stringent automobile diesel and gasoline standards in sulfur. The percentages of vehicle ownership following different standards in 2015 are 7.5% for Standard I and below, 8% for Standard II, 51.6% for Standard III, 30.5% for Standard IV, and 1.4% for Standard V and above. These percentages change to 3%, 4.5%, 19.1%, 42.5%, 30.9%, respectively, in 2018 (MEE (2016b), MEE (2019)). With cleaner fuel and cleaner vehicles, the same amount of vehicles on roads leads to lower sulfur emissions. In a word, vehicle emissions contribute little to the total SO_2 in mainland China.

O_3 is not directly emitted by vehicles but formed through the chemical reaction of NO_x and VOC with heat and sunlight. The two regimes of ozone formation are the NO_x-limited regime and the NO_x-saturated regime (McCubbin and Delucchi (1999)). Salvo and Wang (2017) concludes that the NO_x-limited regime usually applies to the rural areas, the amount of O_3 increases with NO_x, and it is insensitive to VOC; the NO_x-saturated regime usually applies to urban areas, and O_3 increases with VOC and decreases with NO_x. Jin et al. (2017) studies New York, London, and Seoul and finds increasingly longer NOx-limited O_3 chemistry in the warm season. The changing formation regimes of O_3 across spaces and time make it difficult to infer O_3's vehicle share.

The share of the mobile-source PM also varies widely across cities. Mobile source contribution to $PM2.5$ ranges from 45% in Beijing to 13.5% in Hengshui in 2017 (MEE (2017)). Zhang, Lawell, and Umanskaya (2017) finds that 1.8% $PM10$ emissions in the US come from on-road mobile sources, 24% $PM10$ emissions in the UK comes from transportation sources. Therefore, it is also difficult to determine a range of PM's vehicle shares.

C Regression Discontinuity in Time Approach at City Level

Table 30: Influence of Pollution Concentration on the City-specific RDiT Estimation Results

	AQI	CO	NO_2	SO_2	$O3$	$8-hO_3$	$PM2.5$	$PM10$
Pollution	0.0023	0.0001	0.0051	0.0010	0.0024	0.0026	0.0026	0.0019
	(0.0002)	(0.0000)	(0.0004)	(0.0001)	(0.0003)	(0.0003)	(0.0002)	(0.0001)
Obs.	357	357	357	357	357	357	357	357
R-sq.	0.5168	0.3238	0.3539	0.0954	0.2094	0.1873	0.6194	0.5432

Note: Numbers in parentheses are standard errors.

I conduct city-level RDiT analysis and regress the city-level RDiT estimation results on the sets of control variables same as in Section 13.2. The density distributions of the RDiT estimation results are presented in Figure 34. The density distributions look more multimodal than Figure 22 because the number of estimation results is smaller. But we can still observe the negative shifts of the estimation results from air pollution alert triggered policies to traffic-triggered and seasonal and regular policies. The second-stage regression of the RDiT estimation results on the two most influential factors, pollution concentration and policy details, are presented in Table 30 and Table 31 with robust standard errors.

Table 31: Influence of Policy Details on the City-specific RDiT Estimation Results

	AQI	CO	NO_2	SO_2	$O3$	$8-hO_3$	$PM2.5$	$PM10$
				Implementation Criteria				
Traffic	-0.3210	-0.2442	-0.1735	-0.0800	0.1203	0.1602	-0.4627	-0.3465
	(0.0544)	(0.0487)	(0.0406)	(0.0616)	(0.0396)	(0.0396)	(0.0720)	(0.0619)
Seasonal	-0.1235	-0.1139	-0.0190	-0.1145	0.0936	0.1047	-0.1446	-0.1219
	(0.0475)	(0.0429)	(0.0277)	(0.0769)	(0.0516)	(0.0538)	(0.0579)	(0.0582)
Events	-0.5667	-0.4243	-0.2330	-0.4107	-0.0318	-0.0256	-0.7434	-0.6485
	(0.1375)	(0.1196)	(0.1032)	(0.1486)	(0.0936)	(0.0953)	(0.1840)	(0.1673)
				Type				
OE	-0.0274	0.0045	0.0205	-0.0609	0.0030	-0.0033	0.0112	-0.0157
	(0.0438)	(0.0413)	(0.0294)	(0.0456)	(0.0398)	(0.0401)	(0.0573)	(0.0471)
Others	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
				Length				
7-15D	-0.0658	-0.0227	-0.0474	-0.0697	0.0066	0.0228	-0.1115	-0.1251
	(0.0486)	(0.0533)	(0.0387)	(0.0700)	(0.0630)	(0.0660)	(0.0615)	(0.0555)
15-30D	-0.2265	-0.1719	-0.1150	-0.0954	0.0027	0.0338	-0.2761	-0.2550
	(0.0494)	(0.0428)	(0.0286)	(0.0587)	(0.0586)	(0.0598)	(0.0597)	(0.0574)

30-60D	-0.1315	-0.0411	-0.0402	0.1293	-0.0005	0.0039	-0.1846	-0.1529
	(0.0615)	(0.0536)	(0.0333)	(0.0864)	(0.0600)	(0.0618)	(0.0677)	(0.0695)
90-120D	-0.1021	-0.0375	-0.0013	0.0673	-0.0084	-0.0014	-0.1162	-0.0677
	(0.0642)	(0.0774)	(0.0414)	(0.1214)	(0.0650)	(0.0678)	(0.0762)	(0.0742)
$>$ 120D	-0.1843	-0.0752	-0.0575	0.0940	-0.0076	-0.0074	-0.2672	-0.2159
	(0.0691)	(0.0607)	(0.0395)	(0.1032)	(0.0686)	(0.0719)	(0.0846)	(0.0868)
				Hours				
6-15H	0.0703	0.1782	0.0358	0.2991	-0.1225	-0.1012	0.1267	0.1643
	(0.1104)	(0.1245)	(0.0990)	(0.1643)	(0.0621)	(0.0582)	(0.1831)	(0.1038)
$>$ 15H	0.1113	0.2242	0.0602	0.3616	-0.2127	-0.1887	0.1703	0.2422
	(0.1061)	(0.1216)	(0.0983)	(0.1630)	(0.0641)	(0.0612)	(0.1775)	(0.0994)
Obs.	371	371	371	371	371	371	371	371
R-sq.	0.2270	0.1650	0.0994	0.1207	0.0698	0.0774	0.2257	0.2369

Note: Numbers in parentheses are standard errors.

The city-level RDiT results are consistent with the monitor-level RDiT results in Section 13.1 and Section 13.2. In addition, they further verify the conclusions on some pollutants' policy responses and are more consistent with the pooled panel fixed-effect results in Section 13.3. For example, in the implementation criteria section of Table 31, the traffic-triggered policies and seasonal and regular policies do not correlate to more negative estimation results for SO_2 than the air pollution alert triggered policies; only the events-triggered policies correlate to more negative estimation results for SO_2. These results match with the chemical property of SO_2 that it almost does not respond to driving restriction policies but respond to industrial pollution reduction policies applied at the same time of the events-triggered policies. All the coefficients under the OE policies have smaller magnitudes and larger standard errors than in Table 14, verifying further that the OE policies do not have significantly different estimation results than the ODW policies.

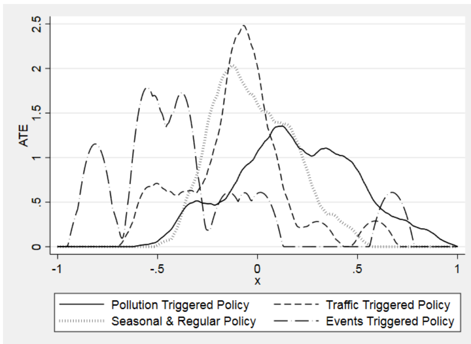
D Block Bootstrap of the Regression Discontinuity in Time Approach

I block bootstrap the full RDiT analysis by resampling the 54 treated cities 50 times. The simulated distribution of second stage coefficients for pollutant concentration and policy implementation criteria are presented in Figure 35 and Figure 36.[52] The results in Section 13.2 are still robust.

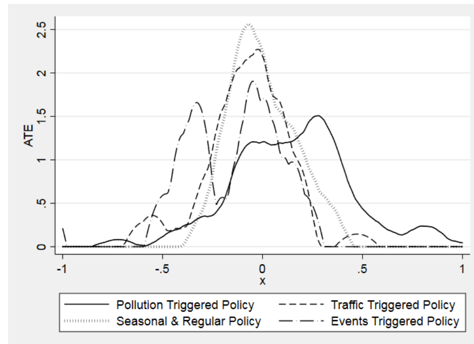
[52]The coefficients on policy implementation criteria are from the regression of policy effects on policy details same as in Section 13.2.

Figure 34: Distribution of City-specific RDiT Estimation Results by Policy Implementation Criteria

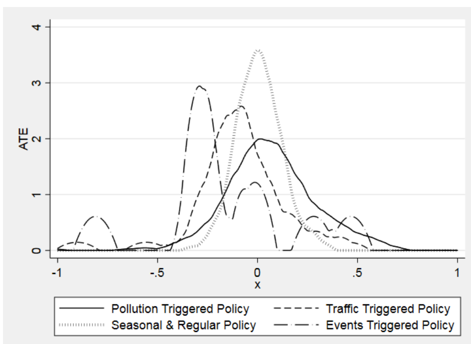
(a) AQI



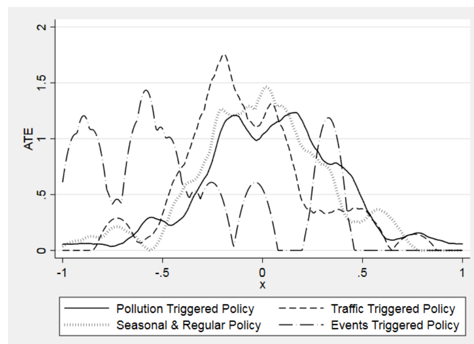
(b) CO



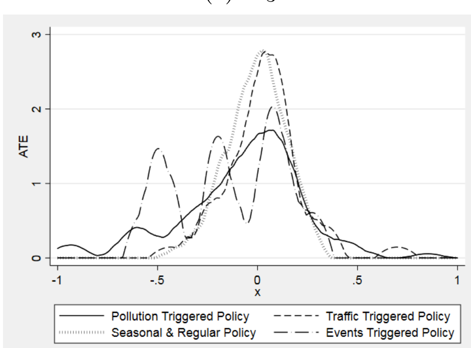
(c) NO_2



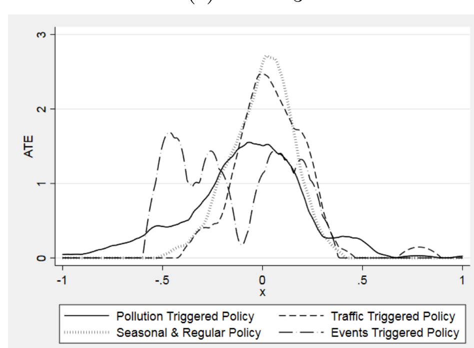
(d) SO_2



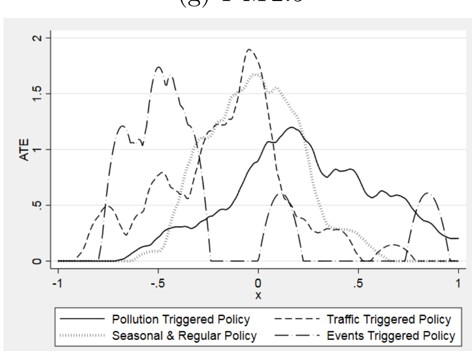
(e) O_3



(f) 8-h O_3



(g) $PM2.5$



(h) $PM10$

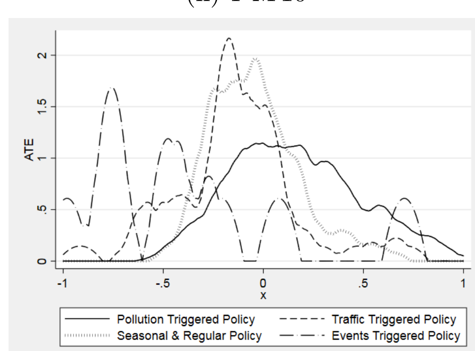
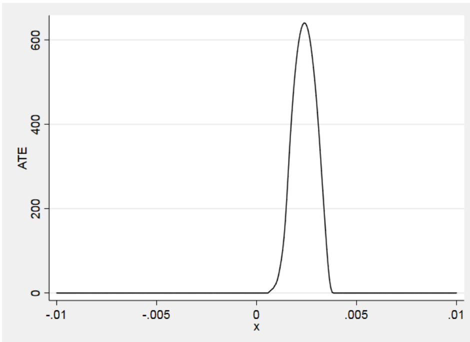
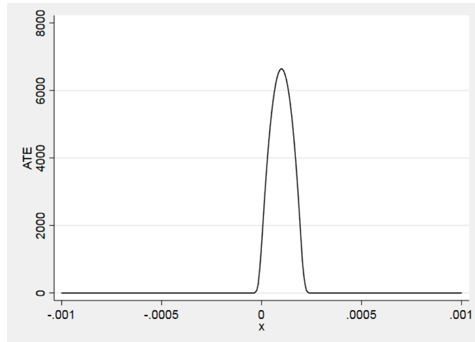


Figure 35: Coefficients of Pollution Concentration

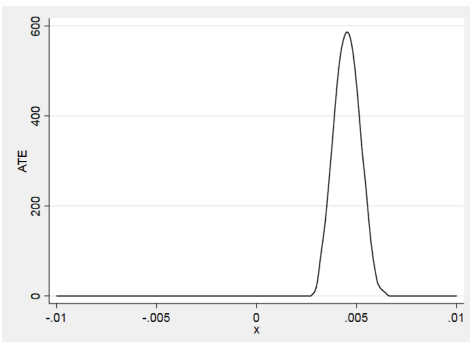
(a) AQI



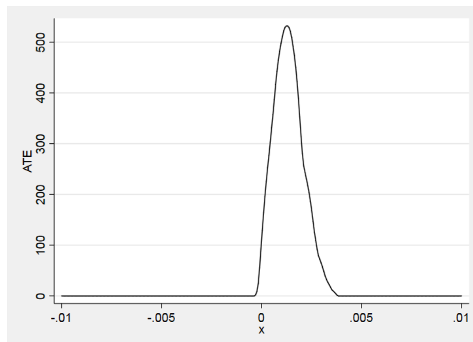
(b) CO



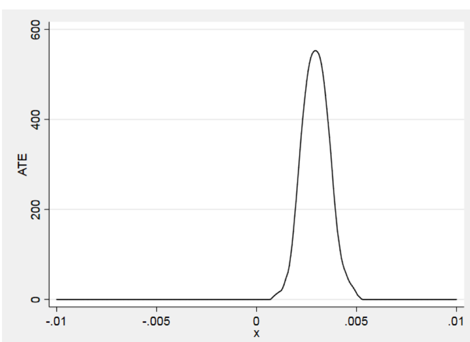
(c) NO_2



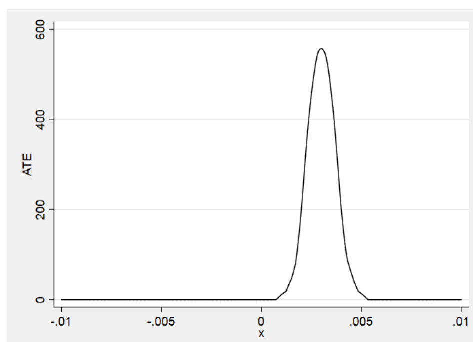
(d) SO_2



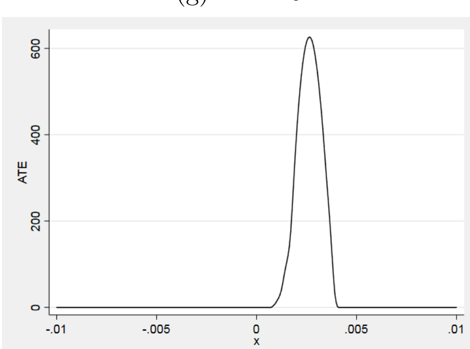
(e) O_3



(f) 8-h O_3



(g) $PM2.5$



(h) $PM10$

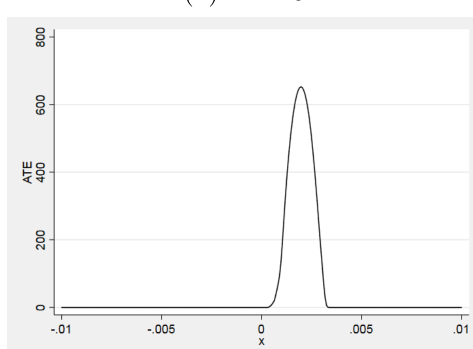
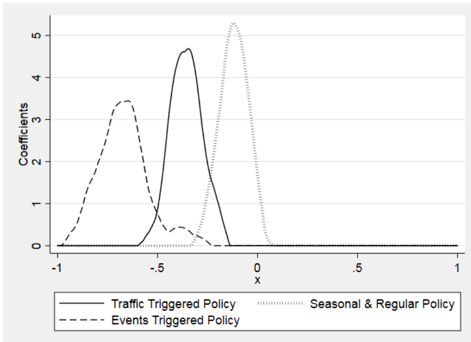
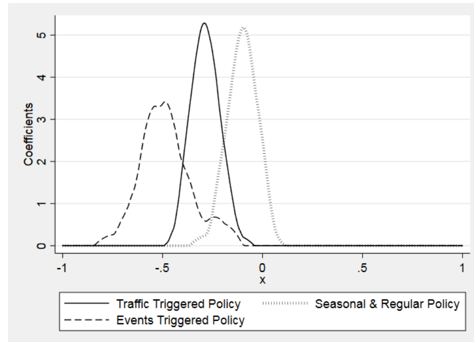


Figure 36: Distribution of the Estimated Coefficients by Policy Implementation Criteria

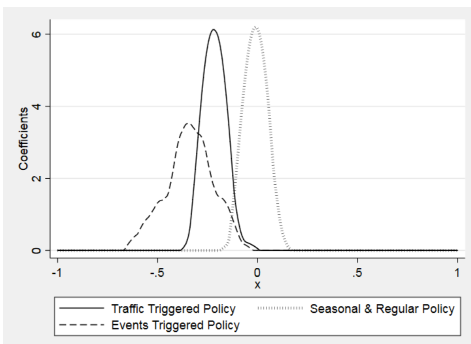
(a) AQI



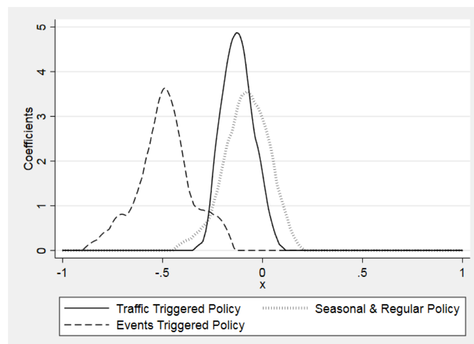
(b) CO



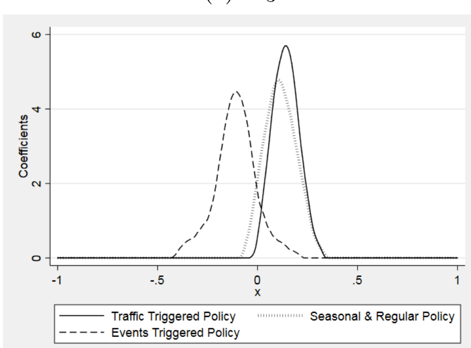
(c) NO_2



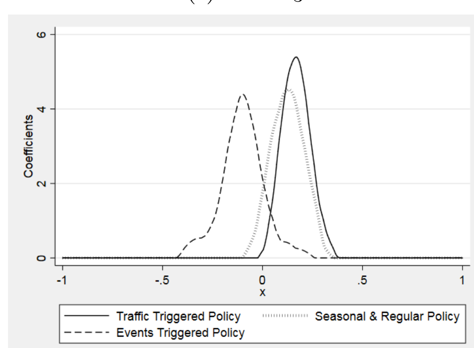
(d) SO_2



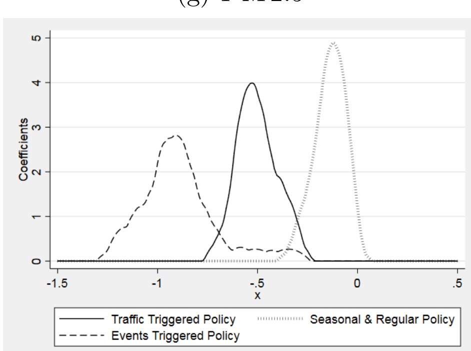
(e) O_3



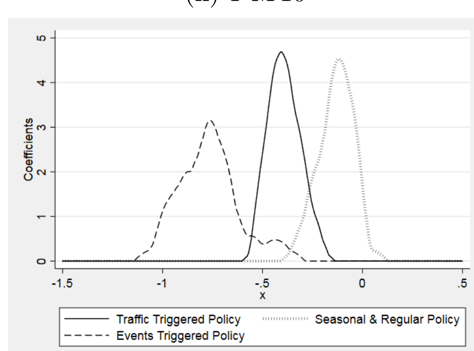
(f) 8-h O_3



(g) $PM2.5$



(h) $PM10$



E Balance Table of the Pooled Panel Fixed-Effect Approach

Crump et al. (2009) characterizes the optimal sub-samples with propensity scores lie between $[0.1, 0.9]$. There are 41 cities in the treatment group and 90 cities in the control group after trimming the sample. The distributions of the propensity scores for the treatment and control cities are drawn in Figure 37, and it shows that the common support is big enough for propensity score matching. Table 32 shows that the mean city characteristics in the treatment group and control group look different from one another in some variables, but the t-statistics for equal means t-test in column 4 show that we do not reject the null hypothesis that the treatment group and the control group have equal means in these variables.[53] To balance these observables further, I reweight the control group by the estimated propensity scores ($weight = \frac{p(D|X)}{1-p(D|X)}$). The reweighted means for the control group are reported in column 3, and the corresponding t-statistics are reported in column 5. Among the 11 variables listed, 9 of them have smaller absolute values of t statistics in column 5 than column 4, showing an improvement in the comparability between the treatment and control groups.

F Welfare Gain under Second-Best Tolling

The variability in marginal external congestion costs documented in Figure 32 suggest tolls based on average values may substantially overstate or understate costs much, if not most, of the time. Therefore, understanding the welfare effects of second-best tolls based on

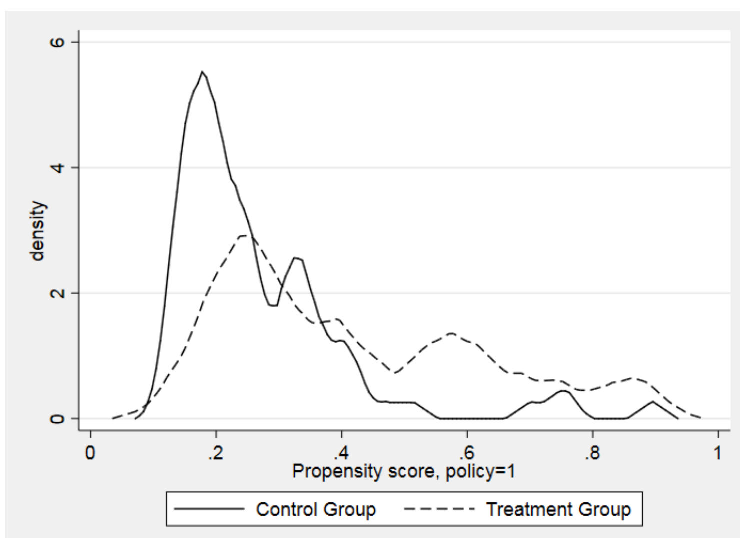
[53]The variable *Location* is not presented because it is only a dummy variable, but it is used in calculating the propensity scores.

Table 32: Mean City Characteristics, 2013

	(1) Treatment	(2) Control	(3) Wgt. Control	(4) t stats	(5) t stats
Population (10,000)	137.4463	113.0000	212.5386	-1.1946	0.5056
Build District Area (sq.km.)	148.2439	119.2778	218.7965	-1.1774	0.5473
GRP (100 mil. CNY)	1214.5400	969.2637	1816.5300	-0.7700	0.5463
Wage (100 CNY)	463.1788	465.3359	535.8149	0.1190	0.4208
Bus (10)	160.8098	109.5911	265.7514	-1.5551	0.5416
Taxi (10)	402.2244	278.4800	657.5765	-1.7883	0.5179
Road area (0.1 sq.km.)	177.5244	150.6733	259.2300	-0.8432	0.5208
Industrial SO_2 (100 t.)	688.0527	606.7736	779.2797	-1.0689	0.3415
Industrial Dust (100 t.)	433.1320	352.7443	559.2419	-1.3269	0.4477
Summer temp. ($^\circ F$)	79.2291	79.4926	87.2294	0.2766	0.3300
Winter temp. ($^\circ F$)	30.5617	32.4153	27.4978	0.6813	-0.6742

Notes: Columns 1-3 report the means of the variables listed in the row headings for the group listed at the top of the column. Column 1 describes the city characteristics for the treated cities. Columns 2 and 3 describe characteristics for untreated cities. In column 3, observations are weighted using propensity weights. Column 4 and 5 report t statistics for tests of equal means in the sub-samples.

Figure 37: Common Support of Propensity Score Matching



historical averages is of critical important for public policy. Recent work by Jacobsen et al. (2020) derives a simple approach for comparing the welfare improvement achieved by second-best policies to welfare under efficient Pigouvian pricing. In their framework, true marginal external costs vary over time, space or over individual agents. Because measuring actual costs may be expensive or impractical in settings with substantial variability, policymakers instead set second-best policies (taxes) based on average values. The authors show how, in some circumstances, the r-squared from a regression of (true) marginal external costs on fixed-effects, where the latter correspond to the levels of second-best average policies, yields an estimate for the ratio of welfare improvement under the second-best instruments compared to welfare under the Pigouvian ideal.

We adopt this procedure here, taking the 2SLS marginal external congestion cost estimates the "true" costs. We use these estimates as the dependent variable in a series of ordinary least-squares regressions. We imagine several policy scenarios, denoting the level at which second-best congestions tolls based on average congestion vary with different sets of fixed effects. The resulting r-squared values are presented in Table 38. The first row imagines a toll that varies by hour of day, but is common to all locations within the district. Given the time-varying features of congestion, *i.e.* strong daily commuting patterns, this policy seems a logical starting point. However, this policy achieves almost no welfare improvement possible under first-best pricing. A policy that ignores the time dimension but instead uses average congestion for each monitor, shown in the second row, performs a little bit better, but still achieving only 5 percent of the potential welfare gain. Combing the two approaches, adding common time of day adjustments to location-specific tolls, do not improve much of the performance. Adding adjustments for day of week, common to all locations, does little to improve performance.

These results are perhaps unsurprising given our intuition that congestion is highly time and place dependent. Therefore, one may ask how much of the welfare gain is achievable with fine scale tolls that vary by both location and time of day. The seventh row in Table 38 presents results with tolls that vary by hour at each monitor. We also include adjustments for day of week and week of year. Here, the second-best toll achieves 8 percent of the welfare gain of the first-best. However, even with relatively fine scale averages, the second-best policy leaves approximately 92 percent of the potential welfare gain on the table. Including weather fixed effects does not improve the performance. This highlights the important role of congestion tolls that vary throughout the day in response to real-time changes in the marginal external congestion costs.

We assume policy makers set tolls based on marginal external congestion cost. This could be done in real-time as we have shown above. In the following section, we will evaluate the distributional effect of the first three sets of second-best congestion tolls as in Table 38. We utilze household demographics from the NHTS, which was administered in 2017. Currently we use the average MECs in 2019 as tolls in 2017, because we do not have the precise trip start and end locations and the route information, eventually we will use the actual MECs.

G Figures

Figure 38: Subwatershed (huc12), Assessment Unit, and Flowlines

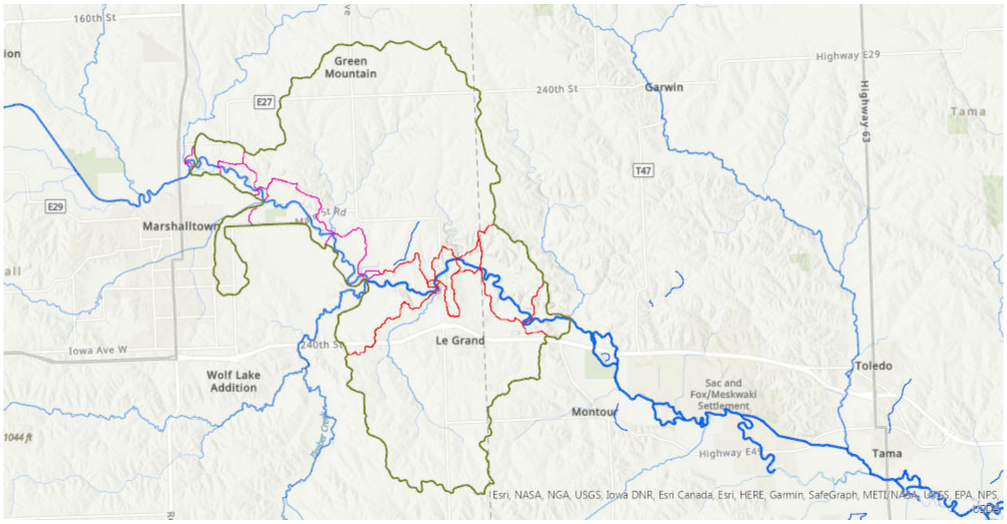
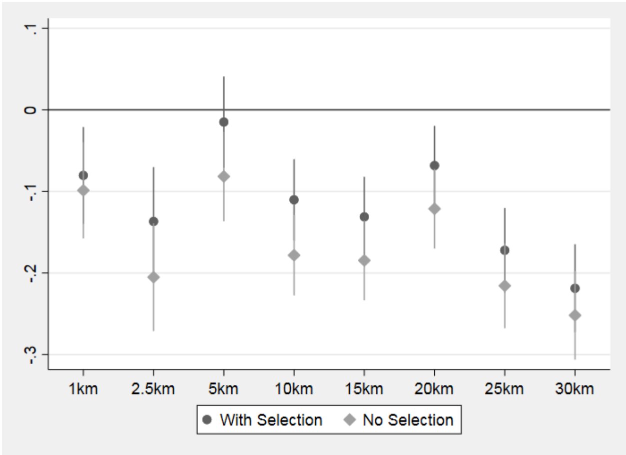


Figure 39: River Borders (NASA)

Figure 40: HUC4 Watershed Boundaries

Figure 41: Estimation Results of the Duration Model without Selection

(a) Distance Bins



(b) Different Portions of the Interstate Rivers

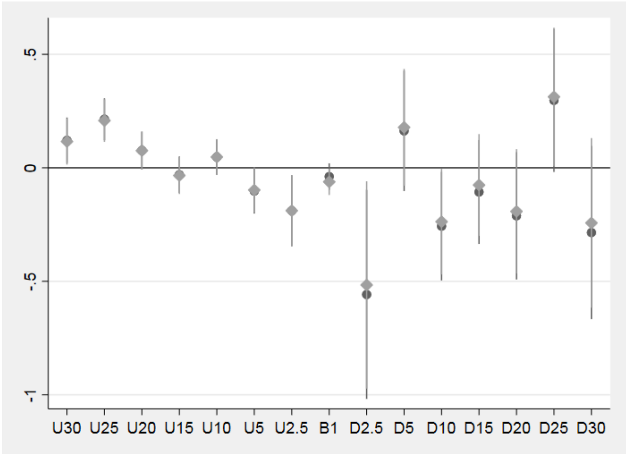
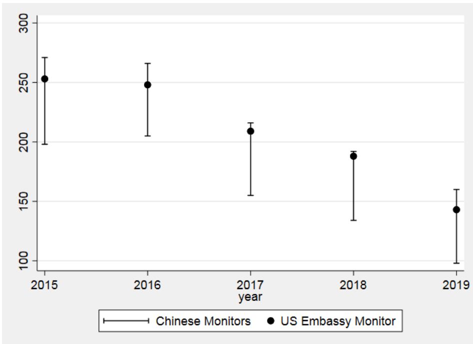
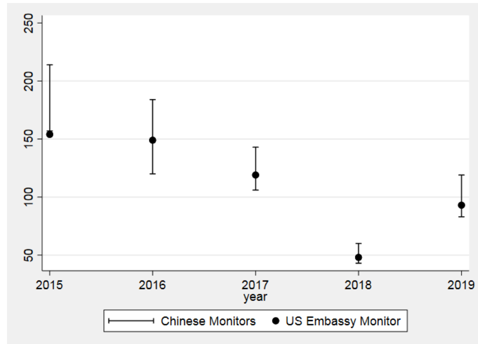


Figure 42: The Number of Unhealthy Days Calculated by $PM2.5$

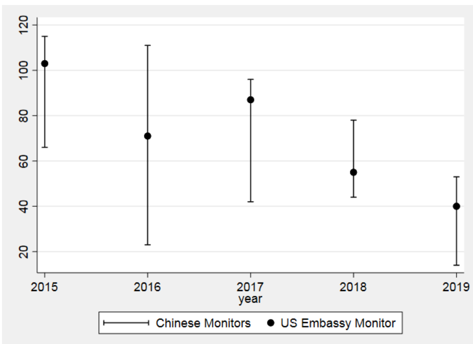
(a) Beijing



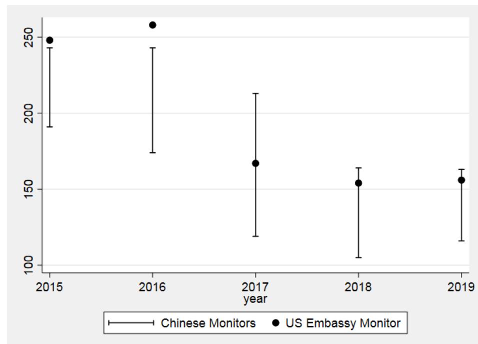
(b) Shanghai



(c) Guangzhou



(d) Shenyang



(e) Chengdu

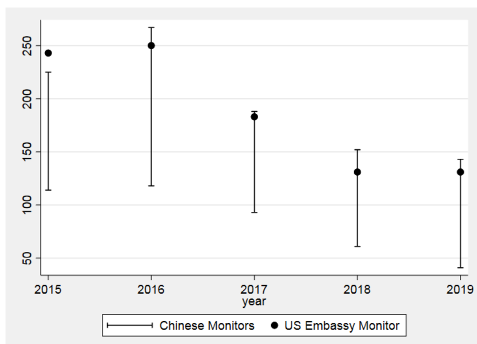


Figure 43: Population Density in 2015

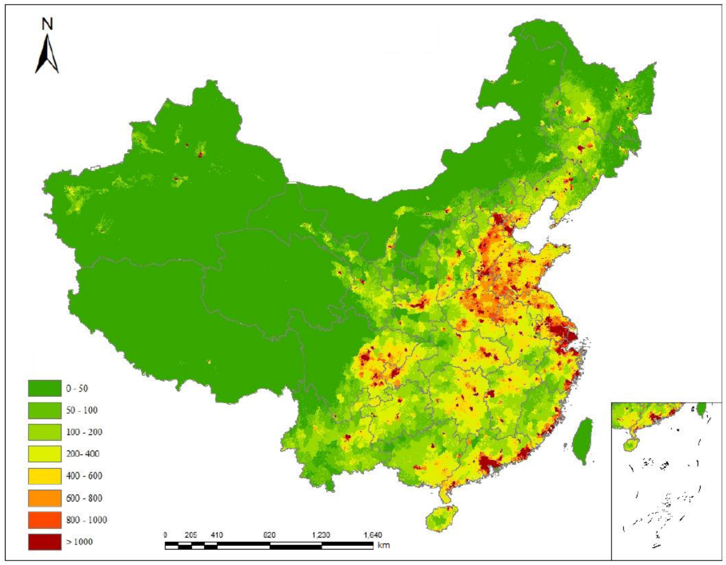


Figure 44: GDP in 2015

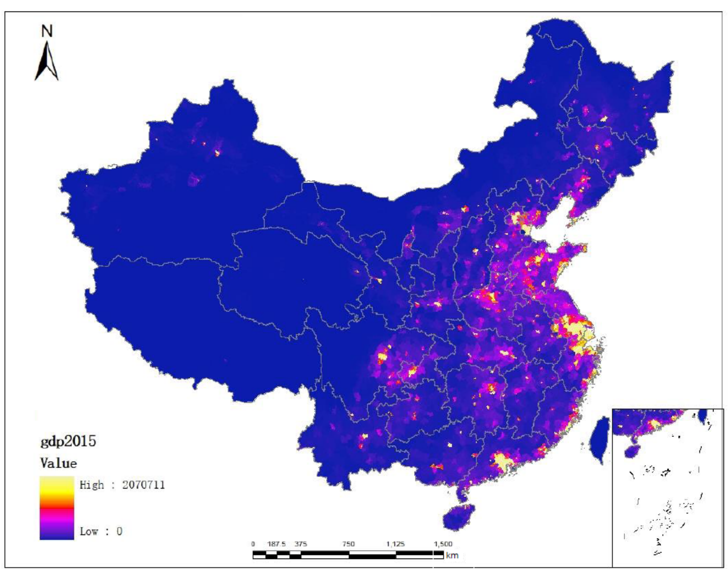


Figure 45: Fuel Standards Upgrading Timeline (Maximum Sulfur Content, ppm)

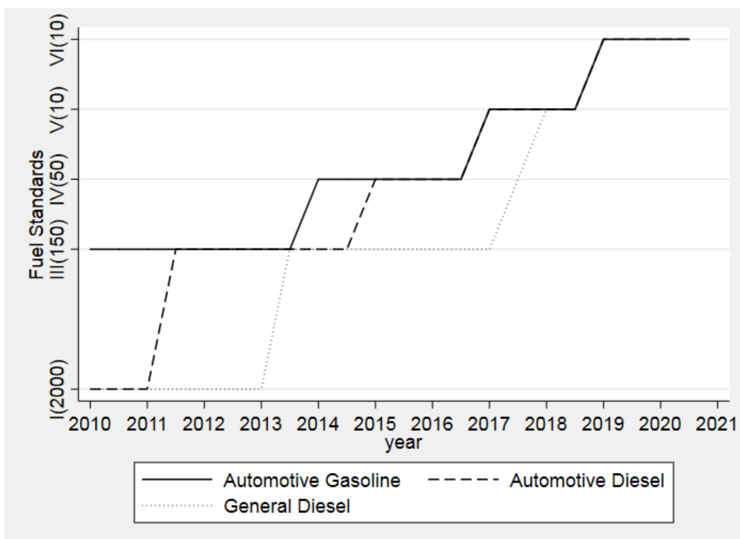
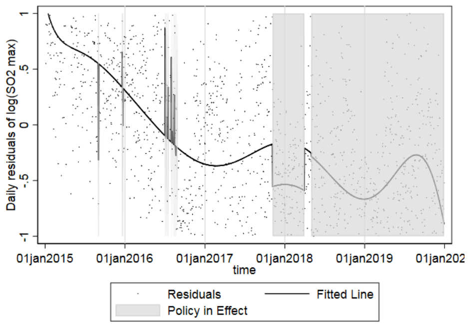
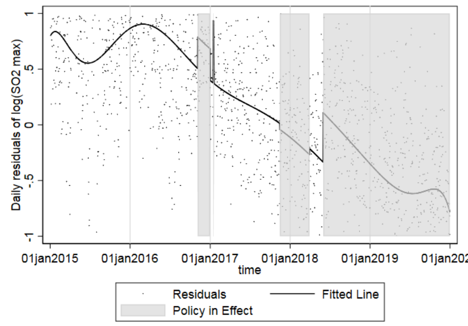


Figure 46: Daily Maximum Pollutant Concentration

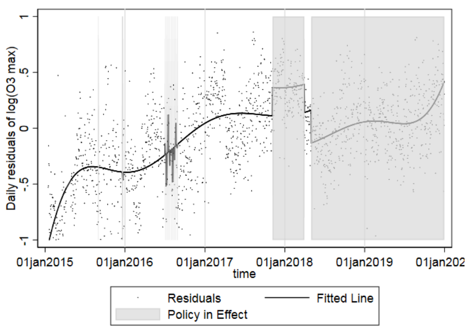
(a) SO_2: Monitor A1042, Qinhuangdao



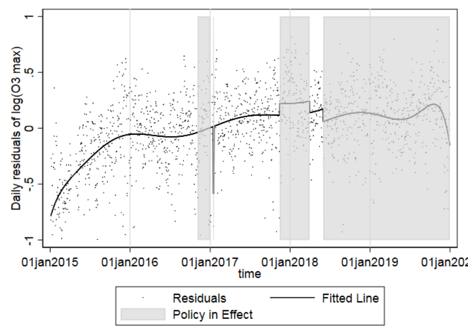
(b) SO_2: Monitor A1820, Anyang



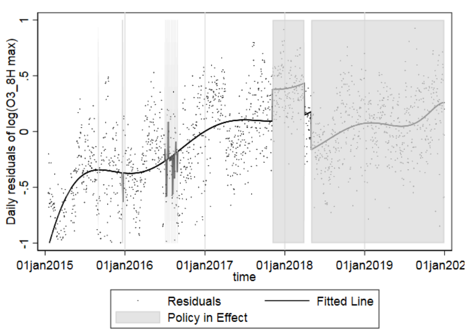
(c) O_3: Monitor A1042, Qinhuangdao



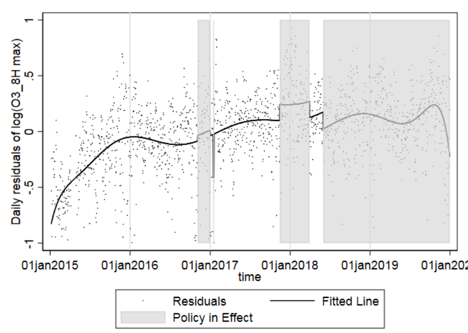
(d) O_3: Monitor A1820, Anyang



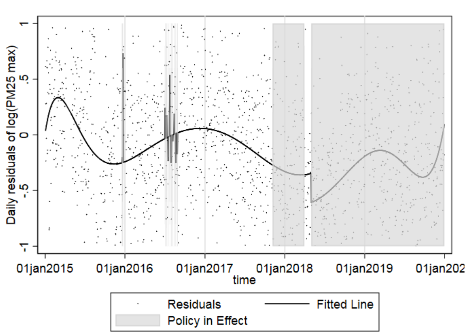
(e) 8-h O_3: Monitor A1042, Qinhuangdao



(f) 8-h O_3: Monitor A1820, Anyang



(g) $PM2.5$: Monitor A1042, Qinhuangdao



(h) $PM2.5$: Monitor A1820, Anyang

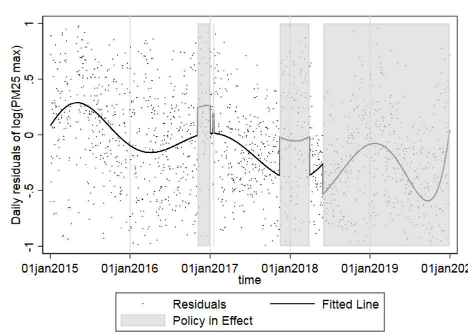
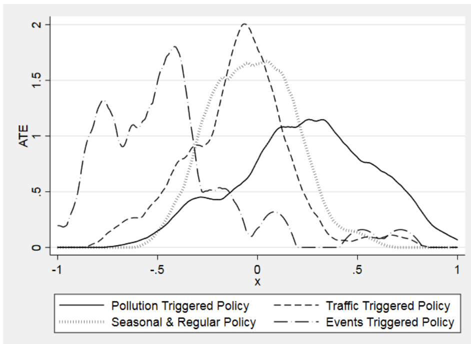
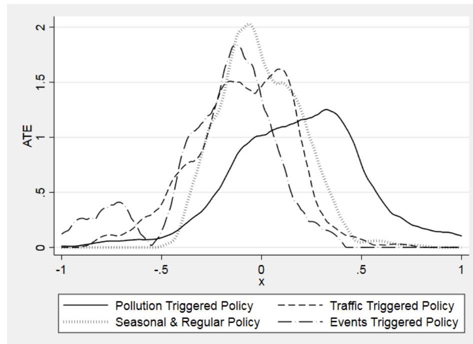


Figure 47: Distribution of the RDiT Estimation Results without Some Weather Variables in Regression

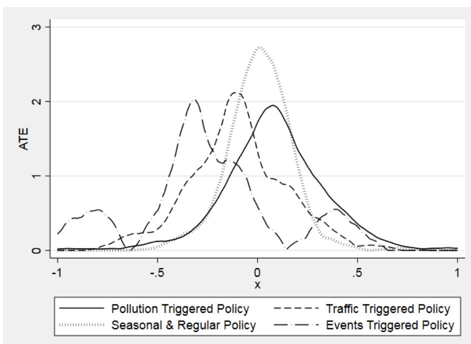
(a) AQI



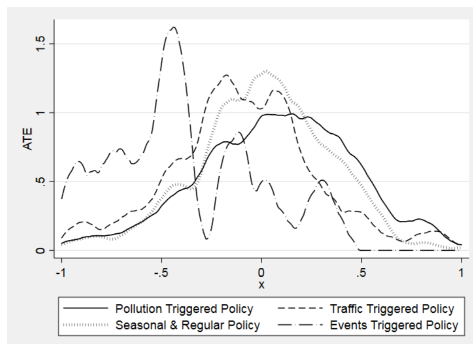
(b) CO



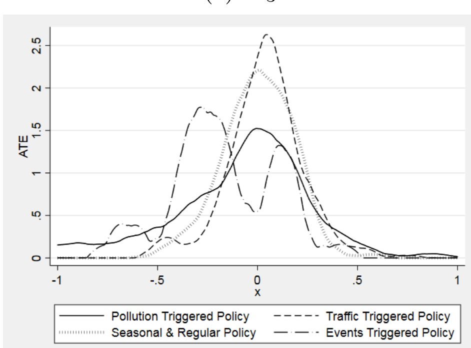
(c) NO_2



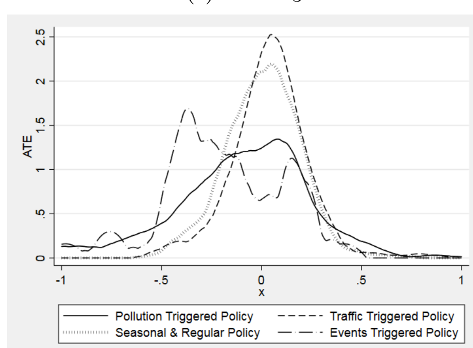
(d) SO_2



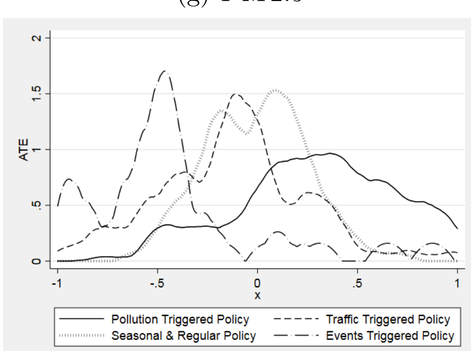
(e) O_3



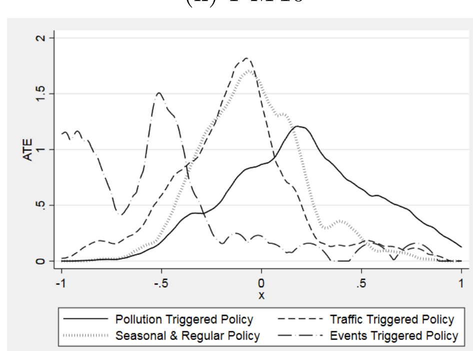
(f) 8-h O_3



(g) $PM2.5$



(h) $PM10$



H Tables

Table 33: Summary Statistics by Year: 2019.7.1 - 2020.6.31

2019							
	Obs.	Mean	Min.	P25	P50	P75	Max.
Speed (mph)	388,477	62.9	3.0	62.0	65.4	67.8	93.3
Flow (veh./hr.)	388,477	2,847.2	0.0	1,055.0	2,446.0	4376.0	16925.0
Density (veh./lane-mi.)	388,477	13.6	0.0	5.0	11.4	18.9	280.7
Lanes	388,477	3.58	1.0	3.0	4.0	4.0	8.0
North	388,477	0.31	0	0	0	1	1
South	388,477	0.31	0	0	0	1	1
East	388,477	0.19	0	0	0	0	1
West	388,477	0.19	0	0	0	0	1
2020							
	Obs.	Mean	Min.	P25	P50	P75	Max.
Speed (mph)	384,834	64.3	3.0	63.0	65.9	67.9	99.0
Flow (veh./hr.)	384,834	2,386.7	0.0	806.0	1,889.0	3583.0	17567.0
Density (veh./lane-mi.)	384,834	10.8	0.0	3.8	8.6	15.3	247.9
Lanes	384,834	3.6	1.0	3.0	4.0	4.0	8.0
North	384,834	0.3	0	0	0	1	1
South	384,834	0.3	0	0	0	1	1
East	384,834	0.2	0	0	0	0	1
West	384,834	0.2	0	0	0	0	1

Table 34: Descriptive Statistics: 2013 City Characteristics

Variable	Unit	Obs.	Mean	Std.Dev.	Min	Max
Gross Regional Product	100 mil. CNY	54	1427.04	2823.97	85.01	14500.23
Total Population at Year-end	10,000	54	148.24	139.24	28.20	686.60
Area of Built Districts	sq.km	54	159.00	193.18	24.00	1024.00
Public Finance Expenditure	100 mil. CNY	54	193.95	315.53	16.14	1690.83
Public Finance Income	101 mil. CNY	54	150.59	294.55	4.34	1731.26
Gross Industrial Output Value	102 mil. CNY	54	2055.29	3936.30	81.47	23095.21
Total Retail Sales of Consumer Goods	103 mil. CNY	54	609.91	1121.26	54.07	6504.89
Investment in Fixed Assets	104 mil. CNY	54	830.54	1088.72	66.49	4391.34
Investment in Real Estate	105 mil. CNY	54	256.15	395.14	10.21	1546.50
Number of Beds of Hospitals and Health Centers	100	54	118.10	119.16	17.18	619.19
Number of Doctors	100	54	61.36	68.54	10.46	366.68
Average Wage of Employed Staffs and Workers	100 CNY	54	466.17	97.94	330.14	777.49
Area of City Paved Roads at Year-end	0.1 sq.km.	54	208.99	243.11	29.40	1149.60
Annual Electricity Consumption	10,000 mwh	54	1168.06	1508.94	86.60	7297.68
Number of Busses and Trolley Busses at Year-end	10	54	223.96	456.34	9.60	3059.00
Number of Taxis at Year-end	10	54	426.18	455.37	40.00	2143.70
Green Covered Area of Completed Area	100 ha.	54	63.96	81.38	9.37	419.83
Volume of Sulphur Dioxide Emission	100 t.	54	743.69	505.08	17.24	2828.06
Volume of Industrial Soot(dust) Emission	101 t.	54	1106.81	4281.34	7.53	31538.22
Spring Max. Temp.	$^{\circ}F$	54	92.58	4.20	75.02	100.04
Spring Min. Temp.	$^{\circ}F$	54	26.32	10.13	-6.70	53.24
Summer Max. Temp.	$^{\circ}F$	54	97.46	4.80	77.00	103.46
Summer Min. Temp.	$^{\circ}F$	54	55.90	5.72	42.26	72.50
Autumn Max. Temp.	$^{\circ}F$	54	89.21	5.03	74.84	97.52
Autumn Min. Temp.	$^{\circ}F$	54	23.41	9.13	3.20	52.70
Winter Max. Temp.	$^{\circ}F$	54	58.96	8.52	29.12	84.20
Winter Min. Temp.	$^{\circ}F$	54	7.65	12.87	-24.88	44.60
Spring Average Temp.	$^{\circ}F$	54	57.78	6.48	37.88	73.27
Summer Average Temp.	$^{\circ}F$	54	78.78	4.36	64.90	84.58
Autumn Average Temp.	$^{\circ}F$	54	57.53	6.37	41.86	77.33
Winter Average Temp.	$^{\circ}F$	54	31.30	10.06	-0.76	62.06

Table 35: Influence of Weather on the RDiT Estimation Results

Temp.	-0.0093	-0.0031	0.0008	0.0050	-0.0004	0.0003	-0.0119	-0.0084
	(0.0055)	(0.0052)	(0.0029)	(0.0054)	(0.0052)	(0.0052)	(0.0075)	(0.0055)
MaxTemp.	0.0014	-0.0008	-0.0011	-0.0040	0.0011	0.0006	0.0026	0.0006
	(0.0027)	(0.0025)	(0.0013)	(0.0025)	(0.0027)	(0.0028)	(0.0036)	(0.0028)
MinTemp.	0.0066	0.0029	-0.0005	-0.0019	-0.0000	-0.0002	0.0077	0.0065
	(0.0029)	(0.0027)	(0.0016)	(0.0030)	(0.0026)	(0.0026)	(0.0040)	(0.0028)
Prcp.	0.0178	0.0020	-0.0016	-0.0028	0.0155	0.0173	0.0268	0.0182
	(0.0107)	(0.0113)	(0.0053)	(0.0151)	(0.0119)	(0.0124)	(0.0163)	(0.0126)
MaxPrcp.	-0.0118	-0.0040	-0.0005	0.0003	-0.0073	-0.0078	-0.0174	-0.0125
	(0.0039)	(0.0039)	(0.0022)	(0.0055)	(0.0048)	(0.0050)	(0.0062)	(0.0049)
MinPrcp.	-0.0043	0.0054	0.0034	0.0050	-0.0087	-0.0101	-0.0071	-0.0038
	(0.0078)	(0.0085)	(0.0039)	(0.0106)	(0.0073)	(0.0077)	(0.0115)	(0.0088)
WindSpeed	-0.0022	0.0017	-0.0013	-0.0002	0.0020	0.0028	-0.0013	-0.0021
	(0.0026)	(0.0022)	(0.0015)	(0.0013)	(0.0014)	(0.0014)	(0.0042)	(0.0028)
SkyCov.	0.0159	-0.0034	0.0033	0.0067	-0.0005	-0.0036	0.0159	0.0136
	(0.0099)	(0.0072)	(0.0050)	(0.0105)	(0.0065)	(0.0073)	(0.0136)	(0.0111)
Obs.	1,618	1,618	1,618	1,618	1,618	1,618	1,618	1,618
R-sq.	0.1961	0.1337	0.0931	0.0592	0.0417	0.0412	0.1825	0.1819

Note: Numbers in parentheses are standard errors.

Table 36: AIC and BIC for Different Orders of Time Polynomials in RDiT Estimation for CO

Order	1	2	3	4	5
AIC	1757.84	1719.26	1705.05	1688.70	1678.58
BIC	1929.06	1895.76	1887.09	1875.95	1871.31
Order	6	7	8	9	10
AIC	1665.34	1654.77	1647.42	1640.88	1633.98
BIC	1863.04	1857.23	1855.28	1854.15	1852.02

Table 37: Influence of 2013 City Characteristics on the RDiT Estimation Results

	AQI	CO	NO_2	SO_2	$O3$	$8-hO_3$	$PM2.5$	$PM10$
GRP	0.0007	0.0004	0.0004	0.0003	0.0004	0.0004	0.0011	0.0007
	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0003)	(0.0002)
Pop.	0.0047	0.0032	0.0019	0.0028	0.0005	0.0004	0.0066	0.0042
	(0.0011)	(0.0009)	(0.0008)	(0.0013)	(0.0009)	(0.0009)	(0.0014)	(0.0013)
Area.	-0.0019	-0.0011	-0.0022	0.0002	-0.0004	-0.0000	-0.0028	-0.0022
	(0.0014)	(0.0011)	(0.0010)	(0.0015)	(0.0018)	(0.0021)	(0.0021)	(0.0015)
PubExp.	-0.0008	-0.0015	-0.0005	0.0013	0.0018	0.0015	-0.0002	-0.0006
	(0.0018)	(0.0015)	(0.0010)	(0.0015)	(0.0012)	(0.0013)	(0.0024)	(0.0020)
PubInc.	0.0040	0.0035	0.0010	-0.0028	-0.0026	-0.0023	0.0047	0.0036
	(0.0025)	(0.0021)	(0.0015)	(0.0022)	(0.0016)	(0.0018)	(0.0033)	(0.0028)
GIOV	-0.0004	-0.0003	-0.0001	-0.0001	-0.0000	0.0000	-0.0005	-0.0004
	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Retail	-0.0013	-0.0008	-0.0005	-0.0005	-0.0008	-0.0006	-0.0018	-0.0011
	(0.0004)	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0005)	(0.0004)
InvFA	-0.0001	-0.0000	-0.0000	-0.0002	-0.0002	-0.0002	-0.0001	-0.0000
	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0002)

InvRealty	-0.0013	-0.0008	-0.0005	0.0005	0.0005	0.0005	-0.0019	-0.0014
	(0.0005)	(0.0004)	(0.0004)	(0.0005)	(0.0003)	(0.0004)	(0.0006)	(0.0005)
HsptBed	0.0003	-0.0006	0.0005	-0.0012	-0.0001	-0.0003	0.0006	0.0008
	(0.0013)	(0.0009)	(0.0009)	(0.0012)	(0.0009)	(0.0010)	(0.0016)	(0.0014)
Doctor	-0.0034	-0.0031	-0.0004	-0.0004	0.0034	0.0035	-0.0042	-0.0034
	(0.0028)	(0.0019)	(0.0018)	(0.0023)	(0.0017)	(0.0018)	(0.0035)	(0.0028)
Wage	0.0008	0.0003	0.0006	0.0005	-0.0000	-0.0001	0.0008	0.0009
	(0.0004)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0005)	(0.0005)
Road	0.0007	0.0009	-0.0003	-0.0006	-0.0006	-0.0007	0.0008	0.0007
	(0.0009)	(0.0006)	(0.0006)	(0.0007)	(0.0004)	(0.0005)	(0.0011)	(0.0009)
Electri.	-0.0003	-0.0002	-0.0001	-0.0001	0.0000	0.0000	-0.0005	-0.0003
	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0001)
Bus	-0.0009	-0.0007	-0.0003	0.0010	0.0000	0.0002	-0.0011	-0.0007
	(0.0006)	(0.0005)	(0.0004)	(0.0007)	(0.0005)	(0.0006)	(0.0008)	(0.0007)
Taxi	-0.0005	-0.0002	-0.0001	-0.0000	-0.0000	-0.0000	-0.0006	-0.0007
	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
GreenCov.	0.0132	0.0095	0.0077	-0.0003	-0.0008	-0.0030	0.0169	0.0140
	(0.0042)	(0.0030)	(0.0027)	(0.0036)	(0.0036)	(0.0041)	(0.0054)	(0.0044)
SO_2	0.0001	0.0000	0.0001	0.0000	-0.0000	-0.0000	0.0002	0.0002

	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Dust	0.0000	0.0000	0.0000	0.0000	-0.0000	-0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
LocNorth	0.2565	0.1918	0.1065	-0.0403	-0.0342	-0.0817	0.3078	0.2839
	(0.1200)	(0.0968)	(0.0771)	(0.1133)	(0.0596)	(0.0669)	(0.1604)	(0.1268)
				Temperatures				
MaxSPR	-0.0509	-0.0304	-0.0478	0.0001	0.0132	0.0189	-0.0659	-0.0550
	(0.0194)	(0.0141)	(0.0121)	(0.0165)	(0.0162)	(0.0196)	(0.0246)	(0.0211)
MinSPR	-0.0007	-0.0117	0.0178	-0.0122	0.0076	0.0010	-0.0017	-0.0020
	(0.0203)	(0.0148)	(0.0125)	(0.0188)	(0.0172)	(0.0184)	(0.0268)	(0.0229)
MaxSMR	-0.0541	-0.0514	-0.0286	-0.0264	0.0119	0.0246	-0.0712	-0.0624
	(0.0276)	(0.0217)	(0.0178)	(0.0219)	(0.0189)	(0.0207)	(0.0361)	(0.0298)
MinSMR	0.0374	0.0350	-0.0088	0.0209	-0.0467	-0.0441	0.0379	0.0564
	(0.0375)	(0.0277)	(0.0180)	(0.0239)	(0.0206)	(0.0225)	(0.0494)	(0.0408)
MaxAUT	0.0685	0.0533	0.0303	0.0323	-0.0091	-0.0141	0.0953	0.0828
	(0.0178)	(0.0132)	(0.0074)	(0.0151)	(0.0117)	(0.0121)	(0.0239)	(0.0192)
MinAUT	-0.0216	-0.0257	-0.0090	-0.0207	0.0183	0.0227	-0.0183	-0.0302
	(0.0164)	(0.0121)	(0.0104)	(0.0131)	(0.0101)	(0.0106)	(0.0218)	(0.0177)
MaxWIN	-0.0183	-0.0063	-0.0055	0.0328	-0.0034	-0.0100	-0.0249	-0.0250

	(0.0134)	(0.0104)	(0.0079)	(0.0132)	(0.0087)	(0.0089)	(0.0169)	(0.0143)
MinWIN	0.0349	0.0395	0.0265	0.0100	0.0184	0.0087	0.0533	0.0294
	(0.0129)	(0.0099)	(0.0075)	(0.0120)	(0.0096)	(0.0110)	(0.0168)	(0.0130)
AveSPR	0.2065	0.1565	0.1383	0.0201	-0.0189	-0.0331	0.2664	0.2110
	(0.0497)	(0.0383)	(0.0324)	(0.0432)	(0.0391)	(0.0466)	(0.0633)	(0.0538)
AveSMR	-0.0932	-0.0866	-0.0275	-0.0807	0.0125	0.0060	-0.1044	-0.0929
	(0.0437)	(0.0346)	(0.0303)	(0.0463)	(0.0338)	(0.0358)	(0.0564)	(0.0481)
AveAUT	0.1443	0.1329	0.0911	0.1175	-0.0038	-0.0168	0.1745	0.1452
	(0.0636)	(0.0481)	(0.0386)	(0.0558)	(0.0423)	(0.0458)	(0.0837)	(0.0713)
AveWIN	-0.2226	-0.1856	-0.1718	-0.0811	-0.0215	0.0137	-0.3140	-0.2158
	(0.0563)	(0.0414)	(0.0345)	(0.0519)	(0.0486)	(0.0554)	(0.0742)	(0.0613)
Obs.	1,618	1,618	1,618	1,618	1,618	1,618	1,618	1,618
R-sq.	0.3188	0.2378	0.1733	0.0566	0.0726	0.0723	0.3974	0.2735

Note: Numbers in parentheses are standard errors.

Table 38: R-square of second best policies using 2SLS MECs as the true MECs

Pricing Regime	R2
Hour of Day Fixed Effects	0.00
Hwy Location (Monitor) Fixed Effects	0.05
Location and Hour of Day Fixed Effects	0.05
Location, Hour, and Week Fixed Effects	0.05
Location, Hour, Week, and DOW Fixed Effects	0.05
Location×Hour, Week, and DOW Fixed Effects	0.07
Location×Hour×DOW, and Week Fixed Effects	0.08
Location×Hour×DOW, week, Rain, and Snow Fixed Effects	0.08