Essays In Health Economics & Economics of Crime

by

Kyutaro Matsuzawa

A dissertation accepted and approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Economics

Dissertation Committee:

Benjamin Hansen, Chair

Edward Rubin, Core Member

Kathleen Mullen, Core Member

Kristen Bell, Institutional Representative

University of Oregon

Spring 2025

DISSERTATION ABSTRACT

Kyutaro Matsuzawa

Doctor of Philosophy in Economics

Title: Essays in Health Economics & Economics of Crime

In Chapter 1, which is forthcoming at the *Journal of Law & Economics*, I exploit an exogenous ban of drunk driving (DUI) checkpoints, which were generally mandated via lengthy Supreme Court cases, to explore whether DUI checkpoints, a salient and targeted enforcement strategy, are effective in reducing drunk driving incidents. Using conventional and new two-way fixed effects models, I find robust evidence that DUI checkpoint bans lead to an approximately 12.4 percent increase in traffic fatalities involving drunk drivers. In addition, I find that DUI checkpoint bans lead to both an increase in DUI arrests as well as self-reported drunk driving behavior. These findings suggest that DUI checkpoints can create a general deterrent effect in curbing drunk driving behavior.

In Chapter 2, I study a reform in the Los Angeles Police Department (LAPD) that aims to reduce pretextual stops, which are types of traffic stops where the officer cites minor traffic violations as a reason to stop and search a driver for contraband. Using regression discontinuity design, I find that immediately following the adoption of this new policy, the number of stops for minor traffic violations – defined as stops for equipment or non-moving violations – reduced. Moreover, I find evidence that this policy may have helped reduce the number of Black drivers getting stopped. On the other hand, I find little evidence of long-term adverse effects in terms of public safety, as defined by traffic accidents and reported crime.

In Chapter 3, my coauthors (David Hall and Ben Hansen) and I replicate, reconcile, and extend two early studies investigating the effect of Oregon and Washington's drug decriminalization law on drug overdoses but reach different conclusions. We document that the conclusions of the two papers differ primarily due to the differences in the outcome investigated and the pre-treatment window utilized. However, these differences do not matter in the long run, as these results converge as we include further post-treatment data. Focusing on the 3-year post-treatment window, we find robust and compelling evidence that drug mortality increased following drug decriminalization.

This dissertation includes published material as well as unpublished coauthored material.

CURRICULUM VITAE

NAME OF AUTHOR: Kyutaro Matsuzawa

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene

San Diego State University, San Diego

DEGREES AWARDED:

Doctor of Philosophy, Economics, 2025, University of Oregon

Master of Science, Economics, 2023, University of Oregon

Master of Art, Economics, 2019, San Diego State University

Bachelors of Art, Economics, 2017, San Diego State University

Bachelors of Science, Statistics, 2017, San Diego State University

AREAS OF SPECIAL INTEREST:

Applied Microeconomics, Public Policy, Labor Economics, Health Economics, Economics of Crime

PROFESSIONAL EXPERIENCE:

Graduate Employee, University of Oregon, 2020-Present

Graduate Affiliate, Center for Health Economics & Policy Studies, 2019 - Present

Graduate Teaching Associate, San Diego State University, 2018-2019

Research Assistant, Institute for Behavioral And Community Health, 2017-2019

GRANTS, AWARDS, AND HONORS:

NAASE Graduate Student Paper Award - 2024

Russell Sage Foundation Dissertation Fellowship - 2024

Kleinsorge Summer Fellowship - 2022, 2024

University oforegon Department of Economics Best Field Paper Award - 2023

San Diego State University Terhune Economics Scholarship - 2017, 2018, 2019

San Diego State University Academic Excellence in Statistics - 2017

PUBLICATIONS:

Matsuzawa, Kyutaro. 2025. "The Deterrent Effect of Salient and Targeted Police Enforcement: Evidence from DUI Checkpoint Bans". Forthcoming. *Journal of Law & Economics*

Matsuzawa, Kyutaro, Daniel I. Rees, Joseph Sabia, and Rebecca Margolit. 2025. "Minimum Wages and Teenage Childbearing in the United States". *Journal of Applied Econometrics*

Dave, Dhaval, Andrew I. Friedson, **Kyutaro Matsuzawa**, Samuel Safford, and Joseph J. Sabia. 2023. "Black Lives Matter Protests and Risk Avoidance The Case of Civil Unrest During a Pandemic". *Journal of Human Resources*

Dave, Dhaval, Andrew I. Friedson, **Kyutaro Matsuzawa**, Drew McNichols, Connor Redpath, and Joseph J. Sabia. 2023. "Sudden Lockdown Repeals, Social Mobility, and COVID-19: Evidence From a Judicial Natural Experiment". *Journal of Empirical Legal Studies*

Anderson, D. Mark, **Kyutaro Matsuzawa**, & Joseph Sabia. 2022. "Marriage Equality Laws and Youth Mental Health". *Journal of Law & Economics* 64(1), 29-51

Dave, Dhaval, Andrew I. Friedson, **Kyutaro Matsuzawa**, Samuel Safford, and Joseph J. Sabia. 2022. "JUE Insight: Were Urban Cowboys Enough to Control COVID-19? Local Shelter-in-Place Orders and Coronavirus Case Growth". *Journal of Urban Economics* 59(1), 29-52

Dave, Dhaval, Andrew I. Friedson, **Kyutaro Matsuzawa**, Drew McNichols, Connor Redpath, and Joseph J. Sabia. 2021. "Risk Avoidance, Offsetting Community Effects, and COVID-19: Evidence From an Indoor Political Rally." *Journal of Risk and Uncertainty* 63:133-167

Dave, Dhaval, Andrew I. Friedson, **Kyutaro Matsuzawa**, and Joseph J. Sabia. 2021. "When Do Shelter-in-Place Orders Fight COVID-19 Best? Policy Heterogeneity Across States and Adoption Time". *Economic Inquiry* 59(1), 29-52

Anderson, D. Mark, **Kyutaro Matsuzawa**, & Joseph Sabia. 2020. "Cigarette Taxes and Teen Marijuana Use". *National Tax Journal* 73(2), 475-510

ACKNOWLEDGEMENTS

I thank Professors Ben Hansen, Ed Rubin, Kathleen Mullen, and Kristen Bell for their advice, mentorship, and unwavering support. I am incredibly grateful to Professor Hansen for his encouragement, kindness, guidance, and support throughout my time as a graduate student. Special thanks are also due to Professors Rubin, Mullen, and Bell, who provided valuable inputs that helped improve my dissertation.

I owe many thanks to all my co-authors for providing helpful inspiration and motivation for my research over the years. I am also thankful to my family, friends, and especially my partner, who believed in me and supported me throughout my journey. Finally, I would like to thank Jono, Paavo, and Tony for providing me with group therapy during times of stress.

This work has been partly supported by the Russell Sage Dissertation Fellowship.

Table of Contents

List of Figures

List of Tables

The Deterrent Effect of Salient & Target Police Enforcement: Evidence from DUI Checkpoint Bans

This chapter is based on a manuscript accepted for publication in the *Journal of Law & Economics* in August 2024.

Abstract: I estimate the causal effect of drunk driving (DUI) checkpoints on traffic fatalities, DUI arrests, and self-reported DUIs. Exploiting quasi-random variation in state-level laws that ban DUI checkpoints, I find a 12.4% increase in DUI-related traffic fatalities within the first five years following a DUI checkpoint ban. I also find a persistent increase in DUI arrests and a short-run increase in self-reported DUI behavior. Together, these findings suggest that targeted, salient police enforcement has a general deterrent effect on dangerous driving. Furthermore, back-of-the-envelope calculations suggest that a federal ban on DUI checkpoints would lead to an annual cost of approximately $6.4 billion in terms of lives lost from DUIs.

1.1 Introduction

Driving under the influence of drugs and/or alcohol (DUI) is a major global public health problem. In the United States, DUIs are responsible for an average of 32 deaths daily and cost society approximately $44 billion annually (The Centers for Disease Control & Prevention 2020), representing about a third of all traffic-related fatalities. DUIs serve as a negative externality, affecting not only the offenders but also other drivers on the road and their families. According to the National Highway Traffic Safety Administration's Fatality Analysis Reporting System (FARS), in 2019, 40% of DUI-related crashes involved multiple vehicles, 32% of DUI-related fatalities were non-drunk drivers (for example, other drivers or passengers), and 27.2% of DUI accidents led to the death of only the external party. Levitt & Porter (2001) calculate that the externality costs associated with each DUI incident, regardless of whether it results in an accident, can be as high as $14,000.[1] Strikingly, while 111 million DUI incidents occur annually, less than 1% result in an arrest (The Centers for Disease Control & Prevention 2020).

[1] Levitt & Porter (2001) initially report $8,000, which translates to $14,000 after being adjusted for inflation.

Enforcement and sanctions are common approaches to deter criminal behaviors. Specific approaches shown to be effective at reducing DUIs include a lower per se legal threshold for DUIs (Dee 2001, Eisenberg 2003, Hansen 2015), administrative license revocation laws (Fell & Scherer 2017, Freeman 2007), minimum legal drinking age laws (Carpenter & Dobkin 2009; 2011, Hansen & Waddell 2018, Voas et al. 2003), 24/7 complete sobriety laws (Midgette et al. 2021), and zero-tolerance DUI laws (Carpenter 2004, Voas et al. 2003). Another common targeted enforcement strategy is DUI checkpoints, where police officers screen drivers on certain roads.[2] These checkpoints not only apprehend violators but also visibly remind drivers about DUI repercussions. While many states conduct DUI checkpoints on a weekly basis, their impact and the consequence of prohibiting their use remain unclear.

In this paper, I measure the deterrent effect of DUI checkpoints. I employ a two-way fixed effects (TWFE) difference-in-differences estimator by leveraging plausibly exogenous spatial and temporal variation in state laws that banned the use of DUI checkpoints, which were largely created by state supreme court decisions. I use three comprehensive national datasets—FARS, Uniform Crime Report (UCR), and Behavioral Risk Factor Surveillance System (BRFSS)—and examine three different DUI-related outcomes: fatalities, arrests, and self-reported behavior. The use of these outcomes helps validate my results more clearly and to examine the mechanisms through which checkpoints might influence DUI behaviors.

Using data from the 1980–2018 FARS, I find a robust 12.4% increase in fatalities involving drunk drivers during the first several years after the ban was imposed. In the longer run, I continue to find an increase of 10%–12%, but the estimates are sometimes imprecisely estimated. Heterogeneity analyses reveal that these effects may be driven by larger states, where the number of DUI checkpoints and police enforcement may be more common and where the policy is expected to have a greater impact. A supplemental analysis examining the impact on non-DUIs suggests that the increase in fatal DUI accidents may be offset by a reduction in fatal speeding accidents. However, these estimates are very imprecisely

[2]DUI checkpoints are also known as DUI checkpoints, DWI (driving while intoxicated) checkpoints, or sobriety checkpoints. In this paper, I refer to them as DUI checkpoints.

estimated.

I also find a persistent 12%–17% increase in DUI arrests using the UCR data. Additionally, using self-reported survey data from the BRFSS, I find evidence of an increase in self-reported DUI behavior in the extensive margin and some evidence of an increase in the intensive margin, an important channel that explains these net increases. Furthermore, back-of-the-envelope calculations suggest that a federal ban on DUI checkpoints would generate \$6.4 billion annual costs from the lives lost from DUIs.[3]

This paper makes several contributions to the literature on DUI checkpoints. First, I estimate their causal effect at the national level. While the existing literature documents an association between the use of DUI checkpoints and reductions in DUI traffic accidents and fatalities in the United States (Bergen et al. 2014, Elder et al. 2002, Evans et al. 1991, Fell et al. 2004; 2005, Kenkel 1993, Lacey et al. 1999, Miller et al. 1998, Peek-Asa 1999, Sanem et al. 2015), their estimates may be biased due to the potential endogenous placement of individual programs in locations with higher DUI rates. Moreover, many of these studies only focus on smaller areas, such as individual states or cities, which raises concerns about external validity.[4]

Banerjee et al. (2019) use a randomized field experiment in Rajasthan, India and find that locations with rotating DUI checkpoints (randomly assigned checkpoints) reduce traffic accidents by 17% and deaths by 25%, relative to locations with no checkpoints. While they find compelling evidence supporting DUI checkpoints, several questions remain unanswered, which I address. First, what is the dynamic effects of conducting DUI checkpoints? Second, is the reduction in traffic accidents occurring through increased deterrence or increased arrests? Understanding the exact mechanism is important when policymakers are deciding on how to efficiently allocate police resources. Third, Banerjee et al. (2019) investigate efficient ways to use checkpoints but do not consider whether removing them and substituting away to other

[3] Further back-of-the-envelope calculations suggest that in states where DUI checkpoints are banned, an estimated \$1.8 billion is lost annually due to lives lost from DUIs.

[4] The magnitude of these studies range from 8% (Bergen et al. 2014) to 70% (Peek-Asa 1999) based on the geography examined, suggesting heterogeneous local treatment effects.

patrol methods are also effective in deterring drunk driving or have spillover effects on other non-drunk drivers.[5]

A working paper by Jones & Morin (2022) investigates the impact of state laws prohibiting the use of DUI checkpoints on traffic fatalities. Using traffic fatality data from 1980 to 2000, they find that laws banning checkpoints lead to a 17% increase in total fatalities. My research complements this paper while providing more robust findings. First, I use an empirical specification that allows me to mitigate biases that may arise from heterogeneous treatment effects due to bigger states experiencing a larger treatment effect than smaller states (Chaisemartin & d'Haultfoeuille 2020, Goodman-Bacon 2021, Sun & Abraham 2021). A concern with the methodology employed by Jones & Morin (2022) pertains to their use of the *overall level* of traffic fatalities. Relying on raw counts rather than rates or logs will result in larger heterogeneous treatment effects due to wide variations in population and driving, potentially introducing bias into their estimates.[6] Second, I use a more saturated model that incorporates various controls for police enforcement, which may be theoretically correlated with DUI behavior and changes in the use of DUI checkpoints.[7] Third, I examine various outcomes to further examine the mechanism in which DUI checkpoints affect DUIs. Fourth, I extend the sample time window and analyze the short-, medium-, and long-run effects of banning DUI checkpoints.[8] Finally, I conduct various sensitivity tests to determine if my estimates are artifacts of researcher choices in specification.

The rest of the paper proceeds as follows. Section 1.2 provides background on DUI checkpoints and the theoretical mechanism. Section 1.3 discusses the data and econometric methods. Section 1.4 presents the main estimates, and Section 1.5 concludes.

[5]A question that remains unanswered is whether the salience of DUI checkpoints or the presence of police is the key to deterring drunk driving.

[6]Assuming that their effect entirely originates from DUI fatalities, their estimates would imply a substantial increase of approximately 50% increase in DUI fatalities, which would be extremely large if true.

[7]Jones & Morin (2022)'s control variables include unemployment rates, per capita income, and the population over 65.

[8]For instance, the effect of DUI checkpoints may be different from two decades ago due to the evolution of technology, which makes communication about DUI checkpoints easier and more salient.

1.2 Background

1.2.1 History of DUI Checkpoints in the United States

DUI checkpoints, first introduced in the 1970s and growing in prevalence during the early 1980s (The Centers for Disease Control & Prevention 2015b), are a form of roadblock that police officers use to detect and deter drunk drivers.[9] At these checkpoints, officers block a street and stop most, if not all, drivers to determine if they are sober. The procedure generally involves the officer speaking with the driver, administering sobriety tests, and conducting additional screening (such as a breath test) if they suspect the driver is under the influence. The screening process at DUI checkpoints is relatively quick and lasts about as long as a red light at an intersection (Mothers Against Drunk Driving 2022).

The main goal of DUI checkpoints is to deter people from driving under the influence rather than to catch drunk drivers. Many checkpoints are salient, publicized, and highly visible to increase a potential drunk driver's perceived risk of getting caught. In many cases, news outlets or police announcements will publicly report the number of arrests made at a DUI checkpoint to remind the public about the consequences of driving under the influence. Additionally, many police departments announce when and where they will conduct the checkpoints (The Centers for Disease Control & Prevention 2015b, Mothers Against Drunk Driving 2022).

The use of DUI checkpoints has sparked controversy, with opponents arguing that they violate the Fourth Amendment because police officers search without probable cause.[10] Additionally, critics claim that the checkpoints can increase police discrimination, for several reasons. First, police departments often disproportionately place the checkpoints in neighborhoods with predominantly more racial minorities (Caputo 2015, Lacombe 2016, Romero

[9] Other types of roadblocks include traffic stops to check for a valid driver's licenses and vehicle registration, as well as to find undocumented immigrants.

[10] The Fourth Amendment states that "the right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."

2016).[11] Second, at a DUI checkpoint, police officers can easily observe the driver's race. This visibility raises concerns that implicit biases of the officers may the influence the decision to conduct further screening of the driver (Goncalves & Mello 2021, Pierson et al. 2020).

In the 1990 U.S. Supreme Court case of *Michigan Department of State Police v. Sitz*, a group of Michigan residents sued the state for conducting DUI checkpoints and violating their individual rights. The Supreme Court concluded, with a ruling of 6-3, that DUI checkpoints are constitutional and meet the standard of "reasonable search and seizure." However, the court also left it to each state to decide whether to allow them.

Between 1980 and 2018, 10 states prohibited the use of DUI checkpoints as illustrated by Table 1.1, which shows these states and the year in when the bans were enacted.,[12,13] States have two methods to banning DUI checkpoints: legislative actions and court decisions. In the legislative approach, a state legislature can ban checkpoints by passing new legislation that is related to traffic stops or vehicle inspections but does not explicitly mention or target DUI checkpoints.[14] In these states, the ban on DUI checkpoints results from legislation unrelated to DUI rates. Consequently, these policies create plausibly exogenous variation, which I use to estimate the causal effect of DUI checkpoints and their prohibition on DUI behavior.

In the second method, defendants arrested for a DUI argue that the evidence obtained at the DUI checkpoint should not be used for legal prosecution. One of their legal defenses is that the evidence was unlawfully obtained due to a search conducted without probable cause.

[11]For example, in Chicago, only 4% of checkpoints between 2010 and 2014 were placed in the city's majority-white police districts despite these neighborhoods accounting for around a quarter of the city's alcohol-related traffic accidents (Caputo 2015).

[12]Montana, although not included in the list, similarly bans DUI checkpoints. Unlike the other states, Montana does not explicitly allow the use of DUI checkpoints but does allow "safety" spot checks instead. These spot checks allow police officers to form roadblocks to check for a valid driver's license or vehicle registration and to screen for drunk drivers. For this paper, I treat Montana as a control state. However, the estimates are robust to whether I treat it as a treatment state or exclude it.

[13]In 2000, Louisiana, which banned DUI checkpoints in 1989, permitted the use of DUI checkpoints. Thus, there are currently 11 states that ban the use of DUI checkpoints.

[14]For example, in 1986, Iowa passed §321K.1, which provides information on when police officers can conduct vehicle roadblocks. An interpretation of §321K.1 led the state to ban DUI checkpoints.

Another argument is that DUI checkpoints are illegal because the state never established an administrative scheme for conducting them.[15] Typically, in these cases the defendant loses or the case gets dropped. However, in some rare instances, the defendant may decide to appeal and bring the case to the Supreme Court or the Court of Appeals. This decision to further argue against the constitutionality made by the defendant or their lawyer may create a plausibly exogenous variation in the timing of when DUI checkpoints might be banned. After long sequences of lower court trials, negotiations, and appeals, a state supreme court judge may rule that the state has no right to conduct DUI checkpoints.[16]

The basis of these decisions come down to whether (i) DUI checkpoints lack statutory constraints on the discretion of the police, (ii) they are intrusive based on the average length of the interaction between the driver and the police officer, and (iii) the invasion of privacy outweighs their social benefit. The third provision allows judge discretion based on their assessment of the trade-offs between public safety and individual privacy. For my analysis, this implies that my estimated effects may be a lower bound of the overall average treatment effect. This inference is drawn from the premise that bans on DUI checkpoints are more likely to occur in jurisdictions where judges feel that the safety impacts are smaller relative to the privacy lost.

1.2.2 Theoretical Effect of DUI Checkpoints and Ban

In Becker (1968)'s model of criminal deterrence, fewer crimes will occur when the expected costs of committing a crime increase—through a higher probability of detection or penalty from detection. The criminology literature classifies deterrent effects as either general deterrence (committing fewer crimes due to future punishments) or specific deterrence (committing fewer crimes due to previous punishments). Theoretically, DUI checkpoints can create either form of deterrence. They can increase a drunk driver's perceived risk of getting

[15] For example, in *State of Oregon v. Boyanovsky* in 1987, the defendant argued that the "roadblock was unlawful because it took place without statutory authority or agency rules allowing and controlling such a procedure."

[16] On average, these court cases take approximately three to four years from the arrest to reach a conclusion.

caught, thereby reducing the likelihood of driving while drunk (general deterrence). They may also dissuade drivers through peer effects when they see their peers getting arrested for DUIs (Billings & Schnepel 2022, Chalfin & McCrary 2017, Leadbeater et al. 2008); dissuade drivers who were previously screened at a DUI checkpoint (Beck & Moser 2004; 2006); and reinforce good behavior by reminding drivers about the potential consequences of DUIs (specific deterrence).

The effect of a DUI checkpoint ban can be asymmetric depending on whether the checkpoint acts as a general or specific deterrence. If checkpoints create a general deterrent effect, then banning them will have similar effects (but opposite in sign) as conducting one. Drivers, without the deterrence method, may perceive a lower risk of getting caught and begin driving drunk more frequently. In this scenario, DUI fatalities and arrests may increase when DUI checkpoints are banned. Conversely, if DUI checkpoints create a specific deterrent effect, then conducting them can have a more persistent effect rather than a temporary one. Therefore, even if DUI checkpoints are banned, the effect of the initial DUI checkpoints may continue, and people may react less to the ban. Under this scenario, the short-run effect of bans may be smaller and asymmetric to conducting one. Instead, the effect of bans on DUI fatalities or arrests may be more dynamic and increase over time.

A change in police behavior and arrests due to DUI checkpoints is also an important channel through which DUI fatalities can change. If DUI checkpoints are effective in arresting drunk drivers, then banning them could lead to fewer DUI arrests, potentially resulting in an increase in traffic fatalities. However, checkpoints may be an inefficient way of catching drunk drivers and might not lead to increased DUI arrests (Greene 2003, Kenney 2018). Their limited geographical coverage and the possibility of drivers avoiding them can reduce effectiveness. Moreover, the testing methods used at DUI checkpoints may be inaccurate or inefficient. Consequently, if DUI checkpoints do not significantly contribute to DUI arrests, their absence might actually reduce fatalities if more effective enforcement strategies are implemented instead.

Another consideration is the displacement effect. If checkpoints primarily result in the arrest of low-risk drivers rather than high-risk ones, they might inadvertently increase traffic fatalities. Under this scenario, banning DUI checkpoints could lead police to substitute away to a more efficient method of arresting drunk drivers.[17,18,19] An increase in the arrest of more and higher-risk drunk drivers following a ban could then lead to a decrease in DUI fatalities.

Finally, changes in traffic and speeding as a result of DUI checkpoints can also affect the rate of DUI fatalities. If checkpoints increase traffic and slow down driving speeds, this change may not necessarily lead to fewer crashes involving drunk drivers. However, it may reduce the severity of such incidents. For example, an accident may be more likely to result in minor collisions, like fender-benders, than serious car accidents. If this mechanism is true, banning DUI checkpoints could have an immediate effect on the day when they would have typically been conducted.

In summary, theoretically, the effect of a DUI checkpoint on traffic fatalities involving drunk drivers and DUI arrests is ambiguous. Similarly, the effect of banning DUI checkpoints is also ambiguous and not necessarily symmetric to conducting or allowing one. Moreover, the effects of these strategies and policies can be dynamic but in an unknown direction. Therefore, this ambiguity motivates both my empirical approach and the variety of measures related to DUIs that I outline in the next section.

[17] Another example of a targeted enforcement strategy is saturation patrol, where a large number of police officers patrol a specific area for a set time to find impaired driving behaviors, such as swerving and speeding. The main difference between a DUI checkpoint and a saturation patrol is that during a saturation patrol, police officers will need probable cause and reasonable suspicion to pull over a driver to screen. Compared to a DUI checkpoint, a saturation patrol may be more effective in arresting drunk drivers because it can cover larger areas and catch riskier drunk drivers (Greene 2003). However, it is less salient than a DUI checkpoint.

[18] For example, Minnesota began Operation NightCAP (Concentrated Alcohol Patrol), a high-profile saturation patrol program, four years after the state's supreme court ruled that DUI checkpoints are unconstitutional.

[19] Other enforcement strategies, such as saturation patrol, can also have spillover effects on non-drunk drivers by detecting other violations like speeding.

1.3 Methods

1.3.1 Data

Data on yearly traffic fatalities come from FARS, which is a national census that provides information on traffic accidents that result in at least one death. FARS, which became operational in 1975, provides a wide variety of information, including the time of day when the accident occurred, the driver's blood alcohol content (BAC), whether the driver was speeding, and demographic information of the people involved in the accident. Using these data, I create a state-by-year panel of total fatalities for all 50 states plus Washington, D.C. between 1980 and 2018. Since Louisiana reversed its ban in 2000, I exclude it from the panel after 2000.[20]

My main outcome of interest is the rate of total fatalities involving at least one driver with a reported positive BAC, a measure used in the previous literature examining DUI fatalities (Adams et al. 2012, D. M. Anderson et al. 2013, M. L. Anderson & Davis 2023, Eisenberg 2003, Sabia et al. 2019).[21] I determine a positive BAC if the result of the BAC test was at least zero or if the reporting officer suspected that alcohol was involved.[22] In addition, I collect information on total non-DUI traffic fatalities and non-DUI speeding fatalities.[23]

Panel (a) of Figure 1.1 plots the total number of DUI fatalities per 100,000 people, stratified by whether a state ever banned DUI checkpoints. In the early 1980s, the DUI fatality rate peaks and has a similar pattern of trends across both the treatment and control groups. Around the late 1980s and 1990s, when most states began to prohibit DUI checkpoints, the fatality rates for both groups begin drastically declining but also begin diverging, with states that banned DUI checkpoints having higher DUI fatality rates. These results suggest that

[20]The estimates are robust to whether I exclude Louisiana entirely from the panel or keep it post-2000 and use it as a control group.

[21]When I experiment with the total number of fatal accidents, rather than the number of fatalities, I continue to find similar patterns of results.

[22]The results are qualitatively similar when I use multiple imputed BAC to determine DUI accidents (Adams et al. 2012).

[23]Because information on speeding is not available until 1982, my analysis for the latter outcome focuses on the time period from 1982 to 2018.

DUI checkpoint bans may have increased the fatality rate involving drunk drivers.

The use of the alcohol measure in FARS may introduce bias due to measurement errors, as the reporting of a driver's intoxication status relies on the discretion of the reporting officer. To address whether measurement error is a concern, I conduct two data checks. First, I decompose each state's trend of DUI traffic fatalities into seasonality and time-series trends and identify anomalies in the time series based on the proportion of time series that are not explained by the two components (Hyndman 2021). This exercise aims to detect inconsistencies in the reported DUI fatality rate. The results show that only 0.4% (9 of 1,989) of the observations have an anomaly, defined as at least three times the interquartile range. The anomalies are spread over time and over the region rather than concentrated during a certain time window or around a certain geography. Furthermore, when I regress the anomaly indicator (equal to one if an observation was an anomaly and zero otherwise) on the treatment variable and state and time fixed effects, I obtain an F-stat of 0.136. This analysis confirms that there is very little or no correlation between potential measurement error and treatment status. These findings suggest that the use of reported DUI fatalities as the outcome is not a severe problem.[24]

Second, I examine the robustness of the estimates to using other traffic fatality outcomes, which could serve as proxies for DUIs. Specifically, I study weekend deaths and weekend or holiday nighttime deaths, a time of the week when DUIs are most likely to occur.

To measure the number of DUI arrests, I use crime data from the 1980–2018 UCR, obtained from Jacob Kaplan's Concatenated Files (Kaplan 2021b). Similar to the FARS data, I create a state-by-year panel for the rate of reported DUI arrests per 100,000 people. It is worth noting that the number of arrests undercounts the total number of offenses because not every crime results in an arrest. However, the use of arrests is appropriate because there is a high correlation between arrest reports from the UCR and actual crimes when data are available for both (Lochner & Moretti 2004).

[24] I find similar patterns of findings when I exclude the anomalies from the main regression estimate.

Although the UCR reports the number of arrests for more than 18,000 police agencies across the United States, reporting is voluntary, which raises concerns about data quality. For example, an agency that reported a crime at time t may not report any crime at time $t-1$. I address this reporting error in the data using several procedures. First, I restrict the sample to agencies that report crimes at least one month of the year, helping me loosen my assumption about missing versus zero crime.[25] Second, when calculating the rate of DUI arrests, I use the sum of the population covered by each agency rather than the state-level population. Using an agency population will guard against any biases that may arise due to agencies dropping out and changing the sample's composition.[26] However, one potential concern when using an agency-level population is that multiple agencies can cover the same geographic region, resulting in some agencies having zero populations (National Archive of Criminal Justice Data). To address this issue, I create a share of the total number of agencies with zero population and include it as a control variable in the regression analysis. Finally, I test if the estimates are sensitive to outliers.

Panel (b) of Figure 1.1 plots the trend in the average DUI arrests per 100,000 people by whether a state ever prohibited the use of DUI checkpoints. The trend of DUI arrests is noisier relative to the trends in DUI fatalities, but it generally mirrors the trend of DUI fatalities. DUI arrest rates drastically decline over time for both the treatment and control groups, but the trend starts diverging around the 1990s. In panel (a), I continue to find slightly higher DUI arrests for the treated group relative to the control group after many states banned the DUI checkpoints. This difference in trends also provides descriptive evidence that DUI arrests increase after the bans.

I also use data from the 1984–2018 BRFSS, a nationally representative health survey conducted by the Centers for Disease Control and Prevention (CDC), to measure self-reported DUI outcomes. This allows me to assess whether DUI checkpoints increase DUI behaviors,

[25]By excluding agencies that do not consistently report arrests, I can safely assume that "zero" reported crime means that no arrests were made rather than missing reports.

[26]I find qualitatively similar results when running the estimates at the agency level, which further guards against any changes in the sample's composition.

an important channel through which DUI checkpoint bans can affect DUI fatalities and arrests. However, the CDC did not extend the BRFSS coverage to all states until the 1990s, which limits the number of policy changes in my sample window to eight, with only four states having at least four years of pre-treatment data. Moreover, some survey years do not provide DUI information, limiting my ability to identify each lead-and-lag coefficient using the same set of treatment states in an event study analysis framework. For these reasons, while the BRFSS estimates are an important part of my analysis, the results should be interpreted with caution.

In the BRFSS, survey respondents are asked if they have consumed any alcoholic beverages during the last 30 days. Those who reported drinking are also asked if they have driven after they "perhaps had too much to drink." Using the answers to these questions, I construct a dummy variable equal to one if an individual reported driving while drunk at least once in the past month and zero otherwise. I also create a similar dummy variable but only focus on the sample of drinkers. Finally, I create a measure based on the log of the number of times an individual drove drunk.

1.3.2 Empirical Strategy

I begin by estimating the following TWFE difference-in-differences model:

$$log(Y_{st}) = \beta_0 + \beta_1 \text{Ban}_{st}^{0-5} + \beta_2 \text{Ban}_{st}^{6-10} + \beta_3 \text{Ban}_{st}^{11+} + \gamma X_{st} + \delta_s + \tau_t + \alpha_s \cdot t + \varepsilon_{st}, \quad (1.1)$$

where Y_{st} represents the outcome variable, which is the rate of DUI fatalities or DUI arrests per 100,000 people in a given state s at time (year) t. To address the skewness of the data and remain consistent with previous literature, I log transform the outcome. Ban_{st}^{0-5} is a treatment indicator denoting 0 to 5 years after the DUI checkpoint was banned, Ban_{st}^{6-10} is a treatment indicator denoting 6 to 10 years, and Ban_{st}^{11+} is a treatment indicator denoting 11 or more years. β_1, β_2, and β_3 are the causal parameters of interest, where β_1 represents the short-run effect, β_2 represents the medium-run effect, and β_3 represents the long-run effect of a DUI checkpoint ban. The vector X_{st} contains time-varying state-level controls, which are indicators for whether a state had a BAC 0.08 law, administrative license suspension law, zero-tolerance DUI law, minimum legal drinking age law, seat belt law, 65 mph and 70 mph speed limit law, and graduated license requirement law; beer tax rate; vehicle miles traveled; police employment; police expenditure; unemployment rate; and GDP per capita.[27] These control variables capture economic condition and policies that are shown to be associated with DUI or traffic fatalities, and they are commonly used in the literature examining the effect of public policy on traffic fatalities (D. M. Anderson & Rees 2015, D. M. Anderson et al. 2024, Carpenter 2004, Carpenter & Dobkin 2009; 2011, Cohen & Einav 2003, DeAngelo & Hansen 2014, Dee 1999; 2001, Eisenberg 2003, Fell & Scherer 2017, Freeman 2007, Hansen 2015, Hansen & Waddell 2018, Lovenheim & Steefel 2011, Voas et al. 2003). δ_s and τ_t are state and time fixed effects, respectively.

I also include a state-specific linear time trend ($\alpha_s \cdot t$) that is, again, used commonly in the literature examining the effect of public policy on traffic fatalities. Including this control aids in reducing omitted variable bias by controlling for unobserved state trends that unfold linearly and are incidentally correlated with both the treatment and outcome. The estimates are weighted by the state population, and standard errors are clustered at the state level (Bertrand et al. 2004).[28,29]

For the BRFSS estimate, I estimate the following model, similar to Equation (1.1):

$$Y_{ist} = \beta_0 + \beta_1 \text{Ban}_{st}^{0-5} + \beta_2 \text{Ban}_{st}^{6-10} + \beta_3 \text{Ban}_{st}^{11+} + \gamma X_{st} + \phi Z_{ist} + \delta_s + \tau_t + \alpha \cdot t + \varepsilon_{ist}, \quad (1.2)$$

where Y_{ist} represents the outcome variable on DUI behavior in the last 30 days. The difference between Equation (1.1) and Equation (1.2) is that since the BRFSS provides individual-level

[27] Appendix Table 1.1 provides summary statistics and sources for each control variable.

[28] Population data come from the Surveillance, Epidemiology, and End Results Program (2022).

[29] When I experiment with inference based on wild-cluster bootstrapping (Cameron et al. 2008, Cameron & Miller 2015), I continue to find similar significance levels as the inference based on clustered standard errors.

survey data, my analysis is performed at the individual level (denoted by i) rather than at the aggregated state level. Because the survey date is available, I include month-by-year fixed effects rather than year fixed effects. Following Carpenter (2004), I also include individual-level controls denoted by Z_{ist}: three indicators of education (high school degree, some college, college or higher), an indicator for white non-Hispanic, an indicator for male, and age. To obtain nationally representative health statistics, I weight the regression using BRFSS-provided survey weights. For ease of interpretation and to remain consistent with Carpenter (2004), Equation (1.2) is estimated using OLS and a linear probability model. Additionally, because I only have eight treatment states that have data on both the pre- and post-treatment period, I report standard errors that are clustered at the state level and wild-cluster bootstrapped.

For Equation (1.1) and Equation (1.2) to be causal, I assume that the parallel trends assumption holds. I observe that the majority of the checkpoint bans (8 out of 12) occurred through states' supreme court rulings, which concluded lengthy (and arbitrarily long) sequences of lower court trials, negotiations, appeals, and decisions. Consequently, these case-driven court rulings plausibly introduce a policy variation independent of trends in DUIs and DUI checkpoints.[30,31] To assess the validity of the parallel trends assumption, I conduct an event study analysis in which I replace the treatment indicators in Equation (1.1) with ζ_{st}^{τ} lead-and-lag treatment indicators:

$$log(Y_{st}) = \beta_0 + \sum_{\tau=-6}^{-2} \beta_{\tau}^{Pre} \zeta_{st}^{\tau} + \sum_{\tau=0}^{12} \beta_{\tau}^{Post} \zeta_{st}^{\tau} + \gamma X_{st} + \delta_s + \tau_t + \alpha \cdot t + \varepsilon_{st}, \tag{1.3}$$

where ζ_{st}^{τ} is equal to one if a state had a DUI checkpoint ban τ years before or after. I use β_{τ}^{Pre} to test whether the policy is endogenous to the outcomes under study or is correlated

[30] For example, in Oregon, the court case that led to the 1987 ban on DUI checkpoints originally began in 1984 but did not reach the state supreme court until January 1987. The court made its final decision in September 1987.

[31] Alaska, Iowa, Wisconsin, and Wyoming are the four states where DUI checkpoints are banned through legislative order. When I exclude them and focus only on treatment that occurred via a judicial order, I still observe similar patterns of results.

with differential trends in outcomes across the treated and non-treated states. I examine β_τ^{Post} to assess for any post-treatment changes.

A recent development in the literature on difference-in-differences emphasizes the bias that could arise from staggered treatment and dynamic and heterogeneous treatment effects (Goodman-Bacon 2021). I argue that because my panel consists of significantly more never-adopters (39 states), the weights of the β estimates are likely to be significantly influenced more by the clean comparison groups, which use never-adopters as a control group, rather than by the bad comparison group that uses already or early adopters as controls. An exercise showing Goodman-Bacon (2021) weights confirms this claim. I find that 91.7% of the weights come from the comparison group of treated versus never treated. Furthermore, I find that the weights assigned to the bad comparison group (late versus early adopters) are relatively low, at only 7.7%. In addition to Goodman-Bacon (2021)'s decomposition, I also test for the presence of negative weights, which could bias the β estimates, using the methods introduced by Chaisemartin & d'Haultfoeuille (2020). The results show that only 14.1% (50 out of 353) of my ATTs receive a negative weight, summing to only a total of 5.9% negative weight.

Together, these exercises suggest that the staggered treatment and heterogeneous and dynamic treatment effects are unlikely to substantially bias the estimates. Nonetheless, as a robustness check, I also re-estimate the main TWFE estimate using the Callaway and Sant'Anna estimator (Callaway & Sant'Anna 2021), the Sun and Abraham estimator (Sun & Abraham 2021), and a stacked difference-in-differences (stacked DD) estimator (Cengiz et al. 2019), which allow me to mitigate such biases. For the Callaway and Sant'Anna estimator, I use not-yet-adopting states as the counterfactual, and for the other two estimators, I use never-adopting states as the counterfactual.[32]

Another key assumption I make is that there is a first-stage effect on the number of DUI checkpoints decreasing after the DUI checkpoint ban. This assumption may be violated,

[32]Because there are some treatment states that are next to each other (for example, Washington and Oregon), not-yet-adopting states may be a better counterfactual than never-adopting states.

for example, if there are some areas that never conducted a DUI checkpoint, resulting in no change in the number of DUI checkpoints. While this assumption cannot directly be tested due to the unavailability of historical data on when and where DUI checkpoints were conducted, I defend it in two ways.

First, I gather information on state supreme court cases using data from casetext.com, a cloud-based legal research platform that provides organized case information about state and federal supreme court cases. Using this source, I identify historic dates when DUI checkpoints were conducted, following two steps. I first search for state supreme court cases involving the keywords "DUI checkpoint," "sobriety checkpoint," or "DWI checkpoint." Next, I conduct a text analysis to determine if the court cases mention when the checkpoints in question were conducted.[33] This process results in 365 unique court cases where the exact date of DUI checkpoints are known.

While these court cases do not encompass the entirety of checkpoints conducted, Appendix Figure 1.1 shows a strong correlation between the number of state supreme court cases and the number of DUI checkpoints conducted between 2008 and 2018.[34] This correlation supports the assertion that this measure can serve as a proxy for the overall amount of DUI checkpoints conducted. Using these court cases, I calculate the yearly number of reported DUI checkpoints for each state. The descriptive trends shown in panel (a) of Figure 1.2 and the event study analysis shown in panel (b) provide descriptive evidence that the number of DUI checkpoints was quite similar between the treatment and control states in the pre-treatment period (for example, the 1980s) but indeed significantly decreased after the ban.[35,36]

[33]In many cases, the court document will include a phrasing along the lines of "on this date, the defendant was arrested at a DUI checkpoint."

[34]The data on the number of reported DUI checkpoints come from DUIBlock.com. DUIBlock.com provide information on when and where a DUI checkpoint was held between 2008 and 2018, encompassing over 33,000 DUI checkpoints. However, these data are only available starting in 2008, meaning I do not have any states contributing to my identification.

[35]I also note that DUI checkpoints started being used in Louisiana after the ban's reversal in 2000.

[36]In Appendix Figure 1.2, I show event study analysis where I examine the number of court cases involving the keyword "Driving Under the Influence." I find no evidence that the number of state supreme court cases arguing about DUIs decreased after the prohibition of DUI checkpoints. This finding suggests that only DUI

Next, to further validate the first-stage assumption, I construct a pseudo-state-by-year panel using only the counties where DUI checkpoints will most likely occur and another pseudo-state-by-year panel using counties where they are unlikely to have a DUI checkpoint.[37] To identify counties likely to conduct a DUI checkpoint, I use the county's population characteristics. In Appendix Figure 1.3, I use data on local DUI checkpoints compiled by DUIBlock.com and find a strong positive association between different population measures and the number of reported DUI checkpoints. I define a county as likely to conduct a DUI checkpoint if (1) its population is greater than 45,000, (2) its weighted population density is greater than 1,000, or (3) its urbanicity is greater than 60%.[38] If my sub-state analysis suggests that the treatment effect is driven by counties where it is ex ante expected that the policy will more likely be binding due to a higher frequency of DUI checkpoints, then this exercise will strengthen my confidence in the first-stage assumption.

1.4 Results

1.4.1 Traffic Fatalities

Table 1.2 shows the TWFE estimates of the effect of DUI checkpoint bans on traffic fatalities involving drivers with a positive BAC level. The point estimate from column (1) implies an 18.5% increase in deaths from DUIs within five years after DUI checkpoints were banned.[39] In the fully saturated model in column (2), where I include various control variables, I continue to find similar patterns of results, though the estimated effect is slightly smaller. The point estimate from column (2) suggests a significant short-run effect of a 12.4% increase in traffic fatalities involving drunk drivers due to a DUI checkpoint ban

court cases involving DUI checkpoints were affected by the policy.

[37] The estimates are qualitatively similar when I conduct the analysis at the county-by-year level instead.

[38] I choose these cutoffs because they are greater than the 90th percentile among counties without DUI checkpoints and around the 50th percentile among counties with DUI checkpoints (Appendix Table 1.2). However, the estimates are robust to using different cutoffs.

[39] $\exp(0.170) - 1 = 0.185$.

and a medium-run (6 to 10 years after the policy ban) effect of a 12.2% increase in such fatalities. However, the latter estimate is only significant at the weakest conventional level. The estimates show that the positive effect continues in the long run (11 or more years after), but the effect size decreases and the estimates become imprecise, making it difficult to draw any definitive conclusions.

Using the estimates from DeAngelo & Hansen (2014), my estimated increase in traffic fatalities is comparable to what would be expected from a 31%–36% decrease in the size of the highway police force, suggesting that the magnitude of this estimated effect is plausible and reasonable. Moreover, this 12.4% increase translates to approximately 474 additional DUI fatalities among the treatment state.[40] This count of 474 additional deaths is reasonable given the sheer number of cars that drive through DUI checkpoints.[41] Because the estimated effects closely match those found in the previous literature examining the original implementation of DUI checkpoints (Banerjee et al. 2019), I can conclude that the start and end of these policies have symmetric effects.

In Appendix Figure 1.4, I further examine which control variables are causing the point estimates to be smaller. First, I find that including the unemployment rate reduces the estimated coefficients by about 16% (0.156 to 0.131). This reduction is not surprising given that previous studies have shown that both police enforcement and mortality may change during the recessionary period (Ruhm 1996, Makowsky & Stratmann 2014). I also find that including other DUI enforcement strategies may be responsible for the change in the estimated effect. States that are likely to ban DUI checkpoints may also be more likely to adopt other DUI policies in lieu of not having checkpoints, which can reduce DUIs. For instance, a state may be more likely to enforce stronger sanctions (for example, administrative license

[40] To calculate this number, I use the total number of DUI fatalities in the period immediately before the ban (3,823), multiplied by 12.4%.

[41] While the exact number of cars that pass through a DUI checkpoint is not available, the numbers are expected to exceed a million cars. For example, just in San Diego, California, over half a million cars were screened between 2012 and 2018 (Payton et al. 2018). Moreover, assuming that approximately 500 cars are screened per night, this translates to over 1.5 million cars being screened annually if we assume that approximately 33,000 DUI checkpoints were conducted between 2008 and 2018 (DUIBlock.com).

suspension law) when it cannot deter DUIs using one of the common enforcement strategies.

In Appendix Figure 1.5, I experiment with using different spatial controls, models, and other functional forms of the outcome variable, and document several findings. First, omitting the state-specific linear time trend does not alter the results. Second, the estimates are robust to including other spatial controls. Third, the results are robust to using (i) inverse hyperbolic sine transformation, (ii) the rate of DUI fatalities per 100,000 people, (iii) the share of total fatalities involving drunk drivers, (iv) the monthly rather than the yearly level, and (v) estimating fatality counts using a Poisson model. However, the estimated effects are slightly sensitive to using fatality levels, an outcome used by Jones & Morin (2022). My findings suggest a slightly larger and more persistent effect size that evolves over time. I argue that the functional form of the outcome used and the possible omitted variable biases, as shown in Appendix Figure 1.4, are likely contributing to the smaller estimated effect observed in comparison to Jones & Morin (2022).

Finally, in the last (very right) estimate in Appendix Figure 1.5, I find that the estimated effects are smaller in magnitude, suggesting relatively large effects for states with larger populations. To address this potential heterogeneity, in Appendix Table 1.3 I show estimated effects categorized by "low-population states" and "high-population states," as suggested by Solon et al. (2015) and following Kelly et al. (2020).[42,43] I find that the estimated effects are larger and driven by high-population states. In Appendix Table 1.4, I note that police enforcement—measured by police employment, police expenditures, and the number of DUI checkpoints conducted—is higher in larger states. These differences are consistent with the hypothesis that higher levels of enforcement can lead to reduced criminality and dangerous driving (Chalfin et al. 2022, DeAngelo & Hansen 2014, Evans & Owens 2007, Mello 2019).

Figure 1.3 presents the results from the event study analysis, in which the treatment indicators are replaced with lead-and-lag treatment indicators. In the pre-treatment period, I

[42] Low-population states are defined as states with a below-median 2018 population, and high-population states are defined as states with an above-median 2018 population.

[43] Due to having only five treatment states in each group, I report standard errors calculated using wild-cluster bootstrapping.

find an insignificant pre-trend centered around zero, suggesting that DUI fatality rates do not differ across the treatment and control states before the implementation of a DUI checkpoint ban.[44] This result instills a degree of confidence that the timing of DUI checkpoints is exogenous to the trends in DUI traffic fatalities. In the post-treatment period, consistent with Table 1.2, I find a positive effect of 10%–12%, which persists for about eight years and then becomes smaller. Additionally, the results are once again robust to excluding (panel (a)) and including (panel (b)) state-specific linear time trends, indicating that the preferred specification is not biased from the potential negative weights that can result from state-specific linear time trends.

Figure 1.4 presents the event study figures from the stacked DD, the Sun and Abraham estimators, and the Callaway and Sant'Anna estimators. Because control variables are critical in my model, I use the residuals obtained after regressing the outcome on all my preferred set of controls as the left-hand side variable for the Callaway and Sant'Anna estimator. This approach allows for the flexible inclusion of various sets of control variables.[45] Across all new difference-in-differences strategies, I find an insignificant pre-treatment trend, followed by a short- to medium-run jump of around 10%–12%. Columns (3) and (4) of Table 1.2 present the overall estimated treatment effects from the Sun and Abraham and stacked DD estimators, respectively. These results continue to show a significant, positive short- and medium-run effect and an imprecisely estimated long-run effect. The exception is the Sun and Abraham estimate, which shows a more precisely estimated effect of similar magnitude. These results suggest that the results are robust to using new difference-in-differences estimators, which guard against biases that could arise due to staggered rollout and heterogeneous

[44] When I test whether the lead coefficients sum to zero, I find a χ^2 of 0.3054 and 0.4825 for panels (a) and (b), respectively.

[45] In Appendix Figure 1.6, I simulate a data-generating process (1,000 times) to show that under omitted variable bias, the Callaway and Sant'Anna estimate, when using residuals as the outcome, produces less biased estimates compared to those obtained by simply using the raw outcome variable. In panel (a) of Appendix Figure 1.7, I present the Callaway and Sant'Anna estimates that use the raw outcome instead of residuals. Consistent with Appendix Figure 1.6, Table 1.2, and the event study figure using TWFE without any controls (panel (b) of Appendix Figure 1.7), the estimated effect is greater, indicating the presence of omitted variable biases. However, my preferred Callaway and Sant'Anna estimates alleviate some of the omitted variable biases.

and dynamic treatment effects.

In Appendix Figure 1.8, I explore the sensitivity of the estimates to using alternative definitions of traffic fatalities, which can be used as a proxy for DUIs or as a falsification test for data quality. Specifically, I present Callaway and Sant'Anna estimates for weekend, weekend nighttime, weekday, and work hour traffic fatalities.[46] The first two categories represent times when DUI fatalities are most likely to occur, while the latter two represent times when they are less likely to occur. Panels (a) and (b) show an insignificant pre-trend, followed by a persistent post-treatment increase in crashes during both weekends and weekend or holiday nights. However, the post-treatment coefficients are sometimes imprecisely estimated. The effect size suggests a 5%–11% increase in these fatalities, which is reasonable considering that the estimated effect on traffic fatalities involving impaired drivers is around 10%–12%. Panels (c) and (d) show no increase in traffic fatalities during the time at which DUIs are less likely to occur, suggesting that the increase in DUI fatalities is not attributed to an increase in traffic fatalities during non-peak DUI hours. Together, the results of this exercise confirm that the results are not driven by potential estimation error.

Table 1.3 explores the effect of a DUI checkpoint ban after excluding subsets of counties based on the perceived treatment intensity. In columns (1) and (2), I construct a pseudo-state-by-year panel using large and small population counties. I find that the estimated effect is driven by counties with a larger population (14.7% increase in the short and medium run), aligning with the observation that larger population counties have a higher rate of DUI checkpoints. The findings from columns (3)–(6) are also consistent with the notion that more urban and denser counties experience larger treatment effects. Together, these results lend support to my assumption that the change in the use of DUI checkpoints is driving the

[46] The weekend is defined as any time from Friday evening at 6:00 pm to Monday morning at 5:59 am. Weekend or holiday nights are defined as the hours from 6:00 pm to 5:59 am the following morning, applicable to weekends, federal holidays, and the evenings before federal holidays. Weekdays are defined as the period starting from 6:00 am on Monday to 5:59 pm on Friday, and work hours are defined as the period from 8:00 am and 5:59 pm from Monday to Friday, excluding any times that fall on federal holidays.

results.

To investigate whether DUI checkpoint bans have an offsetting effect on non-DUI fatalities due to changes in police enforcement or traffic congestion, I estimate their effect on these fatalities. Column (1) of Table 1.4 shows little evidence of total fatalities increasing following the ban on DUI checkpoints. Column (3) indicates that this null effect on total fatalities may be offset by a reduction in speeding-related fatalities; however, the estimates are too imprecise to draw firm conclusions.[47] This reduction in speeding-related fatalities aligns with the hypothesis that police officers, substituting away from DUI checkpoints to other enforcement strategies such as saturation patrols, which can deter speeding, may have spillover effects on other drunk drivers (DeAngelo & Hansen 2014).[48] Finally, column (4) shows no evidence of non-DUI fatalities increasing during the nighttime on weekends, when DUI checkpoints are most likely conducted. This suggests that the change in the flow of traffic or congestion created by checkpoints may not be a predominant channel.

1.4.2 DUI Arrests

The FARS estimates provide robust evidence of a short-run impact and some evidence of a medium-run effect of a DUI checkpoint ban on traffic fatalities involving drunk drivers. Two possible channels exist through which this positive effect can occur. First, if DUI checkpoints are an effective deterrent method, banning them can lead to increased DUIs. Second, if they are effective in catching drunk drivers, banning them can also lead to fewer DUI arrests, which can cause more drunk drivers being involved in a car accident rather than getting caught. To explore these mechanisms, I next explore the impact of a DUI checkpoint ban on DUI arrests.

[47] In Panel I of Appendix Table 1.5, I examine heterogeneous treatment effects by counties for non-DUI speeding-related fatalities. The findings suggest a similar pattern of results: larger short-run treatment effects for larger, more urban, and more densely populated counties. However, the estimates are still imprecisely estimated.

[48] In a supplemental analysis where I regress log police employment capita as the outcome variable, I find no evidence of a reduction in police employment following a DUI checkpoint ban, suggesting that the bans may lead to different types of enforcement rather than to a reduction in police force sizes.

In Table 1.5, I observe a similar pattern to those found for traffic fatalities. Column (1) shows the TWFE estimate without including any of the preferred controls, which suggests a statistically and economically significant increase in reported DUI arrests in the short and medium run after a DUI checkpoint ban. Column (2) shows that the estimated effect is robust to including various control variables, with the preferred specification implying a 12%–17% increase in the reported DUI arrests 10 years after a ban. Columns (3) and (4) report similar results using the new difference-in-differences techniques, suggesting a 12%–20% persistent increase in DUI arrests. The UCR estimates, coupled with the FARS estimates, imply that DUI checkpoints can create a general deterrent effect and the prohibition of DUI checkpoints induces more people to drive drunk.

In Figure 1.5, I present the event study analyses to assess the validity of the parallel trends assumption. Panel (a) uses a standard TWFE estimate, and panels (b)–(d) use alternative difference-in-differences estimators. Across the four panels, I consistently find no evidence of a pre-treatment trend in DUI arrests and an increase of around 10%–20% in the post-treatment period. The results suggest that the estimates can be interpreted as causal.

Appendix Figure 1.9 shows the robustness of the UCR estimates to various data-cleaning strategies. In the first (very left) estimate, I replicate my main preferred estimates found in column (2) of Table 1.5. In the next estimate, I calculate the crime rate using the state-level population as the denominator. The point estimates remain largely unchanged, suggesting that the main results are not driven by the use of agency population as the denominator. In the next four estimates, I account for any state-agency-time-specific unobservables, such as underreporting by agencies, that could bias the estimates.

To identify anomalies in the observations, I experiment with two methods. First, I fit a time-series model for each jurisdiction and exclude any state-by-year observations with significant spikes and drops.[49] I decompose each jurisdiction's trend into seasonality, trend, and unobservable components and classify any observation as an outlier if its unobservable

[49] Due to larger noise at a finer level such as agency level, conducting this exercise at a finer level yields fewer detection. However, the results are qualitatively similar.

exceeds three times the interquartile range. For the second experiment, I use a simple two standard deviation rule to detect for outliers in the data. Any state-by-month-year observations that are classified as an outlier are either excluded from the analysis, replaced with just the trend and seasonality components, or replaced with the mean. The findings from this exercise reveal somewhat larger effect sizes, suggesting that agencies that underreport crimes may be biasing the estimates downward. Nevertheless, these results provide confidence that the positive effect I am finding is not an artifact of poor data quality and that the main UCR effect is the lower bound.

Finally, in the last two estimates in Appendix Figure 1.9, I experiment with stricter sample restrictions. First, I create the state-by-year panel using agencies that report crimes in all 12 months to address any issues that may arise from inconsistent reporting. I then use only primary agencies, defined as agencies with a population greater than zero.[50] I continue to find a positive and persistent effect of 12%–14%, though the estimates are imprecisely estimated for the longer-run coefficients. In summary, these estimates strengthen my confidence that measurement errors that may arise in the UCR are not positively biasing the positive effect found in Table 1.5.

In the remaining UCR regressions, I continue to check for the robustness of the UCR estimates using different specifications. In panel II of Appendix Table 1.5, I explore the effect of a DUI checkpoint ban by removing a subset of counties based on the perceived intensity of treatment. The findings in this table continue to show that the counties that are more likely to conduct a DUI checkpoint observe a larger treatment effect. Appendix Figure 1.10 shows the Callaway and Sant'Anna estimates on arrests from other crimes, serving as a falsification test for the quality of the UCR data. The rationale is that if some unobservable changes in crime reporting are correlated with the treatment, then other crime outcomes should also increase when DUI checkpoints are banned. Furthermore, I can also examine whether the

[50] An example of a zero population agency is a university police, which covers the same jurisdictions as a city police department.

increase in DUI arrests is due to the UCR hierarchy rules.[51]

In panels (a) and (b) of Appendix Figure 1.10, I estimate the effect of a DUI checkpoint ban on index crimes, which are less likely to be affected by a checkpoint and are also considered more severe offenses than DUIs. I do not find any evidence of changes in property crime or violent crime. This null effect implies that the results are not driven by idiosyncratic changes in how agencies report crime. In panel (c), I examine the effect on crimes that are commonly committed along with DUIs to detect any changes in arrests involving multiple offenses.[52,53] I find no evidence of a decline in crimes involving drugs or weapons after the ban on DUI checkpoints. This exercise confirms that the positive effect on DUI arrests is not driven by the change in the composition of the type of offenses being reported to the UCR.

1.4.3 Self-Reported DUIs

I next turn to the BRFSS data to estimate whether the ban on DUI checkpoints leads to more DUI behavior. In columns (1) and (2) of Table 1.6, I find a statistically significant short-run increase of 0.4 percentage points (8% increase relative to the baseline mean) in self-reported DUIs during the last 30 days. These results are consistent with those of Kenkel (1993), who finds that DUI checkpoints lead to a modest decrease in self-reported DUIs. Focusing on people who reported drinking (columns (3) and (4)), I find a larger effect of around 1 to 1.1 percentage points (11%–13%) increase. In the longer run, I continue to find

[51] The UCR uses a hierarchy rule where it reports only the most serious crime in an incident involving multiple offenses. For example, if an offender is arrested for DUIs and murder, then they will be classified under the charge of murder, as it is a more severe crime compared to a DUI. This rule creates a problem if the total number of DUI arrests remains unchanged, whereas the total number of arrests for DUI in conjunction with other crimes (for example, drug offenses) is changing. In this case, the total number of reported DUIs will change despite the number of actual DUI arrests remaining the same.

[52] The offenses I examine are drug offenses, weapon violations, assaults, and vandalism.

[53] Offenders committing multiple crimes may be a concern because according to my calculations using the National Incident-Based Reporting System, in 2018 roughly 30% of offenses resulting in a DUI arrest involved other offenses including drug offenses, vandalism or destruction of property, assaults, and weapon law violations.

a positive effect, but the estimates are smaller and imprecisely estimated.[54] These results suggest that in the extensive margin, DUI checkpoint bans are increasing DUI behavior, consistent with the results that I uncover with the FARS and UCR data.

In columns (5) and (6), I examine the effect on the intensive margin, finding a persistent increase in the frequency of driving under the influence among individuals who reported driving drunk. However, inference using wild-cluster bootstrapping methods cannot conclude any significant effect.[55] During the first 5 years after the ban, the number of times an individual reports driving while drunk increases by 8%–14%. Over time, the effect increases, and the long-run coefficients suggest a 31% increase 11 or more years after the ban. These results are consistent with the possibility that the ban gives an immediate shock to drivers previously deterred by a checkpoint, while those who are driving drunk may progressively increase their frequency of doing so over time. These findings are consistent with the theory of general deterrence, where drivers respond more to current or future DUI checkpoints than to previous checkpoints or arrests.

1.4.4 Heterogeneity

In Table 1.7, I estimate heterogeneous treatment effects by different race groups (white versus black).[56] In columns (1) and (2), I find that the effect on DUI arrests is consistently greater for white individuals than for black individuals. The magnitude of these estimates indicates a weakly significant and persistent 13%–19% increase in DUI arrests involving white offenders but a smaller insignificant 8%–11% increase involving black offenders. However, I cannot rule out with 90% confidence if the effects are significantly greater for white individuals than for black individuals.

Focusing on the estimates using BRFSS data (columns (3)–(6)), I find that, contrary

[54] When using the logit or probit model (columns (1)–(4) of Appendix Table 1.6), I continue to find similar patterns of magnitudes but a statistically significant medium-run effect.

[55] Poisson estimates in column (5) of Appendix Table 1.6 also confirm this result.

[56] Because FARS did not start reporting drivers' races until 1991, I did not conduct this exercise using the FARS data.

to the results from the UCR data, banning DUI checkpoints may lead to an increase in self-reported DUI behaviors among black individuals compared to white individuals. One possible explanation for these heterogeneous treatment effects is the location where police officers decide to conduct a DUI checkpoint. If checkpoints are placed disproportionately and more frequently in neighborhoods with more racial minorities, then they can also have a disproportionate deterrent effect. Together, these point estimates support the notion that prohibiting DUI checkpoints may alleviate racial gaps in DUI arrests and that DUI checkpoints could contribute to greater racial disparities. However, I cannot conclusively reject the hypothesis that DUI checkpoints are equitable.

In Appendix Table 1.7, I consider heterogeneous treatment effects by different age groups. Given the complexity in identifying dynamic effects by cohorts, I report the overall average treatment effects.[57] These findings are generally mixed. Panel I shows the largest effect on DUI behavior for the older group (30–39 years old and 40 or older). For DUI arrests, panel II shows that the effect may be larger for the younger group (29 or younger and 30–39 years old). However, it is uncertain whether these effects are significantly different between different age groups or not. Turning to the BRFSS estimates from panels III and V, I find the largest (and significant) effect among 30- to 39-year-olds. However, there is limited evidence on whether these effects are significantly different between different age groups.

1.5 Conclusion

Since DUI checkpoints were introduced, lawmakers have debated about whether to allow states to continue conducting them. As of 2022, DUI checkpoints are not conducted in 11 states, with some considering implementing bans (Chaduvula 2019, Gutierrez 2021). Although these checkpoints have been in place for several decades, there is limited evidence on their causal effect and the potential consequences of making them illegal. In this paper,

[57] For example, the estimated short-run effects for ages 16–29 will be identified off different samples than the long-run effect for the same age group, since those who are between 16 and 29 within the first few years following a DUI checkpoint ban will become part of the older cohort group after several years.

I fill this gap by exploiting the spatial and temporal variation in state policies banning DUI checkpoints to estimate their impact on DUI fatalities, arrests, and behavior.

Using data from FARS, I find a short- and medium-run effect of DUI checkpoint bans, though the medium-run effects are sometimes imprecisely estimated. My preferred model's point estimate suggests a 12.4% increase in DUI fatalities within the first five years following a ban, equating to an annual cost of approximately $6.4 billion in terms of lives lost from DUIs.[58] These findings are robust to various functional forms and the use of new difference-in-differences approaches that account for treatment effect heterogeneity. Event study analyses and falsification tests further support the causal interpretation of these estimates.

An important channel for the increase in alcohol-related traffic fatalities appears to be an increase in DUI incidents. Analysis of data from the UCR reveals a persistent increase in DUI arrests. Additionally, data from the BRFSS indicate a short-run increase in self-reported DUI behavior on the extensive margin after DUI checkpoints are banned. Together, these findings highlight the effectiveness of salient and targeted enforcement strategies, such as DUI checkpoints, in curbing dangerous driving behaviors across a broad spectrum of drivers, regardless of their sobriety status.

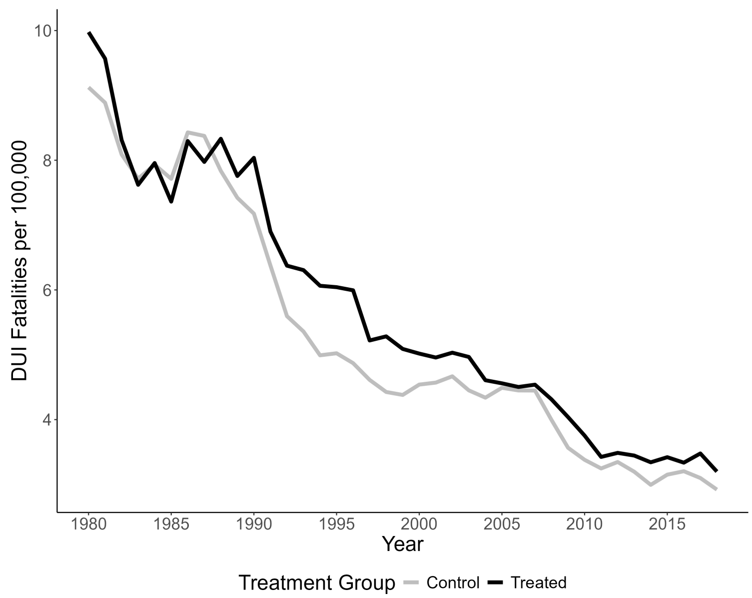
[58] To estimate this number, I take the estimated percentage change from the preferred estimate to calculate the total increase in DUI fatalities among states that allow the use of DUI checkpoints. I then convert the total additional fatalities to total annual costs, assuming that the value of statistical life is approximately $7 million (Banzhaf 2021).

1.6 Table & Figures

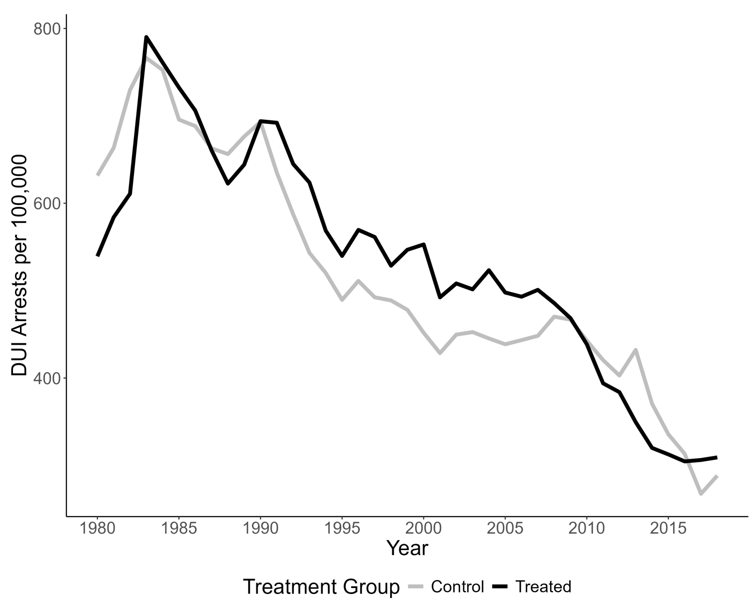
1.6.1 Figures

Figure 1.1: Trend in DUI Fatalities & Arrests

(a) DUI Fatalities by Treatment Status



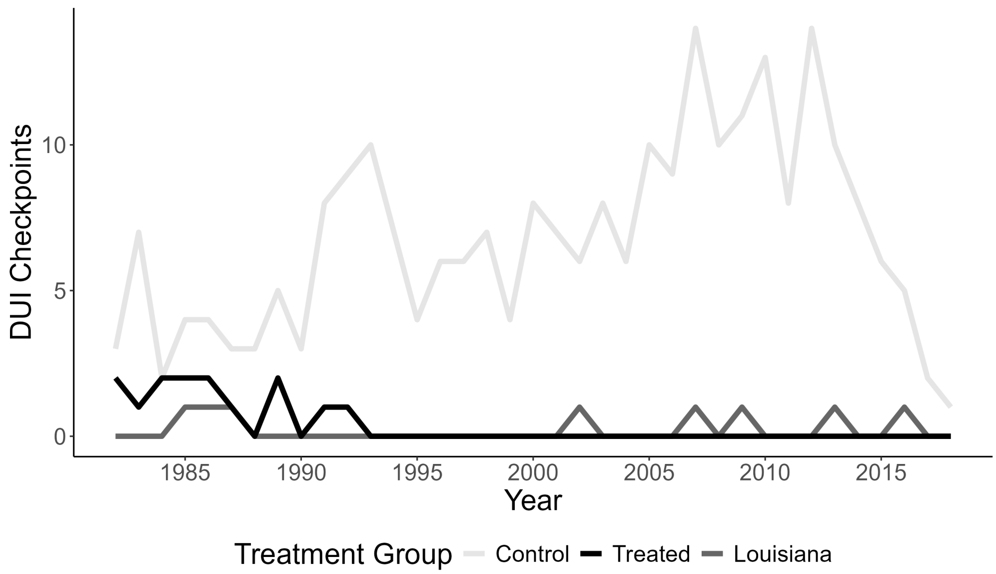
(b) DUI Arrests by Treatment Status



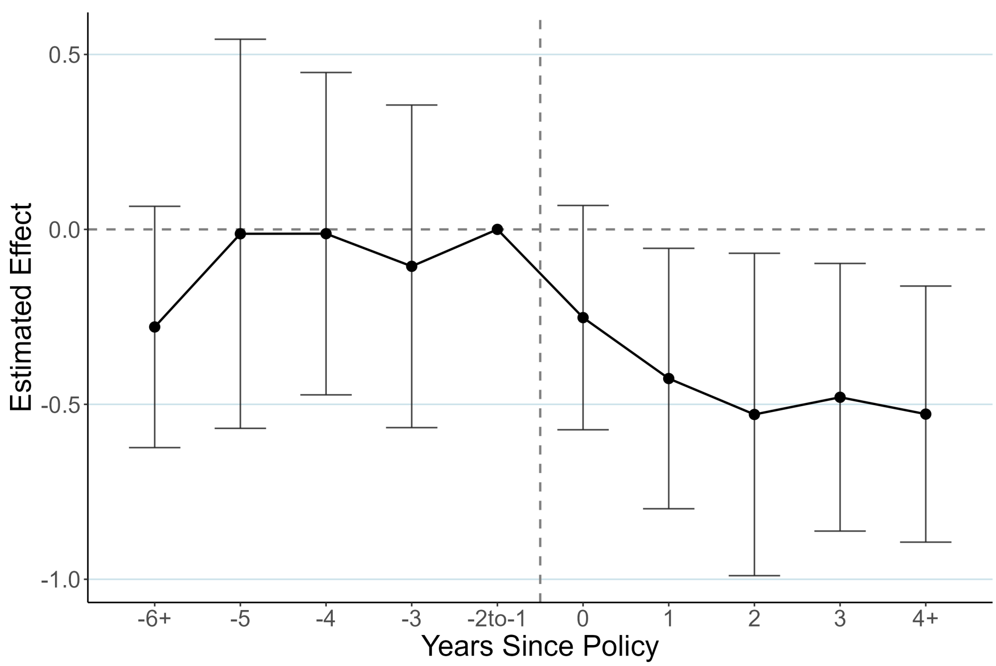
Notes: Panel a uses data from the FARS. Panel b uses data from the UCR. The rate is calculated as the total count of traffic fatalities or DUI arrests divided by the total population for each group.

Figure 1.2: "First Stage" Effect of Changes in State Court Cases Arguing About DUI Checkpoints

(a) Descriptive Trends



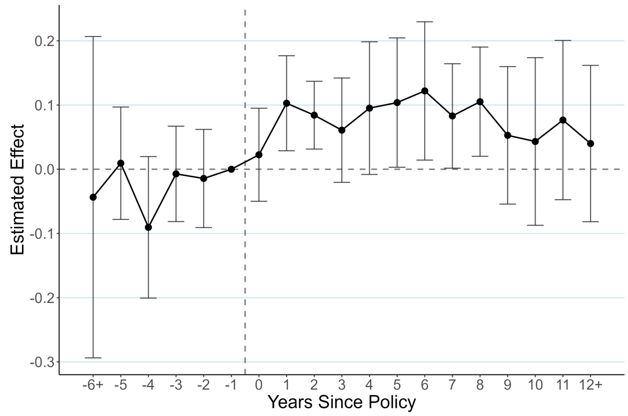
(b) TWFE Estimates



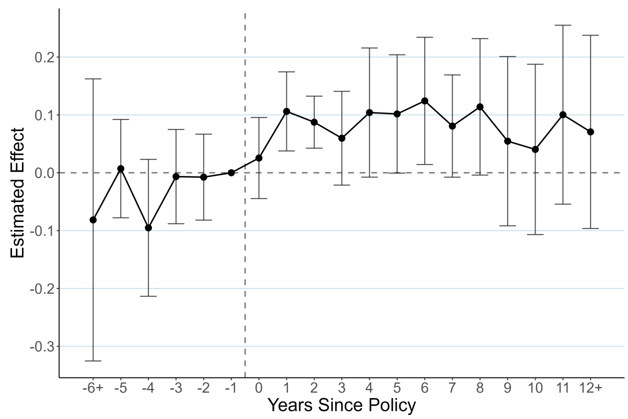
Notes: The data on the number of court cases that mention that a defendant was arrested at a DUI checkpoint are manually collected from casetext.com. Panel a plots the total counts of these court cases. Panel b presents population-weighted TWFE OLS event study estimates. The estimate in panel b includes state and time fixed effects. The bar lines in panel b represent 95% confidence intervals generated using standard errors clustered at the state level.

Figure 1.3: TWFE Event Study Analysis on Drunk Driving Fatalities; FARS 1980-2018

(a) Excluding State-Specific Linear Time Trends



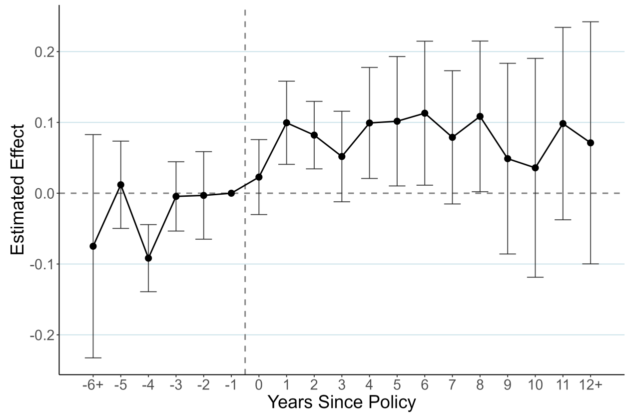
(b) Including State-Specific Linear Time Trends



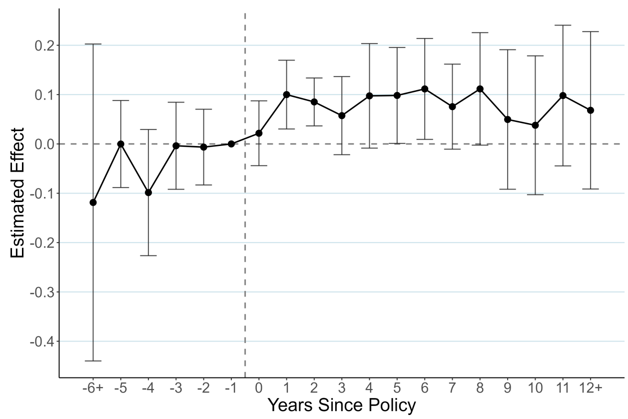
Notes: Population-weighted estimates are generated using data from the 1980-2018 FARS. All estimates include state and year fixed effects and my preferred set of control variables. Panel a excludes state-specific linear time trend and panel b includes state-specific linear time trend. The bar lines represent 95% confidence intervals generated using standard errors clustered at the state level.

Figure 1.4: Robustness to the Use of New Difference-in-Differences Strategy; FARS 1980-2018

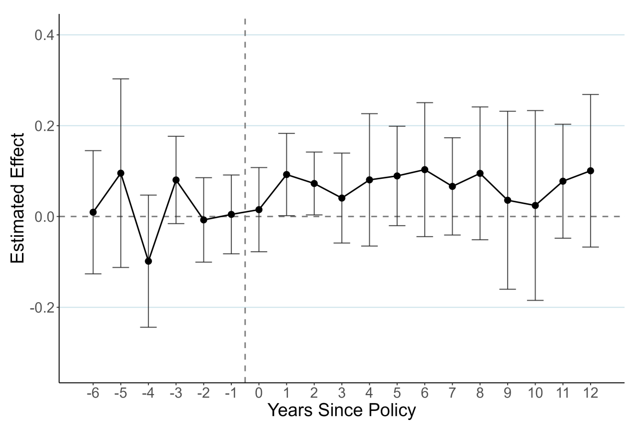
(a) Sun & Abraham



(b) Stacked DD



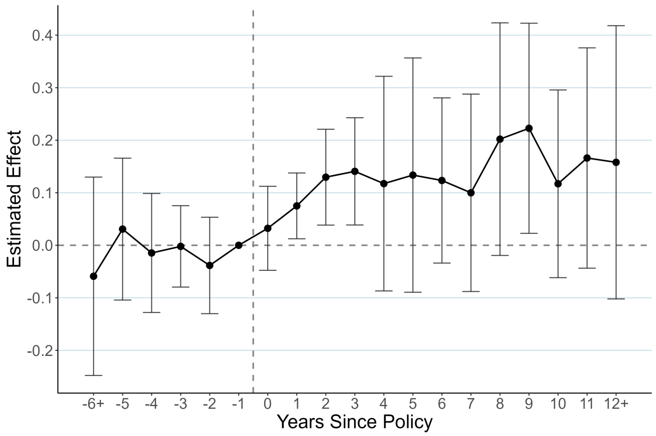
(c) Callaway & Sant'Anna



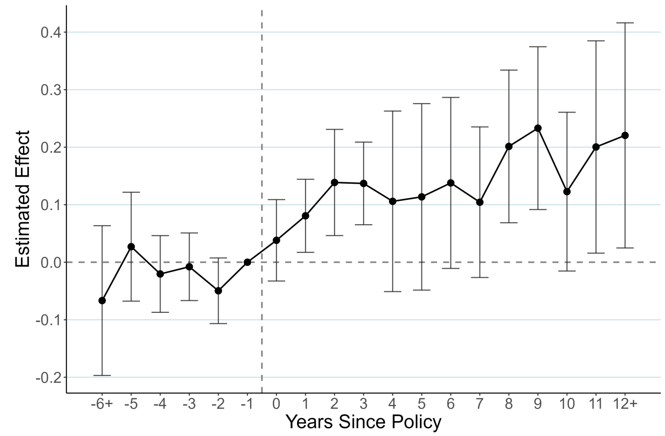
Notes: Population-weighted estimates are generated using data from the 1980-2018 FARS. All estimates include state and year fixed effects and my preferred set of control variables. For panel c, the outcome variable is the residuals after I regress my outcome on my preferred set of controls. The counterfactuals for panel c are restricted to not-yet-adopting states. The bar lines represent 95% confidence intervals generated using standard errors clustered at the state level (panels a and b) or bootstrapped standard errors (panel c).

Figure 1.5: Event Study Analysis on DUI Arrests, UCR 1980-2018

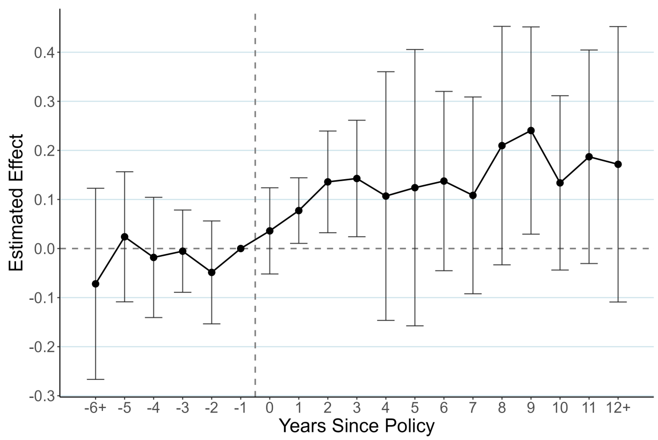
(a) TWFE



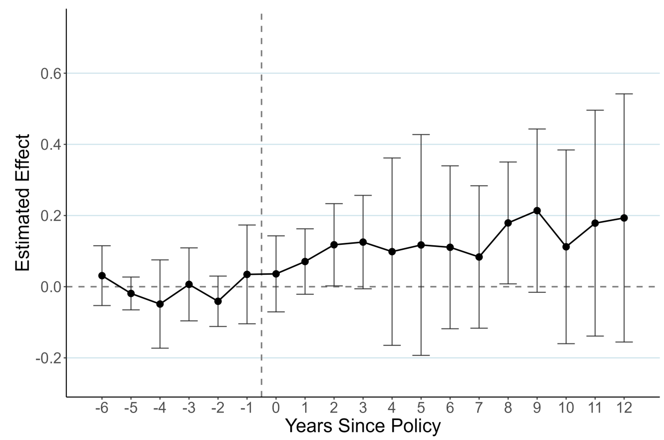
(b) Sun & Abraham



(c) Stacked DD



(d) Callaway & Sant'Anna



Notes: Population-weighted estimates are generated using data from the 1980-2018 UCR. All estimates include state and time fixed effects and my preferred set of control variables. For panel d, the outcome variable is the residuals after I regress my outcome on my preferred set of controls. The counterfactuals for panel d are restricted to not-yet adopting states. The bar lines represent 95% confidence intervals generated using standard errors clustered at the state level (panels a to c) or bootstrapped standard errors (panel d).

1.6.2 Tables

Table 1.1: Year of DUI Checkpoint Bans by State

State	Treament Year
Alaska	Pre-1980
Idaho	1988
Iowa	1986
Louisiana	1989
Michigan	1993
Minnesota	1994
Oregon	1987
Rhode Island	1989
Texas	1991
Washington	1988
Wisconsin	1991
Wyoming	Pre-1980

Table 1.2: Estimated Effect of DUI Checkpoint Bans on Drunk Driving Fatalities, FARS 1980-2018

	(1)	(2)	(3)	(4)
0 to 5 Year After	0.170*	0.117*	0.123**	0.129^{+}
	(0.067)	(0.049)	(0.034)	(0.069)
6 to 10 Years After	0.142^{+}	0.115^{+}	0.125**	0.119^{+}
	(0.084)	(0.058)	(0.039)	(0.062)
11+ Years After	0.046	0.097	0.105^{+}	0.093
	(0.098)	(0.069)	(0.058)	(0.066)
N	1,970	1,970	1,970	14,138
Controls?	No	Yes	Yes	Yes
Model	TWFE	TWFE	Sun & Abraham	Stacked DD

$^{+}$ P < .10; * P < .05; ** P < .01

Notes: Population-weighted estimates are generated using data from the 1980-2018 FARS. All estimates include state and year fixed effects and all of my preferred set of control variables. Standard errors clustered at the state-level are reported inside the parenthesis.

Table 1.3: Estimated Effect of DUI Checkpoint Bans on DUI Fatalities: Using Subset of Counties, FARS 1980-2018

	Population		Urbanicity		Density	
	(1)	(2)	(3)	(4)	(5)	(6)
0 to 5 Year After	0.137*	0.028	0.131**	0.069	0.121*	0.090
	(0.052)	(0.058)	(0.046)	(0.075)	(0.046)	(0.072)
6 to 10 Years After	0.137*	-0.023	0.135*	0.041	0.124*	0.056
	(0.059)	(0.074)	(0.054)	(0.096)	(0.056)	(0.092)
11+ Years After	0.125^{+}	-0.095	0.131*	-0.015	0.109	0.019
	(0.071)	(0.113)	(0.061)	(0.113)	(0.067)	(0.110)
N	1,970	1,719	1,967	1,809	1,970	1,773
	≥45000	<45000	≥60	<60	≥1000	<1000

$^{+}$ P < .10; * P < .05; ** P < .01

Notes: Population-weighted estimates are generated using data from the 1980-2018 FARS. All estimates include state and year fixed effects and all of my preferred set of control variables. Standard errors clustered at the state-level are reported inside the parenthesis. Data are aggregated to a state-by-year panel using specific sets of counties. Column 1 uses counties with at least 45,000 population. Column 2 uses counties with less than 45,000 population. Column 3 uses counties with at least 60% of residents living in urban areas. Column 4 uses counties with less than 60% of residents living in urban areas. Column 5 uses counties with a weighted population density of at least 1000. Column 6 uses counties with a weighted population density of less than 1000.

Table 1.4: Estimated Effect of DUI Checkpoint Ban on Other Traffic Fatalities

	(1)	(2)	(3)	(4)
	All	Non-DUI	Non-DUI: Speeding	Non-DUI: Weekend Night
0 to 5 Year After	0.005	-0.052	-0.111	-0.070
	(0.023)	(0.032)	(0.074)	(0.047)
6 to 10 Years After	0.021	-0.024	-0.006	0.001
	(0.040)	(0.040)	(0.126)	(0.058)
11+ Years After	0.008	-0.039	-0.044	-0.076
	(0.049)	(0.058)	(0.142)	(0.092)
N	1,970	1,970	1,867	1,970

$^{+}$ P < .10; * P < .05; ** P < .01

Notes: Population-weighted estimates are generated using data from the 1980-2018 FARS (columns 1, 3, and 4) and 1982-2018 FARS (column 2). All estimates include state and year fixed effects and my preferred set of control variables. Standard errors, clustered at the state level, are reported inside the parenthesis.

Table 1.5: Estimated Effect of DUI Checkpoint Bans on DUI Arrests: UCR 1980-2018

	(1)	(2)	(3)	(4)
0 to 5 Year After	0.134^{+}	0.118^{+}	0.109*	0.130^{+}
	(0.075)	(0.065)	(0.042)	(0.076)
6 to 10 Years After	0.159^{+}	0.154*	0.158**	0.181*
	(0.087)	(0.071)	(0.057)	(0.078)
11+ Years After	0.121	0.143	0.195*	0.171
	(0.122)	(0.109)	(0.084)	(0.112)
N	1,896	1,896	1,896	13,504
Controls?	No	Yes	Yes	Yes
Model	TWFE	TWFE	Sun & Abraham	Stacked DD

$^{+}$ P < .10; * P < .05; ** P < .01

Notes: Population-weighted estimates are generated using data from the 1980-2018 UCR. All estimates include state and year fixed effects, and my preferred set of control variables. Data are aggregated to state-by-year panel using agencies that report at least once a month. Standard errors, clustered at the state level, are reported inside the parenthesis.

Table 1.6: Estimated Effect of DUI Checkpoint Bans on Self-Reported Drunk Driving, BRFSS 1984-2018

	Drink Drive Yes/No				Drink Drive Frequency	
	(1)	(2)	(3)	(4)	(5)	(6)
0 to 5 Years After	0.0039*	0.0041*	0.0113*	0.0096*	0.1276*	0.0814*
	(0.0018)	(0.0017)	(0.0049)	(0.0040)	(0.0532)	(0.0399)
	{0.083}+	{0.08}+	{0.068}+	{0.073}+	{0.071}+	{0.125}
	[4e-04,0.0076]	[4e-04,0.0079]	[0.0015,0.0217]	[9e-04,0.0184]	[0.0168,0.2418]	[-0.0052,0.1687]
6 to 10 Years After	0.0009	0.0018	0.0060	0.0042	0.1696	0.1335
	(0.0021)	(0.0019)	(0.0064)	(0.0047)	(0.1091)	(0.0914)
	{0.677}	{0.37}	{0.539}	{0.539}	{0.244}	{0.326}
	[-0.0029,0.005]	[-0.0013,0.005]	[-0.0046,0.0175]	[-0.0053,0.0144]	[-0.0372,0.3802]	[-0.0718,0.3492]
11+ Years After	0.0025	0.0017	0.0099	0.0080	0.2690+	0.2695+
	(0.0042)	(0.0037)	(0.0103)	(0.0076)	(0.1551)	(0.1467)
	{0.599}	{0.692}	{0.445}	{0.467}	{0.157}	{0.184}
	[-0.0061,0.0119]	[-0.0058,0.0097]	[-0.0082,0.0295]	[-0.0075,0.0256]	[-0.0277,0.5725]	[-0.0508,0.6103]
Mean of DV	0.051	0.051	0.086	0.086	0.466	0.466
N	4,221,619	4,221,619	2,157,199	2,157,199	78,412	78,412
Sample	Everyone	Everyone	Drinker	Drinker	Drunk Driver	Drunk Driver
Controls?	No	Yes	No	Yes	No	Yes

+ P < .10; * P < .05; ** P < .01

Notes: Survey-weighted estimate is generated using data from the 1984-2018 BRFSS. Every estimates include state and year-by-month fixed effects and all of my preferred set of control variables and demographic controls (gender, race/ethnicity, age, and education). The sample for columns 3 and 4 is restricted to individuals who reported drinking at least one alcoholic beverages in the past 30 days. The sample for columns 5 and 6 is restricted to individuals who reported drinking and driving at least once in the past 30 days. Standard errors clustered at the state-level are reported inside the parenthesis. The p-values obtained using wild-cluster bootstrapping are shown inside the curly bracket and the 90% CIs generated using wild-cluster bootstrapping are shown inside the square bracket.

Table 1.7: Heterogeneous Treatment Effects by Whites vs. Blacks

	Arrests (1)		Drink-Drive Everyone (2)		Drink-Drive Drinker (3)		Drink-Drive Frequency (4)	
	Whites	Blacks	Whites	Blacks	Whites	Blacks	Whites	Blacks
0 to 5 Year After	0.119	0.099	0.004	0.009	0.009*	0.021	0.087^{+}	0.243^{+}
	(0.074)	(0.117)	(0.002)	(0.008)	(0.004)	(0.017)	(0.043)	(0.142)
			{0.175}	{0.492}	{0.1}	{0.388}	{0.099}$^{+}$	{0.196}
			[-0.001,0.008]	[-0.013,0.031]	[1e-04,0.018]	[-0.024,0.069]	[0.002,0.174]	[-0.06,0.559]
6 to 10 Years After	0.178^{+}	0.104	0.002	-2e-04	0.005	-0.001	0.116	0.433*
	(0.091)	(0.159)	(0.002)	(0.008)	(0.004)	(0.019)	(0.088)	(0.202)
			{0.369}	{0.992}	{0.418}	{0.961}	{0.35}	{0.348}
			[-0.001,0.006]	[-0.038,0.038]	[-0.004,0.014]	[-0.082,0.081]	[-0.069,0.309]	[-0.241,1.146]
11+ Years After	0.174	0.076	0.002	-0.003	0.007	-0.001	0.241	0.612**
	(0.135)	(0.159)	(0.004)	(0.008)	(0.006)	(0.018)	(0.147)	(0.227)
			{0.712}	{0.838}	{0.419}	{0.977}	{0.209}	{0.083}$^{+}$
			[-0.006,0.01]	[-0.023,0.017]	[-0.005,0.02]	[-0.048,0.049]	[-0.069,0.567]	[0.065,1.227]
N	1,914	1,903	3,629,480	348,140	1,917,265	133,358	68,868	4,511

$^{+}$ P < .10; * P < .05; ** P < .01

Notes: Weighted estimates are generated using data from the 1980-2018 UCR (columns 1 and 2), and 1984-2018 BRFSS (columns 3 to 8). All estimates include state and time fixed effects, and my preferred set of control variables. Standard errors, clustered at the state level, are reported inside the parenthesis.

Pretextual Stop Restriction and Policing: Evidence from Los Angeles

Abstract: This paper explores the impact of the Los Angeles Police Department (LAPD)'s restriction on pretextual stops on policing behavior and public safety. Using data on all California traffic stops, I find compelling evidence that the policy led to an immediate reduction in stops for equipment or non-moving violations. However, I find little evidence that the overall number of total stops decreased in the short run, potentially due to police substitution behavior. This finding is consistent with my economic framework, which suggests that police officers will respond to increased scrutiny placed on some tasks by shifting their behavior to other tasks. At the same time, I find that this policy led to an approximately 15 percent reduction in the number of racial minorities stopped by police officers. Focusing on traffic stop outcomes, I document that the number of stops resulting in a warning decreased, and conversely, the number of stops resulting in a citation may have increased. Moreover, the policy led to fewer searches and contraband found, but little change in contraband seized. Finally, I find little evidence that the number of reported crimes, arrests, and traffic fatalities increased following the restriction of pretextual stops. Together, my findings imply that the LAPD's pretextual stop restriction achieved its intended goal of reducing racial disparities without diminishing public safety.

2.1 Introduction

With recent social justice movements and incidents of police brutality, the public's trust in the police has fallen recently. For instance, in 2023, confidence in the police fell to an all-time low of 43 percent, down 10 percentage points from 2019 and 21 percentage points from a record high in 2004 (Gallup 2024). Furthermore, another Gallup poll shows that 89 percent of Americans believe policing needs major or minor changes (McCarthy 2022). This decline in trust in the police highlights the ongoing principal-agent problem between the public and the police. In this context, the public (principal) aims to maximize social welfare by minimizing crime and traffic accidents while upholding individual rights and respect. However, police officers (agents) face the challenge of allocating their time between multiple tasks and using their discretion when choosing between their tasks. For example, a police officer may enforce traffic laws or patrol neighborhoods to find and deter criminal behavior. Moreover, an officer may decide to stop and search specific types of individuals over others. Thus, according to multitasking theory, police will focus their effort on measured and rewarded tasks at the expense of other tasks (Dumont et al. 2008; Holmstrom & Milgrom 1991 Hong et al. 2018; Johnson et al. 2015; Knutsson & Tyrefors 2022; Reeves 2024).

This paper first provides empirical evidence supporting police multitasking theory by examining the trade-off between conducting pretextual stops and enforcing traffic laws. Pretextual stops involve stopping a driver for minor infractions, such as equipment or non-moving violations, to investigate other suspected criminal activity. For instance, an officer might pull over a driver for a broken taillight or tinted window and ask, "Do you know why you were pulled over?" The officer asks this question, hoping the driver admits to a more serious crime or violation. While police officers claim these stops are effective in detecting crime, several studies have shown that pretextual stops can lead to higher racial disparities in criminal justice (Makofske 2023; Naddeo & Pulvino 2023; Parker et al. 2024; Rushin & Edwards 2021).

While the U.S. Supreme Court ruled that pretextual stops are constitutional, several cities have enacted policies to reduce them. Los Angeles Police Department (LAPD) is one jurisdiction where the use of pretextual stops is limited. Under LAPD's new policy, officers can only stop a driver who is a severe public threat and must articulate why they stopped an individual. By restricting these stops, this policy increased the cost of a specific task that police officers commonly employ.

Using individual police stop data from California and a combination of regression discontinuity and difference-in-discontinuity approaches, I initially compare immediate changes in different types of traffic stops around the policy reform date and the racial impacts of such changes. In addition, I supplement my analysis using data from the City of L.A. Open Data Portal, California Open Justice, Uniform Crime Reporting, and the Statewide Integrated Traffic Records System to investigate changes in traffic stop outcomes, traffic accidents, or crime rates due to the policy. Because the effects of some of my outcomes may evolve more gradually, I also employ a synthetic difference-in-differences estimator, comparing long-run changes in outcomes in Los Angeles relative to synthetic counterfactuals.

I document an immediate 1.8 per 100,000 reduction in traffic stops due to equipment or non-moving violations. However, in the short-run, the overall number of police-initiated stops stayed the same due to increased stops for other minor moving violations, such as speeding or failure to stop at a stop sign. These findings are consistent with the multi-tasking theoretical model that predicts that this new policy that increases scrutiny placed on some tasks may lead to police substitution behavior such as increased traffic enforcement (Dumont et al. 2008; Holmstrom & Milgrom 1991 Hong et al. 2018; Johnson et al. 2015; Knutsson & Tyrefors 2022; Reeves 2024). Focusing on these effects by race, I find that the policy significantly reduced the number of Black drivers stopped by approximately 10.6 per 100,000 and the number of Hispanic drivers stopped by approximately 2.6 per 100,000. In contrast, there was no change in the number of White drivers stopped.

Focusing on the outcome of traffic stops, I find that the number of warnings has decreased by 2.5 to 2.9 per 100,000 (or 23.8 to 27.0 percent), while the number of citations may have increased. These findings continue to support the potential substitution pattern I am uncovering. I also find that the number of stops that led to searches decreased by 1.2 to 1.3 per 100,000 (or 17.1 to 18.3 percent), and consequently, the number of contraband found decreased by 0.3 per 100,000 (or 16.1 to 16.7 percent). However, I find little evidence that contraband seized decreased. These findings imply that the policy did not lead to significant adverse effects regarding police officers finding less severe contraband.

Finally, I find a statistically insignificant relationship between the pretextual stop restriction and the number of reported crimes in the long term. Moreover, I find no evidence that traffic accidents increased in Los Angeles City or County. These findings suggest that the prohibition of pretextual stops did not adversely impact public safety as measured by the number of traffic accidents or reported crimes.

My paper contributes to the literature in several ways. My first contribution is exploring multitasking behavior and potential trade-offs within the context of police-initiated stops.

Much of the existing literature provides empirical evidence supporting that police officers respond to higher incentives or higher costs (Chalfin & Gonçalves 2023; Kim 2022; Luh et al. 2023; Makowsky & Stratmann 2009; Makowsky et al. 2019). However, few studies investigate whether exerting more effort on one task reduces the quality or quantity of other tasks.[1] Garoupa & Klerman (2002) provides theoretical modeling predicting police multitasking among which types of crime to target. Their findings suggest potential trade-offs between major crimes vs. minor crimes. However, their results depend on several conditions, including the offender's wealth and whether the law enforcement is competitive. Reeves (2024) finds that officers reduce traffic enforcement immediately following a collision response, consistent with the multitasking responsibilities that officers face. In this paper, I empirically show whether raising the cost of one type of stop increases the quantity of other types of stops.

This paper also contributes to the literature on the policies that aim to reduce police discrimination. Several studies have explored whether specific policies such as the adoption of body-worn cameras (Ferrazares 2024), Connecticut's collaborative approach (Parker et al. 2024), Consent Decree (Fagan & Geller 2020), mandatory police training (Dube et al. 2023; Mello et al. 2023), federal oversight (Campbell 2023; Long 2019; Shi 2008), and prosecutorial reform aimed to reduce pretextual stops (Naddeo & Pulvino 2023) are effective in curbing racial disparities in police-encounters. I estimate the causal effect of a unique police department-initiated policy reform that aims to improve racial disparities by reducing pretextual stops. The LAPD's policy is unique in two ways. First, this restriction is a department-wide policy that increases the oversight of police officers and requires a police officer to be held accountable for breaking the policy. For this reason, unlike other policies (e.g., Naddeo & Pulvino 2023), which indirectly affect officers, this policy may directly impact police behavior. Second, this policy does not entirely prohibit police officers from

[1] For instance, Makowsky et al. (2019) finds that officers focus their effort more on arresting for drugs, drunk driving, and prostitution when local fine and forfeiture revenues are higher, but do not study whether this change led to a reduction in other arrests which may not necessarily generate fiscal revenues.

stopping a driver for a specific type of violation, and stops due to equipment or non-moving violations do not entirely disappear. Understanding how this policy impacts racial discrimination can have valuable policy implications, as other jurisdictions (i.e., Denver, CO) are implementing such policies to reduce racial disparities.

The third contribution of this paper is exploring how police enforcement practices affect criminality. While increased police presence reduces criminality and dangerous driving (Chalfin et al. 2022; DeAngelo & Hansen 2014; Evans & Owens 2007; Levitt 2004; Matsuzawa 2022; Mello 2019), some policing strategies may not be as effective as others (Abrams et al. 2023; Banerjee et al. 2019; MacDonald et al. 2016; Tebes & Fagan 2022).[2] Naddeo & Pulvino (2023) and Parker et al. (2024) both leverage a policy change that resulted in fewer pretextual stops and find little evidence that crime or traffic accidents changed, suggesting that pretextual stops may not be a valuable tool to promote public safety. Because my policy led to police substitution behavior where the enforcement of other traffic violations increased, I contribute to this literature by investigating whether marginal pretextual stops are effective or ineffective relative to other traffic enforcement.

Finally, I explore pretextual stops in an unexplored setting. Previous studies, which have shown that pretextual stops are ineffective and inequitable, have focused on various jurisdictions, including Louisville, KY (Makofske 2023), Saint Paul, MN (Naddeo & Pulvino 2023), and state of CT, TX, and WA (Feigenberg & Miller 2023; Parker et al. 2024; Rushin & Edwards 2021). I complement their findings by focusing on a different jurisdiction. Los Angeles is the second largest city in the U.S., has the highest number of Hispanic residents, and is significantly more racially diverse than other cities previously examined. Moreover, Los Angeles is known for its high prevalence of driving. Finally, the neighborhoods within the city vary significantly in terms of their characteristics, including public safety, gang presence, and socioeconomic status.[3]

[2]For instance, Banerjee et al. (2019) use a randomized control trial and document that DUI checkpoints placed in fixed locations do not affect drunk driving rates. However, they find that randomly assigned DUI checkpoints have a deterrent effect.

[3]For example, according to the 2022 Census, median household income in Los Angeles by zip code ranged

The remainder of my paper is formatted as follows: Section 2.2 discusses the background of the policy and theoretical prediction of the policy. Section 2.3 discusses the data. Section 2.4 discusses my identification strategy. Section 2.5 presents the results and Section 2.6 concludes.

2.2 Background

2.2.1 Background on Policies Surrounding Pretextual Stops

Police-initiated stops for suspicious activities have been ongoing for at least 60 years. For instance, in 1963, a police officer suspected that three men were about to commit a robbery, so the officer stopped and detained the suspects. This detainment ultimately led to the police officer finding illegal weapons and arresting two of the suspects. The arrest led to a 1968 *Terry v. Ohio* U.S. Supreme Court case. In this court case, the defendant argued the constitutionality of stopping an individual without a warrant and that it violated the Fourth Amendment rights. The court ruled that police officers can stop and search a person as long as they have "reasonable suspicion."

The U.S. Supreme Court again debated the constitutionality of pretextual stops and stop-and-search after traffic violations in the 1996 *Whren v. United States* court case. In 1993, a police officer stopped a car driving in a high-drug neighborhood for "unreasonable" speed without using their turning signals. The result of the stop was that the officer found drugs and arrested the driver and the passenger. The defendant argued that the traffic stop was a pretext to investigate possible drug crimes without probable cause. However, the U.S. Supreme Court judges unanimously decided that violating any traffic laws constitutes a legitimate reason to stop a driver.

While the U.S. Supreme Court allowed police officers to use pretextual stops and to stop a driver for any traffic violations, recently, many cities and jurisdictions began prohibiting

from $50,000 to more than $200,000.

such traffic stops.[4] These bans occur through several ways. First is through a legislative change where a law explicitly prohibits police officers from stopping drivers for specific traffic violations.[5] Another way is through a more indirect way where prosecutors stopped prosecuting felonies that arise from minor traffic violations. These policy reforms began, especially with the recent rise of the Black Lives Matter movement, because some policymakers and the public believe that pretextual stops contribute to increased racial discrimination without being effective in finding contraband and promoting public safety.

Los Angeles Police Department (LAPD) is one such jurisdiction that began limiting the use of pretextual stops.[6] On February 1, 2022, LAPD announced that they are considering the restriction of pretextual stops because they believe that pretextual stops are ineffective and have undermined public trust in the police. After this announcement, the police chief collected public opinion on this matter via email. Between February 1 and 15, LAPD received 123 emails, of which 78 rejected the initial proposal and demanded more stringent measures of banning pretextual stops.[7] On March 1, 2022, the five-member Los Angeles Police Commission unanimously approved the new policy, which aims to limit such stops without eliminating them entirely.

This new policy, effective immediately, now requires police officers only to stop a driver if the driver is a public threat. When a police officer makes a stop, he must articulate a valid reason for stopping in their body-worn cameras.[8] Unlike other city's policies where a stop for certain traffic violations is prohibited, under this new policy, a police officer can still stop for any traffic violations as long as they can articulate why such traffic violations constitute a public threat. A violation of this new policy – failure to articulate the reason – leads to disciplinary consequences, including mandatory counseling and retraining. This policy was

[4]Some cities that imposed a ban include Philadelphia, Saint Paul, Minnesota, and Pittsburgh.

[5]For example, in Philadelphia, the new law states that police officers can no longer stop drivers for a single brake light or headlight out.

[6]Following the LAPD's policy, on January 1, 2024, California also imposed a statewide restriction in a similar fashion as the LAPD.

[7]Some examples of the emails can be found at https://tinyurl.com/2p9pwu7a.

[8]For instance, a person's race, homeless circumstance, or presence in a high-crime location is not a valid reason.

intended to be a compromise between the public's demand to fully eliminate pretextual stops and the demand of the police union, which still wanted the right to conduct pretextual stops.

2.2.2 Economic Framework

To theoretically predict the effect of LAPD's pretextual stop limitation on policing behavior, I develop a simple model about the officer's decision to stop a driver. According to the models by Anwar & Fang (2006), Abrams et al. (2023), and Feigenberg & Miller (2023), the officer will stop or search a driver if the benefit of such actions outweighs the cost. I build upon these models by adding an additional layer of costs: the opportunity costs of making such stops. Due to time constraints, if an officer stops the driver he currently observes, he has to forgo the next driver he will observe and can potentially stop.

Suppose an officer encounters a random driver and decides whether to stop or let go of that driver. Following the literature, I assume that the objective function of the officer is to maximize his total utility from stopping and searching a driver minus some costs. An example of such utility includes some benefits that the officer receives from writing a ticket or successfully finding contraband, both of which may lead to a higher probability of the officer getting promoted or getting a higher salary (Join LAPD 2023). An example of the costs is the officer's effort when making these stops.

Let driver i be a random driver that the officer observes. For simplicity, suppose that this driver can be grouped into four groups by type of violations, denoted as t, and by race, denoted as r. I suppose that the two types of traffic violations are minor traffic violation, which is more likely to be pretextual stops, (m) or other traffic violation (o), and the two races are white (w) or black (b).[9] The officer observes each group with a probability ϕ_{tr}. Upon observing the driver, the officer also observes an idiosyncratic signal of the total benefit of stopping and searching that driver, denoted as υ_i.[10] I assume that υ_i is distributed by

[9]Minor traffic violation includes equipment or non-moving violation. Other traffic violation includes moving violation such as speeding.

[10]For example, the officer will observe the severity of the offense (e.g., speed). The officer can use this

some random density function that differs between groups, which I denote as $f(\upsilon_i|t, r)$.

The officer's utility for stopping an individual (i) who is committing traffic violation (t) and who is race (r) is as follows:

$$U(\upsilon_i, t, r) = \max_{\substack{\text{stop} \\ \text{let go}}} \{\upsilon_i - c_t, E(U_2)\} \tag{2.1}$$

In Equation (2.1), c_t is the direct cost of making each type of traffic stop.[11] $E(U_2)$ is the expected utility the officer receives from stopping the next random driver that he observes (i.e., focusing his effort on the next task he is provided). I set $\beta = 1$, which means that the officer is indifferent between stopping the same driver right now vs. in the next period (i.e., a few minutes later), ceteris paribus.

In other words, the maximization problem in Equation (2.1) suggests that the officer is choosing between doing the task today (and receiving some benefit from stopping the current driver) or waiting and doing the task in the future (and receiving the expected net benefit from the next driver). Under this maximization problem, I can solve for the officer's decision to stop the driver as follows:

$$\text{Stop if } \upsilon_i > c_t + E(U_2)$$
$$\text{Let go if } \upsilon_i < c_t + E(U_2).$$

The officer will only stop the driver if the net benefit from stopping and searching exceeds a threshold, defined as the total (direct and opportunity) costs of stopping a driver.[12] This finding is consistent to the officer-decision rules shown in Anwar & Fang (2006), Abrams et al. (2023), and Feigenberg & Miller (2023).

information to assess how much potential fines he can give to the driver.

[11]The cost may also vary across individuals (e.g., some individuals may be more violent against an officer than other drivers). I include any of these individual-specific costs in the υ_i term. Moreover, I assume that the costs of making each stop are homogeneous across races. My results are qualitatively similar if I assume heterogeneous costs (and heterogenous policy shocks) across races.

[12]The direct costs (c_t) include the disutility from putting effort into making these stops. The opportunity costs include the foregone benefit from not stopping the next driver.

Using this decision rule, the probability that the officer stops different types of drivers will become the probability that v_i exceeds a specific stop threshold. Mathematically, I can write the probability of the officer stopping an individual who is violation type (t) and race (r) as follows:

$$P(\text{stop}|t, r) = \int_{c_{t^*}}^{\infty} f(v|t, r)dv = 1 - F(c_t^*|t, r),$$

where $c_t^* = c_t + E(U_2)$.

To predict how the probability of stopping a certain group (e.g., $P(\text{stop}|t = m)$) changes as a result of a policy restricting pretextual stops, I make two assumptions regarding the LAPD's policy. First, I assume that the policy increased the direct costs of stopping the driver for minor traffic stops because the officer is required to exert more effort in making minor traffic stops (i.e., articulating the reason), and the policy increased scrutiny. Mathematically, I assume that $c_m^{new} > c_m^{old}$ or $dc_m > 0$. In addition, because the policy increased the costs of stopping the driver for minor traffic stop violations, the expected utility from alternative tasks – which may, in some chance, include observing and stopping another driver violating minor traffic stops– decreases. However, I assume that the rate of decrease is less than one because there is some chance that the next driver may not be committing other traffic violations, in which the costs (and hence the net benefit) remain constant. Mathematically, I assume that $-1 < dE(U_2)/dc_m < 0$. Under these two assumptions, I can predict the change in the probability of stopping a specific type of driver.

Prediction 1: LAPD's pretextual stop limitation will lead to fewer minor traffic stops and more stops for other traffic violations (e.g., substitution from one task to another).

To make this prediction, I first solve for the probability that the officer observes a driver violating a minor traffic (or other traffic) violation and stops that driver. Next, I take the derivatives of each probability with respect to c_m. As shown in Section 4.2.1, mathematically, I have the following:

$$\begin{aligned}\frac{dP(\text{stop}|t=m)}{dc_m} &= \sum_r -\phi_{mr} f(c_t^*|t=m,r)(1+\frac{dE(U_2)}{dc_m}) < 0 \\ \frac{dP(\text{stop}|t=o)}{dc_m} &= \sum_r -\phi_{or} f(c_t^*|t=o,r)(\frac{dE(U_2)}{dc_m}) > 0.\end{aligned} \tag{2.2}$$

Equation (2.2) implies that the policy reduces the probability of stopping a driver for a minor traffic violation. On the other hand, Equation (2.2) also implies that the likelihood of stopping a driver for other traffic violation increases when the cost of minor traffic stops (or alternative tasks) increase. Intuitively, an officer will reduce the number of minor traffic stops as the total costs for these stops increase. However, the higher costs of making minor traffic stops lower the opportunity costs of stopping drivers for other violations, as the benefits from alternative tasks (i.e., stopping another driver for a minor violation) diminish. As a result, an officer is more likely to stop drivers committing other traffic violations instead of waiting for the next driver to stop. This prediction is consistent with multitasking theory, where higher costs for one task shift efforts to alternative tasks.

Prediction 2: The aggregate effect of LAPD's pretextual stop limitation on racial disparity is ambiguous. The sign of the effect depends on (i) the probability density function around the pre-policy threshold and (ii) the probability that the officer observes each violation type.

To make this prediction, I solve for the probability that the officer observes Black drivers and stops that driver. As shown in Section 4.2.1, the derivative of this probability with respect to c_m will become:

$$\begin{aligned}
\frac{dP(\text{stop}|r=b)}{dc_m} &= -\phi_{mb} f(c_m^*|t=m,r=b)(1+\frac{dE(U_2)}{dc_m}) - \phi_{ob} f(c_o^*|t=o,r=b)(\frac{dE(U_2)}{dc_m}) \\
\frac{dP(\text{stop}|r=b)}{dc_m} &< 0 \text{ if } \frac{\phi_{mb} f(c_m^*|t=m,r=b)}{\phi_{mb} f(c_m^*|t=m,r=b) + \phi_{ob} f(c_m^*|t=o,r=b)} < |\frac{dE(U_2)}{dc_m}| \\
\frac{dP(\text{stop}|r=b)}{dc_m} &> 0 \text{ if } \frac{\phi_{mb} f(c_m^*|t=m,r=b)}{\phi_{mb} f(c_m^*|t=m,r=b) + \phi_{ob} f(c_m^*|t=o,r=b)} > |\frac{dE(U_2)}{dc_m}|
\end{aligned} \tag{2.3}$$

The prediction in Equation (2.3) indicates that the probability of Black drivers being stopped depends on four parameters: ϕ_{mb}, ϕ_{ob}, $f(c_m^*|t=m,r=b)$, and $f(c_o^*|t=o,r=b)$. The derivative $\frac{dP(\text{stop}|r=b)}{dc_m}$ decreases as ϕ_{mb} increases and increases as ϕ_{ob} becomes larger. This condition suggests that, all else equal, if an officer is more likely to observe Black drivers committing minor traffic violations, the likelihood of a stop decreases. Conversely, if an officer is more likely to observe Black drivers committing other violations, the probability of a stop increases.

The remaining two parameters, $f(c_m^*|t=m,r=b)$ and $f(c_o^*|t=o,r=b)$, represent the probability density functions for Black drivers violating minor and other traffic violations. The derivative in Equation (2.3) decreases with the former and increases with the latter, implying that if more drivers are near the threshold for being stopped for minor violations than for other violations, the overall number of Black drivers getting stopped may decline.

Intuitively, the overall change in the number of Black drivers stopped is ambiguous due to police substitution behavior, as uncovered in prediction 1. While the number of Black drivers stopped for minor traffic violations will decrease, this decrease may be offset by increased stops for other violations. Consequently, the theoretical predictions suggest that the policy may not fully achieve its intended goal of reducing racial disparities in police traffic stops.

Because I cannot confidently theoretically predict whether LAPD's policy was successful in reducing racial disparities in traffic stops, I next turn to empirical analyses to investigate this question.

2.3 Data

2.3.1 Traffic Stop Data

In 2015, the California legislature passed Assembly Bill 953, the Racial and Identity Profiling Act (RIPA) that requires state and local law enforcement agencies to collect data on all vehicle and pedestrian stops, regardless of the outcome of the stop[13;14] The data collection began and started getting reported in July 2018 from the largest police departments with more than 1,000 peace officers.[15] In January 2019, mid-sized police agencies with 667 to 1,000 peace officers began collecting data. By January 1, 2022, over 500 law enforcement agencies in California were reporting data.

The RIPA traffic stop data for all of California between July 2018 and December 2022 is available from California's Department of Justice (DOJ). While DOJ provides complete information for all cities and agencies that report to RIPA, the reporting comes with a lag, with the most recent data only being updated up to 2022. Moreover, the data provided by the DOJ does not include some publicly available details, such as the location of stops, that other data sources provide. To remedy this concern, I utilize the most updated city-level data and only supplement it with DOJ data when the city level is unavailable. For example, I obtain my primary RIPA traffic stop data from the city of Los Angeles open data portal.[16]

Across all cities, the data from the RIPA is at the individual level and provides various information about each stop. First, officers must provide information on all people's perceived demographics, including gender, race, and ethnicity. Second, officers are required to report the reasoning behind the actions taken, such as the violation code for the basis of a stop and why the officer searched. Finally, they also report the outcome of each stop,

[13] The basis for this law was to identify and eliminate racial biases in traffic stops.

[14] A stop is defined as any police-initiated detention or search.

[15] The eight agencies include California Highway Patrol, City of Los Angeles, San Diego, and San Francisco police departments, and the sheriff departments of Los Angeles, San Bernardino, Riverside, and San Diego Counties.

[16] The data is available at: https://data.lacity.org/Public-Safety/LAPD-RIPA-AB-953-STOP-Person-Detail-from-7-1-2018-/bwdf-y5fe and https://data.lacity.org/Public-Safety/LAPD-RIPA-AB-953-STOP-Incident-Details-from-7-1-20/5gp9-8nrb

including whether they found any contraband during the search and whether they arrested, cited, or warned the driver.

Using RIPA data, I construct police agency-by-day panel data.[17] My first set of outcomes is the total number of traffic stops due to (i) minor traffic violations, (ii) other minor moving violations, and (iii) any violations, which I convert to the rate per 100,000.[18] I define minor traffic violations, which are "likely pretextual stops," as those due to any non-moving or equipment violation. While the raw data indicates whether a traffic violation is classified as a moving, non-moving, or equipment violation, this classification can be inconsistent across officers and agencies. For example, if an officer stops a moving vehicle with a broken taillight, he might record this stop as an equipment violation. However, sometimes, he might record the stop as a moving violation because the car was in motion. To address this inconsistency, I use the most frequent classification for each vehicle violation code to determine whether I should categorize it as a moving, non-moving, or equipment violation.[19] Some common examples of likely pretextual or minor traffic stops are broken taillights and expired, missing, or non-visible registration tags.[20] I also define "other minor moving violations" as any moving infraction violation, such as failure to stop at a stop sign or speeding. Officers generally consider these violations less severe and use more discretion when deciding whether to initiate a stop. Finally, stops due to any violations are simply the total number of stops made on a given day.[21]

My next set of outcomes is the rate of non-Hispanic White, non-Hispanic Black, and

[17]I also experiment with weekly and monthly levels. Because treatment happens on Tuesday, for the weekly level, I define a week as starting on Tuesday and ending on Monday.

[18]For a city police agency, I use that city's population as the denominator. For county sheriff's offices, I use that county's population, and for state highway patrol, I use the state's population to calculate the rate per 100,000.

[19]For example, if a particular violation is recorded as an equipment violation 70 percent of the time and as a moving 30 percent of the time, I classify it as an equipment violation.

[20]Appendix Table 2.B.1 provides a full list of vehicle violation codes and descriptions of offenses classified as minor traffic stops.

[21]All traffic stops include the two types of stops and also other types of stops, such as major traffic violations, drunk driving, or stops due to suspicious activities. These violations are severe, and police officers will likely stop a driver upon observing such violations. My results show that LAPD's policy did not affect these stops.

Hispanic people stopped per 100,000 people.[22] In addition, I also measure various traffic stop outcomes. First, I count the total number of stops by whether the stop led to (i) a warning, (ii) a citation, (iii) use-of-force, or (iv) arrest, and convert that to rate per 100,000 population. Second, I measure the number of person or vehicle searches per 100,000. Third, I also calculate the total number of stops that led to discovering contraband and the total number of stops that resulted in seizing contraband.[23;24] I also convert these outcomes to the rate per 100,000 population. Fourth, I calculate the hit rate, defined as the total number of searches that yielded at least one contraband discovered (or seized) divided by the total number of searches. Finally, I also calculate the average stop time per stop, measuring the time police officers spend per stop.

In Appendix Table 2.B.2, I present the descriptive pre-treatment (October 1, 2021, to February 28, 2022) characteristics of these two types of traffic violations. I document a few findings. First, I find that minor traffic stops account for approximately 31 percent of all traffic stops. Second, Black or Hispanic individuals are more subject to stops for minor traffic violations than other traffic violations (85.2 vs. 69.0 percent). Third, minor traffic violations are more likely to end up in a warning (76.8 vs. 34.1 percent) or searching (27.8 vs. 12.9 percent) and less likely to end up with a citation (28.3 vs. 66.3 percent). These findings imply that the officer's objective function when conducting these two stops may differ. Finally, while the search rate is different across the two types of violations, contraband discovery rates are relatively similar between the two types of traffic stops (23.6 vs. 25.6 percent), implying that the efficiency of searches may not necessarily be different.

[22] The total number of people stopped is not equivalent to the total number of stops, and that is why I am using a different phrase between the stop and race outcome. For instance, if a stop involves two individuals, I have one total stop but two people stopped. My results are qualitatively similar when my outcome is counting the total number of individuals stopped.

[23] One downside with the RIPA data is that it only indicates whether officers seized any contraband. This shortcoming limits me to count the number of contrabands discovered or seized precisely.

[24] Only about 13 percent of stops with contraband discovered also lead to property seized.

2.3.2 Other Datasets

To supplement my analysis of traffic stops, I obtain other information from various datasets. The city of Los Angeles' open data portal offers valuable datasets about the city. First, I obtain two types of crime data: (i) incident-level crime data for reported offenses and (ii) arrest data.[25] The former dataset provides information on crime that LAPD reported, regardless of whether the case was solved, but only is restricted to severe crime.[26] The latter dataset provides information on all arrests made by the LAPD and on less severe crimes, such as drunk driving. Using these datasets, I construct daily time series data of the reported crimes and arrests.

To supplement my crime analysis for Los Angeles, I also obtain crime data from the California Data Justice Open portal and the Uniform Crime Report (UCR) to increase the size of my control jurisdictions.[27] Both datasets provide monthly counts of reported index crimes and cleared crimes.[28] Using these data, I construct a police agency-by-month and county-by-month panel of reported crime and cleared crime.[29] My sample window starts in June 2021 to ensure a sufficient pre-treatment period without contaminating it with the COVID-19 period and covers up to December 2022. For my agency-level analysis, I restrict my sample to large police agencies covering more than 100,000 people and report crime in all 19 months during my sample window to ensure I have a balanced panel and can interpret zero crime as truly zero crime happening and not missing. For my county-level analysis, I first restrict my police agency to those reporting all 19 months for the same reason stated earlier and to ensure that the sample within each county stays constant over time. Furthermore, after the aggregation into county-level data, I restrict my sample to large counties where the

[25] These data are available at https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8 and https://data.lacity.org/Public-Safety/Arrest-Data-from-2020-to-Present/amvf-fr72.

[26] For instance, drunk driving is not included as part of the offense.

[27] Ideally, I want to focus on using the UCR. However, the LAPD does not consistently report to the UCR.

[28] Index crimes are violent crime (aggravated assault, murder, robbery, and rape) and property crime (arson, burglary, larceny-theft, and motor vehicle theft).

[29] Reported crime is any known crime to the police. Cleared crime is any crime the police solved through arrest or exceptional means.

total population covered by the police agency in my sample exceeds 100,000. My primary sample includes 371 police agencies and 412 counties across the U.S.

I collect information about traffic accidents from the California Statewide Integrated Traffic Records System (SWITRS) to measure any changes in traffic safety. The SWITRS is a database that collects and processes data gathered from a collision scene and provides incident-level information on traffic accidents in California. Using SWITRS, I construct a police agency-by-month and county-by-month panel of total traffic collisions, traffic collisions from speeding, and other vehicular violations such as improper turning.[30] I follow the same sample cuts as the UCR outlined earlier. My primary sample includes 85 police agencies and 35 counties across California.[31] Because SWITRS data covers up to 2023, my sample window includes 2023 to ensure a longer post-treatment window.

2.4 Methods

2.4.1 Short-run Estimation

I begin by estimating a regression discontinuity in time (RDiT) model. The main idea of RDiT is to compare the changes in my outcomes right around the treatment window of March 1, 2022. One advantage of using an RDiT is that it only requires data for the treatment unit. Thus, RDiT will be the only feasible estimator for certain outcomes, such as arrests, where the data for counterfactuals may not exist.

To estimate an RDiT, I implement a two-step augmented local linear methodology proposed by Hausman & Rapson (2018). First, using daily data from the full sample period between July 1, 2018, and December 31, 2022, I identify important regressors by estimating

[30] I note that not all traffic accidents are due to speeding or other vehicular violations because, for example, accidents due to equipment failure can happen when the driver still obeys traffic rules.

[31] Because many traffic accidents occur on highways, the information reported by California highway patrol may be helpful. Hence, for accidents reported by California highway patrol, I split my sample into California highway patrol agencies that cover Los Angeles and those that do not.

the following model:

$$Y_t = \alpha_0 + \alpha X_t + \gamma_d + \rho_m + \phi_y + SIPO_t + u_t. \tag{2.4}$$

In Equation (2.4), Y_t is my outcome variable of interest at the daily level (t). X_t is a vector of control variables, including average daily temperature, average daily precipitation, and total vehicle miles driven.[32] γ_d is the day of the week fixed effects, ρ_m is the month fixed effects, and ϕ_y is the year fixed effects. Because my full sample includes 2020, when COVID-19 was happening, I also include controls for California's COVID-19 stay-at-home orders, defined as $SIPO_t$.[33]

Using the residuals from Equation (2.4), I estimate the main regression discontinuity specification:

$$e_t = \beta_0 + \beta_1 Post_t + \beta_2 \, t + \beta_3 Post_t \cdot t + \varepsilon_t. \tag{2.5}$$

In Equation (2.5), e_t is the residual. $Post_t$ is a dichotomous treatment variable denoting whether the LAPD limitation went into effect. t and $Post_t \cdot t$ are linear functions of the running variable, in which the slopes vary before and after the treatment. To account for autocorrelation in the treatment and outcome, I conduct my statistical inference using Newey-West standard errors (Newey & West 1987; Newey & West 1994).[34] To ensure consistency across my various outcome variables, I use a bandwidth of 2 months.[35]

The advantage of using augmented local linear specification is that it increases statistical power relative to a traditional local linear approach. With a traditional local linear approach, I need to separately identify the "Wednesday" effect from the treatment effect of interest. Moreover, because traffic stops can have seasonality, this approach helps me net out

[32] The data on temperature and precipitation come from NOAA, and the data on vehicle miles come from CalTrans.

[33] My findings are robust to: (i) excluding time window during COVID-19 stay-at-home-order was in place, (ii) excluding 2020, and (iii) restricting the sample to just 2019.

[34] My estimates are more precise when I estimate heteroskedastic-robust standard errors, which is mathematically equivalent to clusterring around the running variable (time) (Lee & Lemieux 2010)

[35] Appendix Figure 2.B.1 plots the optimal bandwidth for various outcome variables I examined. These bandwidths ranged from 35 to 100 days.

March seasonality effects, which is impossible with a traditional local linear because since the treatment occurs on March 1, including March fixed effects with only 2022 data will soak up all my variation. Finally, with augmented local linear, I do not need to worry about including higher-order polynomials for my running variable, which can increase the chances of overfitting (Gelman & Imbens 2019).

In Equation (2.5), my main parameter of interest is β_1, which can be interpreted as the immediate effect of a policy change. The underlying assumption for β_1 to be causal is that there are no other changes in the disturbance term (ε_t) at the cutoff date. Because I aggregate my data to the daily level and the treatment occurs at the beginning of the month, one potential threat to identification can be any changes in enforcement at the beginning of the month vs. the end of the month. Another potential threat to identification is anticipation effects. For instance, I cannot have police officers change their behavior due to the anticipation of such a policy. I conduct several tests to rule out these concerns. First, I experiment with aggregating my data to a lower frequency level, such as the weekly or monthly level.[36] This aggregation helps me soak up any cyclical day-to-day variation. Second, I also experiment with a donut RD where I exclude the 30 days around the threshold, which includes the time window between policy announcement and policy enactment where the anticipations, if any, are most likely to happen.[37]

In addition to a regression discontinuity model, I also experiment with a difference-in-discontinuities estimate (Hansen et al. 2020). The main idea of a difference-in-discontinuities approach is that I compare how discontinuity in LAPD is different relative to a placebo group using other California police agencies.[38;39] To estimate a difference-in-discontinuities model,

[36] Because treatment starts on March 1, 2022 (Wednesday), I define a week starting from Wednesday and ending on Tuesday.

[37] In addition, because the Super Bowl happened in Los Angeles on February 13, 2022, a donut RD can also test if the Super Bowl effect is driving my results.

[38] For my main placebo group, I use 7 police agencies where the RIPA data is available back to 2018.

[39] My results are qualitatively similar when I use LAPD, defining treatment happening in 2019 as my counterfactual.

I estimate the following equation:

$$\begin{aligned} e_{it} =& \beta_0 + \beta_1 Post_{it} + \beta_2 t + \beta_3 \ Post_{it} \cdot t \\ &+ \alpha_1 Treat_i \cdot Post_t + \alpha_2 Treat_i \cdot t + \alpha_3 Treat_i \cdot \ Post_t \cdot t \ + \gamma_i \ + \upsilon_{it}. \end{aligned} \tag{2.6}$$

In Equation (2.6), e_{it} is the residuals from estimating Equation (2.4) for police agency i in time t.[40] $Treat_i$ is an indicator variable, which is one for LAPD and zero for other police agencies. γ_i is unit fixed effects. α_1 is the parameter of interest where I compare how the discontinuity differs between LAPD and other placebo units. In this equation, I allow the slope of the running variable to differ between pre-treatment period, post-treatment period, and by treatment unit vs. control unit. This approach will net out any changes around the cutoff, assuming these changes are common across California. I continue my inference using Newey-West standard errors.

2.4.2 Long-run Estimation

One downside of a regression discontinuity approach is that the Local Average Treatment Effect (LATE) is only specific to the immediate impact around the treatment time. Thus, this model will fail to detect any time-varying, long-run, and dynamic effects of the treatment. While changes in policing behavior may be immediate, changes in some outcomes, such as crime deterrence, may take some time to happen. For this reason, I will next leverage panel data with untreated jurisdictions to estimate the longer-run impact of the policy. Intuitively, I compare how the outcomes in LAPD or Los Angeles changed in the post-treatment period relative to the counterfactual units. Mathematically, I estimate the following model:

$$\underset{\beta_0,\beta_1,\gamma_i,\gamma_t}{\arg\min}\{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \beta_1 LA_i \cdot Post_t - \gamma_i - \rho_t)^2 \omega_i \lambda_t)\}. \tag{2.7}$$

[40]To allow for seasonalities or the effects of weather to vary across agencies, I estimate Equation (2.4) for each police agencies separately and store their residuals. My baseline estimates are qualitatively similar when I estimate one Equation (2.4) for all police agencies with agency fixed effects included instead.

In Equation (2.7), Y_{it} is the outcome for unit (police agency or county) i in month t. LA_i is a dichotomous variable indicating if the police agency (or county) is LAPD (or Los Angeles County).

A natural question posed when estimating Equation (2.7), especially with only one treatment unit, is what the counterfactuals should be and how much weight each control unit (ω_i) and each time period (λ_t) should receive. To analytically determine the optimal weights for counterfactuals, I estimate Equation (2.7) using Synthetic Difference-in-Differences (SDiD) where the optimal unit and time weights are computed by minimizing the sum of squared errors between the observed and predicted values (Arkhangelsky et al. 2021). To ensure that the choice of my counterfactuals does not drive my estimates, I also experiment with an unweighted two-way fixed effects (TWFE) difference-in-differences estimator. Mathematically, this estimation is the same as regressing Equation (2.7) where I use equal unit and time weights for all observations.

Because there is only one treatment unit, I conduct my inference for these estimations using permutation-based p-values. To implement the procedure, I estimate placebo β_1 estimates by estimating additional regressions in each case, replacing the treatment with an indicator for one of the other control units. To calculate the p-values, I can rank the treatment effect and find the relative rankings of Los Angeles. I note that this is a very demanding test, especially when the number of control units is low.[41] For this reason, I supplement my inference using a standard t-test, calculating the standard errors by taking the standard deviations of placebo estimates. This inference assumes homoskedasticity of error term across units (Arkhangelsky et al. 2021).[42] In a supplemental test, I also experiment with conducting hypothesis testing using a rearrangement test (Hagemann 2020). The benefit of such a test is that it allows for inference when I only have one treatment unit and heteroskedasticity of

[41] For example, if I have only 30 control units, to achieve statistical significance, this test would require the treatment estimate to be at least 3rd largest in absolute magnitude.

[42] A procedure proposed by Arkhangelsky et al. (2021) draws 500 placebo estimates and taking the standard deviations of the estimates. However, in the case of one treated unit, 500 unique placebo units are not plausible because I can only estimate the placebo for the number of controls I have.

unknown form. However, a downside of such a test is that this test applies to unweighted TWFE estimates, so I cannot conduct inference for synthetic difference-in-differences.

2.5 Results

2.5.1 Immediate Impact of Pretextual Stop Restriction on Number of Stops

I first focus on the immediate impact of LAPD's pretextual stop limitation on the change in the number of likely pretextual stops or stops for minor traffic violations. In panel (a) of Figure 2.1, I find a large discontinuity in the number of minor traffic stops. The point estimates from column (1) of Table 2.1 suggest a statistically significant reduction in minor traffic stops by 1.8 per 100,000 (or approximately 29 percent relative to the baseline mean). Using a difference-in-discontinuity approach (panel II) and netting out the March 2022 effect, I continue to find that the number of "likely pretextual stops" fell by 2.1 per 100,000 (or 32.5 percent). I note that my estimated "first-stage" effect may be smaller than previously reported (Naddeo & Pulvino 2023). However, the magnitude of my estimated effect is plausible due to the nature of the policy I am examining. Because LAPD did not explicitly ban the use of pretextual stops and instead allowed officers to use them under strict conditions, this policy did not completely eliminate the use of pretextual stops, and the estimated effect on likely pretextual stops may have been smaller.

In contrast to minor traffic violations, I find little evidence of discontinuity in the overall number of police-initiated stops. The point estimate from panel I, column (2) of Table 2.1 suggests that with 95 percent confidence, I can rule out an immediate decrease in all stops by 3.1 per 100,000 (or 15 percent). I note, however, that panel b of Figure 2.1 suggests that there is a structural break in the trend after the policy enactment, potentially driven by longer-run changes in the number of stops for minor traffic violations (panel a of Figure 2.1). This finding implies that the policy may have reduced the overall number of traffic stops in the long-run. Nonetheless, the fact that there is no discontinuity for all stops while there is a large discontinuity in minor traffic violations is an economically significant finding. In column (3) of Table 2.1 and panel (c) of Figure 2.1, I document that a potential increase in other minor traffic enforcement by 1.5 per 100,000 may explain the small change in the overall number of stops. This finding is consistent with the possibility of police substitution behavior, which I predict in my theoretical model.

In Appendix Figure 2.C.1, I conduct robustness tests for my regression discontinuity design. In the very left estimate, I show my baseline specification for comparison. In the second estimate, I show that my estimates are robust when using higher-order polynomials. In the next four estimates, I experiment using a different bandwidth of (i) a narrower bandwidth of 45 days, a wider bandwidth of 75 and 90 days, and (iii) optimal bandwidth. In the following estimate, I experiment using a donut RD, excluding 30 days before and after treatment. In the final two estimates, I experiment with weekly and monthly aggregates rather than daily aggregations. These figures continue to show a large statistically significant reduction in minor traffic stops, offsetting increases in stops for minor moving violations, and little overall changes in the number of traffic stops. These findings confirm that the manual selection of my bandwidth, anticipation, or time-varying treatment effects are not driving my results.[43]

In Appendix Figure 2.C.2, I explore which type of vehicle violation is driving my results. To maintain clarity and readability, I focus on the ten most common types of vehicle violations for minor violations (panel a) and minor moving violations (panel b) and report the percent changes relative to the baseline mean. I find that the reduction in minor violations is primary driven by stops for window obstruction (41.2 percent), illegal parking (39.1 percent), lack of vehicle registration (33.6 percent), headlight violations (33.0 percent), and improper display of registration (23.9 percent). I also observe a sizable reduction in stops for taillamp violations, but this result is imprecisely estimated. For minor moving violations, I find that the increase in stops is driven by failure to obey turn signs (67.5 percent), speeding (39.7

[43] In Appendix Figure 2.C.3, I conduct a regression discontinuity assigning March 1, 2019 as the treatment date. This placebo test confirms little March 1st effect happening in LAPD.

percent), and unsafe lane change (34.2 percent).[44]

Another potential explanation for the police substitution behavior I am uncovering is the changes in how police officers are recording the stops. For instance, a police officer may still stop the driver for a minor traffic violation but may record the reason for the stop as a more severe traffic violation. To address that this reporting issue is not a concern, in Appendix Table 2.C.1, I examine spatial and temporal heterogeneity in the estimated effect of the LAPD's pretextual stop limitation. In columns (1) and (2), I find that the increase in other traffic violations only happens during the daytime, where pretextual stops are less likely to happen, but stops for other traffic violations are more likely to occur during (Appendix Figure 2.B.2). Moreover, in columns (3) to (6), I find that the increase in other traffic stops is larger in zip codes with a higher number of traffic violations or in zip codes with a higher number of traffic accidents, where I may expect police officers to be enforcing traffic more often.[45] These findings suggest that police officers may enforce traffic stops instead of continuing with minor traffic stops but reporting them as speeding infractions.

In Appendix Table 2.C.2, I examine the spatial distribution of traffic stops by neighborhood income and race. In panel I, I find evidence that this policy has decreased stops for minor traffic violations or likely pretextual stops at a higher rate in low-income zip codes (28.7 vs. 22.9 percent) and in zip codes with a higher share of racial minorities (28.6 percent vs. 23.0 percent), where more pretextual stops were happening during the pre-treatment period. Focusing on heterogeneous treatment effects for the number of stops for all other types of violations (panel II), I do not find heterogeneous treatment effects across high vs. low-income neighborhoods. However, I document that the increase in other traffic enforcement is larger in neighborhoods with more racial minorities.

I investigate the racial impacts of pretextual stop restrictions in Table 2.2 and Figure 2.2. In panel I, columns (1) to (3) of Table 2.2 and panels (a), (c), and (e) of Figure 2.2, I find that

[44] Stops for unsafe turns also increased by approximately 22 percent relative to the baseline mean, but this increase is not statistically significant.

[45] Appendix Figure 2.B.3 shows the zip codes above and below the median for the number of traffic accidents and other traffic violations.

the total number of people stopped for likely pretextual stops significantly fell by 10.6 per 100,000 (or 37.6 percent) and 2.6 per 100,000 (or 30.7 percent) for non-Hispanic Black and Hispanic, respectively, but little decrease for non-Hispanic White.[46] In panel II columns (1) to (3) of Table 2.2, when using a difference-in-discontinuities approach comparing LAPD to other placebo California police agencies, I continue to find a large reduction in stops for minor traffic violations involving racial minorities. While the difference-in-discontinuities estimate now shows a statistically significant decrease in the number of non-Hispanic White people getting pulled over, the magnitude of the estimated effect, both in levels and percentage points, still shows that racial minorities may have benefitted more from this policy.

In columns (4) to (6) of Table 2.2 and panels (b), (d), and (f) of Figure 2.2, I focus on the total number of people stopped for any violation. I find a similar pattern of results as minor traffic violations. The number of Black individuals stopped significantly fell by 11.4 to 13.0 per 100,000 (or 15.0 to 17.2 percent), and the number of Hispanic individuals stopped may have reduced by 1.6 per 100,000. However, the estimates for Hispanic individuals are statistically insignificant. In contrast, I find little change in the number of non-Hispanic White people getting stopped. Taken together, these estimates suggest that the policy limiting the use of pretextual stops achieved one of its intended goals of reducing racial disparities in traffic stops. However, these estimates also suggest that racial disparities persist, as the rate at which officers stop Black individuals remains higher than that of White individuals (64.3 vs. 14.0 per 100,000).

In Appendix Table 2.C.3, I focus on whether there are racial differences in the effects of stops for minor moving violations. While my estimates are imprecisely estimated, my point estimates support the idea that racial disparities regarding the number of traffic stops have reduced. While the number of stops for minor moving violations is increasing for all groups, I find a larger increase for non-Hispanic White individuals (7.4 to 16.6 percent) than Black

[46] I note that the total number of people stopped are slightly different from the total number of stops, which I have been examining thus far. On average, the number of people involved per stop does not change between pre- and post-period (1.046 vs. 1.051).

(1.9 to 5.2 percent) or Hispanic (7.6 to 14.0 percent) individuals. Furthermore, the point estimates suggest that all other stops may have fallen for non-Hispanic Black individuals but not non-Hispanic White individuals; however, none of these estimates are statistically significant.

One concern with examining the racial impact of such policy is that the observed reduction in Black or Hispanic individuals getting stopped can be due to the misreporting of race. For instance, with this new pretextual stop limitation and increased officer accountability, police officers may be more incentivized to misreport the driver's race (Luh 2022). To address this concern, first, I argue that the fact that I do not find large increases in White individuals getting stopped may suggest that police officers are not recording non-White drivers as White drivers. Second, in Appendix Table 2.C.4, I compare neighborhoods with higher versus lower frequencies of stops involving racial minorities before the policy implementation.[47] I find that the reduction in stops for Black and Hispanic individuals is more pronounced in neighborhoods where police officers previously stopped racial minorities at a higher rate. In contrast, there is no significant difference in the rate of change for White stops between these neighborhoods. These differential patterns suggest that the observed reduction in Black or Hispanic individuals getting stopped is unlikely to be due to misreporting of race.

Finally, in Table 2.3 and Figure 2.3, I investigate heterogeneous treatment effects across gender. While female and male individuals are being stopped less frequently for minor traffic violations, the reduction is larger for males than females (32.8 to 34.9 percent vs. 20.0 to 27.0 percent). When examining the total number of traffic stops (columns 3 and 4), I observe some evidence that the overall number of stops for male civilians may have decreased, but no changes in the number of female civilians getting stopped. This finding is consistent with the possibility that for male drivers, who are more subject to investigatory traffic stops, LAPD's policy can have heterogeneous effects relative to female drivers (Roach et al. 2022; Smith et al. 2006).

[47] Panel c of Appendix Figure 2.B.3 shows the zip codes above and below the median.

2.5.2 Immediate Impact of Pretextual Stop Restriction on Stop Outcomes

In the next set of analyses, I investigate whether the effectiveness of traffic stops changed. In columns (1) and (2) of Table 2.4, and panels (a) and (b) of Figure 2.4, I first examine whether the number of warnings and the number of traffic citations changed. I document that the number of stops that ended with a warning significantly decreased by 2.5 to 2.9 per 100,000 (or 23.8 to 27.0 percent relative to the baseline mean). On the other hand, I also find that the number of citations issued increased by 1.0 to 1.6 per 100,000 (or 12.9 to 20.3 percent). However, my estimate is only statistically significant at the weakest conventional level for regression discontinuity design (panel I).

These findings from columns (1) and (2) of Table 2.4 suggest a few valuable insights. First, these findings continue to show that police substitution behavior is indeed occurring, rather than the officer reporting minor traffic stops as moving violations. Because most minor or likely pretextual traffic stops (48.9 percent) end up with a warning, and most non-pretextual stops and other traffic violations are more likely (53.2 percent) to end up in a citation, the number of citations should increase if police officers enforce other traffic violations. Finally, these findings suggest that while restricting pretextual stops reduced racial disparities, it may had a disproportionate effect on other groups of individuals, such as low-income people, who are more adversely affected by increased citation (Mello 2021).

In columns (3) to (5) of Table 2.4 and panels (c) to (e) of Figure 2.4, I examine if search behavior and contraband discovery changed as a result of the policy. I document that police searches decreased significantly by 17.1 percent (or 1.2 per 100,000). Consequently, I also find that the number of contraband the officer found decreased by approximately 16.2 percent. However, I find no evidence that the number of contraband seized decreased. These findings may suggest that the reduction in the number of searches may have little adverse effect because the officer is still finding severe contraband that may pose a serious threat to the community.[48]

In columns (6) and (7) of Table 2.4 and panels (f) and (g) of Figure 2.4, I explore the efficiency of search behavior. I find little overall changes when I focus on contraband discovery rates, which I define as the rate of finding something conditional on searches. With 95 percent confidence, I can rule out an increase larger than 4.4 percentage points (or 16.6 percent relative to the baseline mean). On the other hand, I find that contraband seizure rates significantly increased by 33.2 percent. These findings imply that police officers are not efficient in finding contraband but may be more efficient in finding more severe contraband, such as firearms.

In the remainder columns and panels of Table 2.4 and Figure 2.4, I examine other traffic stop outcomes. In column (8) of Table 2.4 and panel (h) of Figure 2.4, I document little change in the average minutes of total officer-civilian interaction. With 95 percent confidence, I can rule out a reduction and increase of more than 3.5 minutes in total time an officer spends during a traffic stop. This null finding suggests that police substitution behavior did not change the time police officers spend on each traffic stop (i.e., from writing more citations), limiting the total hours spent on other enforcement. In panels (i) and (j) of Figure 2.5 and columns (9) and (10) of Table 2.5, I investigate whether use-of-force and arrest after police-initiated stops changed. While my estimated effect is consistent with the possibility that the LAPD's pretextual restriction reduced police use of force and arrests, my estimates are imprecisely estimated to draw a firm conclusion.

The analysis, thus far, focused on the change in the number of traffic stops across all types of stops. In Appendix Table 2.C.6, I investigate how the outcome of the traffic stop changed for the different types of traffic stop violations. Overall, minor traffic stops led to a larger reduction in warnings, citations, searches, contraband discoveries, use-of-force, and arrests than other traffic violations. These reductions may be mechanical due to the

[48] In Appendix Table 2.C.5, I disaggregate the contraband that the officer found into firearms, drugs, and other types of contraband. I find that the reduction is driven by drugs (17.6 to 20.8 percent) and other types of contraband (22.1 to 22.4 percent) rather than illegal firearms (2.4 to 6.9 percent).

total number of minor traffic stops, but the number of other traffic violations increased. An economically significant result from these exercises is that the number of warnings, searches, and contraband discovered during other moving traffic stops decreased.

Given that I am uncovering some racial heterogeneity in the number of stops, in Appendix Table 2.C.7, I re-analyze Table 2.4 columns (1) to (10) by the three race groups. In column (1), I find that while all three groups are experiencing a reduction in the number of warnings, I find that the number of warnings reduced at a higher rate, in both absolute and relative magnitude, for Black (12.0 per 100,000 or 29 percent) and Hispanic people (2.7 per 100,000 or 22.6 percent) than White individuals (0.8 per 100,000 or 20.5 percent). This finding is also consistent with the findings that racial minorities experienced a larger reduction in minor traffic stops and thus have experienced a larger reduction in stops that resulted in a warning. Focusing on the number of citations (column 2), I find that the number of citations may have increased at a higher rate for White individuals (23.1 percent) than Black individuals (9.4 percent), though my estimates are imprecisely estimated.

In column (3) of Appendix Table 2.C.7, I also document some evidence of heterogeneous treatment effects for the number of searches performed. I find that racial minorities experienced a larger reduction in the number of searches performed. This finding also suggests that LAPD's policy improved racial disparities by not just reducing the number of racial minorities getting stopped but also lowering the number of racial minorities being subject to search. In the remaining columns of Appendix Table 2.C.7, I focus on the efficiency of traffic stops. I find some evidence that the officers are discovering less contraband possessed by racial minorities, which is consistent with the earlier findings that the number of searches reduced for racial minorities. Moreover, I find some evidence that the hit rates for White individuals had a larger change (19 percent increase) than racial minorities (5.4 percent increase). However, I cannot rule out if these estimated effects are statistically different across these groups. In the final three columns, I find little significant or meaningful heterogenous treatment effects on average stop time, use-of-force, or arrest rate across races.

2.5.3 Impact of Pretextual Stop Restriction on Reported Arrests & Traffic Accidents

Given that the number of searches fell, one may be concerned that this policy may lead to fewer detection of crime and less deterrence. In panels (a) to (d) of Figure 2.5 and columns (1) to (4) of Table 2.5, I investigate whether restricting pretextual stops affected the number of reported crimes. I find little evidence that crime or arrest rates increased following the policy reform. I find little evidence that the number of arrests and reported crimes increased in Los Angeles in the short run. With 95 percent confidence, I can rule out an 6.8 and 1.8 percent increase in reported property crime and violent crime, respectively.

In panels (e) to (g) of Figure 2.5 and columns (5) to (7) of Table 2.5, I examine whether the policy led to an immediate change in traffic accidents. I find no immediate change in traffic accidents following the policy reform. This null effect on traffic accidents suggests that the driver's behavioral changes (i.e., drivers driving more recklessly due to the policy change) did not drive my increase in other traffic violation stops and continues to show support for the police substitution pattern.

While these short-run results provide valuable insights, the analysis thus far captures only part of the picture of whether public safety changed due to restricting pretextual stops. Theoretically, the change in how police officers enforce traffic and the increased number of citations can reduce speeding accidents. However, the effects may be more dynamic and happen in the long term rather than the week of policy implementation. Similarly, criminal behaviors may change over time as people learn about such policies. Thus, I next turn to a difference-in-differences estimator to investigate the dynamic, longer-run impact of restricting pretextual stops.

In Table 2.6, I present my preferred synthetic difference-in-differences estimates where I find optimal unit weights for each control jurisdiction.[49] In columns 1 to 3, I compare the

[49] The unit and time weights used for the synthetic difference-in-differences estimations are available upon request.

changes in crime rates reported by the LAPD to those reported by other police agencies. In columns 4 to 6, I compare the changes in crime reported in Los Angeles counties to other counties. In addition, in Appendix Table 2.C.8, I present my TWFE difference-in-differences where my time weights and unit weights are equal across my observations.

I note a few findings. First, in panels I and II, I find similar results when I use all of California vs. the rest of the country, suggesting that the data artifacts from combining two different crime data sources are not driving my results. Second, my results are qualitatively similar between synthetic difference-in-differences and TWFE, implying that the data-driven choice of counterfactual units is not driving my results. Third, I note that my estimated coefficient is positive, which is consistent with the possibility that restricting pretextual stops may harm public safety and increase crime, such as burglary or theft, because criminals may have less fear of getting stopped and searched. Another potential explanation for these increases is coincidental increases in crime post-pandemic. Nonetheless, the estimated effect is small (at most a 4.1 percent increase and upper bound of 95 percent confidence interval of 16 percent increase) and statistically insignificant across different inference techniques (p-values ranging from 0.333 to 0.877).[50] Taken together, these findings suggest that restricting marginal pretextual stops had little public safety consequence in terms of increased reported crime in the long run.

In Appendix Table 2.C.9, I present the synthetic difference-in-differences estimates for crime clearance. Another unintended consequence of reducing pretextual stops is that the number of cleared crimes (i.e., solved crimes) can decrease because police officers are not finding conclusive evidence. I find some (7.2 percent) reduction in the total number of clearances for Los Angeles County, but this reduction is not apparent when I focus on LAPD (column 1). Moreover, my estimated effects are imprecise to draw firm conclusions.

One threat to identification is the violation of the parallel trends assumption. In Figure 2.6, I present synthetic difference-in-differences event studies using the rest of the U.S.

[50] Inference for Appendix Table 2.C.8 using a re-arrangement test (Hagemann 2020) continues to fail to reject the null hypothesis of no effect under all possible maximum relative heterogeneity parameters.

as my counterfactuals (Clarke et al. 2023).[51] These figures show little divergence in pre-treatment trends between Los Angeles and synthetic counterfactuals, providing evidence supporting the common trends assumption.

Another potential threat to the null effect can be due to long-run changes in traffic stops. For instance, while minor traffic stops sharply declined immediately, this number may have converged back to the pre-treatment levels in the long run. In Appendix Figure 2.B.4, I rule out this possibility. I show that the raw trend in the number of minor traffic stops has reduced and remained low in the post-treatment period, whereas the number of all police-initiated stops remained similar over time.[52]

In Table 2.7, I present the long-run effect on traffic accidents using synthetic difference-in-differences where I compare LAPD (or Los Angeles County) to other police jurisdictions (or counties) in California.[53;54] Moreover, in Figure 2.7, I present the synthetic difference-in-differences event study figures. The pre-treatment trend is flat and statistically indistinguishable from zero, suggesting the validity of my parallel trends assumption. Moreover, in the post-treatment window, the estimated effect is negative and somewhat large, implying approximately an 8.0 percent reduction in traffic accidents in Los Angeles County and an 11.6 percent reduction in traffic accidents reported by LAPD. These estimated effects are consistent with the possibility of a potential deterrence effect from changes in police traffic enforcement. However, my estimates from Table 2.7 are imprecisely estimated, so I cannot draw a firm conclusion on whether deterrence occurred. Nonetheless, these estimates con-

[51] Appendix Figure 2.C.4 show the event studies using only California as my counterfactuals.

[52] Estimating a longer-run effect of the change in traffic stops is intriguing. However, I note that the monthly trend in traffic stops in LAPD may be unique relative to other police agencies in California, where the number of traffic stops significantly increases during the summer (panel a of Appendix Figure 2.B.5). I also note that my control variables (i.e., vehicle miles traveled or weather) do not explain these unique summer increases (panel b of Appendix Figure 2.B.5). Thus, a longer-run difference-in-differences style estimation may not be feasible because of the lack of valid counterfactuals and the violation of parallel trends. Nonetheless, in Appendix Figure 2.C.5, I estimate a medium-run effect of LAPD's pretextual restrictions on traffic stops using weekly data from January to April and synthetic difference-in-difference. My findings support the idea that the effect may be persistent rather than temporary.

[53] The unit and time weights used for the synthetic difference-in-differences estimations are available upon request.

[54] Appendix Table 2.C.10 shows the results using TWFE.

tinue to imply little adverse effect of LAPD's pretextual stop limitation regarding increased dangerous driving. With 95 percent confidence, I can rule out an 11.3 percent and 12.5 percent increase in traffic incidents reported by LAPD and traffic incidents in Los Angeles County, respectively.

2.6 Conclusion

While many police departments utilize pretextual stops, a considerable number of policymakers question their effectiveness and fairness. Advocates of pretextual stops argue that these stops are indispensable tools for detecting and preventing crime. Thus, they are concerned that removing pretextual stops can increase crime. Conversely, critics argue that these stops contribute to racial discrimination as police officers may disproportionately target racial minorities during these stops.

In this paper, I shed light on this debate by focusing on a unique policy reform that took place in the Los Angeles Police Department (LAPD). This reform restricted the use of pretextual stops, thereby providing a valuable case study for our understanding of the intended and potential unintended consequences of limiting this practice.

Using stop-level data for all police-initiated stops in California and a regression discontinuity estimator, I find that following LAPD's limitation of pretextual stops, there was an approximately 30 percent reduction in minor traffic stops, which are likely pretextual stops. However, I find little evidence that the number of all stops decreased, potentially offset by changes in police enforcement and an increase in other traffic stops. These findings are consistent with the multitasking theory, where higher relative costs of making one particular cost lead to a change in who the officers stop. Focusing on the racial impacts of such policy, I find evidence that this policy led to a 15 to 17 percent reduction in police-initiated stops involving Black civilians but a statistically insignificant reduction in police-initiated stops involving White civilians.

Examining the impact of pretextual stop limitations on traffic stop outcomes and public safety, I document several findings. First, I find evidence that the number of stops that resulted in a warning decreased by 23.8 to 27.0 percent, and conversely, some evidence that the number of citations has increased. Second, I find that the number of searches and contraband found decreased. However, I also find that this reduction in searches led to increased contraband seizure rates. Moreover, I document that this policy change and changes in policing behavior did not lead to statistically significant changes in reported crimes. Finally, I find little evidence that this policy change increased dangerous driving and traffic accidents.

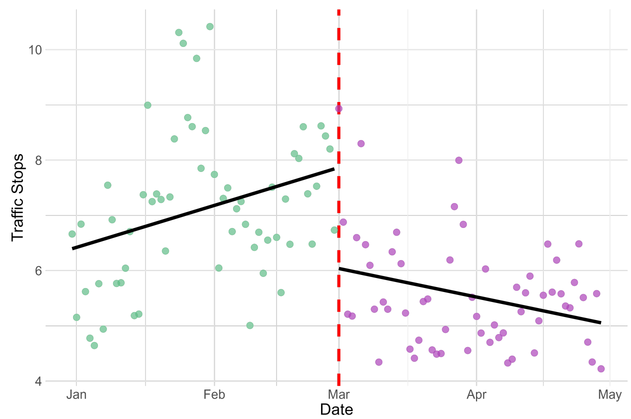
In conclusion, the unique policy introduced by LAPD that restricted the use of pretextual stops achieved some of its goal of reducing, but not entirely eliminating, racial disparities without having many adverse consequences regarding public safety. These findings imply that pretextual stops are more inequitable but not more effective than other police stops, such as enforcing traffic. Taken together, this paper sheds empirical evidence of how increasing scrutiny on some tasks can lead to more socially efficient outcomes.

2.7 Tables & Figures

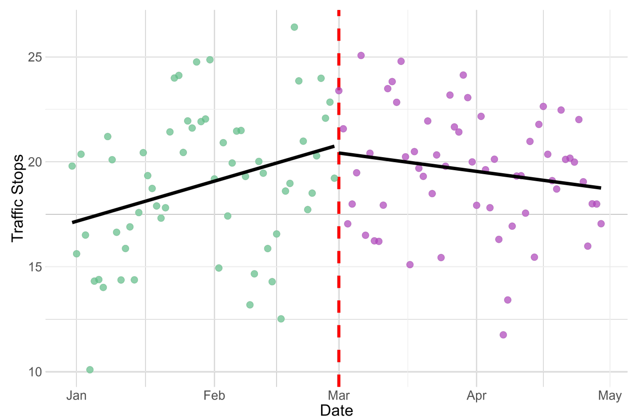
2.7.1 Figures

Figure 2.1: RDiT Estimate: Police-Initiated Stops

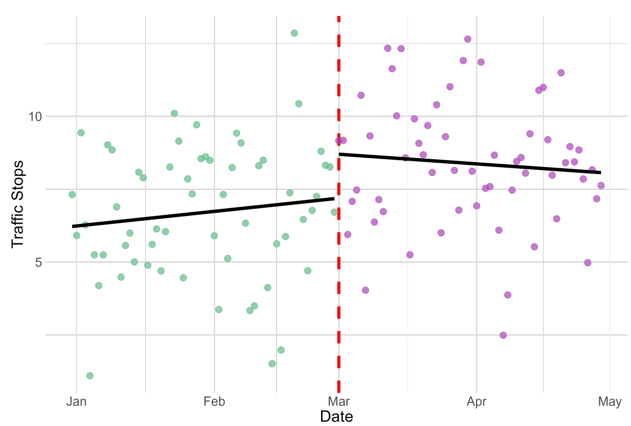
(a) Minor Traffic Violation



(b) All Stops



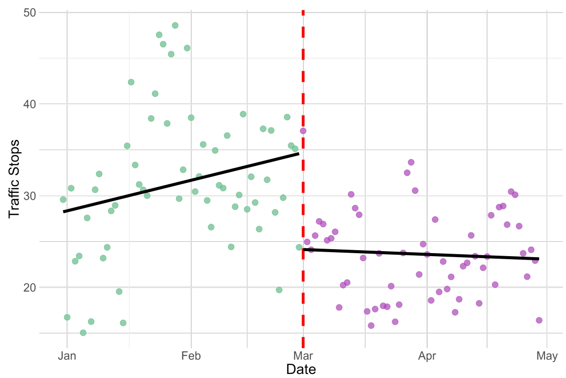
(c) Other Minor Traffic Violation



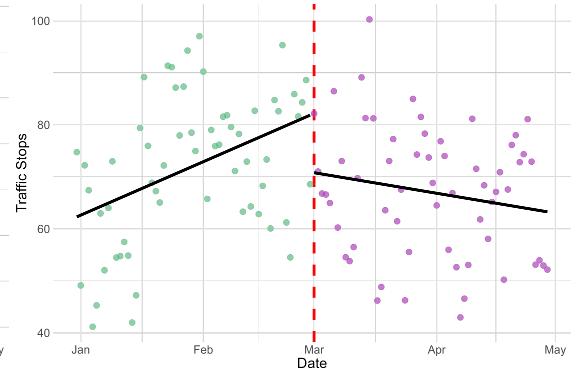
Notes: The residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022 are shown. For ease of interpretation, I added the residuals with the baseline mean of the outcome. The green color points represent my observation in the pre-treatment period. The purple color points represent my observation in the post-treatment period. The black line represents linear fit for both the pre- and post-treatment periods.

Figure 2.2: RDiT Estimate by Race

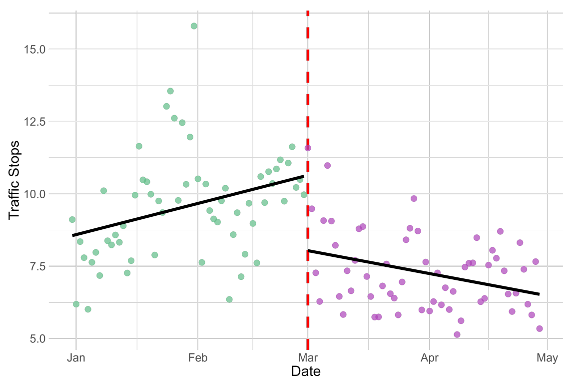
(a) Black: Minor Traffic



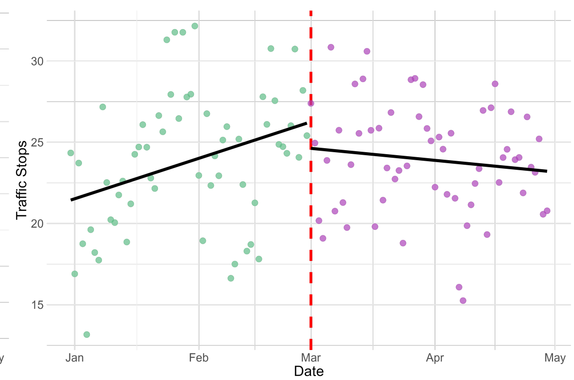
(b) Black: All



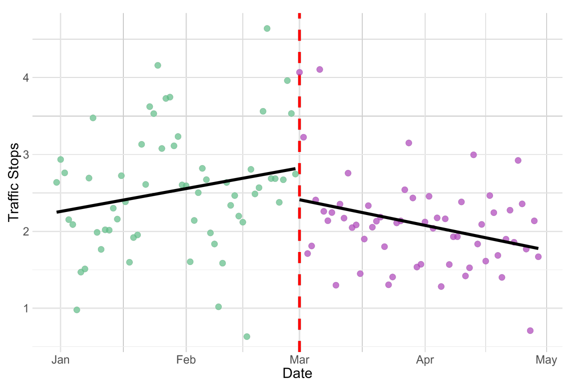
(c) Hispanic: Minor Traffic



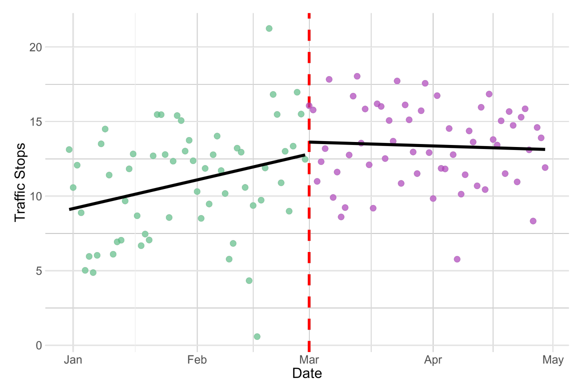
(d) Hispanic: All



(e) Non-Hispanic White: Minor Traffic



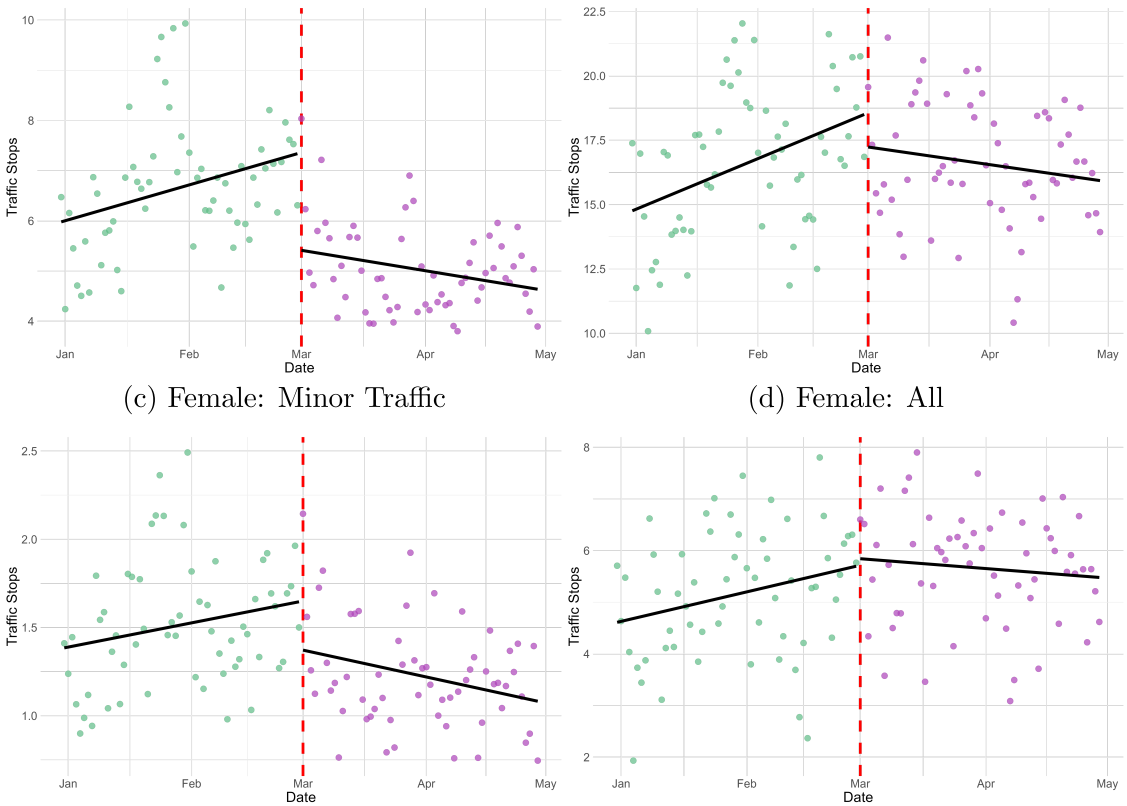
(f) Non-Hispanic White: Minor Traffic



Notes: The residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022 are shown. For ease of interpretation, I added the residuals with the baseline mean of the outcome. The green color points represent my observation in the pre-treatment period. The purple color points represent my observation in the post-treatment period. The black line represents linear fit for both the pre- and post-treatment periods.

Figure 2.3: RDiT Estimate by Gender

(a) Male: Minor Traffic (b) Male: All

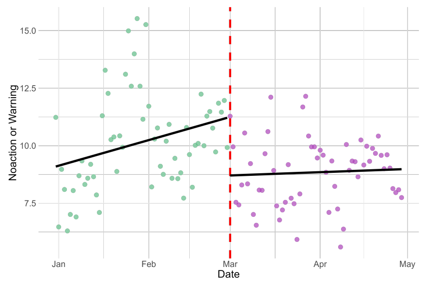


(c) Female: Minor Traffic (d) Female: All

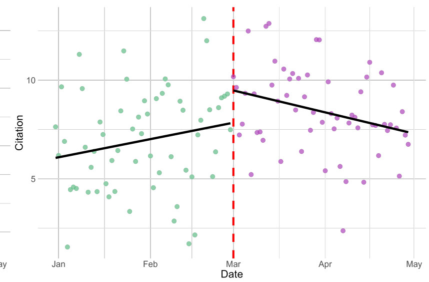
Notes: The residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022 are shown. For ease of interpretation, I added the residuals with the baseline mean of the outcome. The green color points represent my observation in the pre-treatment period. The purple color points represent my observation in the post-treatment period. The black line represents linear fit for both the pre- and post-treatment periods.

Figure 2.4: RDiT Estimate Stop Outcomes: All Stops

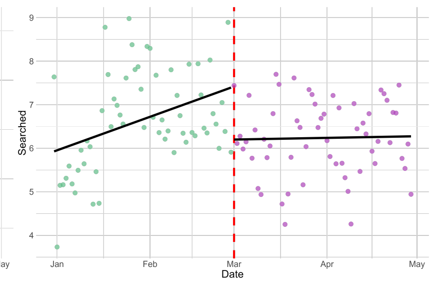
(a) Warnings



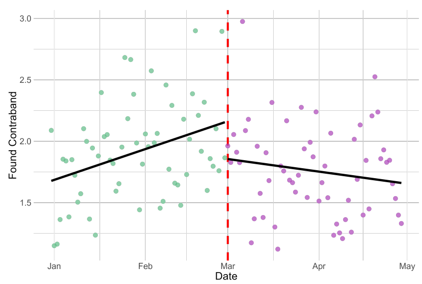
(b) Citations



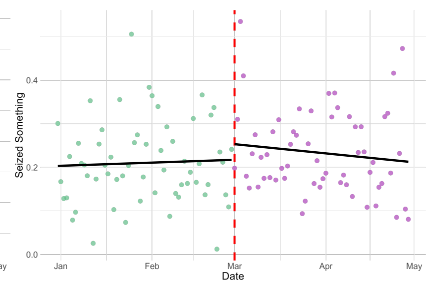
(c) Searches



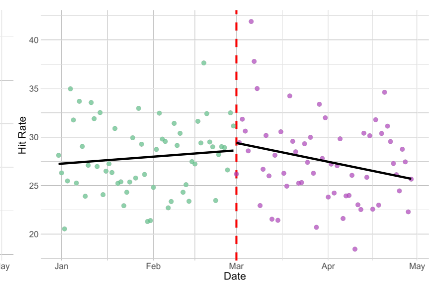
(d) Contraband Found



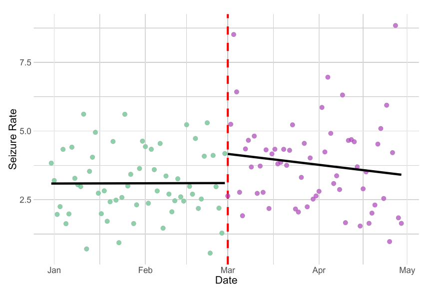
(e) Contraband Seized



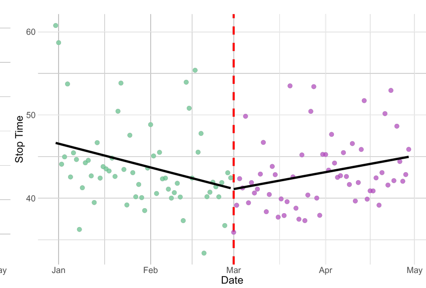
(f) Contraband Discovery Rates



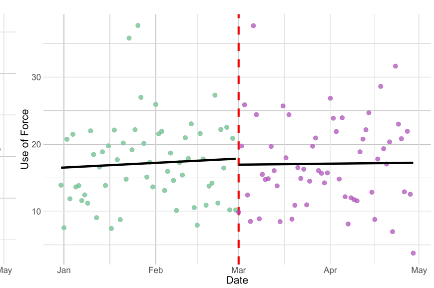
(g) Contraband Seizure Rate



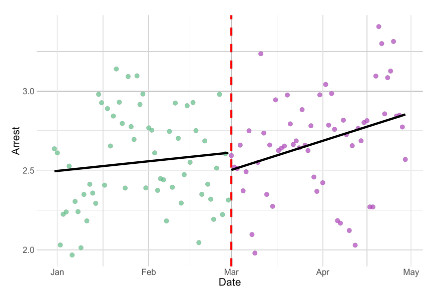
(h) Stop Time



(i) Use-of-Force



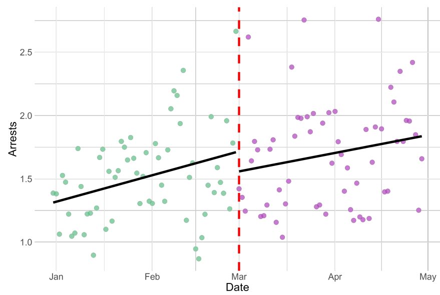
(j) Arrest



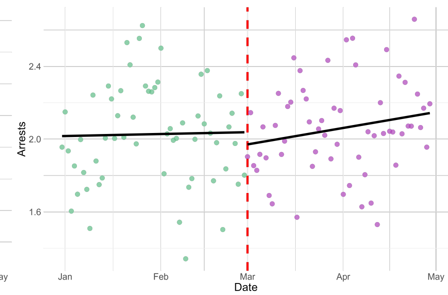
Notes: The residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022 are shown. For ease of interpretation, I added the residuals with the baseline mean of the outcome. The green color points represent my observation in the pre-treatment period. The purple color points represent my observation in the post-treatment period. The black line represents linear fit for both the pre- and post-treatment periods.

Figure 2.5: RDiT Estimate: Arrests, Crime, & Traffic Accidents

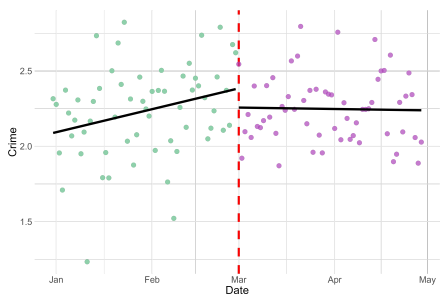
(a) Arrest: Misdemeanor



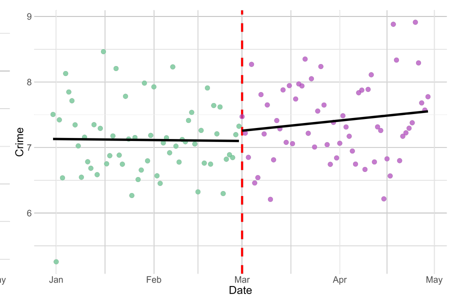
(b) Arrest: Felony



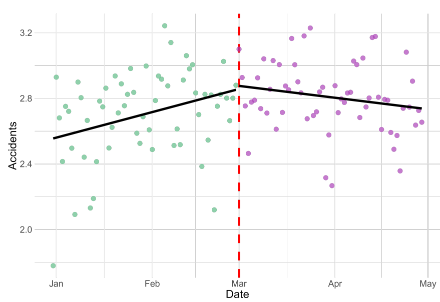
(c) Violent Crime



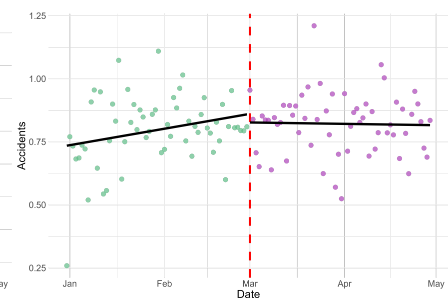
(d) Property Crime



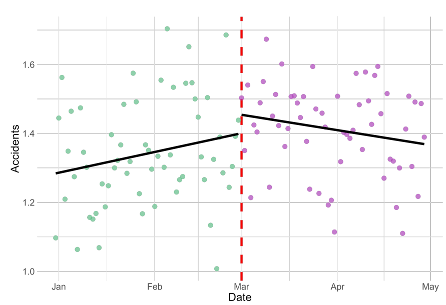
(e) Traffic Accidents



(f) Speeding Accidents



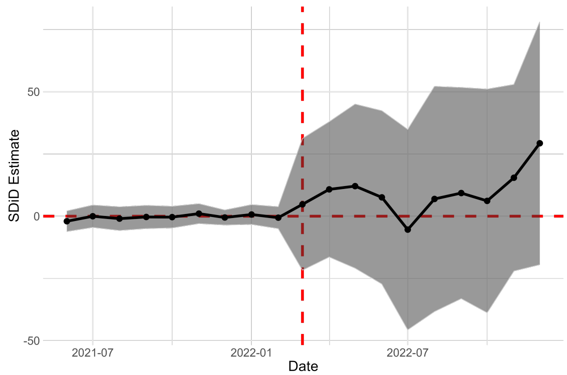
(g) Traffic Accidents from Other Vehicle Violations



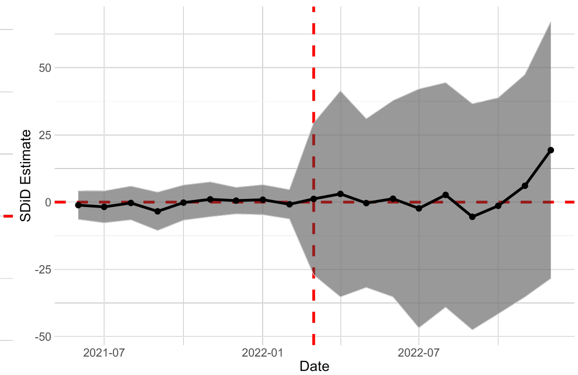
Notes: The residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022 are shown. For ease of interpretation, I added the residuals with the baseline mean of the outcome. The green color points represent my observation in the pre-treatment period. The purple color points represent my observation in the post-treatment period. The black line represents linear fit for both the pre- and post-treatment periods.

Figure 2.6: SDiD Estimate: Crime Comparing Los Angeles to the Rest of U.S.

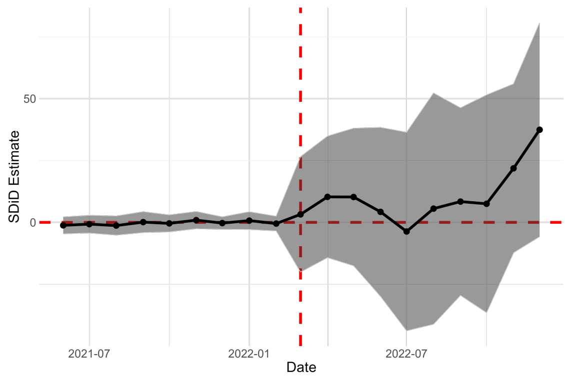
(a) All Part I: Agency-Level



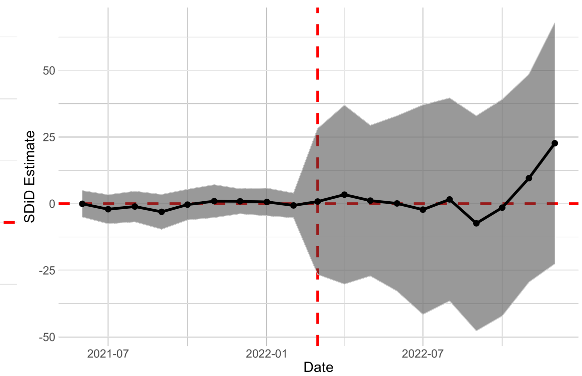
(b) All Part I: County-Level



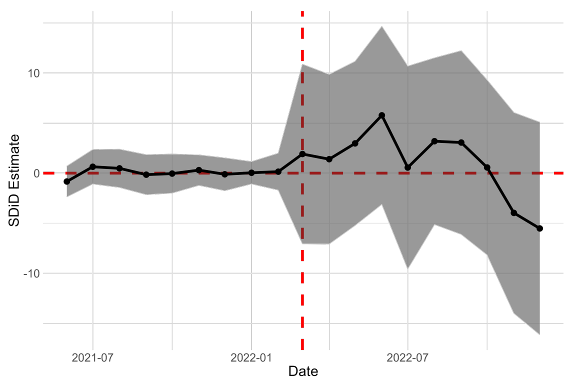
(c) Property: Agency-Level



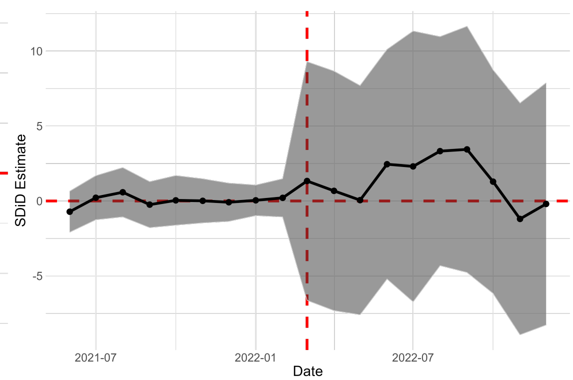
(d) Property: County-Level



(e) Violent: Agency-Level



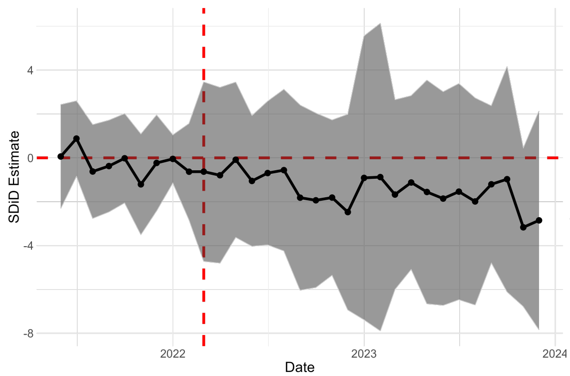
(f) Violent: County-Level



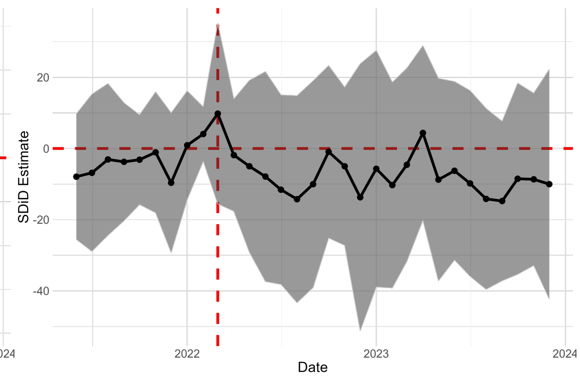
Notes: The synthetic difference-in-differences model is estimated. The left side of the panel uses agency-level data and compares LAPD (treated) to other large police agencies not in Los Angeles County. The right side of the panel uses county-level data and compares the county of Los Angeles (treated) to other large counties. The gray area represents 95 percent confidence intervals generated using placebo-based standard errors. The sample is restricted to June 2021 to December 2022.

Figure 2.7: SDiD Estimate: Accidents

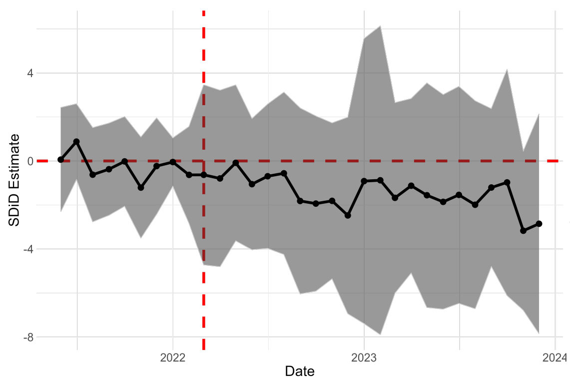
(a) All Accidents: Agency-Level



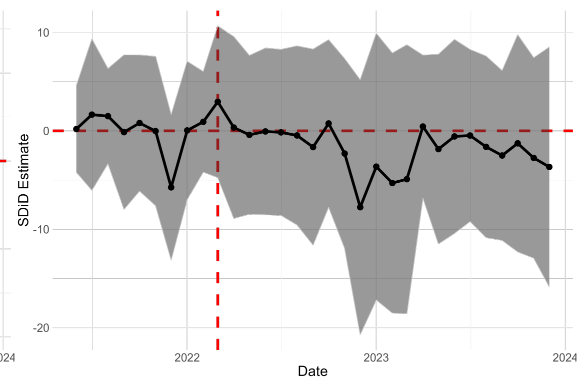
(b) All Accidents: County-Level



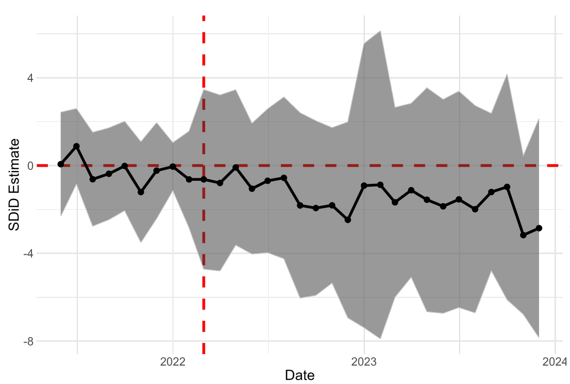
(c) Speeding Accidents: Agency-Level



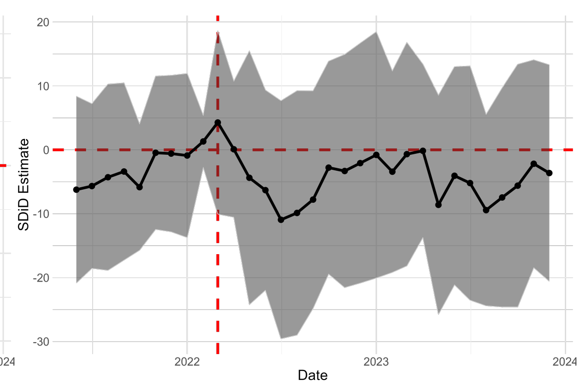
(d) Speeding Accidents: County-Level



(e) Accidents from Other Traffic Violation: Agency-Level



(f) Accidents from Other Traffic Violation: County-Level



Notes: The synthetic difference-in-differences model is estimated. The left side of the panel uses agency-level data and compares LAPD (treated) to other large police agencies not in Los Angeles County. The right side of the panel uses county-level data and compares the county of Los Angeles (treated) to other large counties. The gray area represents 95 percent confidence intervals generated using placebo-based standard errors. The sample is restricted to June 2021 to December 2023.

2.7.2 Tables

Table 2.1: RDiT: Number of Stops

	Minor Traffic (1)	All Stops (2)	Minor Moving (3)
	Panel I: RDiT		
Post	-1.827***	-0.386	1.508*
	(0.500)	(1.286)	(0.911)
	Panel II: Differences in Discontinuities		
Post*LAPD	-2.077***	-1.429	0.910
	(0.510)	(1.323)	(0.935)
Mean of DV	6.390	20.799	8.440

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis.

Table 2.2: RDiT: Number of People Stopped by Race

	Minor Stops			All Stops		
	Black (1)	Hispanic (2)	White (3)	Black (4)	Hispanic (5)	White (6)
	Panel I: RDiT					
Post	-10.569***	-2.606***	-0.414	-11.398**	-1.633	0.799
	(2.681)	(0.695)	(0.289)	(4.640)	(1.546)	(1.396)
	Panel II: Differences in Discontinuities					
Post*LAPD	-10.669***	-2.741***	-0.722**	-13.052***	-2.414	-0.474
	(2.722)	(0.708)	(0.319)	(4.736)	(1.599)	(1.451)
Mean of DV	28.041	8.492	2.464	75.763	26.030	13.246

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis.

Table 2.3: RDiT: Number of People Stopped by Gender

	Minor Stops		All Stops	
	Male (1)	Female (2)	Male (3)	Female (4)
	Panel I: RDiT			
Post	-1.949***	-0.278**	-1.328	0.127
	(0.464)	(0.129)	(1.042)	(0.461)
	Panel II: Differences in Discontinuities			
Post*LAPD	-2.069***	-0.377***	-1.919*	-0.277
	(0.471)	(0.135)	(1.064)	(0.479)
Mean of DV	5.934	1.394	18.155	5.777

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis.

Table 2.4: RDiT: Traffic Stop Outcomes

	Warning	Citation	Searched	Found Some-thing	Seized Some-thing	Discovery Rate	Seizure Rate	Avg. Stop Time	Use-of-Force	Arrest
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel I: RDiT									
Post	-2.543***	1.634*	-1.219***	-0.308**	0.036	0.762	1.060**	-0.019	-0.903	-0.110
	(0.717)	(0.904)	(0.350)	(0.147)	(0.035)	(1.878)	(0.431)	(1.781)	(1.801)	(0.105)
	Panel II: Differences in Discontinuities									
Post*LAPD	-2.889***	1.038	-1.297***	-0.319**	0.036	0.601	2.711	0.417	-1.420	-0.129
	(0.734)	(0.930)	(0.356)	(0.148)	(0.038)	(2.317)	(2.340)	(1.850)	(2.047)	(0.111)
Mean of DV	10.684	8.049	7.100	1.907	0.226	26.756	3.190	40.922	18.467	2.648

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis.

Table 2.5: RDiT: Arrest, Crime, & Accidents

	Arrests		Crime		Accidents		
	Misdemeanor	Felony	Property	Violent	All	Speeding	Other Violation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post	-0.159	-0.067	0.157	-0.128	0.019	-0.034	0.053
	(0.192)	(0.094)	(0.166)	(0.086)	(0.089)	(0.042)	(0.052)
Mean of DV	1.679	2.078	7.115	2.234	2.704	0.797	1.342

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis.

Table 2.6: SDiD: Reported Crime

	LAPD vs. Other Agencies			LA vs. Other Counties		
	All	Property	Violent	All	Property	Violent
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel I: California Only					
Post*LA	9.694	10.510	0.994	2.416	2.811	1.349
	(17.642)	(16.758)	(3.062)	(17.038)	(15.804)	(2.734)
Rank Based P-Value	{0.381}	{0.333}	{0.690}	{0.833}	{0.829}	{0.640}
Mean of DV	276.529	207.754	68.776	244.913	195.107	49.805
	Panel II: All					
Post*LA	9.370	8.925	0.271	10.129	11.441	0.981
	(17.671)	(15.974)	(4.547)	(12.765)	(12.195)	(5.807)
Rank Based P-Value	{0.477}	{0.464}	{0.877}	{0.455}	{0.438}	{0.676}
Mean of DV	276.529	207.754	68.776	244.913	195.107	49.805

* P-val $<$ 0.1; ** P-val $<$ 0.05; *** P-val $<$ 0.01

Notes: The synthetic difference-in-differences model is estimated using monthly data. Columns 1 to 3 use an agency-by-month panel and compare the LAPD (treated) to other large police agencies not in Los Angeles County. Columns 4 to 6 use a county-by-month panel and compare the county of Los Angeles (treated) to other large counties. The standard errors estimated using placebo-based tests are reported inside the parenthesis. Rank-based p-value where the p-value is computed using a relative ranking of the average treatment effect is shown inside a curly bracket. The sample is restricted to June 2021 to December 2022, and the counterfactual includes the rest of California (panel I) or the rest of the U.S. (panel II).

Table 2.7: SDiD: Reported Traffic Accidents

	LAPD vs. Other Agencies			LA vs. Other Counties		
	All	Speeding	Other Violation	All	Speeding	Other Violation
	(1)	(2)	(3)	(4)	(5)	(6)
Post*LA	-3.304	-1.436	-1.138	-7.162	-1.676	-4.289
	(3.332)	(1.549)	(2.237)	(9.354)	(3.207)	(6.518)
Rank Based P-Value	{0.281}	{0.276}	{0.280}	{0.314}	{0.490}	{0.308}
Mean of DV	28.483	7.215	12.943	89.045	27.586	43.774

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: The synthetic difference-in-differences model is estimated using monthly data. Columns 1 to 3 use an agency-by-month panel and compare the LAPD (treated) to other large police agencies not in Los Angeles County. Columns 4 to 6 use a county-by-month panel and compare the county of Los Angeles (treated) to other large counties. The standard errors estimated using placebo-based tests are reported inside the parenthesis. Rank-based p-value where the p-value is computed using a relative ranking of the average treatment effect is shown inside a curly bracket. The sample is restricted to June 2021 to December 2023, and the counterfactual includes the rest of California.

The Impact of Drug Decriminalization on Overdoses: Evidence from Measure 110

This chapter is co-authored with David Hall and Ben Hansen

Abstract: In this paper, we replicate and extend two recently published papers that find conflicting evidence on whether Oregon and Washington's drug decriminalization led to an increase in drug overdose. We first document that slight differences in outcomes and pre-treatment window create different counterfactual weights, which led to different conclusions regarding the first-year impacts. We next document that with an extended time window, the differences in the methodologies between the two papers do not matter, and we find robust evidence that drug overdose, particularly after the second year, significantly increased by 0.6 to 0.7 per 100,000. Finally, we conduct a few analyses to confirm that the arrival of the fentanyl crisis on the West Coast is not driving our results.

3.1 Introduction

The overdose epidemic has been one of the most devastating public health crises in recent history. The United States saw more than 100,000 overdose deaths in each year from 2021-2024. Unintentional overdose mortality is the leading cause of external mortality (Center for Disease Control and Prevention, National Center for Health Statistics 2022) surpassing transportation accidents ($\sim 47,000$) and firearm discharge-related ($\sim 46,000$).

Prior research highlights the need to study more than overdose mortality in the context of this epidemic. In particular, involvement in the criminal legal system generates substantial social and economic costs. The criminalization of drug use imposes burdens, including reduced employment opportunities for individuals with criminal records, lower lifetime earnings, and deterrence from seeking emergency care during overdose events (Agan & Starr 2018, Holzer et al. 2006, Mueller-Smith 2015, Byles et al. 2024). Decriminalization may mitigate these harms, but the net benefit depends on whether the gains from decriminalization outweigh the costs of continued criminalization. In this paper, we estimate the impact of decriminalizing drug possession on one such potential benefit: drug-related overdose mortality.

We study the impact of decriminalization on the fatal overdose rate in the context of Oregon's Measure 110 and Washington's Blake Fix. Enacted in February of 2021, Measure 110 is a harm reduction policy primarily intended to reduce racial inequities within the incarceration system via decriminalizing the possession of small amounts of drugs (Anthony Johnson et al. 2019). In particular, Measure 110 reduced the penalty for small amounts of drug possession from a Class A Misdemeanor to a Class E Violation carrying a fine of $100. Similarly, Washington also experimented with a similar drug decriminalization law beginning in March 2021.

Our main empirical specification uses the synthetic control method. Synthetic Control is preferred for its ability to create a counterfactual in case studies with just one treated unit, and also due to the transparency of the design (Abadie 2021). The weights used to create Synthetic Oregon are a direct result of implementing the model. Similar studies use the Synthetic Control model to estimate the causal impact of Measure 110 on overdose-related mortality; however, the findings are mixed. Joshi et al. (2023) and Spencer (2023) provide early estimates of the causal impact, with one suggesting a statistically insignificant effect and one suggesting a significant increase. Zoorob et al. (2024) use a generalized Synthetic Control with controls for fentanyl to estimate the causal impact of Measure 110 on longer-term overdose-related mortality but find little long-run effects. The goal of this paper is to replicate these findings, explain methodological differences among the designs, extend prior estimates, and produce our preferred specification.

We begin by reconciling the differences between point estimates from Joshi et al. (2023)

and Spencer (2023). Although both studies use the synthetic control method matching on state-year outcomes only (as suggested by Ferman et al. (2020)), we find that the estimates diverge for two main reasons. First, the authors use a different definition of overdose mortality; Joshi et al. (2023) uses all overdose mortality (both intentional and unintentional), whereas Spencer (2023) uses only unintentional overdose deaths. Second, Spencer (2023) defines January 2018 to December 2021 as his pretreatment window, whereas Joshi et al. (2023) include another month (January 2021) for their treatment window.

We extend the methodologies of Spencer (2023) and Joshi et al. (2023) through the end of 2023 and provide updated estimates from our preferred specification. Across all model variations, we find that the implementation of Measure 110 led to an increase of between 0.79 to 0.85 per 100,000 overdose-related deaths per month. This corresponds to approximately 1125 to 1213 additional deaths—an increase of 38 to 46% relative to the baseline mean. Measure 110 was a multifaceted policy that combined decriminalization of drug possession with investments in treatment access and support services for people who use drugs. We present evidence suggesting that the primary driver of the observed mortality increase was a rise in fentanyl supply and trafficking caused by decreased justice system interaction.

We provide several robustness tests to ensure that our estimated effect is not biased from either unobservable factors or design choices in synthetic control estimation. We show results from various extensions of synthetic control, including the augmented synthetic control (Ben-Michael et al. 2021), generalized synthetic control (Xu 2017), and synthetic difference-in-differences (Arkhangelsky et al. 2021). Results from these alternative specifications are in line with our main results. We also show results from an exercise where we switch the weights of Spencer (2023) and Joshi et al. (2023) while using their respective definitions.

While this results in poor pre-treatment matches, the conclusion of an increase in overdose-related mortality remains. To test whether particular donor units are driving our results, we perform a remove-one robustness check.

A particular concern is whether our results reflect the effects of COVID-19 or are driven by broader West Coast trends rather than Measure 110 specifically. The COVID-related concern stems from potential structural breaks in the donor pool – states that served as good counterfactuals for Oregon before 2020 may no longer be valid comparisons post-pandemic. To address this, we implement a sensitivity check by restricting the synthetic control match to data through February 2020, allowing us to isolate pre-pandemic trends and test whether the estimated effects are confounded by state-specific COVID-19 impacts on substance use. To assess the possibility of a regional effect, we construct placebo synthetic control estimates for California, Nevada, and Idaho. Finally, we evaluate the robustness of our findings across various definitions of overdose mortality – including opioid-related, non-opioid, and non-synthetic deaths – to assess whether Measure 110 disproportionately influenced synthetic opioid outcomes or affected drug use more broadly.

Finally, we replicate Zoorob et al. (2024). The methodology of Zoorob et al. (2024) differs from Spencer (2023) and Joshi et al. (2023) in several ways. First, they utilize bi-annual data and expand their pre-treatment window to include earlier waves of the opioid crisis. Secondly, they consider fentanyl market saturation as a time-varying confounder for the overdose crisis in Oregon. We begin by outlining the assumptions underlying the matrix completion approach, particularly those related to the strict exogeneity assumption, and the rationale for including fentanyl supply as a control variable. We then replicate the change-point analysis of fentanyl seizures and extend their approach by incorporating fentanyl-related overdose mortality and legal import data to assess whether synthetic saturation confounds the estimated effect of Measure 110 in Oregon.

This paper contributes to two areas of research. First, it adds to the literature on opioid abatement strategies. Traditional approaches often involve state-level policies such as doctor shopping restrictions and prescription drug monitoring programs (Maclean et al. 2020). In contrast, decriminalization is part of a broader harm reduction framework, alongside interventions like syringe exchange programs (SEPs) and expanded access to naloxone. For example, Packham (2022) finds that while SEPs effectively reduce the spread of HIV and HCV, they may unintentionally increase overdose mortality. Compared to these strategies, decriminalization has received relatively little causal analysis, particularly outside the context of Oregon's Measure 110, due to limited quasi-experimental opportunities. The primary evidence to date comes from Portugal's 2001 decriminalization, where Félix et al. (2017) estimate a significant reduction in drug-related deaths following the policy change. Beyond its policy relevance, empirical analysis of decriminalization also tests classical economic theory. Our results speak to the Economic Theory of Illegal Goods proposed by Becker et al. (2004). Becker et al. (2004) argue that public expenditure on apprehension of illegal good suppliers — such as narcotics traffickers — depends on the social value of consuming that illegal good. Under plausible estimates of demand elasticity, they show that criminal penalties for drug use may be welfare-reducing. We contribute to this theoretical framework by examining one critical welfare-relevant outcome of decriminalization: overdose mortality.

We also contribute to ongoing efforts to verify published results in science more generally and economics more specifically (Hillary & Medaglia 2020, Camerer et al. 2018, Chang & Li 2015). We replicate prior work estimating the overdose mortality effects of Measure 110, synthesize methodological differences, and provide our estimates transparently. In using the synthetic control method, we ensure reproducibility by making the donor weights used to construct our counterfactual explicit and matching on outcome only, following Ferman et al. (2020). We document this approach in both the paper and our replication package. In doing so, we underscore how synthetic control model can help align empirical research with the broader goals of transparency and replicability in applied economics.

The remainder of the paper is structured as follows. Section 3.2 provides context on Measure 110 as well as Washington's "Blake Fix" decision. Section 3.3 reviews prior literature estimating the impact of Measure 110 on overdose mortality, as well as other evaluations of the policy's broader effects. Section 3.4 shows the results of our replication and extensions of Spencer (2023), Joshi et al. (2023), and Zoorob et al. (2024). We conclude that section with robustness tests to ensure that potential threats to causal identification — including COVID-19 and the timing of fentanyl saturation – are not driving our findings. Section 3.5 concludes.

3.2 Background

3.2.1 Oregon's Measure 110

Oregon voters passed Ballot Measure 110 on November 3, 2020.[1] The bill passed by 8.46 percentage points and was enacted on February 1, 2021. The primary component of the policy was the decriminalization of all scheduled narcotics at the user-level of possession. The bill adjusts drug possession violations depending on the categorized amount of drugs

[1]Measure 110 is also known as the Drug Addiction and Recovery Act.

an individual is found in possession of. Individuals found with ”user amounts” of drugs are subject to a Class E Violation as opposed to a Class A Misdemeanor (Anthony Johnson et al. 2019).[2] This particular Class E violation entails a $100 fine which can be waived if the individual calls a helpline that provides information on treatment services. Individuals possessing more than the ”user” threshold are subject to a Class A misdemeanor rather than a Class B or C felony [3]. Table 3.1 lists how the classification changed for each drug type and threshold amount.

Original provisions of the bill did not include fentanyl nor its analogues.[4] Synthetic opioids, such as fentanyl, are the primary driver of overdose deaths nationwide, and have been so since the start of 2016. On July 27, 2023, the state adjusted the violations of fentanyl possession under House Bill 2645 (HB2645). Before the passage of this bill, an individual found in possession of less than one gram of fentanyl would be charged with a Class A Misdemeanor. Fentanyl is lethal at a dose of as little as 2mg. An individual possessing 1g of fentanyl could have 500 user doses of fentanyl and be charged for a misdemeanor before HB2645.

The bill's goal was to shift legal attitudes surrounding addiction (Oregon Health Justice Recovery Alliance 2021). Supporters of the bill, including the Drug Policy Alliance and Mark Zuckerburg Foundation advocated that substance use disorders (SUDs) should be treated as a disease rather than as a crime, thus an individual with a SUD should receive treatment rather than incarceration. This would alleviate concerns about capacity constraints in jails. Black

[2]A class A Misdemeanor is subject to a maximum of up to 364 days of imprisonment and up to $6,250 in fines.https://oregon.public.law/statutes/ors_161.635

[3]Consequences of a Class B (C) felony are up to 10 (5) years imprisonment and a $250,000 ($125,000) fine.

[4]Moving forward, we use “fentanyl” to encompass all fentanyl analogues as well.

people who use drugs (PWUD) are also more likely to be incarcerated for drug possession, and Measure 110 would reduce racial inequities within the Oregon incarceration system.

Critics of the bill included law enforcement and legislative opposition (Vote No on Measure 110 2020). Oregon was already progressive regarding drug decriminalization; those found in possession of drugs were already given opportunities to seek treatment through the legal system via drug courts, for instance. Another concern was that the increase in treatment funding was overstated. Treatment providers were already at capacity and providing more beds required immediate funding that was not provisioned in the bill.[5]

Due to the limited timeframe in which the full decriminalization policy was active, little published research has been conducted.[6] Davis et al. (2023) find a statistically significant decrease in drug possession arrests following the enactment of Measure 110. They also find a significant increase in "displaced" arrests, providing evidence of a substitution pattern among law enforcement agents. Smiley-McDonald et al. (2023) find evidence of asymmetric information among law enforcement agents and the public regarding the entirety of Measure 110 provisions. Law enforcement officers (LEOs) were unaware of treatment funding provided by Measure 110. LEOs were also hesitant to interact with PWUD after M110's enactment due to perceptions that the policy lacks teeth (Henderson et al. 2023). Decreased interaction with law enforcement agents may be the reason for limited use of the statewide referral hotline (Russoniello et al. 2023). Those found in possession of user amounts of drugs were given a $100 fine or the option to call the statewide telephone hotline to have the fine removed and be referred to treatment services (Russoniello et al. 2023). By the end of 2023, only 577

[5] The bill does include provisions for funding treatment services, but the process of funding health providers would not begin until more than a year post-Measure 110 enactment.

[6] We acknowledge there is active and ongoing unpublished as-of-yet research, particularly that of researchers who attended Comagine Health's Measure 110 Research Symposium in January of 2024.

people had called into the hotline (Gaitán 2023).

After just three years since the enactment of the bill, Governor Tina Kotek signed House Bill 4002, repealing much of the progress made in the way of decriminalization (Representative Jason Kropf & Senator Kate Lieber 2024). Advocates of recriminalization cite increasing overdose rates in the state of Oregon; the state itself declared an emergency in June of 2024 to address the fentanyl crisis in Portland (A. B. C. News 2024). On September 1, 2024, the repeal became effective, making the possession of drugs back as a misdeamenor offense.

3.2.2 Washington's "Blake Fix"

On February 5, 2021 - just four days after Oregon's decriminalization went into effect - the Washington Supreme Court found the state's possession of a controlled substance statute unconstitutional.[7] In the case of *State v. Blake*, the court ruled that "convictions for Possession of Controlled Substance aka Violation of Uniform Controlled Substances Act (VUCSA), were not lawful," and made it possible for those charged to vacate the conviction (State of Washington 2021). This ruling decriminalized the possession of scheduled substances until WA lawmakers enacted SB5476 ("Blake Fix"). On May 13, 2021 — with an effective date of 7/25/21 — Washington recriminalized possession but lowered sentencing guidelines for simple possession (Manka 2021).[8]

[7] Washington's controlled substance statute did not require prosecution to prove that the defendant was aware they possessed such a substance.

[8] Simple possession is similar to personal use as defined by the Oregon law. SB5476 was effective until mid-2023. In May of 2023, WA Governor Jay Inslee signed SB5536, a bill that increased sentencing on simple possession.

3.3 Literature Review

3.3.1 Two Contradicting Early Evidence of Drug Decriminalization on Drug Overdose

Two papers published in 2023 investigate the causal effect of drug decriminalization on overdoses using a synthetic control estimator (Joshi et al. 2023, Spencer 2023). Both teams estimate the average treatment effect using monthly data on overdoses directly from the CDC multiple cause of death files. Despite nearly identical settings for the analysis, Spencer (2023) finds a statistically significant increase in the overdose rate of approximately 0.39 in the state of Oregon, whereas Joshi et al. (2023) find no statistically significant change in the rate of drug deaths (ATE of 0.268 with a p-value of 0.26). For the state of Washington, Spencer (2023) also finds a statistically significant increase in drug mortality, and Joshi et al. (2023) find an increase, which is only statistically significant at the 10 percent level.

A synthetic control is a linear combination of observations from non-treated (donor) units with weights derived by minimizing the pre-treatment prediction error. Therefore, the synthetic control model is sensitive to the weight and inclusion of donor units. Table 3.2 provides the resulting weights from each of the authors' analyses. Spencer (2023)'s estimate weighs more heavily Kansas and Maryland, while Joshi et al. (2023) weigh Mississippi and Alaska more heavily. Both synthetic controls rely heavily on Montana as well (0.178 for Spencer (2023), 0.272 for Joshi et al. (2023)). We document that there are several differences in methodologies that drive the difference in these weights and the estimate. Table 3.3 summarizes the key differences between the two papers, and we discuss these differences in the subsequent paragraphs.

Definitions

Users of the CDC WONDER database can filter underlying causes of death using ICD (International Cause of Death) codes.[9] Spencer (2023) filters by unintentional overdoses (X40-44) whereas Joshi et al. (2023) filter by all overdoses (X60-64, X85, Y10-14, multiple cause of death T36-50).

As categorized by the World Health Organization, unintentional overdoses are external causes of morbidity and mortality. Specifically (X40-X49) is defined as "accidental overdose of drug, wrong drug given or taken in error, and drug taken inadvertently"; "accidents in the use of drugs, medicaments and biological substances in medical and surgical procedures;" and "(self-inflicted) poisoning, when not specified whether accidental or with intent to harm." The subset Spencer (2023) uses, X40-X44, is categorized under "Accidental posioning by and exposure to noxious substances." X60-64 includes intentional self-harm from poisoning by drugs. X85 is classified as homicidal poisoning. Y10-14 covers all drug poisonings with undetermined intent. Finally, T36-50 is classified as "Poisoning by drugs, medicaments and biological substances" and would be included in a multiple cause of death download of the series.[10]

The average difference between total overdoses (nationally) is 692 per month (nationwide). The average using Joshi et al. (2023)'s definition is 7710, and the average using Spencer (2023)'s definition is 7019. The standard deviation is 1539 and 1543, respectively.

[9] At this time, the tenth revision of these codes (ICD-10) is being used. Codes can be searched using the ICD-10-CM Browser Tool https://www.cdc.gov/nchs/icd/icd10cm_browsertool.htm

[10] Filtering by "multiple cause of deaths" allows the user to subset even overdoses into a specific drug used.

In the states of Oregon and Washington, the averages using Joshi et al. (2023)'s definition are 89 and 176, respectively. For Spencer (2023)'s definition, the monthly averages are 79 in Oregon and 162 in Washington. Figure 3.1 plots the national time trend for these different definitions. We note that the majority of drug deaths are due to accidental overdoses (panel a), and these two outcomes follow a similar trend over time (panel b).[11]

There are reasonable arguments for using different outcome measures in each study. Drug decriminalization may affect mortality through a demand-side channel—that is, by increasing the number of users and/or the frequency or dosage of use—rather than through increases in suicide or homicide. As a result, researchers may have more statistical power when focusing specifically on accidental overdoses. However, analyzing all drug-related mortality may help mitigate potential classification or measurement errors, since sometimes the data can misclassify unintentional overdoses as suicides or homicides.

Timing

The synthetic control method requires selecting a pre-treatment period over which to match outcomes to determine the optimal synthetic weights As Abadie et al. (2010) notes: "In any research design that exploits time variation in the outcome variable to estimate the effect of an intervention, synthetic control estimators may be biased if forward-looking economic agents react in advance of the policy intervention under investigation." Thus, researchers need to use economic theory to justify their decisions on which periods to minimize the sum of squared prediction errors and determine the correct weights.

[11] In Appendix Figure 3.2, we present the same figure, but focusing on Oregon and Washington.

Spencer (2023) matches on 36 monthly pre-treatment periods between January 2018 and January 2021.[12] While Spencer (2023)'s pre-treatment sample window ends one month before the implementation of Measure 110, he argues that he chose this decision because he wanted consistency across outcomes. Some of his outcomes are at the annual level, meaning the synthetic control can only match until December and define treatment starting in January. Joshi et al. (2023) also begin their sample in January 2018 but utilize all 37 pre-treatment periods, including January 2021, citing they "followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for observational studies."

Beyond ensuring consistency across outcomes, the correct specification between the two models depends on whether anticipation effects are relevant. For instance, if there was anticipation prior to implementation (e.g., in January 2021), where overdoses had already begun to change, researchers should exclude the anticipatory window for matching. While we do not statistically verify whether and when (if any) anticipation happened, in our preferred specification, we will experiment with ending our matching period in December 2021 and January 2022 for robustness.

Finally, the authors differ in the number of observations in the post-treatment period. Spencer runs their analysis through the end of 2021, while Joshi et al. (2023) run theirs through the end of March, 2022.

[12] Spencer (2023) makes a note of the policy being implemented on February 1, 2021, in the paper, but chooses January as the treatment unit in coding and chart output. In his replication package, this is identified by unit 37 (t=37) being selected as the treatment period, which in a monthly dataset beginning (t=1) in January of 2018, corresponds to January of 2021.

Data Set

For reasons of security and privacy, the CDC suppresses the number of deaths for an observation if the number of deaths is below ten. Spencer (2023) uses this restricted version of the data and imputes integers from a random uniform distribution on the interval [0,9] to fill suppressed values. Joshi et al. (2023) had access to the unrestricted version of the data.

In Appendix Figure 3.3, we re-estimate our synthetic control 1,000 times using different imputed values and plot the distribution of possible point estimates. The findings from this exercise confirm that the point estimates reported by Spencer (2023) are reasonable and not driven by artifacts of imputed data.

In our replication and extension of each work, we utilize the restricted data.

3.3.2 Longer-run Effect of Drug Decriminalization

Zoorob et al. (2024) extend the time period analyzed in Spencer (2023) and Joshi et al. (2023), focusing on the second-year update of Measure 110, though their methodologies differ slightly. They aggregate their panel data into half-year periods rather than monthly, and they use a longer pre-treatment window starting in January 2008. In addition, they adopt a matrix completion approach to synthetic control. Initially, they find a statistically significant biannual increase in drug mortality of 1.83 per 100,000 (equivalent to a monthly increase of 0.3 per 100,000). However, in their preferred specification—which controls for fentanyl supply as measured by the percentage of all drug seizures involving fentanyl—they find no association between Measure 110 and drug mortality.

We raise two methodological concerns with the approach taken by Zoorob et al. (2024).

In their preferred specification, the authors explicitly control for changes in the supply of illicitly manufactured fentanyl. If the saturation of fentanyl in the drug market coincided with the passage of Measure 110, synthetic control estimates could be biased upward—unless the donor pool units used to construct the synthetic Oregon also experienced similar supply-side shocks. To address this, the authors include a direct control for fentanyl availability, using the ratio of fentanyl-related seizures to the total number of drug seizures in Oregon.

There are two issues with this approach. One relates to the assumption underlying the inclusion of this control; the other pertains to the specific measurement strategy used. First, the authors implicitly assume that Measure 110 had no effect on the fentanyl supply chain itself. We find this assumption limiting. Consider a rational agent engaged in illicit fentanyl trafficking, as modeled in Becker's (1968) economic theory of crime (Becker 1968). If such agents weigh the costs of apprehension and conviction, decriminalization may reduce their perceived risk and lead to increased offending. In other words, if drug-related arrests fell post-Measure 110—as they did—then the assumption that fentanyl supply remained unaffected by the policy change is hard to justify. This raises concerns about treating the supply of fentanyl as exogenous in the model.

Second, the specific control used—the ratio of fentanyl seizures to all drug seizures—may be a poor proxy for fentanyl supply. Aside from documented concerns about the accuracy of NFLIS data compared to actual illicit supply (Pitts et al. 2023), this ratio can vary either because of changes in fentanyl prevalence or because of changes in the seizure patterns of other drugs.[13] If this ratio changes due to the changes in the seizure patterns of other drugs,

[13]In Appendix Table 3.1, we confirm that the total number of drugs seized and as a result the ratio of fentanyl seizures to all drug seizures increased, even though the number of fentanyl seizures did not significantly change.

which is induced by the treatment, then this ratio is considered a bad control and bias the treatment effect (Caetano et al. 2024). If the goal is to capture a structural shift in fentanyl availability, then more direct measures—less affected by the broader seizure landscape—would be preferable.

A separate empirical concern involves the decision to use biannual data and to extend the pre-treatment window back to 2008. The authors aim to address potential temporal confounding between Measure 110 and the transition from the heroin wave to the fentanyl wave of the opioid epidemic. However, by extending the pre-period to 2008, they allow the synthetic control to span across multiple regime shifts—most notably, the period when overdose deaths shifted from being driven by prescription opioids to heroin following the reformulation of OxyContin (Evans et al. 2018, Ciccarone 2019). Prior research suggests that these transitions occurred idiosyncratically across states (Ciccarone 2019, Shover et al. 2020, West et al. 2021), raising concerns about the plausibility of achieving valid synthetic matches—not just for the fentanyl era but for each of the preceding waves as well.

3.4 Results

3.4.1 Replication & Extension of Spencer (2023) and Joshi et al. (2023)

We begin our estimation by replicating both papers using a traditional synthetic control estimate. More specifically, for each donor state (j $\in \{2,\ldots, J\}$), we find a donor weight w_j which minimizes the pre-treatment (t $\in \{1, \ldots, T^{pre}\}$) differences in the outcome Y_{jt} between the treatment state (j=1) and synthetic control group which is a weighted combination of all donor states. Mathematically, we find w_j that solves the following optimization problem.

$$\min_{w_s} \sum_{t=1}^{T^{pre}} (Y_{1t} - \sum_{j=2}^{J} w_j Y_{jt})^2$$
$$\text{s.t.} \sum_{j=2}^{J} w_j = 1$$
$$w_j \geq 0 \quad \forall j$$

Once we solve for w_j, we use these donor weights to construct a synthetic counterfactual during the post-treatment period, which we assume to be the outcome of interest in the treatment states had neither Measure 110 nor The Blake Fix been enacted. The average post-treatment difference between the actual treatment state and the synthetic counterfactual is the causal effect of the policy.

We conduct our inference using permutation-based p-values. For each donor state $j \in \{2, ..., J\}$, we assign treatment and re-estimate the synthetic control estimate. Then, we calculate the ratio of the pre- and post-mean squared prediction errors (MSPE), which we define as the sum of the squared differences between the actual treatment state and the synthetic counterfactual. A high MSPE implies a better pre-treatment fit as compared to the post-treatment fit (i.e., large treatment effects). In contrast, a low MSPE can be interpreted in two ways: either as a poor pre-treatment fit that carries over into the post-treatment fit, or as a tight post-treatment fit (i.e., null treatment effects). We rank the ratio of MSPE and use the relative ranking (in percentage terms) for our p-value for inference.[14]

[14]For instance, if Oregon's MSPE ratio was ranked 1 out of 50, then the MSPE will be 0.02 (1/50).

In column of panel I, and column 4 of panel II of Table 3.4, we successfully replicate both Joshi et al. (2023) and Spencer (2023) results for Measure 110 in Oregon. However, we note that our point estimates are slightly different than those reported by the papers. One explanation for this difference is the difference in the Stata command that we utilize.[15]

In the remaining columns of Table 3.4, we experiment with changing the specification to be more similar to each other's specification, including changing the definitions of drug overdose (panels I and II), changing the treatment start month (columns 1 and 3 vs. 2 and 4), and sample window (columns 1 and 2 vs. 3 and 4). We document a few findings. First, ceteris paribus, our estimated effects are smaller when we focus on all drug-related mortality instead of unintentional drug overdoses. Second, our point estimate and precision are sensitive to whether we begin our treatment in January or February. For instance, when we focus on unintentional drug overdoses and end our sample in March 2022 (panel A columns 3 and 4), our estimated coefficient decreases by 0.057 percentage points and becomes no longer statistically significant. Third, the results are robust to the inclusion of three more months of post-treatment data.

We repeat the same exercise in Table 3.5; however, this time we focus on Washington. Similar to Oregon, we find that the differences in the two studies' conclusions stem from differences in the definitions of drug overdoses and when the treatment date began (and when the synthetic control was matched until).

In panel I of Appendix Table 3.2, we re-estimate Spencer (2023) and Joshi et al. (2023)'s preferred model and add 90 percent prediction intervals Cattaneo et al. (2021). In panel II, we also calculate the 90 percent confidence intervals using t-tests following Chernozhukov

[15]We use Synth package, whereas Spencer (2023) uses Synth2 and Joshi et al. (2023) use synth_runner.

et al. (2024) and using K=3 cross-fold validation. We note that our point estimates and statistical significance, reported in Appendix Table 3.2, are slightly different from those of our replication results. This discrepancy may be due to slight differences in methodologies and weights assigned to each donor state. Nevertheless, we note that the 90 percent confidence and prediction interval overlaps between Spencer (2023)'s preferred specification and Joshi et al. (2023)'s main model. The results from this exercise confirm that the estimated treatment effects in the two papers may not be statistically different.

In Figure 3.2, we extend our synthetic control estimates for Oregon to include up to November 2023. For simplicity, we report synthetic control estimates using Spencer (2023) and Joshi et al. (2023)'s specifications. We note that the gap between Oregon and synthetic Oregon becomes larger in late 2022, suggesting a potential dynamic effect that was not explored in either study. In Table 3.6, we report the average treatment effects and the p-values. Regardless of which specification we use, we find a statistically significant increase of 0.79 to 0.85 per 100,000 in Oregon, driven by increases in late 2022 and 2023. Relative to the synthetic counterfactual, these point estimates suggest a roughly 37 to 47 percent increase relative to the baseline (synthetic counterfactual) mean. This finding from Table 3.6 suggests that the difference between the two papers may also be due to imprecision problems caused by an insufficient post-treatment window.

In Figure 3.3, we expand our synthetic control estimate for Washington. Consistent with Oregon, we continue to find a second and third-year effect. However, this finding is surprising given that Washington only had decriminalization for five months. In Appendix Figure 3.4, we experiment with estimating a synthetic control model using crime and arrest data from the National Incident-Based Reporting System (NIBRS). Our results indicate that in Washington, even after the re-criminalization of drug possession, the arrest rate for drug overdose did not revert to the pre-treatment period, which is similar to that of Oregon, suggesting that the change in police enforcement may explain our results.

3.4.2 Robustness

We experiment with several robustness tests to ensure that our estimated effect is not biased from other unobservables or choices of our synthetic control model (Ferman et al. 2020).

First, we experiment with other extensions of the synthetic control model. In columns (1) to (6) of Appendix Table 3.3, we find that our estimates continue to show a significant increase in drug overdose rate when we use augmented synthetic control (Ben-Michael et al. 2021), synthetic difference-in-differences (Arkhangelsky et al. 2021), or generalized synthetic control methods (Xu 2017). In columns (7) to (8), we find that our point estimates are qualitatively similar when we estimate a two-way fixed effects difference-in-differences estimator. However, our estimates are more imprecise and have larger p-values.

In the next set of estimations shown in Appendix Figure 3.5, we experiment with keeping the same weight variables for both outcomes. While we have poor pre-treatment matches when we force the same weights, we continue to find an uptick relative to synthetic control estimation.[16] Moreover, in Appendix Figure 3.6, we experiment with excluding one donor state at a time. We find that no particular control states are driving our increase in drug

[16] In Appendix Table 3.4, we also experiment with estimating difference-in-differences using the same synthetic control weights, which allows the outcome level to be different across the treated unit and the counterfactual. We continue to find a large and statistically significant increase of 0.77 to 1.09 per 100,000 increase in drug overdose rate in Oregon, and 0.76 to 0.83 per 100,000 increase in drug overdose rate in Washington.

overdose rate. Our results from this exercise suggest that the drug overdose rate increased by at least 0.77 per 100,000 in Oregon and 0.71 per 100,000 in Washington. Finally, in Appendix Table 3.5, we conduct different inference procedures (Chernozhukov et al. 2024, Arkhangelsky et al. 2021). With the exception of Washington using placebo-based standard errors (column 4 panel I), we continue to find a statistically significant increase in drug mortality. This finding confirms that our findings are not due to false positives that can stem from a specific inference procedure.

In Appendix Table 3.6, we experiment with matching our Synthetic Control only up to February 2020, allowing for our donor pool to be based on the pre-pandemic trend of drug-related fatalities. The main idea of this exercise is to ensure that our choice of the donor pool and, consequently, our results are driven by state-specific dynamic effects of COVID-19 on substance use. Our findings remain qualitatively similar, suggesting that the impact of COVID-19 on biasing our estimates is little.

In Appendix Table 3.7, we experiment with estimating a placebo synthetic control for California, Nevada, and Idaho. One potential concern with our estimate is the fact that the timing of Measure 110 could be correlated with the arrival of fentanyl from the East Coast to the West Coast. We find little evidence that drug overdose rates increased in these nearby states, suggesting that the coincidental arrival of fentanyl in Oregon is not driving our estimates. In addition, along the same lines, we experiment with three alternative tests to ensure fentanyl is not driving our results. First, we estimate a two-way fixed effects model using only these three states as our counterfactual. While inference is a bit tricky due to the small cluster problem, our estimated treatment effects suggest increases of 0.77 to 0.79 in Oregon and 0.76 to 0.78 in Washington, which are qualitatively similar to those in our main specification. Next, we experiment with conducting structural break tests for the onset of fentanyl. We find that in Oregon, fentanyl may have arrived in February 2020 instead of around the time of Measure 110. Finally, our main results are robust to controlling for the timing of the fentanyl crisis for each state.

3.4.3 Replication of Zoorob et al. (2024)

In Panel I of Table 3.8, we present results from a replication of Zoorob et al. (2024).[17] The purpose of this exercise is to assess the robustness of key modeling choices made by the authors—specifically, the use of a fentanyl control and the extension of the pre-period to 2008. To evaluate the robustness of the fentanyl control, we compare models without any fentanyl control to two specifications: one using the original control employed by the authors (i.e., the share of fentanyl seizures relative to all drug seizures, denoted as "Share"), and a second using fentanyl seizures per capita.

We include the per-capita measure as an alternative proxy for fentanyl supply, as it captures overall prevalence while being independent of the frequency with which other drugs are seized. The primary justification for using the fentanyl seizure share, rather than raw counts, is that it is less susceptible to variation in overall law enforcement activity (Zibbell et al. 2023). However, while our per-capita measure is potentially influenced by changes in police behavior, we argue that such behavioral changes are themselves a consequence of Measure 110—again, our preferred specification omits any control variables due to changing behavior of those associated with the criminal justice system. Furthermore, if Zoorob et al.'s

[17] We use the exact data and code provided by Zoorob et al. (2024), available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/JT4CDS

intent in controlling for fentanyl supply was to isolate shifts in law enforcement activity, then any regime switching—as previously discussed—would itself induce endogenous changes in police behavior, as enforcement efforts are likely to focus on drugs most associated with overdose mortality.

Columns 1 through 3 of Table 3.8 present replication estimates for Oregon. Panels I through III replicate the analysis in Zoorob et al. (2024), with Panel I using biannual observations and a pre-treatment period beginning in January 2008, and Panel II using the same biannual structure but a shortened pre-period beginning in 2014. By shortening the pre-period to 2014, we isolate the fentanyl wave of the opioid epidemic, thereby limiting confounding from earlier regime shifts that could distort share-based measures of fentanyl supply. Column 2 across these panels provides a direct replication of the authors' preferred specification, which includes a control for the share of fentanyl seizures. In this specification, the estimated impact of Measure 110 on overdose mortality is consistently statistically insignificant.

In contrast, when we replace the share-based fentanyl control with an alternative proxy—fentanyl seizures per capita—the estimated treatment effect remains positive and statistically significant. These estimates (ATTs of 2.301 and 2.924 in Panels I and II, respectively) are also much closer in magnitude to the estimates produced when no fentanyl control is used at all (1.83 and 2.374, respectively).

Columns 4 through 6 show a similar pattern for Washington. Models that control for the share of fentanyl seizures yield statistically insignificant estimates. However, when using the per-capita fentanyl control, we observe significant increases in overdose mortality (ATTs of 2.035 and 2.658), both significant at the 1% level. Again, these results closely mirror those from models without any fentanyl control (ATTs of 1.341 and 2.124). Taken together, these findings suggest that the estimated impact of Measure 110 on overdose mortality is highly sensitive to the choice of fentanyl control variable.

Panel III of Table 3.8 presents replication estimates using the difference-in-differences (DiD) specification from Spencer (2023), following the approach outlined in Table 2 of Zoorob et al. (2024). As with the synthetic control models, controlling for the share of fentanyl seizures attenuates the estimated effect of Measure 110 to statistical insignificance. However, when we instead control for fentanyl seizures per capita, the estimated effect remains positive and statistically significant, closely aligning with results from the uncontrolled specification. These findings reinforce the conclusion that estimates of Measure 110's impact on overdose mortality are susceptible to how fentanyl supply is measured, with substantively different results depending on whether a relative (share-based) or absolute (per-capita) control is used.

Panels IV and V further shorten the pre-period while maintaining the biannual structure, replicating the estimation strategies of Spencer (2023) and Joshi et al. (2023), respectively. In both panels, we again compare specifications with no fentanyl control, a share-based fentanyl control, and a per-capita control. Notably, the attenuation observed in prior panels disappears —across all specifications, the estimated effects of Measure 110 on overdose mortality remain positive and statistically significant, regardless of the fentanyl control used.

By further shortening the pre-period—and accepting the conclusion from the change-point analysis in Zoorob et al. (2024) that Oregon's fentanyl saturation coincided with the implementation of Measure 110—we would expect the inclusion of a fentanyl control to substantially attenuate the estimated treatment effect. In this restricted sample, the donor pool is primarily composed of states that had already experienced full fentanyl saturation, making it difficult to separately identify the effects of Measure 110 from those of increased fentanyl exposure. Under this logic, controlling for fentanyl supply should absorb some of the increase in overdose mortality and thus reduce the estimated impact of M110. However, the fact that the point estimate changes little when using the share-based control—and in fact increases when using a per-capita measure—raises two possibilities. Either fentanyl saturation occurred well before the policy change, leaving no substantive shift to control for, or the variables used to proxy fentanyl supply—particularly the share-based measure—are inadequate and fail to capture meaningful variation in fentanyl prevalence.

3.5 Conclusion

The overdose epidemic is among the most devastating public health crises in recent memory. In an effort to reduce overdose deaths, many legislators have turned to decriminalization of drug possession as a potential solution. In February 2021, Oregon became the first state to implement such a policy through Measure 110. In this paper, we replicate and extend short-run findings on the causal impact of Measure 110 on overdose mortality (Spencer 2023, Joshi et al. 2023), and critically examine the empirical strategy of the only longer-term evaluation to date (Zoorob et al. 2024).

Our replication results show that small differences in the implementation of the synthetic control method can lead to contradictory short-run findings regarding Measure 110's impact on overdose mortality. While there were some differences in timing specifications and data access, the primary source of divergence stems from how overdose mortality was defined—

specifically, unintentional overdoses versus all-cause overdose deaths—which affects the donor weights that compose the synthetic control. We also find that the synthetic control method is relatively underpowered to detect small effect sizes, such as those estimated in Joshi et al. (2023). When extending the analysis to cover a longer horizon, we find that regardless of model specification, the long-run (2+ year) results consistently indicate a statistically significant increase in overdose mortality. The estimated average treatment effect on the treated (ATT) grows year-over-year from 2021 to 2023, suggesting a compounding effect in the second and third years post-implementation. Our estimates imply that Measure 110 is associated with a monthly increase of 0.78 to 0.85 deaths, corresponding to approximately 1125 to 1213 additional deaths during our sample window—an increase of 38 to 46% relative to the baseline mean.

While the magnitude of the estimated effect may be striking, we acknowledge several limitations in our empirical approach. Most notably, the pre-treatment period overlaps with the COVID-19 pandemic, meaning that donor states that served as suitable counterfactuals for Oregon pre-pandemic may not have remained so afterward. To address this, we exclude the COVID period from our pre-treatment window and find that the estimated effect remains positive and statistically significant, with an ATT of 0.88 to 0.89.

Additionally, concerns have been raised that the fentanyl wave of the opioid epidemic may confound estimates of Measure 110's impact (Zoorob et al. 2024). We replicate, extend, and conduct sensitivity analyses on the findings in Zoorob et al. (2024) to assess (a) whether the saturation of fentanyl into the illicit drug market biased M110 estimates, and (b) whether their biannual specification was robust to alternative pre-treatment periods. We find that their use of fentanyl's share of all drug seizures is highly sensitive to law enforcement priorities across substances. As an alternative, we control for fentanyl seizures per capita—acknowledging that this measure is also influenced by enforcement intensity—and observe that the attenuation effect found using the fentanyl share disappears. Beyond sensitivity to the fentanyl measure itself, we also test the robustness of these results to variations in the length of the pre-treatment window. Across shorter pre-treatment periods, the estimated effect remains positive and statistically significant.

We contribute to the harm reduction and substance use disorder intervention literature by providing credible evidence of externalities associated with decriminalization policies. However, it is important to acknowledge that Measure 110 included many components beyond decriminalization—several of which are still ongoing—such as expanded funding for treatment services. We also draw attention to the so-called "fentanyl loophole," which was not closed until June 2023. While our findings suggest that Measure 110 increased Oregon's overdose mortality rate, further research is needed—and is already underway—to identify the specific mechanisms driving this increase. Finally, our analysis is necessarily limited to the unique context of Oregon's implementation of decriminalization, including its imperfections. Our results reflect the policy as it was enacted in Oregon, rather than decriminalization in the abstract.

Any evaluation of Oregon's outcomes must also consider parallel evidence from Washington's Blake Fix. Our analysis indicates that the brief period of decriminalization under the Blake Fix was likewise associated with an increase in overdose mortality. Thus, any proposed mechanism driving Oregon's outcomes must also be plausible in light of Washington's distinct implementation. Further research is needed to uncover shared pathways through which decriminalization may contribute to overdose increases in both states. One such candidate is reduced enforcement, as proxied by declining arrest rates—a pattern we illustrate in Figure 3.4.

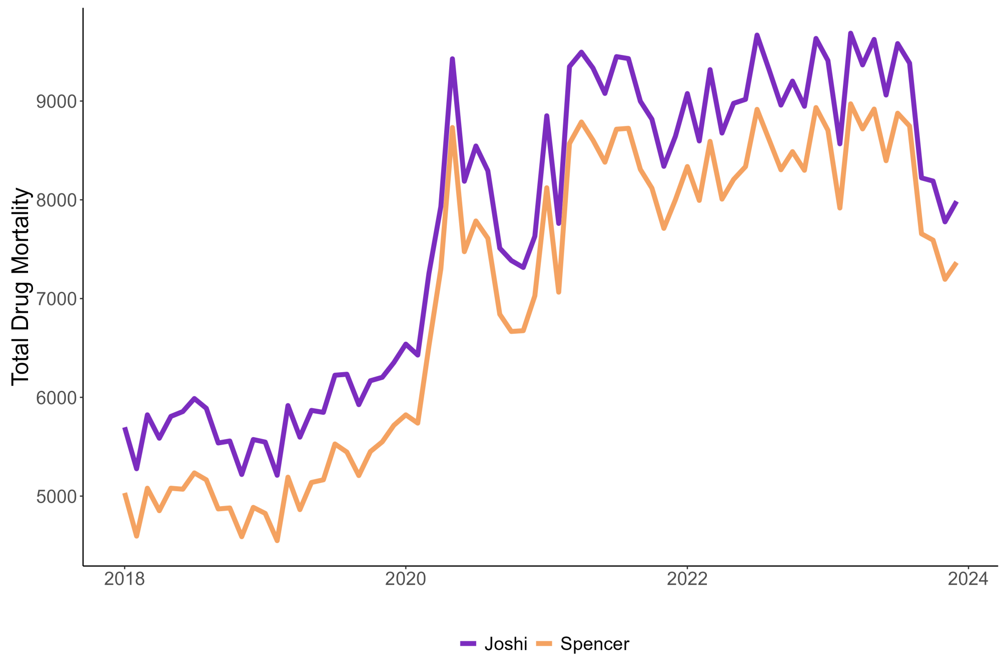
Further research is also needed to assess the broader impacts of Measure 110 beyond overdose mortality. As Netherland et al. (2022) note: “The success or failure of Measure 110 has the potential to shape drug policy in the USA for decades to come, but ‘success’ or ‘failure’ is entirely dependent on what outcomes are being measured, how the data are gathered, and whether the findings are understood within the broader context of what is happening on the ground in Oregon.” A comprehensive evaluation of the policy must therefore move beyond a narrow focus on mortality and engage with the full spectrum of ways in which substance use intersects with health, criminal justice, and social systems.

3.6 Tables & Figures

3.6.1 Figures

Figure 3.1: A chart of the overdoses nationwide using different definitions of overdose.

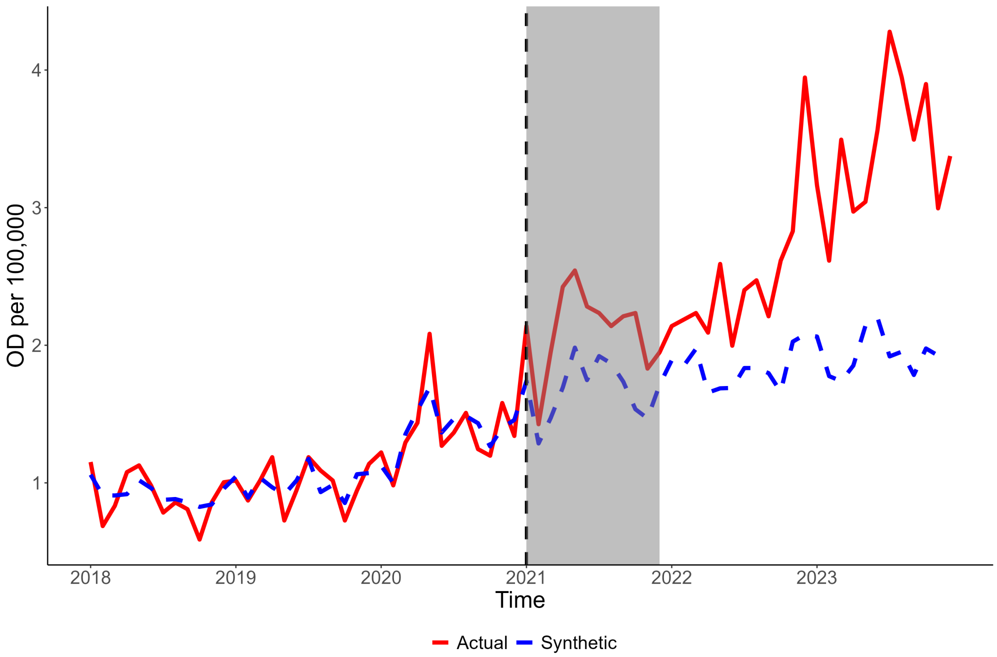
(a) By Definition



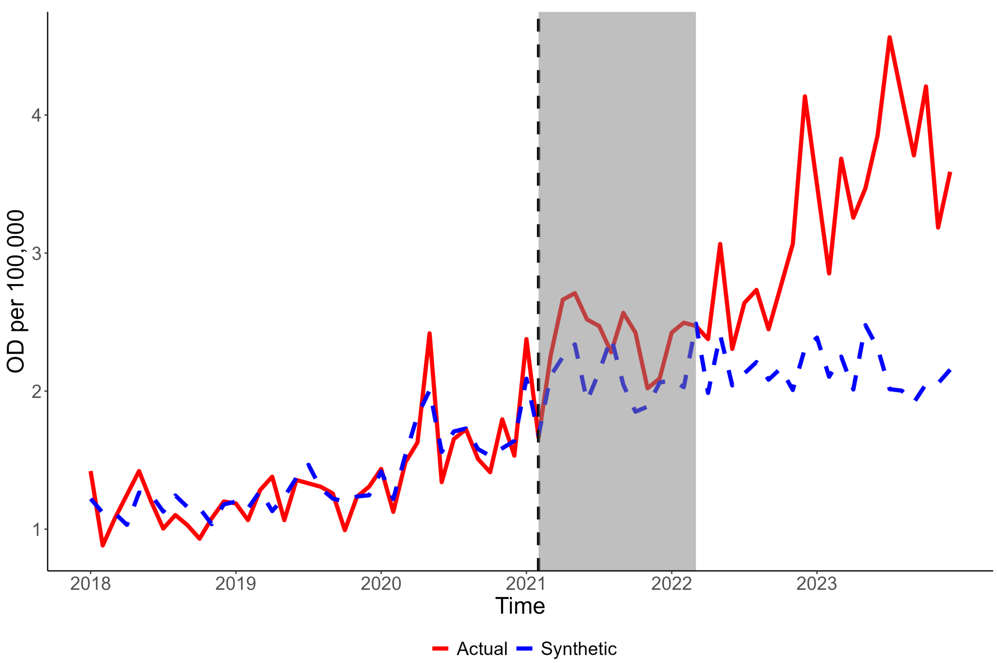
(b) Differences Between Spencer and Joshi et al. Definition

Figure 3.2: Oregon vs. Synthetic Oregon in Drug Overdose Rate, Longer Time Window

(a) Spencer (2023) Specification



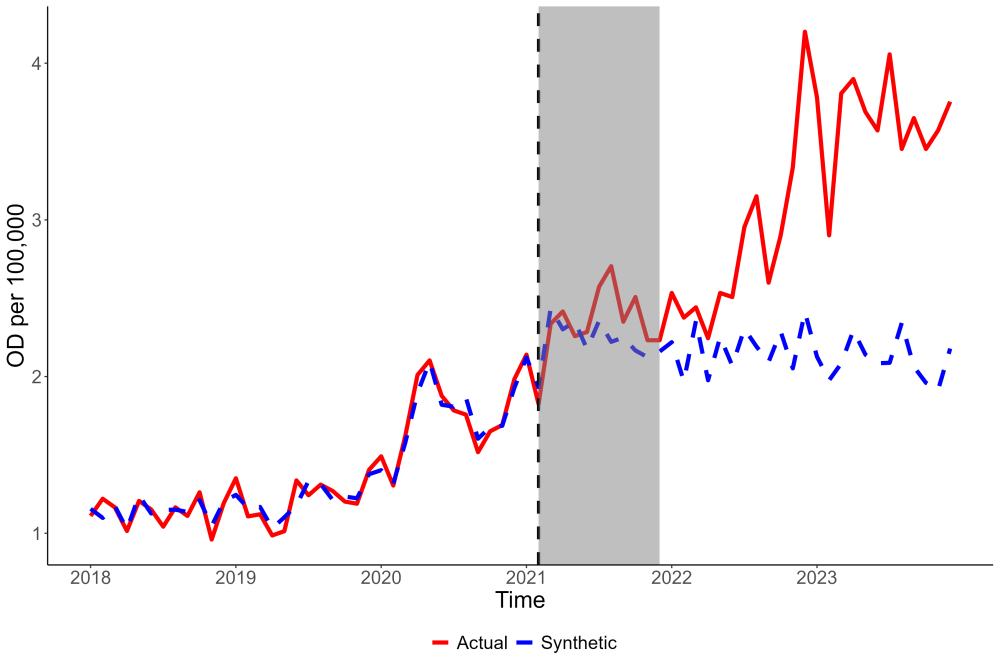
(b) Joshi et al. (2023) Specification



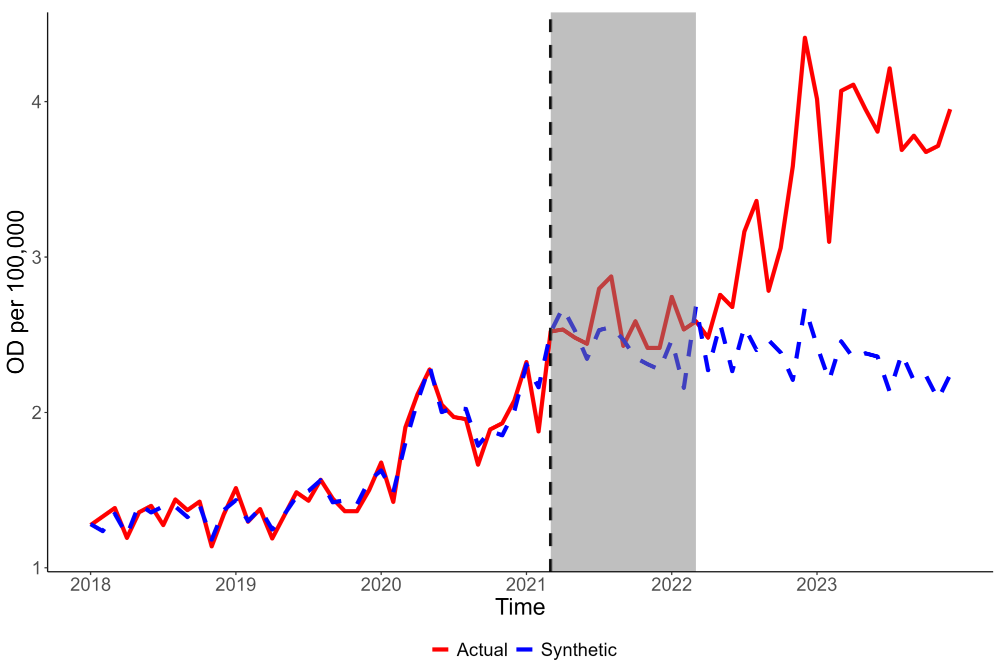
Notes: Synthetic Control Estimates using Spencer (2023)'s specification (panel a) and Joshi et al. (2023)'s specification (panel b) are shown. In panel a, the outcome is unintentional drug overdose, and the pre-treatment window is from Jan 2018 to Dec 2020. In panel b, the outcome is all drug overdose and the pre-treatment widnow is from Jan 2018 to Jan 2021. The gray shaded area represents each paper's post-treatment window.

Figure 3.3: Washington vs. Synthetic Washington in Drug Overdose Rate, Longer Time Window

(a) Spencer (2023) Specification



(b) Joshi et al. (2023) (2023) Specification



Notes: Synthetic Control Estimates using Spencer (2023)'s specification (panel a) and Joshi et al. (2023)'s specification (panel b) are shown. In panel a, the outcome is unintentional drug overdose, and the pre-treatment window is from Jan 2018 to Jan 2021. In panel b, the outcome is all drug overdose, and the pre-treatment widnow is from Jan 2018 to Feb 2021. The gray shaded area represents each paper's post-treatment window.

3.6.2 Tables

Table 3.1: Changes from M110 for possession above or below "user threshold."

Drug:	Section:	Threshold Amt.:	Pre: Below	Post: Below	Pre: Above	Post: Above
LSD	11	40 u.u.	Class A Misdemeanor	Class E Violation	Class B Felony	Class A Misdemeanor
Psilocybin/Psilocin	11	12g	Class A Misdemeanor	Class E Violation	Class B Felony	Class A Misdemeanor
Methadone	12	40 u.u.	Class A Misdemeanor	Class E Violation	Class C Felony	Class A Misdemeanor
Oxycodone	13	40 tab.	Class A Misdemeanor	Class E Violation	Class C Felony	Class A Misdemeanor
Heroin	14	1g	Class A Misdemeanor	Class E Violation	Class B Felony	Class A Misdemeanor
MDMA/Ecstacy	15	1g/5 tab.	Class A Misdemeanor	Class E Violation	Class B Felony	Class A Misdemeanor
Cocaine	16	2g	Class A Misdemeanor	Class E Violation	Class C Felony	Class A Misdemeanor
Methamphetamine	17	2g	Class A Misdemeanor	Class E Violation	Class C Felony	Class A Misdemeanor

Table 3.2: Synthetic control weights in each authors' analysis by state (abbreviated).

	Weight	
State Abbr.	**Spencer (2023)**	**Joshi et al. (2023)**
AK	0.026	—
AL	—	0.111
CO	0.072	0.045
DC	0.036	—
IA	0.061	—
KS	0.216	0.058
MD	0.282	—
MN	—	0.027
MS	0.012	0.149
MT	0.178	0.272
NC	0.04	0.107
NE	—	0.038
OK	—	0.089
SD	0.032	—
VT	0.021	—
WV	—	0.01
WY	0.023	0.044

Table 3.3: An overview of the differences between the two replicated works.

Author	Joshi et al. (2023)	Spencer (2023)
ICD-10 Codes	X40-44, X60-64, X85, Y10-14, T36-50	X40-44
NVSS Access	Unrestricted	Restricted
Treatment Date	Feb. 2021	Jan. 2021

Table 3.4: Replication: Synthetic Control Estimate for Oregon

	Panel I: Spencer (2023) Outcome			
ATT	0.434**	0.381**	0.403**	0.346**
P-Val	{0.02}	{0.02}	{0.02}	{0.04}
	Panel II: Joshi et al. (2023) Outcome			
ATT	0.329*	0.272	0.32	0.27
P-Val	{0.08}	{0.16}	{0.16}	{0.26}
Sample	2018-21	2018-21	2018-22	2018-22
Matching Until	Dec	Jan	Dec	Jan

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Synthetic control estimates are shown. All estimates use January 2018 to either December 2020 or January 2021 to find the synthetic weights. Panel I focuses on all unintentional drug overdoses, which is the outcome used by Spencer (2023). Panel II focuses on all drug mortality, which is the outcome used by Joshi et al. (2023). The p-values, calculated via a permutation test, are provided inside the curly brackets.

Table 3.5: Replication: Synthetic Control Estimate for Washington

	Panel I: Spencer (2023) Outcome			
ATT	0.113**	0.135**	0.147**	0.166**
P-Val	{0.02}	{0.02}	{0.02}	{0.02}
	Panel II: Joshi et al. (2023) Outcome			
ATT	0.06**	0.095	0.087**	0.116
P-Val	{0.02}	{0.14}	{0.02}	{0.1}
Sample	2018-21	2018-21	2018-22	2018-22
Matching Until	Jan	Feb	Jan	Feb

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Synthetic control estimates are shown. All estimates use January 2018 to either January 2021 or February 2021 to find the synthetic weights. Panel I focuses on all unintentional drug overdoses, which is the outcome used by Spencer (2023). Panel II focuses on all drug mortality, which is the outcome used by Joshi et al. (2023). The p-values, calculated via a permutation test, are provided inside the curly brackets.

Table 3.6: Extension: Synthetic Control Estimate for Oregon, Include Longer Post-treatment Period

	Spencer (2023)		Joshi et al. (2023)	
ATT Overall	0.848**	0.787**	0.796**	0.806**
P-Val	{0.02}	{0.02}	{0.02}	{0.02}
ATT 2021	0.434**	0.381**	0.329*	0.272
P-Val	{0.02}	{0.02}	{0.08}	{0.16}
ATT 2022	0.644**	0.567**	0.563**	0.582**
P-Val	{0.02}	{0.02}	{0.02}	{0.02}
ATT 2023	1.466**	1.378**	1.496**	1.521**
P-Val	{0.02}	{0.02}	{0.02}	{0.02}
Marching Until	Dec	Jan	Dec	Jan

* P-val $<$ 0.1; ** P-val $<$ 0.05; *** P-val $<$ 0.01

Notes: Synthetic control estimates are shown. All estimates use January 2018 to either December 2020 or January 2021 to find the synthetic weights. Panel I focuses on all unintentional drug overdoses, which is the outcome used by Spencer (2023). Panel II focuses on all drug mortality, which is the outcome used by Joshi et al. (2023). The p-values, calculated via a permutation test, are provided inside the curly brackets.

Table 3.7: Extension: Synthetic Control Estimate for Washington, Include Longer Post-treatment Period

	Spencer (2023)		Joshi et al. (2023)	
ATT Overall	0.776**	0.802**	0.752**	0.782**
P-Val	{0.02}	{0.02}	{0.02}	{0.02}
ATT 2021	0.113**	0.135**	0.06**	0.095
P-Val	{0.02}	{0.02}	{0.02}	{0.14}
ATT 2022	0.633**	0.633**	0.586**	0.586**
P-Val	{0.02}	{0.02}	{0.02}	{0.02}
ATT 2023	1.528**	1.528**	1.551**	1.551**
P-Val	{0.02}	{0.02}	{0.02}	{0.02}
Marching Until	Jan	Feb	Jan	Feb

* P-val $<$ 0.1; ** P-val $<$ 0.05; *** P-val $<$ 0.01

Notes: Synthetic control estimates are shown. All estimates use January 2018 to either January 2021 or February 2021 to find the synthetic weights. Panel I focuses on all unintentional drug overdoses, which is the outcome used by Spencer (2023). Panel II focuses on all drug mortality, which is the outcome used by Joshi et al. (2023). The p-values, calculated via a permutation test, are provided inside the curly brackets.

Table 3.8: Zoorob (2024) Replication & Extension

	Oregon			Washington		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel I: Zoorob SC Replication, 2008-2022					
ATT	1.83***	-0.505	2.301***	1.341**	-0.868	2.035***
P-Val	{0}	{0.407}	{0}	{0.044}	{0.132}	{0.001}
	Panel II: Zoorob SC Replication, 2014-2022					
ATT	2.374***	0.415	2.924***	2.124***	-0.031	2.658***
P-Val	{0}	{0.473}	{0}	{0}	{0.961}	{0}
	Panel III: Zoorob DiD Replication					
ATT	0.248***	0.094	0.352***	0.197**	-0.123	0.301***
	(0.072)	(0.106)	(0.063)	(0.074)	(0.158)	(0.065)
	Panel IV: Spencer (2023) SC Estimate					
ATT	0.848***	0.725***	0.883***	0.776***	0.576***	0.358***
P-Val	{0.02}	{0.02}	{0.02}	{0.02}	{0.02}	{0.04}
	Panel V: Joshi et al. (2023) SC Estimate					
ATT	0.806***	0.82***	0.79***	0.782***	0.572***	0.615***
P-Val	{0.02}	{0.02}	{0.02}	{0.02}	{0.02}	{0.02}
Fentanyl Control	None	Share	PC	None	Share	PC

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

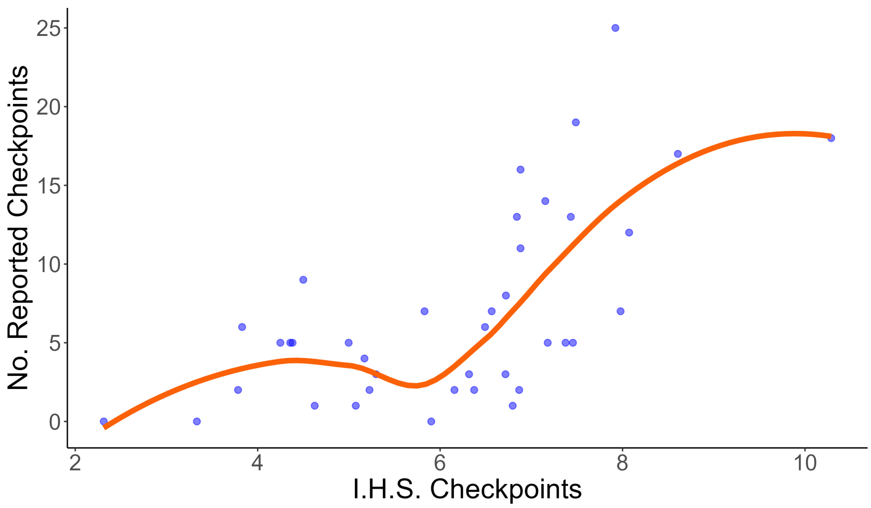
Notes: Panel I replicates and extends Zoorob et al.'s preferred synthetic control method. Panel II replicates and extends Zoorob et al.'s preferred specification using a narrower pre-treatment window. Panel III replicates and extends Zoorob et al.'s replication of Spencer (2023). Panels IV and V provide robustness tests for the inclusion of fentanyl control in our preferred synthetic control estimation. In columns 2,3, 5, and 6 of panels IV and V, we use bi-annual pre-treatment rate of drug overdose rates as well as bi-annual fentanyl controls from 2018-2022 to find our synthetic weights.

Appendices

4.1 Appendix for Chapter 1

4.1.1 Chapter 1. Appendix Figures

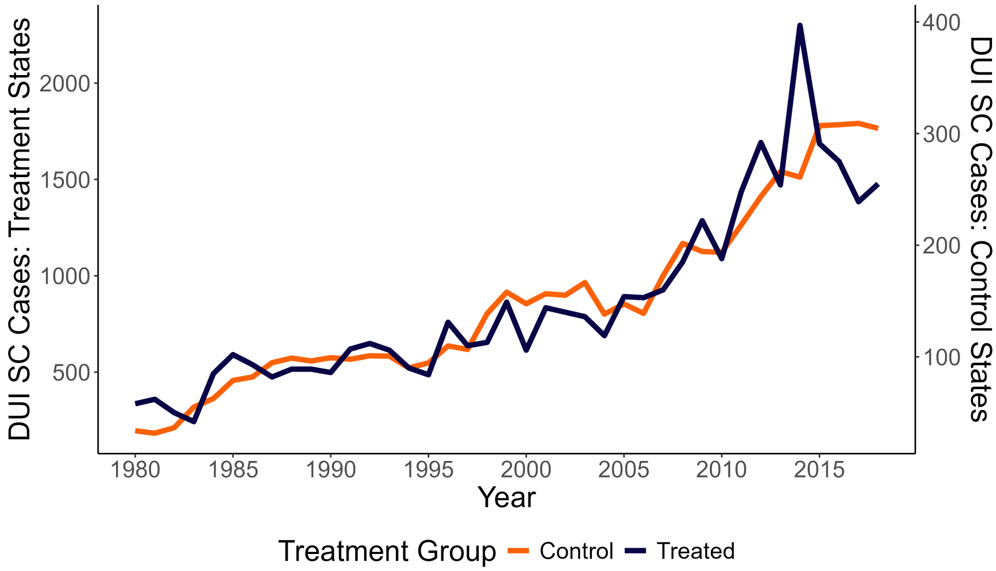
Appendix Figure 1.1: Correlation between Number of DUI Checkpoints & Number of Court Cases Involving DUI Checkpoints, 2008-2018



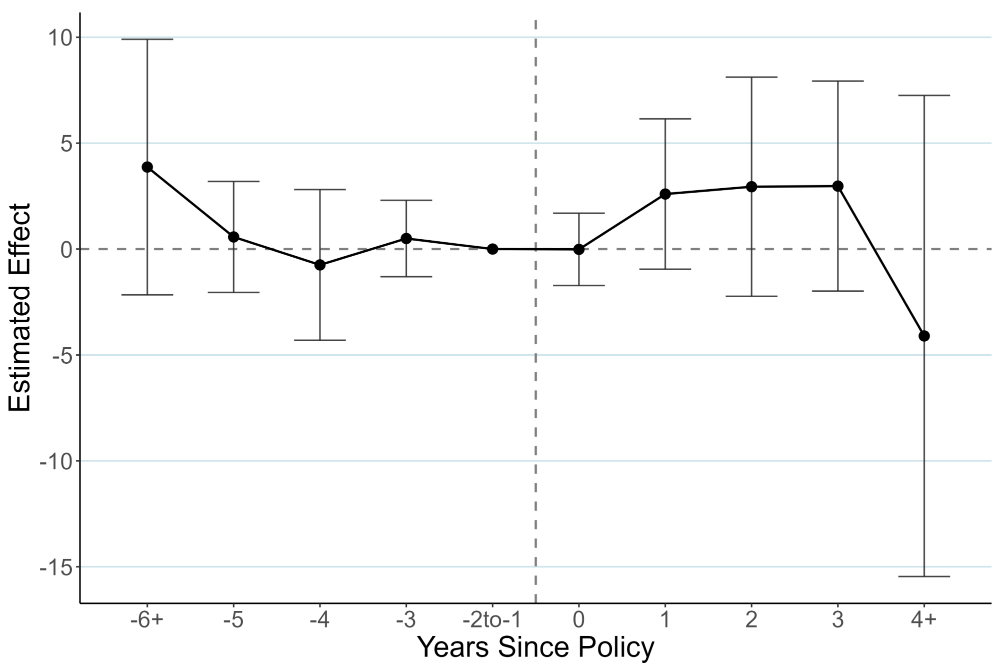
Notes: The scatterplot shows the correlation between the reported number of DUI checkpoints and the number of state supreme court cases in which a defendant was arrested at a DUI checkpoint. The sample is restricted to states that conduct DUI checkpoints. The number of DUI checkpoints is transformed using the inverse hyperbolic sine transformation. The orange line shows locally estimated scatterplot smoothing line.

Appendix Figure 1.2: "First Stage" Effect of Changes in State Court Cases Arguing About DUIs

(a) Descriptive Trends



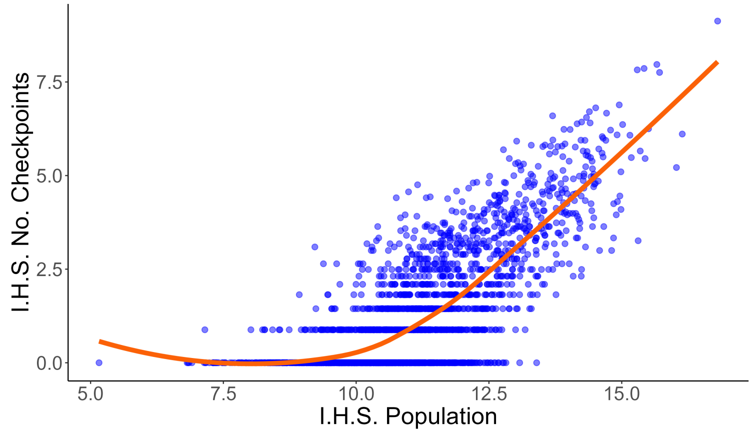
(b) TWFE Estimates



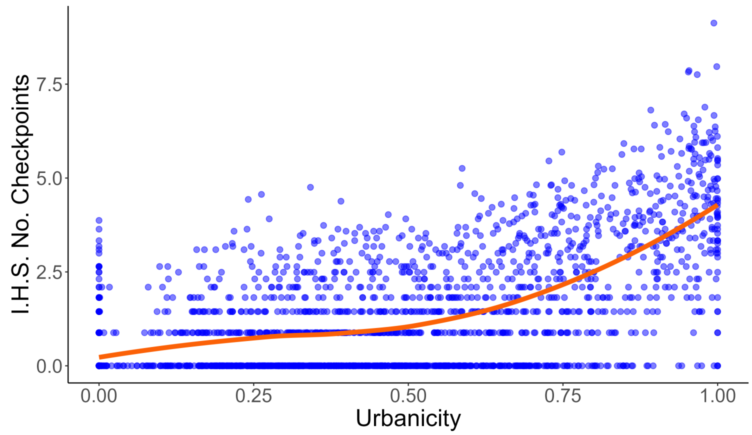
Notes: The data are manually collected from casetext.com. The outcome is the count of court cases that mention the keyword DUI. Panel a presents the total counts of court cases for each group. Panel b presents the population-weighted TWFE OLS event study estimates. The estimate in panel b includes controls for state and year fixed effects. The bar lines in panel b represent 95% confidence intervals generated using standard errors clustered at the state-level.

Appendix Figure 1.3: Correlation between Population & Number of DUI Checkpoints, 2008-2018

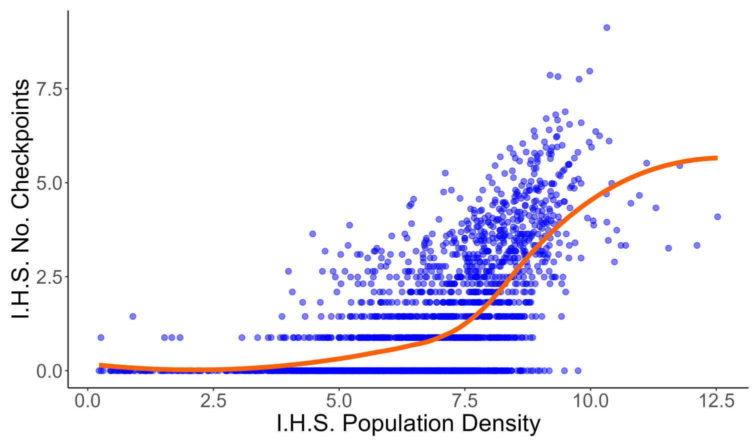
(a) Population



(b) Urbanicity

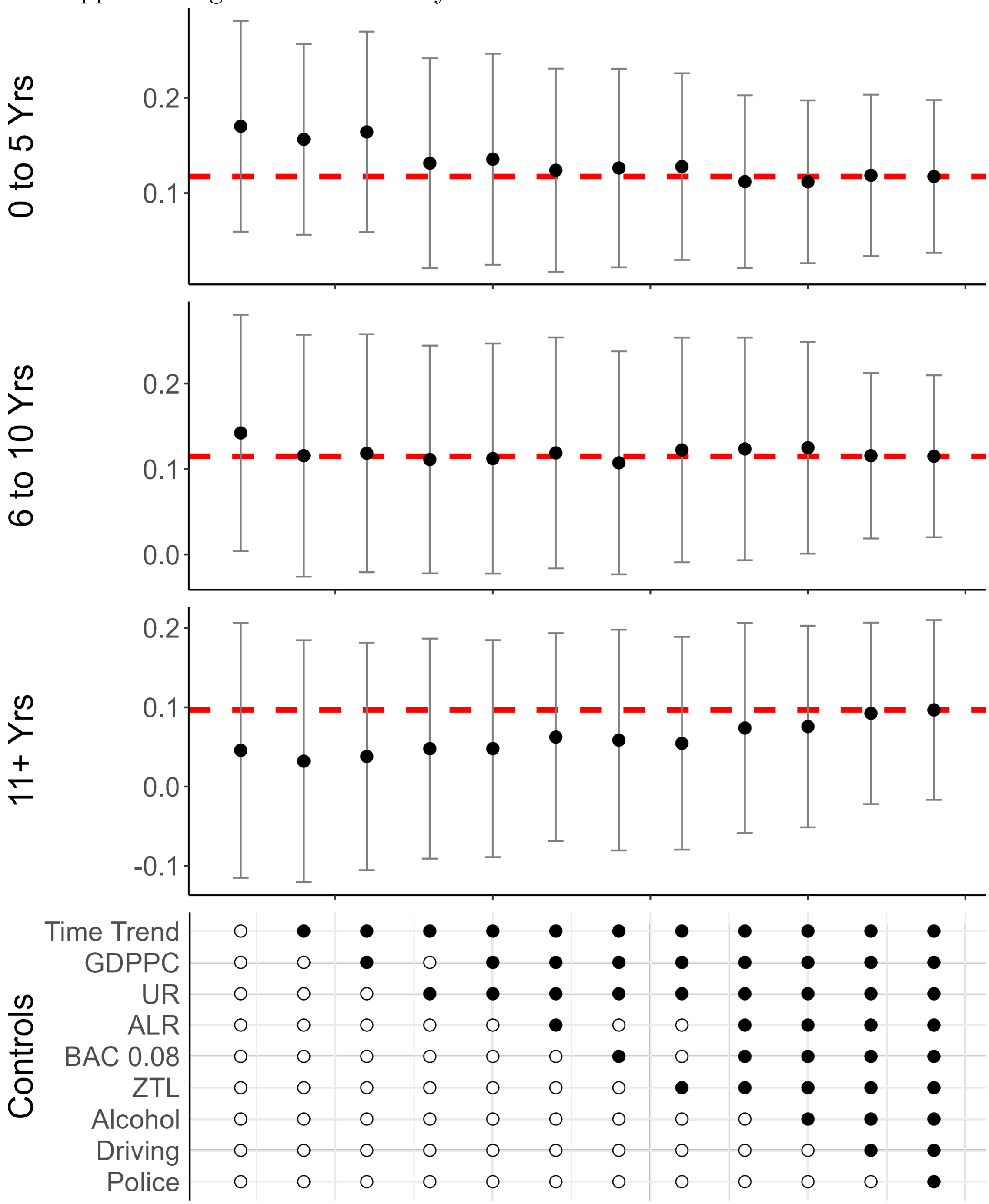


(c) Population Densitiy



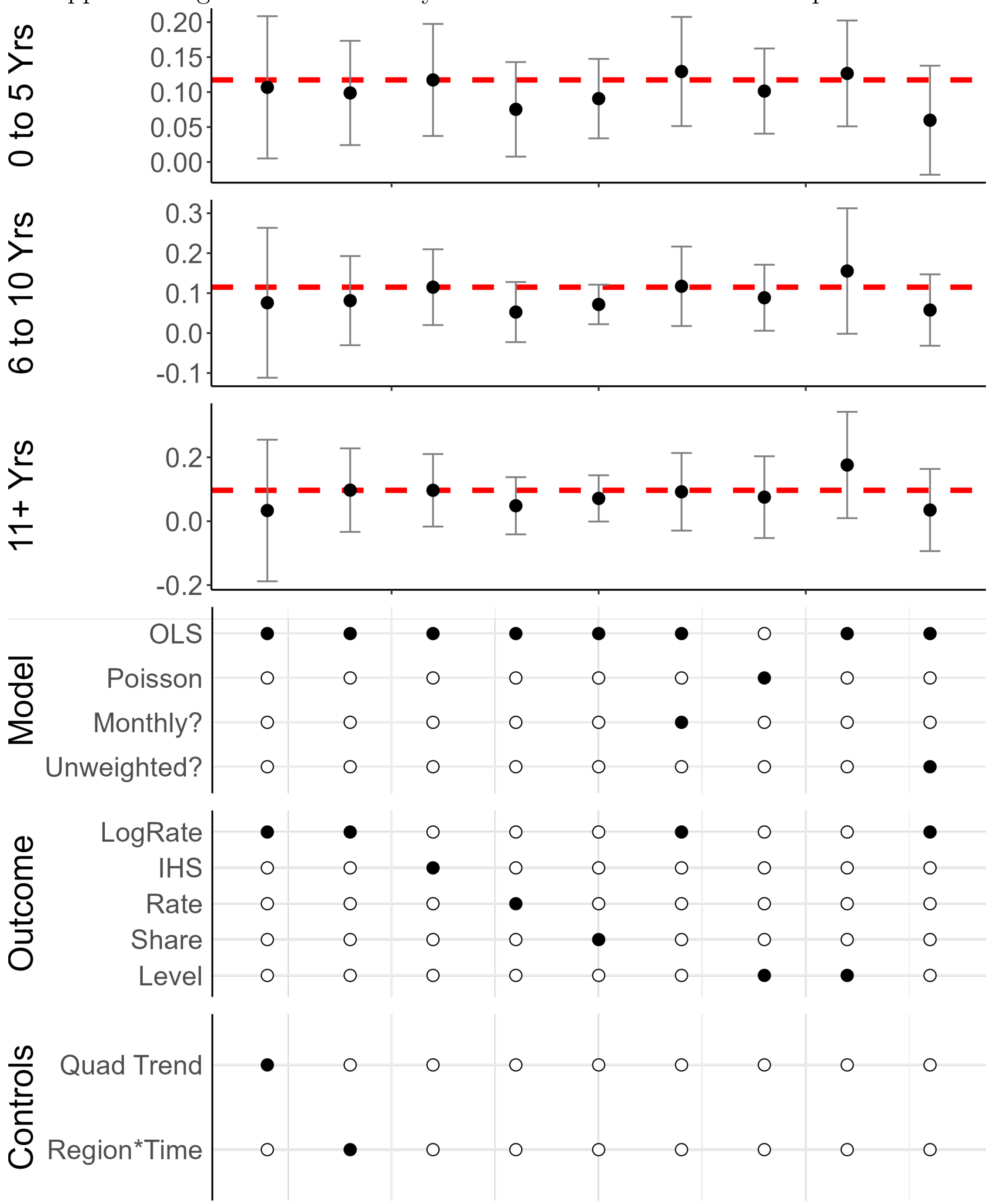
Notes: The scatterplots show the correlation between population, urbanicity, population density, and each county's reported number of DUI checkpoints. The sample is restricted to counties in states that conduct DUI checkpoints. Population, population density, and the number of DUI checkpoints are transformed using the inverse hyperbolic sine transformation. The orange line shows locally estimated scatterplot smoothing line.

Appendix Figure 1.4: Sensitivity of FARS Estimates to Control Variables



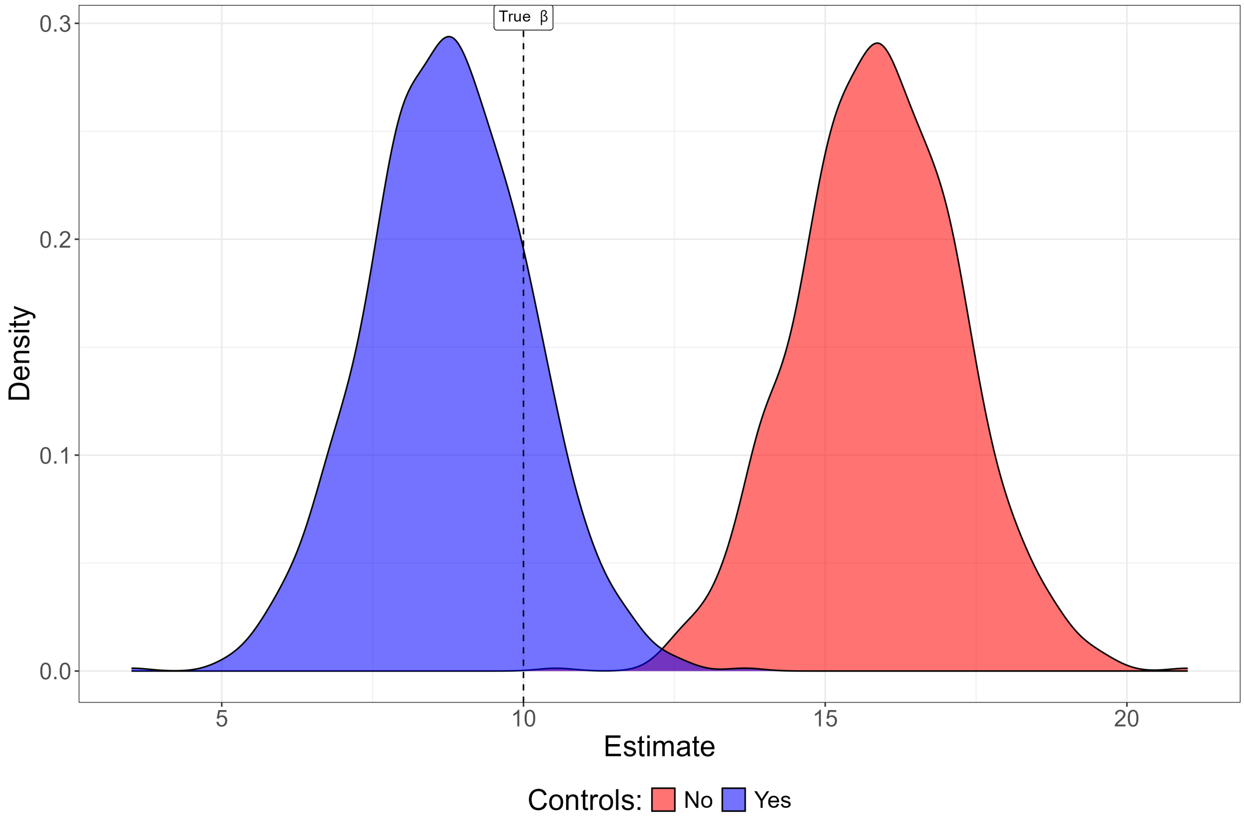
Notes: The top row presents sensitivity to estimated coefficients for 0 to 5 year post-ban coefficients. The second row presents sensitivity to estimated coefficients for 6 to 10 year post-ban. The third row presents sensitivity to estimated coefficients for 11 or more years post-ban. The bar lines represent 90% confidence intervals generated using standard errors clustered at the state-level. The red dashed line represents the point estimate from my preferred specification. All estimates are estimated using population-weighted OLS. Time trend control is a state-specific linear time trend. Alcohol controls include MLDA laws and beer tax rates. Driving controls include speed limit, GDL, total vehicle miles driven, and seatbelt law. Police controls include police expenditure and employment per 100,000.

Appendix Figure 1.5: Sensitivity of FARS Estimates to Model Specification



Notes: The top row presents sensitivity to estimated coefficients for 0 to 5 year post-ban coefficients. The second row presents sensitivity to estimated coefficients for 6 to 10 year post-ban. The third row presents sensitivity to estimated coefficients for 11 or more years post-ban. The bar lines represent 90% confidence intervals generated using standard errors clustered at the state-level. The red dashed line represents the point estimate from my preferred specification. The beta coefficients and confidence intervals for OLS estimates, where the outcomes are rate, share, and level, are divided by the pre-treatment mean to represent the percent change in the outcome variable.

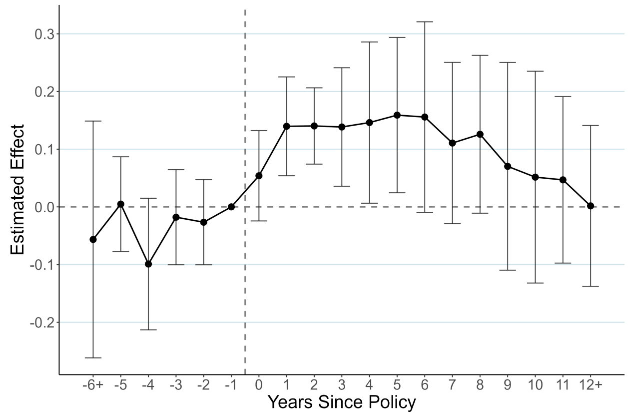
Appendix Figure 1.6: 1000 Simulations of Callaway & Sant'Anna Estimates when Omitted Variable Bias is Present



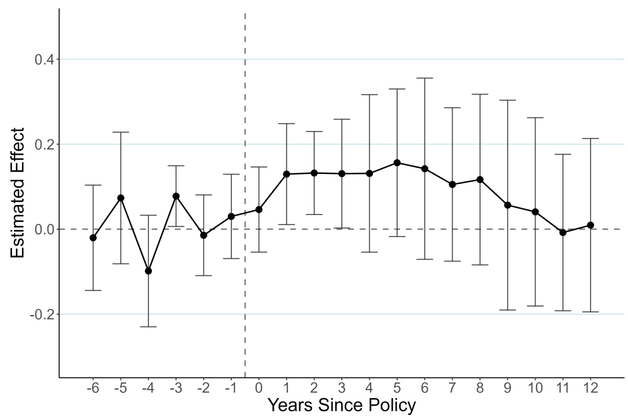
Notes: In this simulation, I generate a data-generating process where I generate a panel of 50 states + D.C. from 1980 to 2018. My outcome is defined as $Y = \beta Treat + \alpha X + \gamma_s + \delta_t + \varepsilon$, where Treat is my main policy (DUI checkpoint-ban) variable I am using in my paper; X is a vector of set of controls I include in my preferred model; γ and δ are my state and time fixed effects; and ε is my random error term distributed normally with mean 0 and standard deviation of 1. I set my β as equal to 10. The red shaded area represents my 1,000 estimated ATTs from Callaway Sant'Anna, where I use Y as my outcome. The blue shaded area represents my 1,000 estimated ATTs from Callaway Sant'Anna, where I use residuals after I regress my Y on X (but not on my policy variable).

Appendix Figure 1.7: Event Study Analysis Without Covariates

(a) Callaway Sant'Anna: No Controls



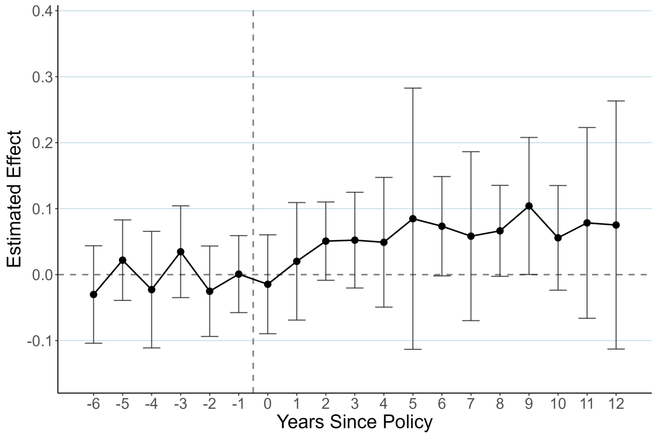
(b) TWFE: No Controls



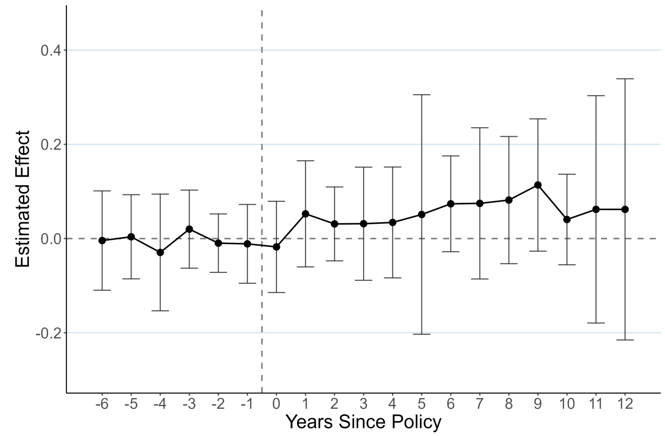
Notes: The population-weighted estimates are generated using data from the 1980-2018 FARS. Counterfactuals for panel a are restricted to not-yet-adopting states. The bar lines represent 95% confidence intervals generated using bootstrapped (panel a) and clustered (panel b) standard errors.

Appendix Figure 1.8: Callaway Sant'Anna Estimates on Other Traffic Fatalities, FARS 1980-2018

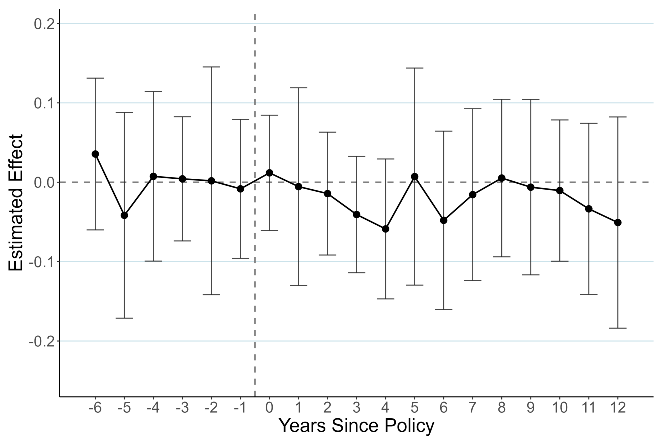
(a) Weekend Fatalities



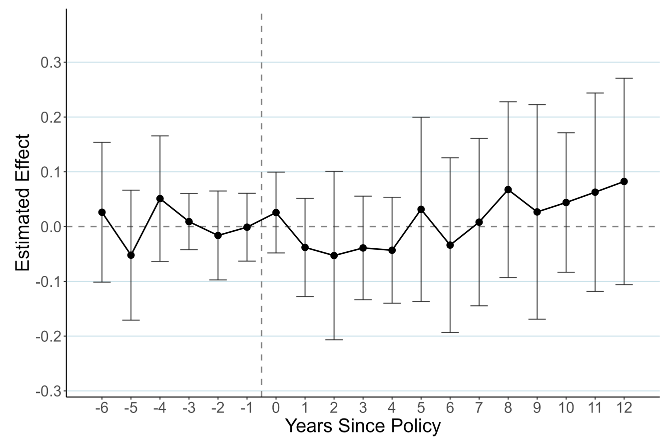
(b) Weekend or Holiday Night Fatalities



(c) Week Day Fatalities

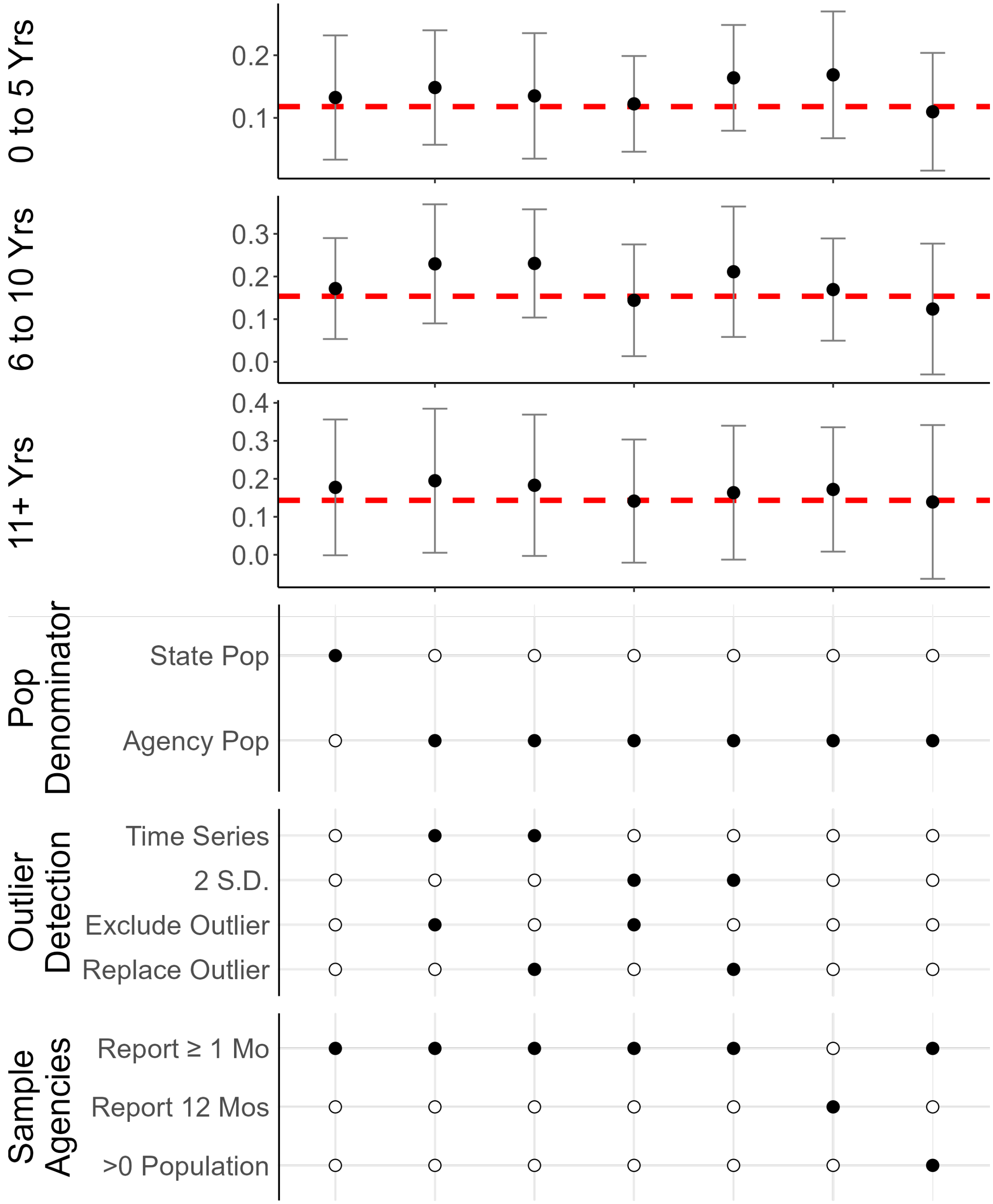


(d) Work Hour Fatalities



Notes: Population-weighted estimates are generated using data from the 1980-2018 FARS. The outcome variables are the residuals after I regress my outcomes on my preferred set of controls. The counterfactuals are restricted to not-yet-adopting states. The bar lines represent 95% confidence intervals generated using bootstrapped standard errors.

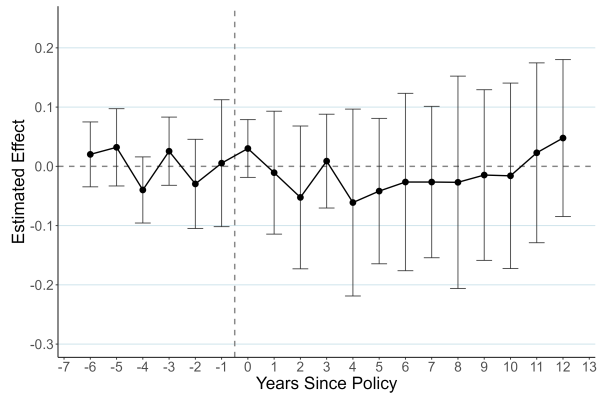
Appendix Figure 1.9: Sensitivity of UCR Estimates to Various Data Cleaning & Samples



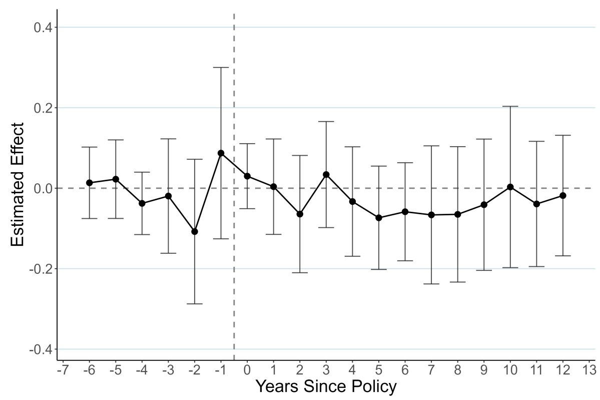
Notes: The top row presents sensitivity to estimated coefficients for 0 to 5 year post-ban coefficients. The second row presents sensitivity to estimated coefficients for 6 to 10 year post-ban. The third row presents sensitivity to estimated coefficients for 11 or more years post-ban. The bar lines represent 90% confidence intervals generated using standard errors clustered at the state-level. The red dashed line represents the point estimate from my preferred specification. All estimates are estimated using population-weighted OLS. For time-series outlier detection, I decompose each state's time series into seasonality, trends, and unobservables. Any unobservables three times the IQR are defined as outliers and replaced with the mean. For 2 S.D. outlier detection, I calculate each state's mean and standard deviation. Any observation over two times the standard deviation away from the mean is classified as an outlier and replaced with the mean.

Appendix Figure 1.10: Callaway Sant'Anna Estimates on Other Arrests, UCR 1980-2018

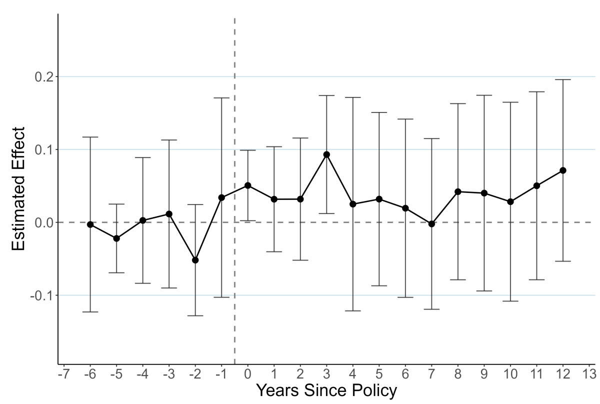
(a) Property Crime



(b) Violent Crime



(c) DUI Complement Crime



Notes: Population-weighted estimates are generated using data from the 1980-2018 UCR. The main outcome variables are the residuals after I regress my outcomes on my preferred set of controls. The counterfactuals are restricted to not-yet-adopting states. The bar lines represent 95% confidence intervals generated using bootstrapped standard errors. Property crimes include arson, burglary, larceny, and motor vehicle theft. Violent crimes include aggravated assault, forcible rape, murder, and robbery.

DUI complement crimes include drug possession, weapon law violation, aggravated assault, simple assault, and vandalism.

4.1.2 Chapter 1. Appendix Tables

Appendix Table 1.1: Summary Stats of Control Variables

Variable	Mean	S.D.	Source
Unemployment Rate	6.238	2.078	University of Kentucky Centers for Poverty Research (2022)
GDP per capita*	50381.787	12164.079	University of Kentucky Centers for Poverty Research (2022)
Zero Tolerance Drunk Driving Laws	0.662	0.473	National Highway Traffic Safety Administration (2001)
0.08 BAC Law	0.569	0.495	Freeman (2007)
Administrative License Revocation	0.658	0.475	Fell & Scherer (2017)
Minimum Legal Drinking Age	0.927	0.260	National Highway Traffic Safety Administration (2001)
Beer Tax Rate*	1.075	0.314	Beer Institute (2021)
65 MPH Maximum Speed Limit	0.350	0.477	FARS
70 MPH Maximum Speed Limit	0.454	0.498	FARS
Graduated Driver's Licensing Law	0.566	0.496	Argys et al. (2019)
Total Miles Driven per Capita*	9104.669	1723.422	U.S. Department of Transportation (2020)
Seatbelt Law - Secondary Enforcement	0.846	0.361	The Centers for Disease Control & Prevention (2015a)
Seatbelt Law - Primary Enforcement	0.471	0.499	The Centers for Disease Control & Prevention (2015a)
Total Police Employment per 100,000*	322.175	82.150	Kaplan (2021a)
Police Expenditure per 100,000*	21174.250	11245.021	Urban Institute

* denotes variables that are log transformed in my regression analysis.

Notes: The mean and standard deviations are population-weighted using data from 1980-2018.

Appendix Table 1.2: County Population Statistics by Reported DUI Checkpoints

	Mean	Median	75th Percentile	90th Percentile	Max
			Counties w/ DUI Checkpoints		
Urbanicity	0.555	0.565	0.806	0.958	1.000
Population Density	459.198	109.745	324.926	1080.479	18486.078
Population	198732.325	58658.000	172895.000	488449.200	9943046.000
			Counties w/o DUI Checkpoints		
Urbanicity	0.263	0.232	0.456	0.621	1.000
Population Density	64.834	27.498	50.881	93.548	4585.080
Population	21735.136	14267.500	25600.250	44449.500	329331.000

Notes: The unweighted means are reported. The sample is restricted to states with full coverage, as defined by DUIBlock.com. Population data is from SEER and represents the value for the year 2020. Urbanicity and population density data are from the Census and represent the values for the year 2010.

Appendix Table 1.3: Estimated Effect of DUI Checkpoint Bans by State Population, FARS 1980-2018

	(1)	(2)
0 to 5 Years After	0.152**	0.025
	(0.052)	(0.092)
	{0.044}*	{0.835}
	[0.027,0.292]	[-0.179,0.25]
6 to 10 Years After	0.137^{+}	0.023
	(0.072)	(0.089)
	{0.164}	{0.788}
	[-0.024,0.305]	[-0.138,0.174]
11+ Years After	0.134	-0.063
	(0.078)	(0.121)
	{0.178}	{0.636}
	[-0.031,0.305]	[-0.275,0.163]
Sample	Pop $\geq$ Median	Pop < Median
N	995	975

$^{+}$ P < .10; * P < .05; ** P < .01

Notes: Population-weighted estimates are generated using data from the 1980-2018 FARS. All estimates include state and year fixed effects and all of my preferred set of control variables. Standard errors clustered at the state-level are reported inside the parenthesis. The p-values obtained using wild-cluster bootstrapping are shown inside curly brackets, and the 90% CIs generated using wild-cluster bootstrapping are shown inside the square brackets.

Appendix Table 1.4: Average Police Enforcement by Population Size

	Police Per Capita	Police Expenditure per Capita	No. DUI Checkpoints[a]
Pop $\geq$ Median	32.74	65.41	1453.2
Pop $<$ Median	28.44	62.09	175.0

[a] The number of DUI checkpoints is based on the total reported DUI checkpoints between 2008 and 2018 among states allowing he use of DUI checkpoints.

Notes: Population weighted means are reported.

Appendix Table 1.5: Estimated Effect of DUI Checkpoint Bans on Non-DUI Speeding Fatalities & DUI Arrests: Using Subset of Counties

	Population		Urbanicity		Density	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel I: Non-DUI Speeding Fatalities, FARS 1980-2018					
0 to 5 Year After	-0.143	-0.061	-0.141^{+}	-0.041	-0.132	-0.081
	(0.086)	(0.091)	(0.082)	(0.069)	(0.086)	(0.078)
6 to 10 Years After	-0.026	-0.052	-0.059	0.048	-0.050	0.020
	(0.130)	(0.128)	(0.128)	(0.150)	(0.130)	(0.147)
11+ Years After	-0.050	-0.155	-0.090	-0.015	-0.065	-0.089
	(0.142)	(0.218)	(0.149)	(0.211)	(0.149)	(0.205)
N	1,866	1,595	1,858	1,694	1,865	1,666
	Panel II: DUI Arrests, UCR 1980-2018					
0 to 5 Year After	0.134*	0.140	0.154*	0.060	0.147**	0.061
	(0.052)	(0.097)	(0.062)	(0.093)	(0.055)	(0.112)
6 to 10 Years After	0.182*	0.102	0.198*	0.071	0.191*	0.054
	(0.077)	(0.131)	(0.095)	(0.128)	(0.081)	(0.155)
11+ Years After	0.178	0.109	0.200	0.076	0.190	0.043
	(0.109)	(0.153)	(0.119)	(0.149)	(0.117)	(0.176)
N	1,896	1,663	1,896	1,765	1,896	1,687
	$\geq$45000	<45000	$\geq$60	<60	$\geq$1000	<1000

$^{+}$ P < .10; * P < .05; ** P < .01

Notes: Population-weighted estimates are generated using data from 1980-2018 FARS (panel I) and UCR (panel II). All estimates include state and year fixed effects and my preferred set of control variables. Standard errors, clustered at the state level, are reported inside parentheses. Data are aggregated to a state-by-year panel using specific sets of counties. Column 1 uses counties with at least 45,000 population. Column 2 uses counties with less than 45,000 population. Column 3 uses counties with at least 60% of residents living in urban areas. Column 4 uses counties with less than 60% of residents living in urban areas. Column 5 uses counties with a weighted population density of at least 1000. Column 6 uses counties with a weighted population density of less than 1000.

Appendix Table 1.6: Estimated Effect of DUI Checkpoint Bans on Self-Reported Drunk Driving, BRFSS 1984-2018

	Drink Drive Yes/No				Drink Drive Frequency
	(1)	(2)	(3)	(4)	(5)
0 to 5 Year After	0.1135**	0.2689**	0.1175**	0.2546**	0.1629**
	(0.0214)	(0.0481)	(0.0313)	(0.0614)	(0.0581)
6 to 10 Years After	0.0798*	0.2006**	0.0740^{+}	0.1682^{+}	0.2215*
	(0.0318)	(0.0720)	(0.0428)	(0.0894)	(0.1061)
11+ Years After	0.0561	0.1451	0.1087	0.2497^{+}	0.4527*
	(0.0564)	(0.1282)	(0.0662)	(0.1391)	(0.1871)
N	4,221,419	4,221,419	2,157,106	2,157,106	78,412
Sample	Everyone	Everyone	Drinker	Drinker	Drunk Driver
Model	Probit	Logit	Probit	Logit	Poisson

$^{+}$ P < .10; * P < .05; ** P < .01

Notes: Survey-weighted estimates are generated using data from the 1984-2018 BRFSS. All estimates include state and year fixed effects and my preferred set of control variables and demographic controls (gender, race/ethnicity, age, and education). The sample for columns 3 and 4 is restricted to individuals who reported drinking at least one alcoholic beverage in the past 30 days. The sample for column 5 is restricted to individuals who reported drinking and driving at least once in the past 30 days. Odds ratio is reported for columns 1 to 4. Standard errors, clustered at the state level, are reported inside parentheses.

Appendix Table 1.7: Heterogeneous Treatment Effects by Age Group

	(1)	(2)	(3)
	< 30	30-39	40+
		Panel I: DUI Fatalities, FARS	
Policy Ban	0.107*	0.134**	0.132*
	(0.052)	(0.047)	(0.063)
		Panel II: DUI Arrests, UCR	
Policy Ban	0.150*	0.148*	0.122*
	(0.064)	(0.070)	(0.057)
		Panel III: Drunk Driving: Everyone, BRFSS	
Policy Ban	-1e-04	0.014^{+}	0.004^{+}
	(0.005)	(0.007)	(0.002)
Mean of DV	0.107	0.056	0.018
		Panel IV: Drunk Driving: Drinker, BRFSS	
Policy Ban	0.003	0.022^{+}	0.008^{+}
	(0.008)	(0.012)	(0.004)
Mean of DV	0.155	0.084	0.035
		Panel V: Drunk Driving Frequency, BRFSS	
Policy Ban	0.057	0.082*	0.035
	(0.049)	(0.041)	(0.054)
Mean of DV	0.459	0.427	0.549

$^{+}$ P < .10; * P < .05; ** P < .01

Notes: Weighted estimates are generated using data from the 1980-2018 FARS (panel I), 1980-2018 UCR (panel II), and 1984-2018 BRFSS (panels III to V). All estimates include state and time fixed effects, and my preferred set of control variables. In addition, all estimates in panels III to V include demographic controls (gender, race/ethnicity, age, and education). Standard errors, clustered at the state level, are reported inside the parenthesis.

4.2 Appendix for Chapter 2

4.2.1 Chapter 2. Appendix A. Proof

Proof of Prediction 1

Given that drivers from two different racial groups (Black and White) are violating each traffic violation, the probability of stopping someone for a traffic violation t is the probability of observing Black drivers violating t and stopping that driver or observing White drivers violating t and stopping that driver. Mathematically, I can express the probability of stopping someone for a traffic violation as follows:

$$
\begin{aligned}
P(\text{stop}|t=m) =& \underbrace{\phi_{mb}P(\text{stop}|t=m, r=b)}_{\substack{\text{Observe \& Stop} \\ \text{Violation m, Race b}}} + \underbrace{\phi_{mw}P(\text{stop}|t=m, r=w)}_{\substack{\text{Observe \& Stop} \\ \text{Violation m, Race w}}} \\
=& \phi_{mb}[1 - F(c_m^*|t=m, r=b)] + \phi_{mw}[1 - F(c_m^*|t=m, r=w)] \\
=& \phi_{mb} + \phi_{mw} - \phi_{mb}F(c_m^*|t=m, r=b) - \phi_{mw}F(c_m^*|t=m, r=w) \\
P(\text{stop}|t=o) =& \underbrace{\phi_{ob}P(\text{stop}|t=o, r=b)}_{\substack{\text{Observe \& Stop} \\ \text{Violation o, Race b}}} + \underbrace{\phi_{ow}P(\text{stop}|t=o, r=w)}_{\substack{\text{Observe \& Stop} \\ \text{Violation o, Race w}}} \\
=& \phi_{ob}[1 - F(c_o^*|t=o, r=b)] + \phi_{ow}[1 - F(c_o^*|t=o, r=w)] \\
=& \phi_{ob} + \phi_{ow} - \phi_{mb}F(c_o^*|t=o, r=b) - \phi_{ow}F(c_o^*|t=o, r=w)
\end{aligned}
\tag{A.1}
$$

Taking the derivative of Equation (A.1) with respect to c_m, I can mathematically show prediction 1.

$$
\begin{aligned}
\frac{dP(\text{stop}|t=m)}{dc_m} =& -\phi_{mb}\frac{dF(c_m^*|t=m,r=b)}{dc_m} - \phi_{mw}\frac{dF(c_m^*|t=m,r=w)}{dc_m} \\
=& -\phi_{mb}\frac{dF(c_m^*|t=m,r=b)}{dc_m^*}\cdot\frac{dc_m^*}{dc_m} - \phi_{mw}\frac{dF(c_m^*|t=m,r=w)}{dc_m^*}\cdot\frac{dc_m^*}{dc_m} \\
=& -\phi_{mb}f(c_m^*|t=m,r=b)\cdot\frac{d(c_m+E(U_2))}{dc_m} \\
& -\phi_{mw}f(c_m^*|t=m,r=w)\cdot\frac{d(c_m+E(U_2))}{dc_m} \\
=& -\phi_{mb}f(c_m^*|t=m,r=b)(1+\frac{dE(U_2)}{dc_m}) \\
& -\phi_{mw}f(c_m^*|t=m,r=b)(1+\frac{dE(U_2)}{dc_m}) \\
=& \sum_r -\underbrace{\phi_{mr}}_{>0}\underbrace{f(c_t^*|t=m,r)}_{>0}\underbrace{(1+\underbrace{\frac{dE(U_2)}{dc_m}}_{\in(-1,0)})}_{\in(0,1)} < 0 \qquad \blacksquare
\end{aligned}
$$

$$
\begin{aligned}
\frac{dP(\text{stop}|t=o)}{dc_m} =& -\phi_{ob}\frac{dF(c_o^*|t=o,r=b)}{dc_m} - \phi_{ow}\frac{dF(c_o^*|t=o,r=w)}{dc_m} \\
=& -\phi_{ob}\frac{dF(c_o^*|t=o,r=b)}{dc_m^*}\cdot\frac{dc_o^*}{dc_m} - \phi_{ow}\frac{dF(c_o^*|t=o,r=w)}{dc_m^*}\cdot\frac{dc_o^*}{dc_m} \\
=& -\phi_{ob}f(c_o^*|t=o,r=b)\cdot\frac{d(c_o+E(U_2))}{dc_m} \\
& -\phi_{mw}f(c_o^*|t=o,r=w)\cdot\frac{d(c_o+E(U_2))}{dc_m} \\
=& -\phi_{ob}f(c_o^*|t=o,r=b)(\frac{dE(U_2)}{dc_m}) - \phi_{mw}f(c_m^*|t=o,r=b)(\frac{dE(U_2)}{dc_m}) \\
=& \sum_r -\underbrace{\phi_{mr}}_{>0}\underbrace{f(c_t^*|t=m,r)}_{>0}(\underbrace{\frac{dE(U_2)}{dc_m}}_{\in(-1,0)}) > 0 \qquad \blacksquare
\end{aligned}
$$

Proof of Prediction 2

To make this prediction, I first find the probability of stopping a Black driver, defined as the probability of observing a Black driver committing a minor traffic violation and stopping them or the probability of observing a Black driver committing other traffic violation and stopping them.

$$
\begin{aligned}
P(\text{stop}|r=b) =& \underbrace{\phi_{mb} P(\text{stop}|t=m, r=b)}_{\substack{\text{Observe \& Stop} \\ \text{Violation m, Race b}}} + \underbrace{\phi_{ob} P(\text{stop}|t=o, r=b)}_{\substack{\text{Observe \& Stop} \\ \text{Violation o, Race b}}} \\
=& \phi_{mb}[1 - F(c^*_m|t=m, r=b)] + \phi_{ob}[1 - F(c^*_o|t=o, r=b)] \\
=& \phi_{mb} + \phi_{ob} - \phi_{mb} F(c^*_m|t=m, r=b) - \phi_{ob} F(c^*_o|t=o, r=b)
\end{aligned}
$$

Taking the derivative of this probability function with respect to c_m, I can derive the following:

$$
\begin{aligned}
\frac{dP(\text{stop}|r=b)}{dc_m} =& -\phi_{mb} \frac{dF(c^*_m|t=m, r=b)}{dc_m} - \phi_{ob} \frac{dF(c^*_o|t=o, r=b)}{dc_m} \\
=& -\phi_{mb} \frac{dF(c^*_m|t=m, r=b)}{dc^*_m} \cdot \frac{dc^*_m}{dc_m} - \phi_{ob} \frac{dF(c^*_o|t=o, r=b)}{dc^*_o} \cdot \frac{dc^*_o}{dc_m} \\
=& -\phi_{mb} f(c^*_m|t=m, r=b) \cdot \frac{d(c_m + E(U_2))}{dc_m} \\
& - \phi_{ob} f(c^*_o|t=o, r=b) \cdot \frac{d(c_o + E(U_2))}{dc_m} \\
=& -\phi_{mb} f(c^*_m|t=m, r=b)(1 + \frac{dE(U_2)}{dc_m}) - \phi_{ob} f(c^*_o|t=o, r=b) \frac{dE(U_2)}{dc_m}
\end{aligned}
\tag{A.2}
$$

Unlike in Prediction 1, where the sign of the derivative is clearly one direction, the sign of the derivative in Prediction 2 is more complex and cannot be determined directly. This is because in Prediction 2, the stopping probability for Black drivers depends on many parameter values. To determine the conditions under which this derivative is positive or negative, I derive the optimal condition as follows:

$$\begin{aligned}
&\frac{dP(\text{stop}|r=b)}{dc_m} > 0 \\
\Rightarrow & -\phi_{mb} f(c_m^*|t=m,r=b)(1+\frac{dE(U_2)}{dc_m}) - \phi_{ob} f(c_o^*|t=o,r=b)(\frac{dE(U_2)}{dc_m}) > 0 \\
\Rightarrow & -\phi_{mb} f(c_m^*|t=m,r=b) - \phi_{mb} f(c_m^*|t=m,r=b)\frac{dE(U_2)}{dc_m} > \phi_{ob} f(c_o^*|t=o,r=b)\frac{dE(U_2)}{dc_m} \\
\Rightarrow & -\phi_{mb} f(c_m^*|t=m,r=b) > \phi_{ob} f(c_o^*|t=o,r=b)\frac{dE(U_2)}{dc_m} + \phi_{mb} f(c_m^*|t=m,r=b)\frac{dE(U_2)}{dc_m} \\
\Rightarrow & -\phi_{mb} f(c_m^*|t=m,r=b) > [\phi_{ob} f(c_o^*|t=o,r=b) + \phi_{mb} f(c_m^*|t=m,r=b)]\frac{dE(U_2)}{dc_m} \\
\Rightarrow & -\underbrace{\frac{\phi_{mb} f(c_m^*|t=m,r=b)}{\phi_{ob} f(c_o^*|t=o,r=b) + \phi_{mb} f(c_m^*|t=m,r=b)}}_{>0} > \underbrace{\frac{dE(U_2)}{dc_m}}_{<0} \\
\Rightarrow & \frac{\phi_{mb} f(c_m^*|t=m,r=b)}{\phi_{ob} f(c_o^*|t=o,r=b) + \phi_{mb} f(c_m^*|t=m,r=b)} > |\frac{dE(U_2)}{dc_m}| \qquad \blacksquare
\end{aligned}$$

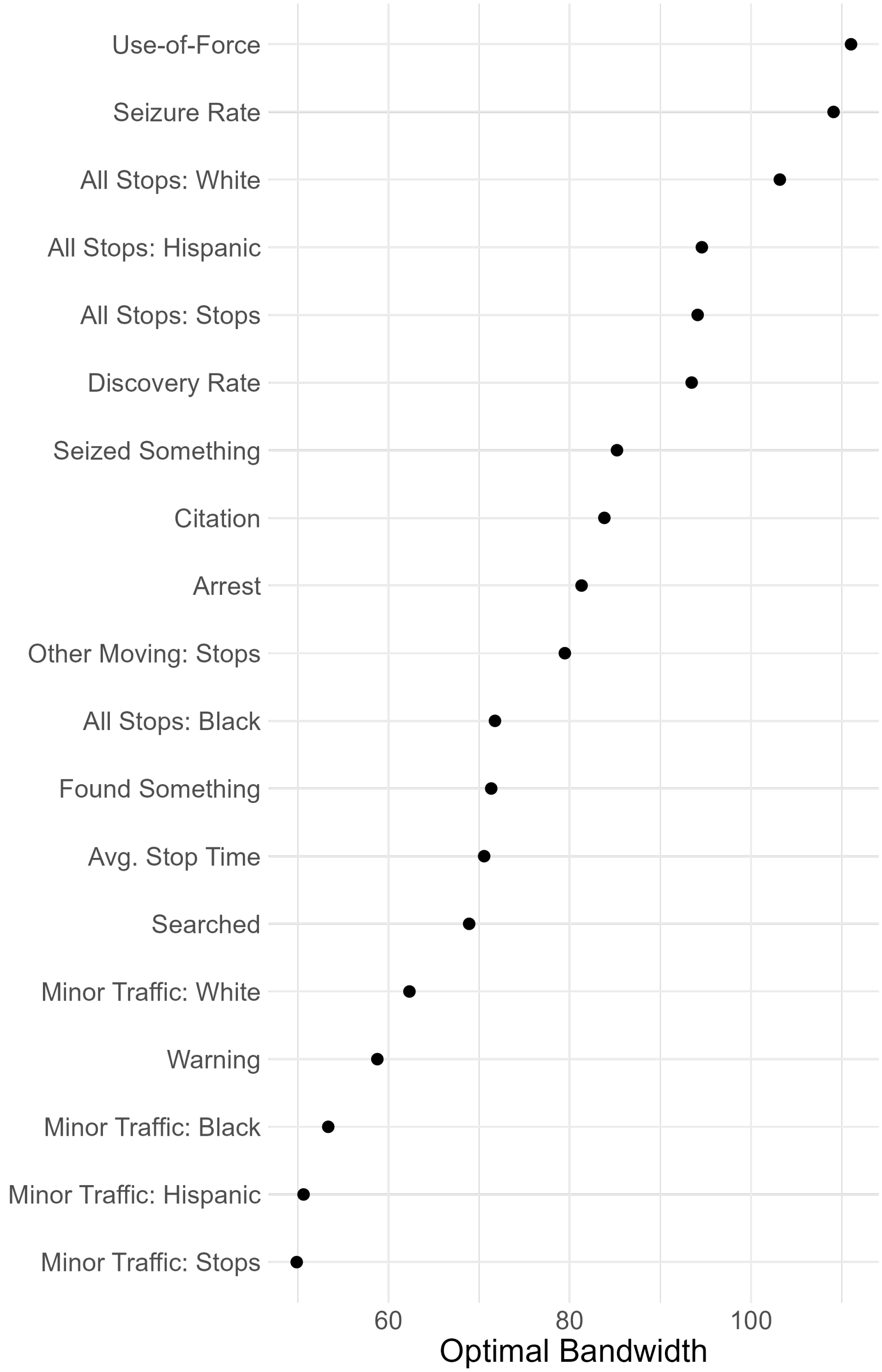
This result implies that the sign of the change in the total number of drivers getting stopped depends on ϕ_{mb}, ϕ_{ob}, $f(c_o^*|t=m,r=b)$ and $f(c_o^*|t=o,r=b)$. To find how these parameters, I will find the cross derivatives to determine the sign of each:

$$\begin{aligned}
\frac{dP(\text{stop}|r=b)}{dc_m d\phi_{mb}} &= -f(c_m^*|t=m,r=b)(1+\frac{dE(U_2)}{dc_m}) < 0 \\
\frac{dP(\text{stop}|r=b)}{dc_m d\phi_{ob}} &= -f(c_o^*|t=o,r=b)\frac{dE(U_2)}{dc_m} > 0 \\
\frac{dP(\text{stop}|r=b)}{dc_m df(c_m^*|t=m,r=b)} &= -\phi_{mb}(1+\frac{dE(U_2)}{dc_m}) < 0 \\
\frac{dP(\text{stop}|r=b)}{dc_m df(c_o^*|t=o,r=b)} &= -\phi_{ob}\frac{dE(U_2)}{dc_m} > 0
\end{aligned}$$

The cross derivatives above imply that the change in the probability that the officer stops a Black driver will more likely be negative (in sign) as ϕ_{mb} and $f(c_m^*|t=m,r=b)$ increases. On the other hand, the cross derivatives also suggest that this change in the probability will more likely be positive as ϕ_{ob} and $f(c_o^*|t=m,r=b)$ increases.

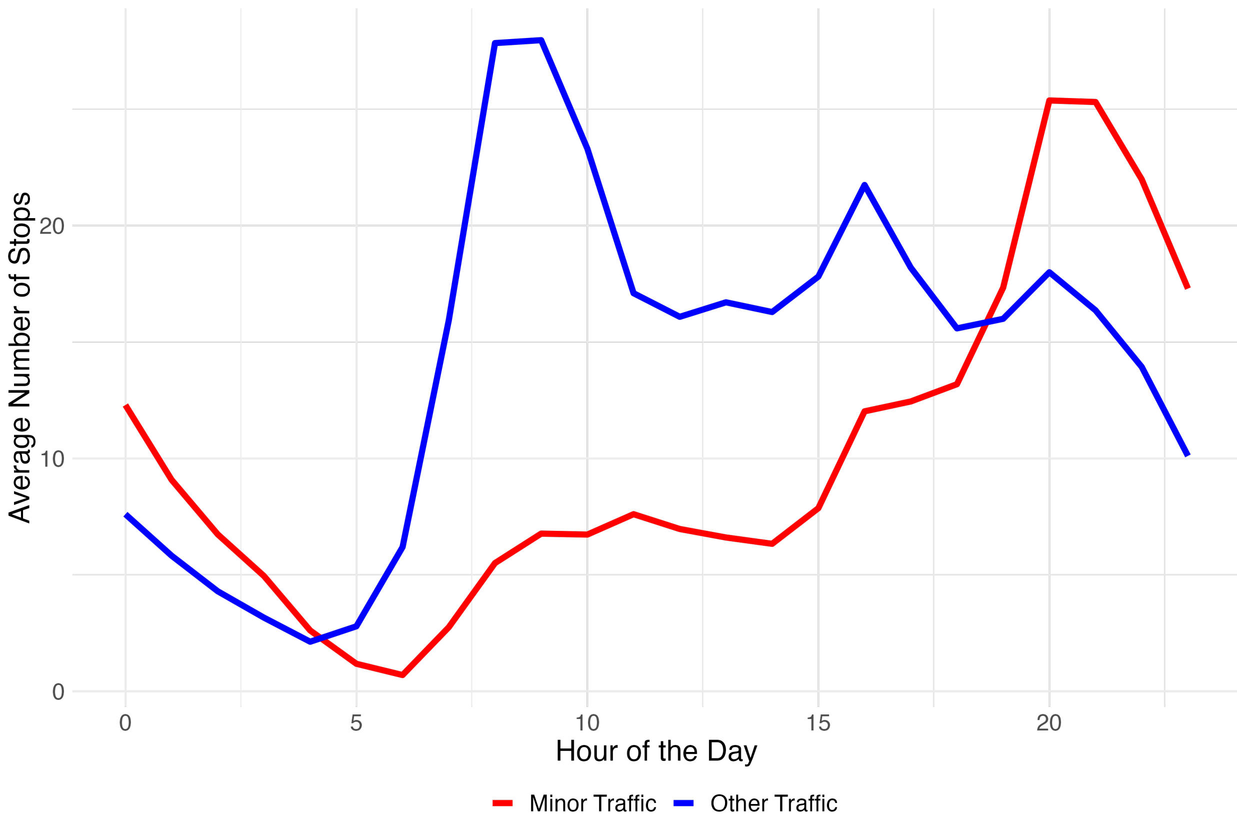
4.2.2 Chapter 2. Appendix B. Descriptive Figures & Tables

Appendix Figure 2.B.1: Optimal RDiT Bandwidth by Outcome



Notes: This figure plots the MSE-optimal bandwidth for each RIPA outcome I examined. The optimal bandwidth is determined using procedures laid out by Calonico et al. (2015) and R package `rdrobust`.

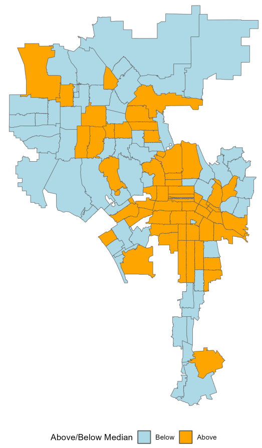
Appendix Figure 2.B.2: Traffic Stops by Hour



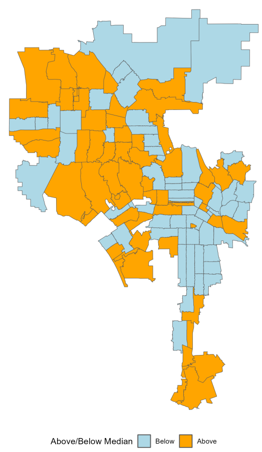
Notes: The average number of stops by each hour of the day is shown. Minor Traffic is defined as any equipment or non-moving violations. Other traffic is defined as any other traffic stops.

Appendix Figure 2.B.3: Spatial Distribution of Accidents & Stops by Zip Code

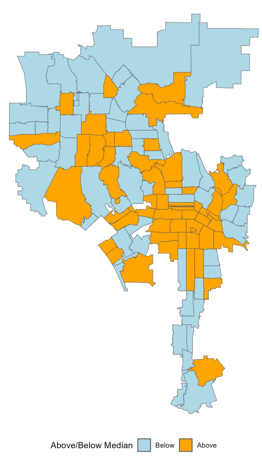
(a) Accidents per Capita



(b) Percent of Stops that are Minor Infraction

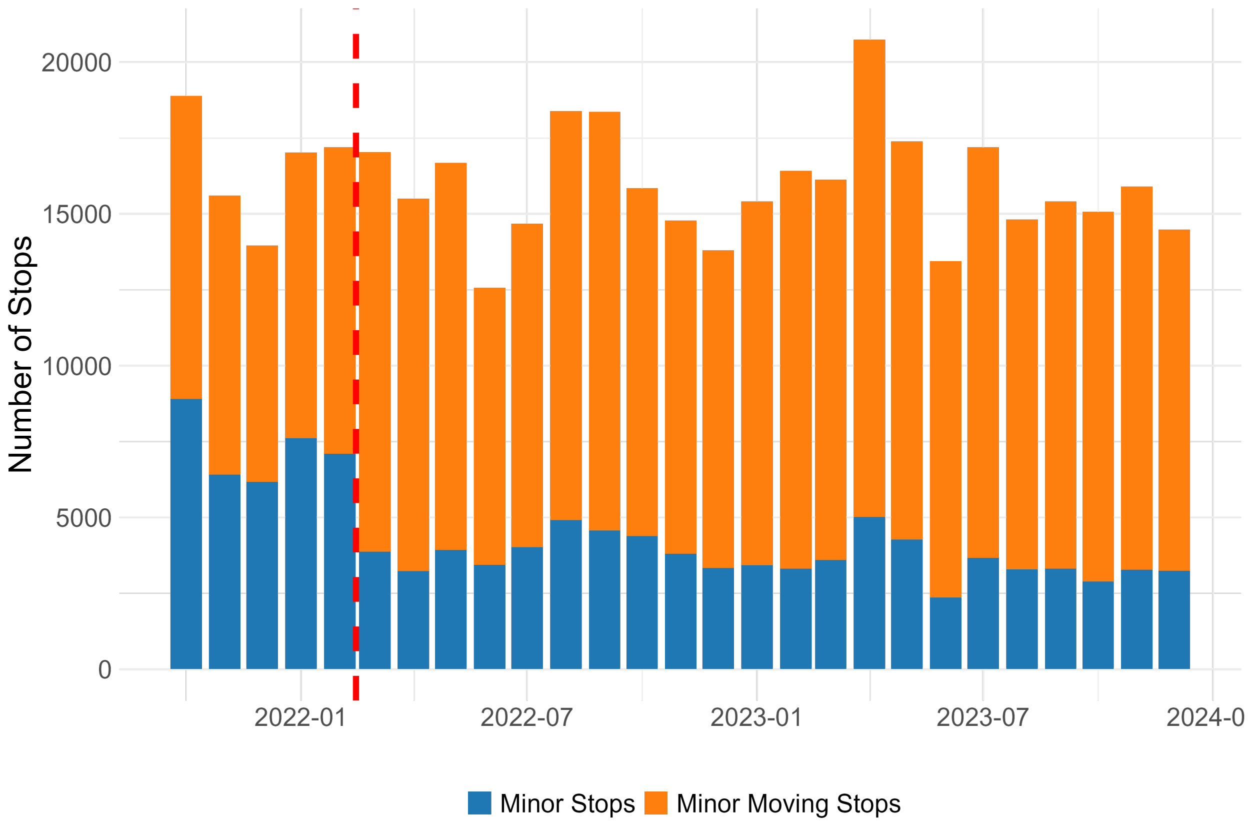


(c) Stops Involving Racial Minorities



Notes: These figures show whether each zip code is above or below the median accident per capita (panel a), stops for minor infractions per capita (panel b), and any stops involving racial minorities per capita (panel c). I use the pre-treatment window (July 2018 to Feb 2022) to measure the total count.

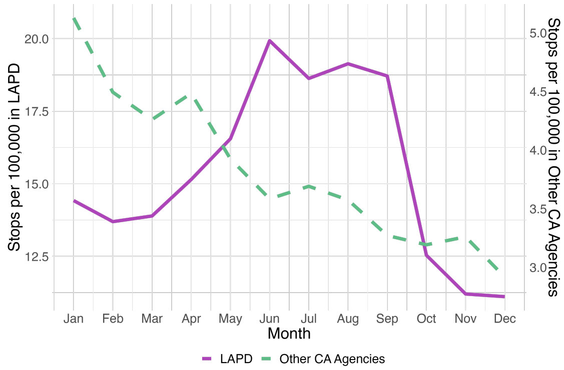
Appendix Figure 2.B.4: Time Series Plot: Number of Stops



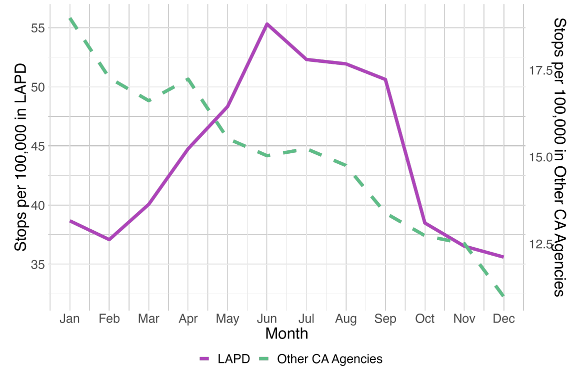
Notes: This figure shows the total monthly counts of stops made by the LAPD.

Appendix Figure 2.B.5: Monthly Traffic Stop

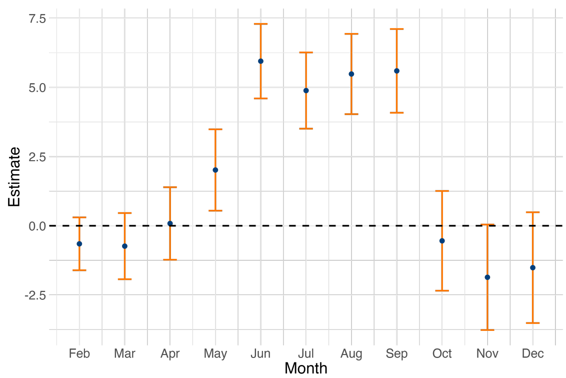
(a) Raw Trend: Minor Traffic Stops



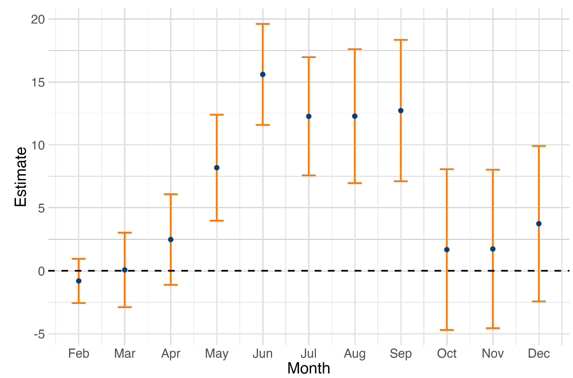
(b) Raw Trend: All Stops



(c) Monthly Coefficients Interacted w/ LAPD: Minor Traffic Stops



(d) Monthly Coefficients Interacted w/ LAPD: All Stops



Notes: Panels (a) and (b) present the aggregate count of traffic stops by month. Panels (c) and (d) present the point estimate and 95 percent confidence intervals generated using standard errors clustered at the agency level for the interaction between monthly coefficients (relative to January) and the dummy variable for LAPD. In panels (c) and (d), estimates are estimated at the daily level and control for daily temperature, precipitation, vehicle miles traveled, and agency and day-of-the-week fixed effects.

Appendix Table 2.B.1: Offense Classified as Minor Traffic (Non-Moving & Equipment) Violations

Code	Offense	Code	Offense
25063	SHO ON VEH/GIV PO FLS DOC	54443	FRONT RUNNING LAMPS VIOL
25107	PASS FALSE CLEAN AIR STKR	54444	IMPROPER: UNDOC VESSEL
42123	DRNK ALC/MARIJ AS PASSNGR	54451	VEHICLE HORN VIOLATION
42125	ALC/MARIJ AS PASSENGR HWY	54457	UNLAW DIRCT DRV OPR/HWY
44054	IMPERSONATE CHP OFFICER	54458	OPR VEH/ETC W/O LIC TYP
48054	FALSE REP THEFT TO DECEIV	54468	VEH BRK SYS VIOL:MTR FAIL
54013	UNLAWFUL OPERATION OF VEH	54469	HIRE VEH/ETC:ID DSPL VIOL
54014	HEADLAMP:OPR/AMT/SIZE:VIO	54471	TOW TRK/ETC ID INFO VIOL
54015	WINDOW INSTAL/ETC MAT VIO	54473	UNREG CA BASED VEHICLE
54016	FAIL NOTE DMV:OWN TRANSFR	54474	IDENT PLATE DISPLAY VIOL
54018	NO PROOF:FINANCE RESP:VEH	54478	TOW TRUCK TAILLAMP VIOL
54090	POSS OPEN CONTAINER:DRIVE	54479	TOW TRK TAIL/STOP LMP VIO
54099	NO REG:VEH/TRAILER/ETC	54480	HEADLAMP VIOLATION
54101	FAIL PROVE FIN RSP:PO REQ	54483	FAIL NOTE DMV:VEH SEL/ETC
54102	NO EVID:ID/INS/ETC:ACCDNT	54493	INOP SIGNL LAMP:ARM SIGNL
54109	FAIL MAINT VEH LITE EQUIP	54499	REAR VEH REFLECTOR VIOL
54110	FAIL MAINT LIC PLATE LAMP	54500	REF:VEH MFG/REG A/1-1-65
54116	INADEQUATE MUFFLERS	54504	FAIL OBEY TRAF LANE SIGN
54138	DEFECTIVE WINDSHIELD/ETC	54510	EXHAUST PIPE VIOLATION
54140	DRIVE W/O VALID LICENSE	54513	UNAUTH VEH IN FIRE AREA
54141	BIKE HEADLIGHT/ETC VIOL	54515	LAMP VOLT:85 PER REQ VOLT
54142	NO WINDSHIELDS	54516	TIRE N/CONFORMANCE W/REG
54143	BRAKE SYS CONDITION VIOL	54518	STOPLAMPS VIOL:N/VISIBLE
54144	STOPLAMPS VIOL:SPEC VEH	54528	OBSTRUCT OF LIC PLATE
54148	FAIL PROVIDE VEH REG:PO	54531	MISUSE INSTRUCTION PERMIT
54150	OPR VEH:VIOL LIC RESTRCTN	54533	SINGLE BEAM:PROPER ADJUST
54165	ILL MOD EXHAUST SYS:NOISE	54534	NOT EQUIPED W/SMOG DEVICE
54168	EXPIRED TABS/FAIL DISPLAY	54536	REAR PROJECTION VIOLATION
54171	VEH LAMPS/ETC COLOR VIOL	54540	IDENTIFICATION PLATE VIOL
54172	FOREIGN COM VEH:NO PERMIT	54543	FAIL COMPLY:MOUNTING REQ
54190	STORE OPEN/ETC ALC IN VEH	54545	CHILD 6- ALONE IN VEHICLE
54193	TAILLAMP VIOLATIONS	54548	TRESP W/VEH ETC:PUB GRNDS
54194	STOPLAMP VIOLATIONS	54549	METAL TIRE:EXCESS 6MPH
54195	NO LAMP/FLAG/ETC EXT LOAD	54552	OBST DRIVER VIEW/CONTROL
54204	WRONG COLOR:WINDO/ETC MAT	54553	SELL/USE UNAP LIGHT EQUIP
54205	TIRE TREAD DEPTH VIOL	54571	OPR VEH:WINDOW OBSTRUCTED
54206	FAIL REG FOREIGN VEH:CA	54572	TRNSP 10/MORE USED TIRES
54208	NO REGISTRATION IN VEH	54574	STOPLAMPS:VEH MUST HAVE
54211	LICENSE PLATE DISPLAY VIO	54584	PERSON FAIL TO PAY TOLL
54214	SPEC VEH FENDER/ETC VIOL	54586	OP MOTORSCTR:DRK HWY:LAMP
54216	LICENSEE POSS/ETC 1+ CDL	54587	OP MOTRSCTR:DRK HWY:R/REF
54221	PARK IN SPACE FOR DISABLE	54591	UNAUTH FLASHING BLU LIGHT
54222	AUXILIARY DRIVE LAMP VIOL	54594	DIFFUSED LIGHT:NO RED
54223	VEHICLE BUMPER VIOLATION	54604	LAMP REQUIREMENT VIOLATIO
54226	VEH SIDEVIEW MIRROR VIOL	54605	MOTORCYC H/LAMP:1REQ/2PRM
54227	LOST/ETC DMV REG/ETC VIOL	54608	FRM LABR VEH:WOUT SEATBLT
54229	FUEL TANK CAP VIOLATION	54612	ILLEGAL FLASHING LIGHTS
54230	MOTORCYCLE HEADLAMP VIOL	54614	OPR VEH:WINDOW OBSTRUCTED
54233	SAFETY GLAZING MATRL VIOL	54617	UNLAWFUL DISPLAY ID PLATE
54234	LICENS PLATE POSITION VIO	54618	INADEQUATE BRAKE SYSTEM

Appendix Table 2.B.1: Offense Classified as Minor Traffic (Non-Moving & Equipment) Violations *(continued)*

Code	Offense	Code	Offense
54300	GASTIGHT EXHAUST SYS VIOL	54619	FAIL TO DISPLAY WGT DECAL
54301	HITCH/ETC MOUNT VIOLATION	54626	OPR BIKE W/O BRAKES
54305	FAIL NOTE DMV ADD CHG:CDL	54628	FAIL UNLOCK LIMO:EMERGNCY
54307	BACKUP LAMPS VIOLATION	54644	DISPLAY LIC PLATES WRONG
54308	REARVIEW MIRROR VIOLATION	54645	DISPLY ONE LIC PLATE WRNG
54311	ABANDON VEHICLE ON HIWAY	54649	STOPLAMPS:VEH 2 REQUIRED
54313	NONRES:DRIVE W/O MED CERT	54657	NO REG:VEH/TRAILER/ETC
54314	SERVICE BRAKES VIOLATION	54663	NO PARK/STOP ETC FIRE LN
54316	VIOL VISBLTY REQ:TURN SIG	54666	UNREG COMM MTR VEH 10000+
54321	SELL/ETC UNLAWF EQUIPMENT	54667	COMMERC VEH WGHT FEES DUE
54324	EXCESSIVE EXHAUST VIOL	54670	EMPL ALLW DRIVER COMM VEH
54330	PARKING/ETC VIO:SPEC CIRC	54672	AUX LMPS NOT COVERD W/DRV
54331	MODIFIED VEH RIM HGT VIOL	54675	5TH WHL CONNECT DEVC REQ
54334	FAIL TO REP WT ALTER/ETC	54676	5TH WHL LOCKING DEVC REQ
54338	DISPLAY ALTERED LIC PLATE	54683	UNINTERRUPTED TOWS
54340	WINDSHIELD WIPER VIOL	54685	WARNING LAMPS ON TOW TRK
54345	DRIVE W/O COMERCL VEH LIC	54687	FLOOD LAMP EXCEEDS 75 FT
54346	PARK/ETC BY FIRE HYDRANT	54688	TEMP LIC PLATE NOT RPLCED
54347	DRIVE W/PARK LIGHTS ONLY	54691	BREAKAWAY DEVICE REQ VEH
54349	OPR CARRIER:NO ID NUMBER	54693	OPERATE AFTER NTC BY OFCR
54351	VENDING ON/NEAR FREEWAY	54695	VEH PNEU TIRES EXCD WIDTH
54354	NO PROOF $ RESP:ACCIDENT	54697	VEH/LOAD EXCEDS HGT 14 FT
54357	EQUIP MODIFY DEVICE VIOL	54698	BOOM/MAST NOT SECURE
54358	FT/COMPLY:INSPEC RULE/REG	54699	VEHICLE EXCEEDS LEN 40 FT
54359	REG/ETC:SMOG CERTS VIOL	54702	GROSS VEH WGHT VIOL COMBO
54376	VEH WITH UNLAWFUL LAMPS	54706	WRN SIG NOT RMVED W/O LD
54377	FOG TAILLAMPS VIOL	54707	IMPRP USE WRN LAMPS PILOT
54379	PARALLEL PARKING VIOL	54708	FLR/DISP COMP NM PILOT CR
54393	OPR UNDOC VES W/O NUMBERS	54709	FLR/DISP ID SIG PILOT CAR
54396	OPR VEH W/O LIC:PARK LOT	54712	NME/TRDMK ON FOR-HIRE VEH
54399	LIGHT DIMMER SWITCH VIOL	54713	VIS SIGN VEH LIQ PET/GAS
54407	LOWBEAM GLARE VIOLATION	54714	LGHT REFL TRK/TRLR FR/SID
54408	FRONT FENDER/ETC LAMP VIO	54715	LT REFL 30+TRK/TRLR FR/SD
54410	OPR/ETC GROSS POLLUTER	54716	FLR DISP REFL MAT ON TRLR
54412	O-O-S VEH:NO REG/SMOG CRT	54717	CLEAR/SIDE MARKR LAMP REQ
54413	OPR UNAUTH POLICE VEHICLE	54718	BRAK REQ FOR TRLERS 6000+
54414	LITE ON HWY:IMPAIR VISION	54720	INADEQUATE PARKING BRAKE
54415	BIKE HANDLEBAR ABV SHLDS	54721	VEH W/AIRBRKS PRES GA REQ
54436	FOG LAMPS VIOL	54722	AIR PRESS WARNING DEV REQ
54438	NO COMMERCIAL VEH LIC/ETC	54723	OBSTRUCT OF LIC PLATE
54441	LIC LOST/ETC:DESTROY ORIG	54726	LIC PLT:LVSTCK TRLR VIOL
54442	SIDE LAMPS VIOL	-	-

Notes: All traffic offense names and codes for violations that are considered minor traffic stops (equipment or non-moving) violations are listed. I define a specific violation as minor traffic stops if more than a third of the total offense was classified as non-moving or equipment violation by the police officer.

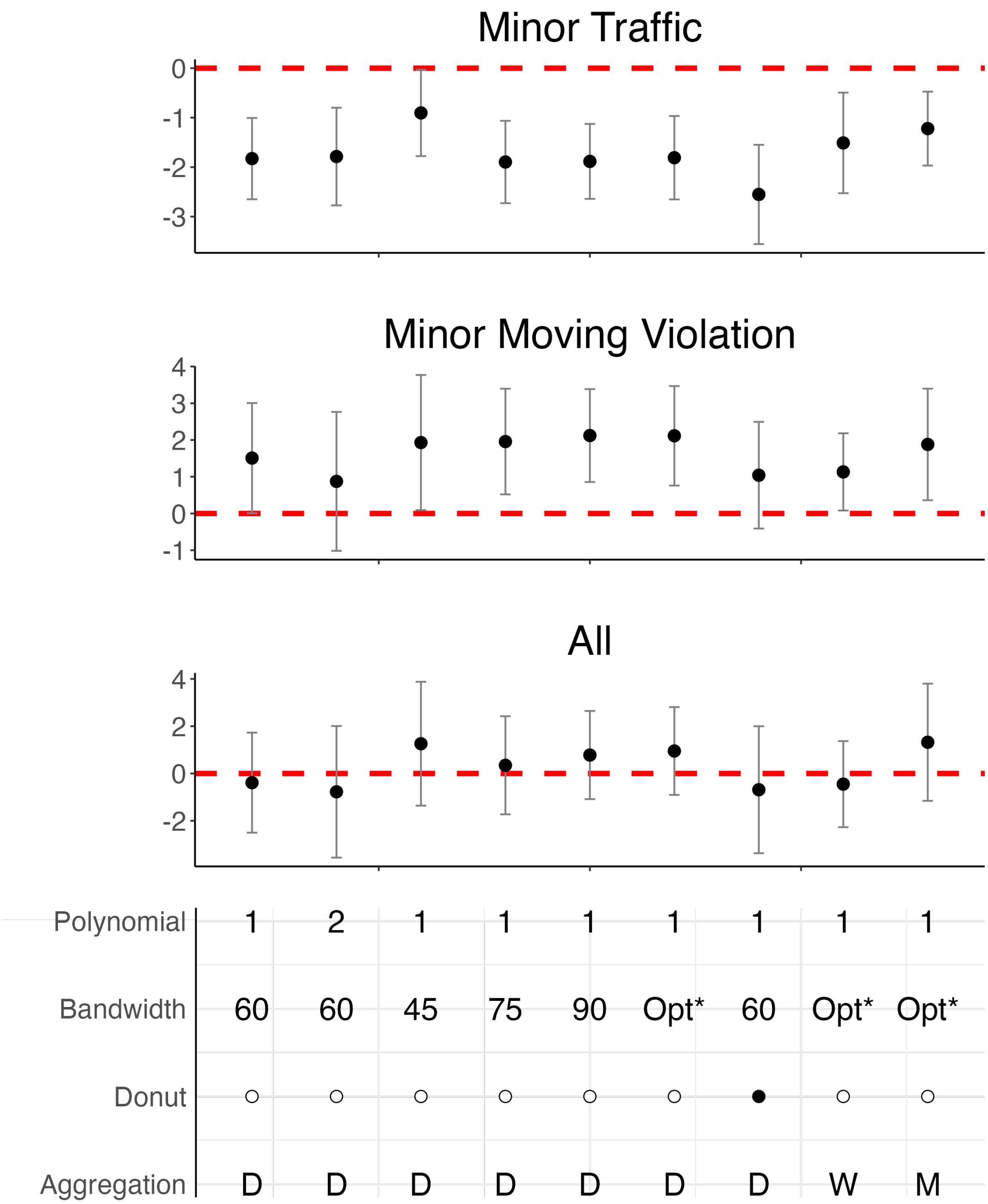
Appendix Table 2.B.2: Characteristics of Minor & Other Minor Moving Violation

	Minor Stops	Other Minor Moving
Share of All Stops	31.027	44.144
Percent Black	30.631	20.408
Percent Hispanic	54.598	48.557
Percent Warning	67.431	31.759
Percent Citation	24.855	61.739
Percent Searched	24.406	12.023
Contraband Discovery Rate	23.596	25.508

Notes: The sample is restricted to October 1, 2021, to February 28, 2022. Column 1 presents the characteristics of minor, likely pretextual stops, while Column 2 shows the characteristics of stops due to other minor moving violations. The share of all stops is calculated by dividing the number of each type of traffic stop by the total number of police-initiated stops. The percentage of Black and Hispanic individuals stopped is calculated by dividing the number of stops involving Black or Hispanic individuals by the total number of individuals stopped. The percentages of stops resulting in a warning, citation, or search are calculated by dividing the number of each outcome by the total number of stops. The contraband discovery rate is defined as the total number of contraband discoveries divided by the total number of searches performed.

4.2.3 Chapter 2. Appendix C. Supplemental Analysis

Appendix Figure 2.C.1: Robustness: RDiT Model Specification

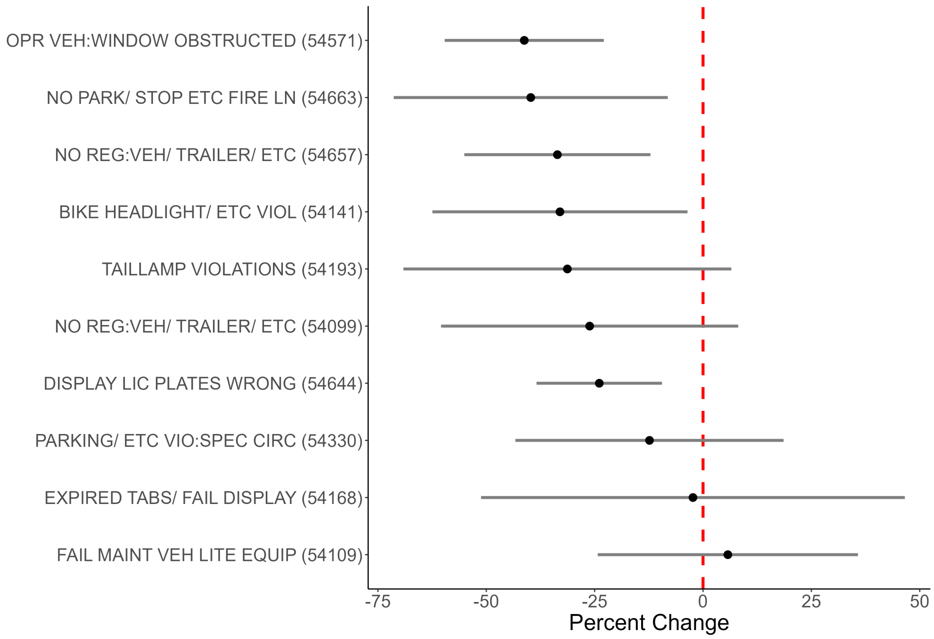


* Opt denotes optimal bandwidth estimated using procedures laid out by Calonico et al. (2015).

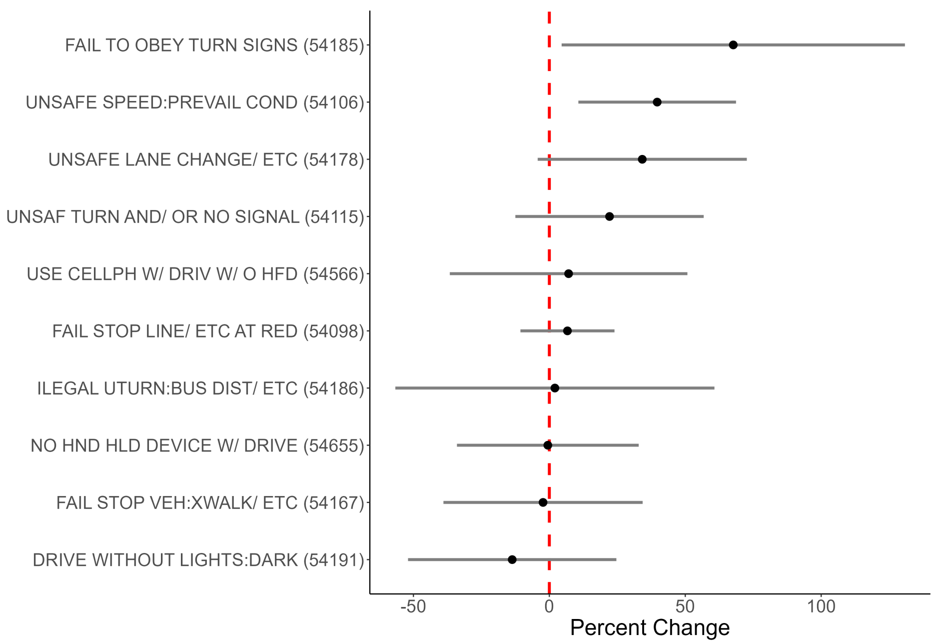
Notes: This figure shows robustness to the use of quadratic polynomial (2nd estimate), varying bandwidth, donut RD where I exclude the sample window of January 30 to March 30 (7th estimate), and aggregating the data to the weekly or monthly level (8th and 9th estimate). The first estimate (very left) shows my preferred estimate for comparison. The black dots represent my estimated treatment effect, and the bar plots represent the 95 percent confidence interval generated using Newey-West standard errors.

Appendix Figure 2.C.2: RDiT Estimate by Violation Type

(a) Minor Traffic Violation



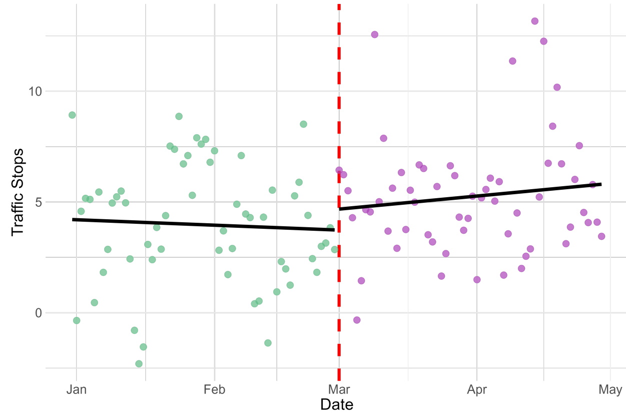
(b) Other Moving Violation



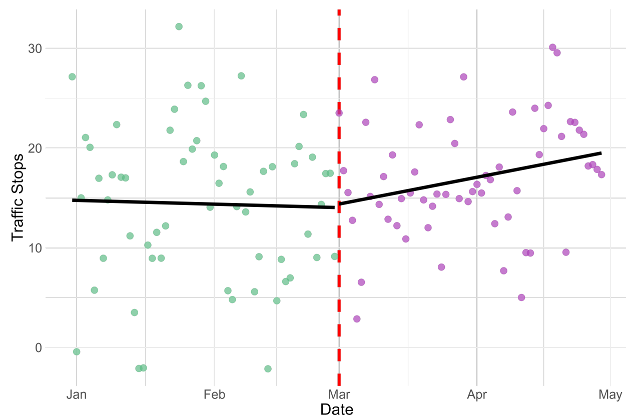
These figures show RDiT estimates for the 10 most common violation type. The bar plots represent 95 percent confidence intervals generated using Newey-West standard errors. To standardize comparisons across outcomes, the estimates are scaled by the pre-treatment mean, representing percentage changes.

Appendix Figure 2.C.3: RDiT Estimate: Police-Initiated Stops Using March 1, 2019 Treatment Date

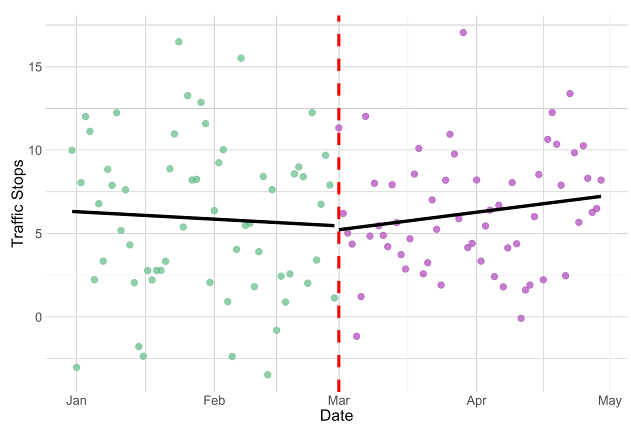
(a) Minor Traffic Violation



(b) All Stops



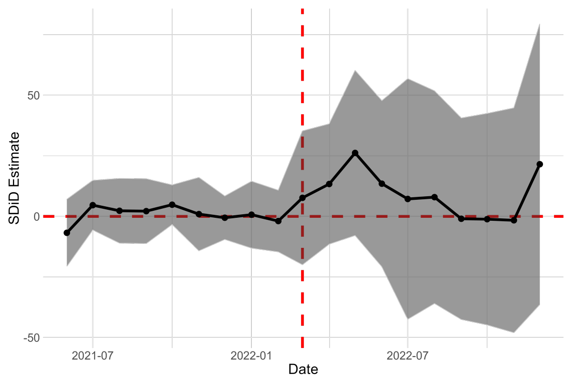
(c) Other Minor Traffic Violation



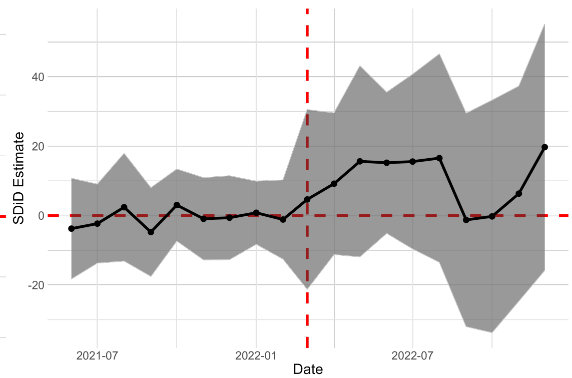
Notes: The residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022 are shown. For ease of interpretation, I added the residuals with the baseline mean of the outcome. The green color points represent my observation in the pre-treatment period. The purple color points represent my observation in the post-treatment period. The black line represents linear fit for both the pre- and post-treatment periods.

Appendix Figure 2.C.4: SDiD Estimate: Crime Comparing Los Angeles to the Rest of California

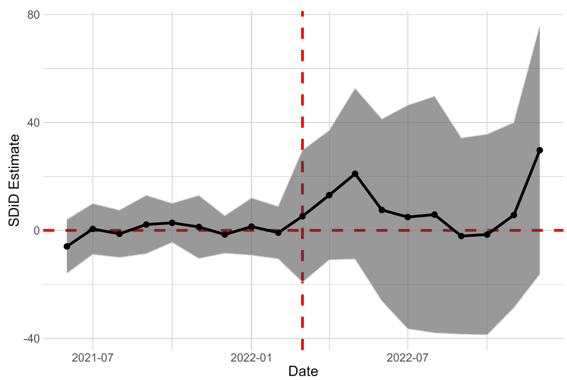
(a) All Part I: Agency-Level



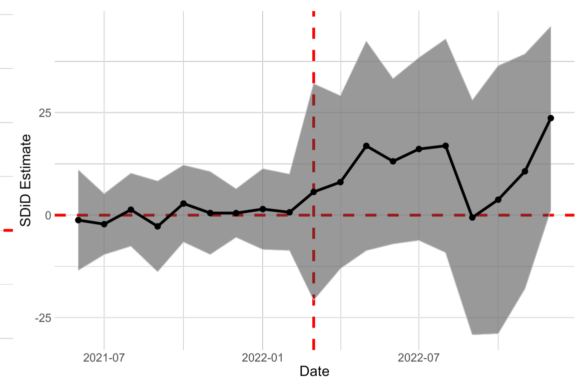
(b) All Part I: County-Level



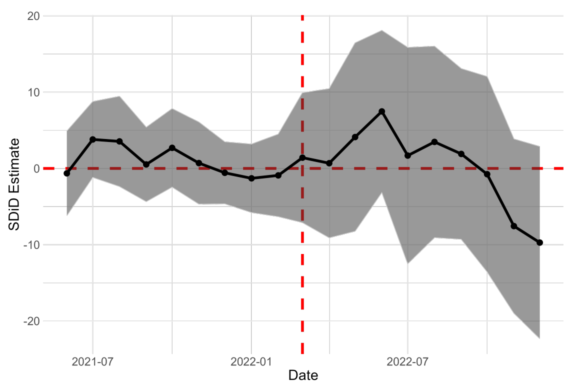
(c) Property: Agency-Level



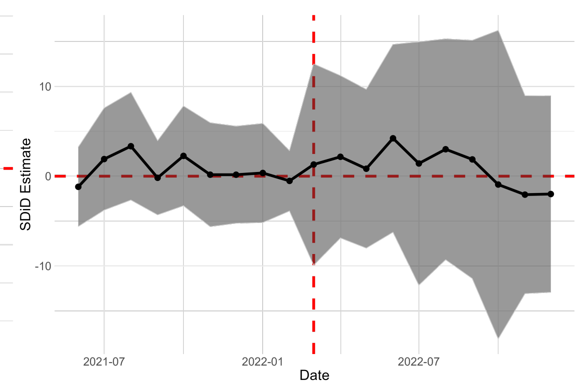
(d) Property: County-Level



(e) Violent: Agency-Level



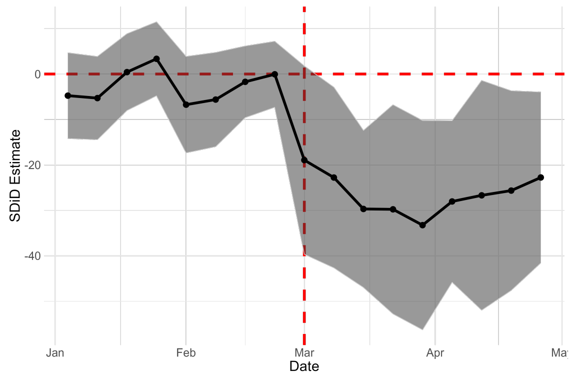
(f) Violent: County-Level



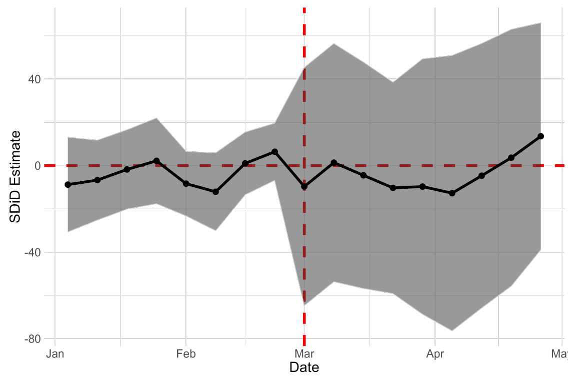
Notes: The synthetic difference-in-differences model is estimated. The left side of the panel uses agency-level data and compares LAPD (treated) to other large police agencies not in Los Angeles County. The right side of the panel uses county-level data and compares the county of Los Angeles (treated) to other large counties. The gray area represents 95 percent confidence intervals generated using placebo-based standard errors. The sample is restricted to June 2021 to December 2022.

Appendix Figure 2.C.5: SDiD Estimate: Stops per 100,000

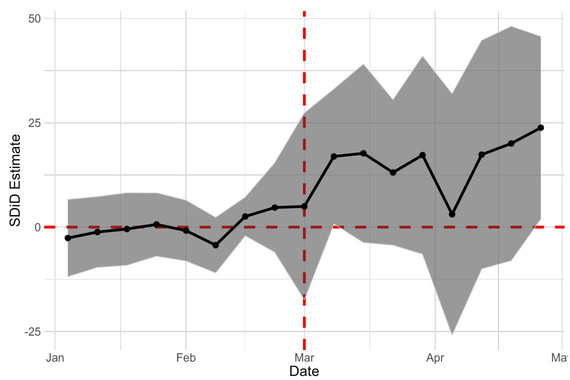
(a) Minor Traffic



(b) All Stops



(c) Other Minor Traffic Violations



Notes: The synthetic difference-in-differences model is estimated using weekly data from January 1 to April 30, 2022. The gray shaded area represents 95 percent confidence intervals generated using placebo-based standard errors.

Appendix Table 2.C.1: RDiT: Heterogenous Effects on Infractions Across Time & Space

	Night v. Day		Non-Minor Traffic Stops		Traffic Accidents	
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-0.158	1.474*	1.999	0.784	2.231	0.414
	(0.188)	(0.858)	(1.237)	(0.984)	(1.441)	(0.752)
Mean of DV	2.928	9.878	13.297	12.424	17.435	8.248
Sample	Night	Day	$\geq$ Median	$<$ Median	$\geq$ Median	$<$ Median

* P-val $<$ 0.1; ** P-val $<$ 0.05; *** P-val $<$ 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis. Columns (1) and (2) focus on nighttime stops defined as 9 pm to 6 am and daytime stops defined as 6 am to 9 pm, respectively. Columns (3) and (4) split the sample by whether the zip code had pre-treatment (July 2018 to February 2022) total minor moving stops per capita that were above or below the median. Columns (5) and (6) follow the same logic as columns (3) and (4) but split the sample by traffic accidents per capita. Zip code level traffic accident data is obtained from the city of Los Angeles open data portal.

Appendix Table 2.C.2: RDiT: Heterogeneous Effects by Zip Code

	Income		Minority Population	
	(1)	(2)	(3)	(4)
	Panel I: Minor Traffic Violation			
Post	-0.552**	-2.217***	-2.094***	-0.665**
	(0.243)	(0.634)	(0.617)	(0.304)
Mean of DV	2.409	7.727	7.334	2.897
	Panel II: All Other Violation			
Post	1.112	1.429	1.911*	0.360
	(0.992)	(1.192)	(1.101)	(1.011)
Mean of DV	9.494	14.974	14.209	10.524
Sample	$\geq$ Median	$<$ Median	$\geq$ Median	$<$ Median

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis. Columns (1) and (4) split the sample by whether the zip code's 2022 median household income is above or below the median. Columns (3) and (4) follow the same logic as columns (1) and (2) but split the sample by 2022 share of Hispanic and Black population. Both median household income and population are collected from the 2020 census.

Appendix Table 2.C.3: RDiT: Number of People Stopped by Race, Other Minor Moving Traffic Stops

	Black	Hispanic	White
	(1)	(2)	(3)
	Panel I: RDiT		
Post	1.152	1.258	1.118
	(2.380)	(0.892)	(1.049)
	Panel II: Differences in Discontinuities		
Post*LAPD	0.416	0.686	0.504
	(2.438)	(0.932)	(1.079)
Mean of DV	21.948	8.977	6.738

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis.

Appendix Table 2.C.4: RDiT: Heterogeneous Treatment Effects on All Stops By Pre-Treatment Stops Involving Racial Minorities

	All		Black		Hispanic		White	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-1.453	0.783	-15.302***	0.165	-3.122	0.409	0.533	0.850
	(1.655)	(1.075)	(5.836)	(2.969)	(1.985)	(1.159)	(1.775)	(1.453)
Mean of DV	26.771	11.486	82.657	32.399	31.753	15.634	16.598	8.580
Sample	$\geq$ Median	$<$ Median	$\geq$ Median	$<$ Median	$\geq$ Median	$<$ Median	$\geq$ Median	$<$ Median

* P-val $<$ 0.1; ** P-val $<$ 0.05; *** P-val $<$ 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis. The sample is split by whether each zip code's pre-treatment (July 2018 to February 2022) sum of racial minorities stopped per 100,000 is above or below the median.

Appendix Table 2.C.5: RDiT: Contraband Found by Types

	Any	Firearms	Drugs	Other Contraband
	(1)	(2)	(3)	(4)
	Panel I: RDiT			
Post	-0.308**	-0.031	-0.185	-0.091
	(0.147)	(0.077)	(0.129)	(0.105)
	Panel II: Differences in Discontinuities			
Post*LAPD	-0.319**	-0.011	-0.219*	-0.089
	(0.148)	(0.078)	(0.130)	(0.107)
Mean of DV	1.907	0.451	1.053	0.402

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis.

Appendix Table 2.C.6: RDiT: Traffic Stop Outcomes for Minor Traffic Stop Violations

	Warning	Citation	Searched	Found Some-thing	Seized Some-thing	Discovery Rate	Seizure Rate	Avg. Stop Time	Use-of-Force	Arrest
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel I: Minor Traffic Violation									
Post	-1.801***	-0.188	-0.814***	-0.237***	-0.012	-2.751	0.134	1.071	-0.137	-0.076***
	(0.338)	(0.190)	(0.132)	(0.055)	(0.009)	(2.959)	(0.757)	(1.977)	(0.629)	(0.021)
Mean of DV	5.351	1.452	1.954	0.453	0.032	22.999	1.607	32.631	2.083	0.180
	Panel II: Other Moving Infraction Violation									
Post	-0.505**	1.885**	-0.187**	-0.079*	-0.005	-0.694	-0.308	-0.306	0.785	-0.016
	(0.226)	(0.740)	(0.087)	(0.041)	(0.011)	(3.914)	(1.287)	(1.365)	(0.538)	(0.022)
Mean of DV	2.902	5.849	0.911	0.239	0.019	25.841	1.972	20.877	1.300	0.137

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis.

Appendix Table 2.C.7: RDiT: Traffic Stop Outcomes by Race

	Warning	Citation	Searched	Found Something	Seized Something	Discovery Rate	Seizure Rate	Avg. Stop Time	Use-of-Force	Arrest
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel I: Black										
Post	-12.034***	1.601	-5.445**	-1.487*	0.423*	1.657	2.038***	-0.255	1.094	-0.274
	(3.295)	(2.043)	(2.386)	(0.860)	(0.228)	(2.462)	(0.729)	(2.279)	(0.944)	(0.652)
Mean of DV	41.420	16.976	29.021	8.808	0.766	30.384	2.788	48.264	5.433	9.231
Panel II: Hispanic										
Post	-2.681***	1.553*	-1.608***	-0.459***	-0.006	-0.470	0.497	0.658	-2.432**	-0.190
	(0.803)	(0.875)	(0.374)	(0.170)	(0.051)	(1.924)	(0.617)	(1.448)	(1.199)	(0.161)
Mean of DV	11.861	8.161	8.054	2.047	0.242	25.293	2.985	41.086	10.617	2.986
Panel III: White										
Post	-0.841*	1.466	-0.003	0.149	0.039	4.334	0.527	-1.723	-0.058	-0.006
	(0.479)	(1.041)	(0.161)	(0.116)	(0.035)	(4.237)	(1.346)	(3.541)	(0.543)	(0.098)
Mean of DV	4.087	6.328	2.267	0.517	0.124	22.784	5.449	37.358	1.767	1.219

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Regression discontinuity in time model is estimated using daily data from January 1 to April 30, 2022. The outcome used is the residuals after regressing the outcome on daily temperature, precipitation, vehicle miles traveled, and day of the week, month, and year fixed effects using data from July 2018 to December 2022. All estimates include controls for the linear function of the running variable, where the slope varies between the pre-treatment and post-treatment periods. Newey-West standard errors are reported inside the parenthesis.

Appendix Table 2.C.8: TWFE: Reported Crime

	LAPD vs. Other Agencies			LA vs. Other Counties		
	All	Property	Violent	All	Property	Violent
	(1)	(2)	(3)	(4)	(5)	(6)
Post*LA	12.157	12.207	-0.050	6.271	4.871	1.401
	(22.151)	(22.494)	(3.358)	(21.825)	(20.855)	(2.770)
Rank Based P-Value	{0.441}	{0.412}	{0.962}	{0.668}	{0.729}	{0.534}
Mean of DV	276.529	207.754	68.776	244.913	195.107	49.805

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: The difference-in-differences model is estimated using monthly data. Columns 1 to 3 use an agency-by-month panel and compare the LAPD (treated) to other large police agencies not in Los Angeles County. Columns 4 to 6 use a county-by-month panel and compare the county of Los Angeles (treated) to other large counties. The standard errors estimated using placebo-based tests are reported inside the parenthesis. Rank-based p-value where the p-value is computed using a relative ranking of the average treatment effect is shown inside a curly bracket. The sample is restricted to June 2021 to December 2022, and the counterfactual includes the rest of the U.S.

Appendix Table 2.C.9: SDiD: Crime Clearance

	LAPD vs. Other Agencies			LA vs. Other Counties		
	All	Property	Violent	All	Property	Violent
	(1)	(2)	(3)	(4)	(5)	(6)
Post*LA	1.081	-0.055	0.981	-2.599	-0.505	-1.011
	(5.210)	(3.347)	(2.561)	(6.632)	(2.826)	(3.020)
Rank Based P-Value	{0.842}	{0.991}	{0.652}	{0.556}	{0.882}	{0.583}
Mean of DV	38.770	12.185	26.585	35.935	14.618	21.317

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: The difference-in-differences model is estimated using monthly data. Columns 1 to 3 use an agency-by-month panel and compare the LAPD (treated) to other large police agencies not in Los Angeles County. Columns 4 to 6 use a county-by-month panel and compare the county of Los Angeles (treated) to other large counties. The standard errors estimated using placebo-based tests are reported inside the parenthesis. Rank-based p-value where the p-value is computed using a relative ranking of the average treatment effect is shown inside a curly bracket. The sample is restricted to June 2021 to December 2022, and the counterfactual includes the rest of the U.S.

Appendix Table 2.C.10: TWFE: Reported Traffic Accidents

	LAPD vs. Other Agencies			LA vs. Other Counties		
	All	Speeding	Other Violation	All	Speeding	Other Violation
	(1)	(2)	(3)	(4)	(5)	(6)
Post*LA	-3.608	-1.000	-1.664	-2.886	-1.544	-0.817
	(2.986)	(1.389)	(2.398)	(11.169)	(3.533)	(6.464)
Rank Based P-Value	{0.203}	{0.211}	{0.259}	{0.691}	{0.574}	{0.815}
Mean of DV	28.483	7.215	12.943	89.045	27.586	43.774

* P-val $<$ 0.1; ** P-val $<$ 0.05; *** P-val $<$ 0.01

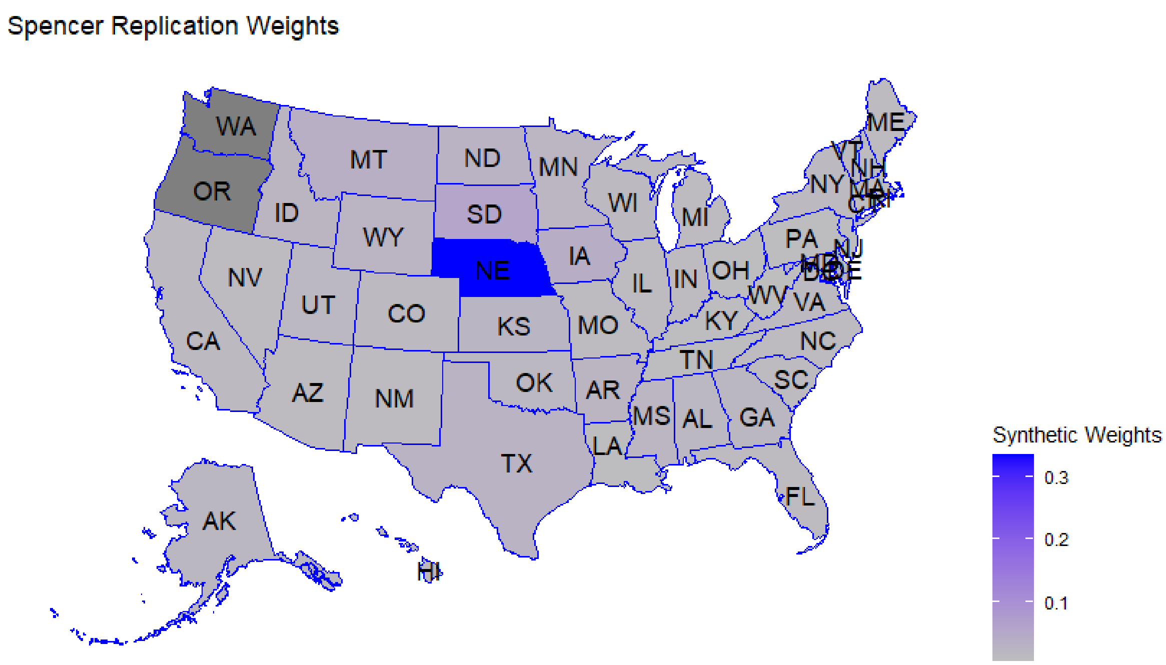
Notes: The difference-in-differences model is estimated using monthly data. Columns 1 to 3 use an agency-by-month panel and compare the LAPD (treated) to other large police agencies not in Los Angeles County. Columns 4 to 6 use a county-by-month panel and compare the county of Los Angeles (treated) to other large counties. The standard errors estimated using placebo-based tests are reported inside the parenthesis. Rank-based p-value where the p-value is computed using a relative ranking of the average treatment effect is shown inside a curly bracket. The sample is restricted to June 2021 to December 2023, and the counterfactual includes the rest of California.

4.3 Appendix for Chapter 3

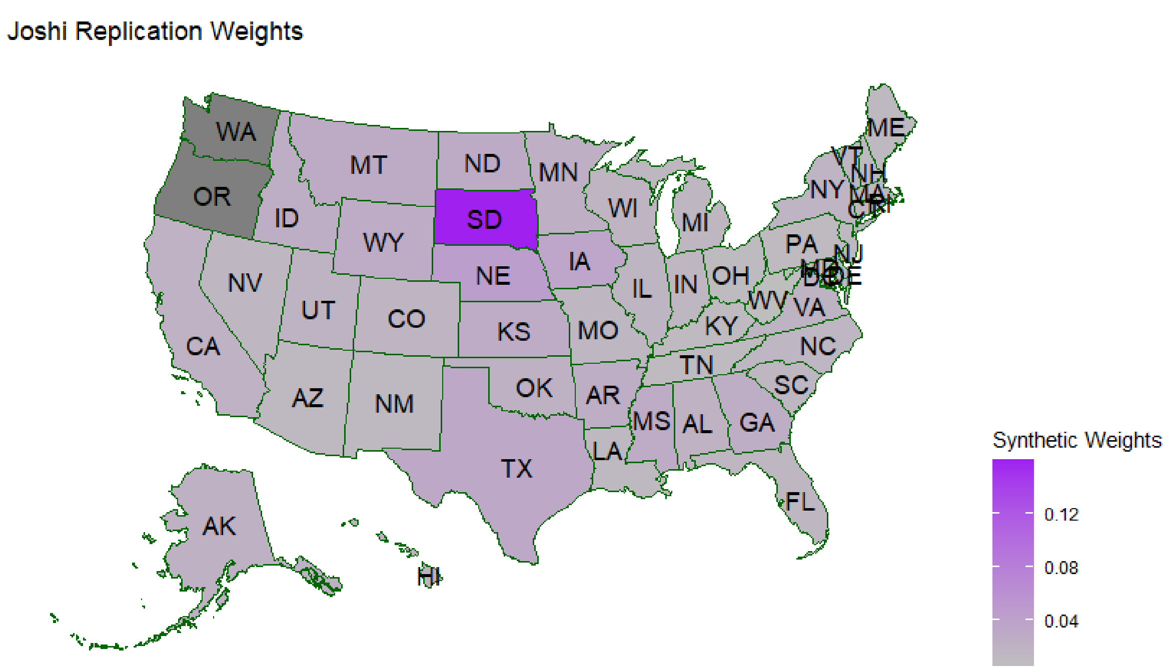
4.3.1 Chapter 3. Appendix Figures

Appendix Figure 3.1: Maps of the replication weights used in the studies of Spencer and Joshi et al.

(a) A map of the weights used in the replication of Spencer's work.

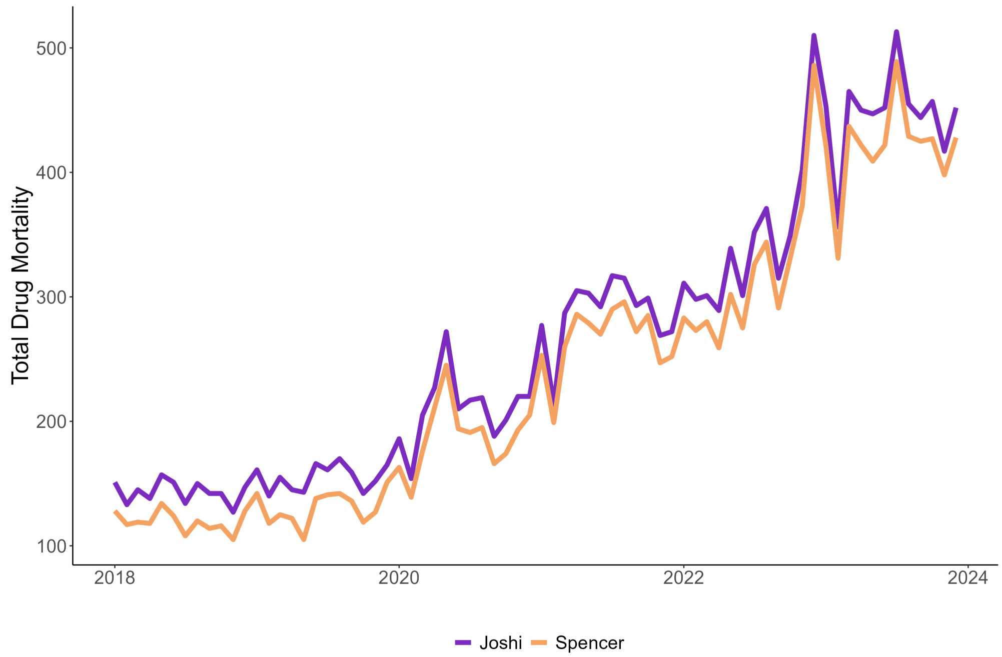


(b) A map of the weights used in the replication of Joshi et al. work.

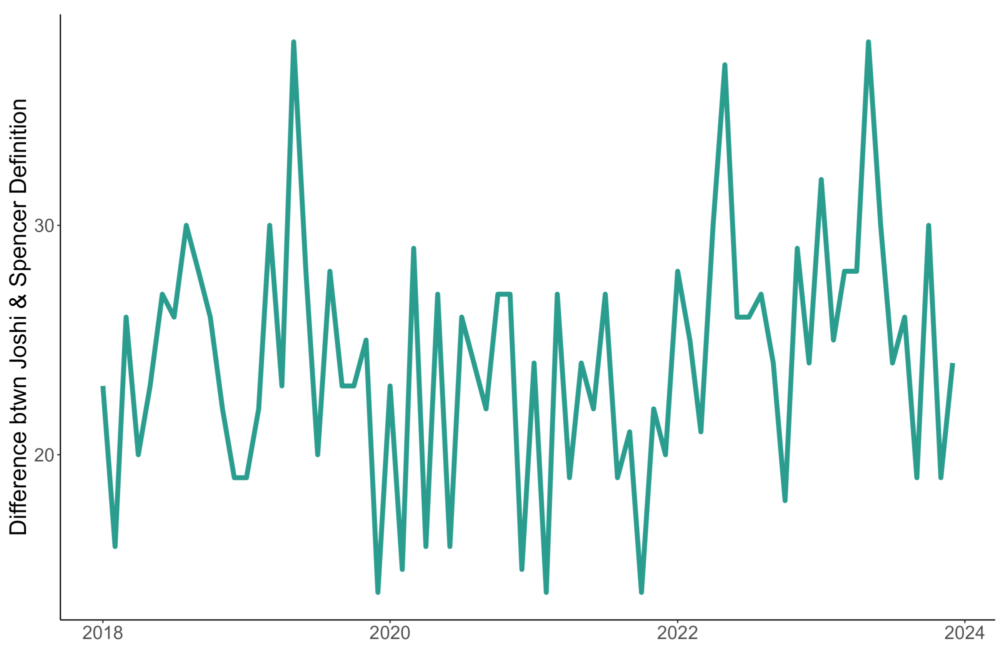


Appendix Figure 3.2: A chart of the overdoses in oregon and washington using different definitions of overdose.

(a) By Definition

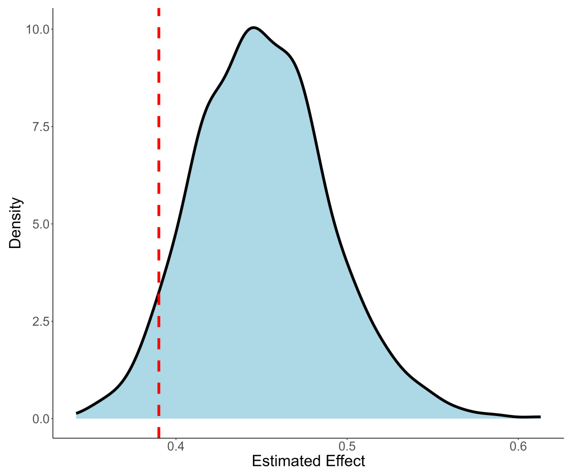


(b) Differences btwn Definition

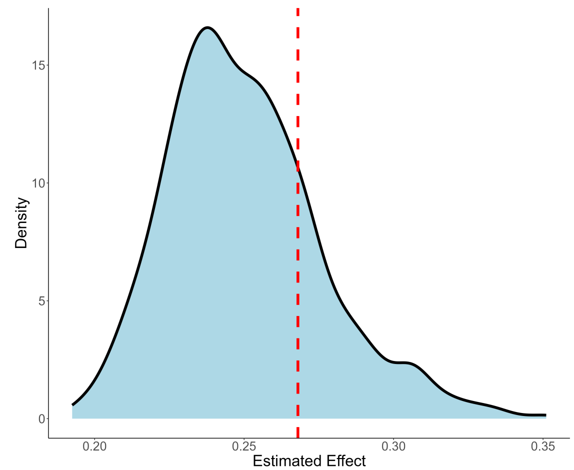


Appendix Figure 3.3: Simulated Effect (1000 times) of Random Imputation of Overdose

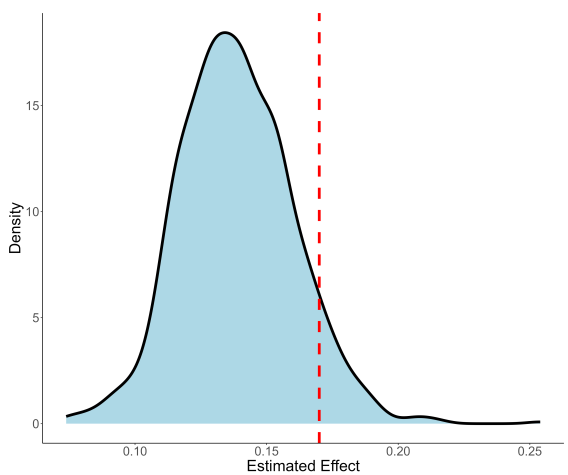
(a) Spencer Specification: Oregon



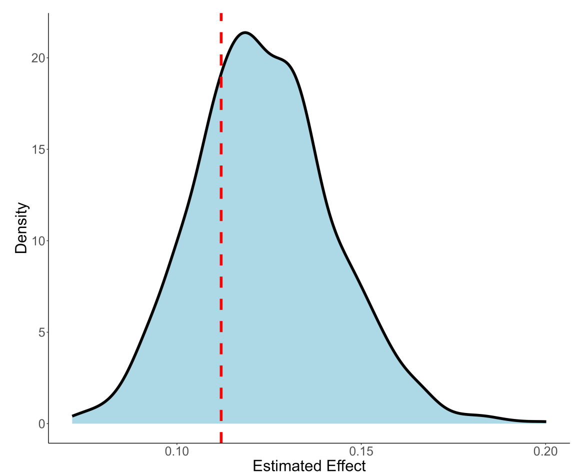
(b) Joshi et al. Specification: Oregon



(c) Spencer Specification: Washington



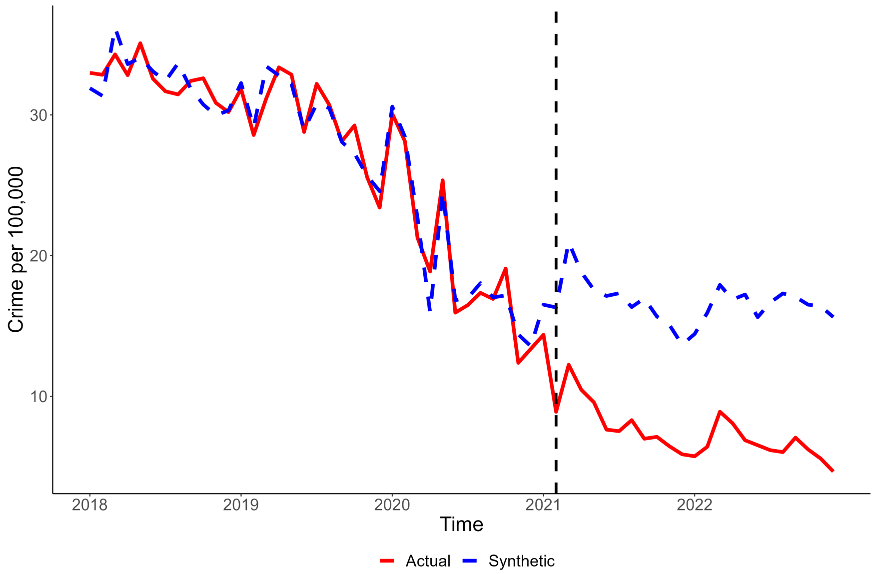
(d) Joshi et al. Specification: Washington



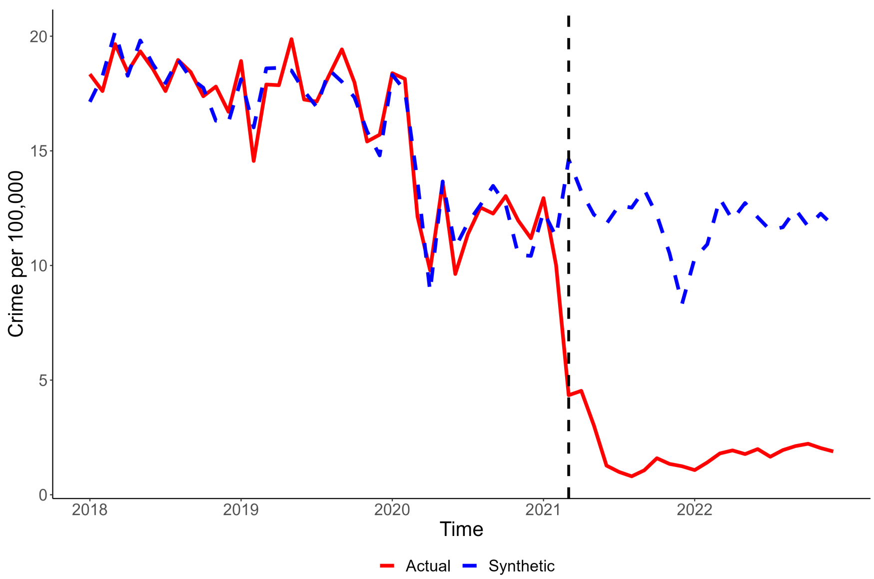
Notes: These figures show simulated ATEs for 1,000 different imputed values for state-month where fewer than 9 overdoses were reported. The red dashed line shows the average treatment effect reported by Spencer (2023) or Joshi et al. (2023).

Appendix Figure 3.4: NIBRS Estimate, Arrests

(a) Oregon



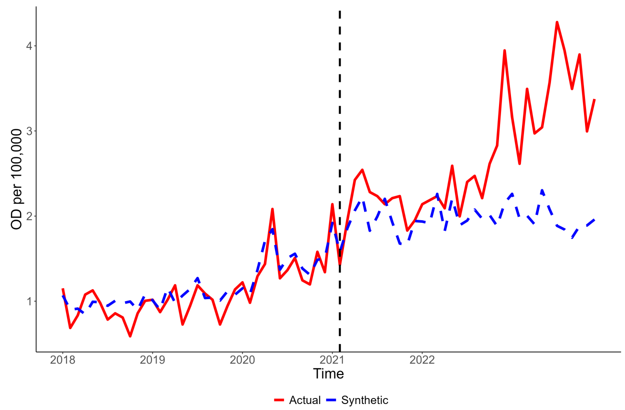
(b) Washington



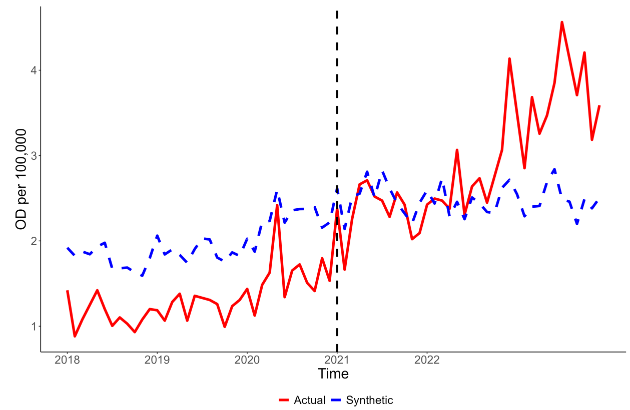
Notes: Synthetic Control Estimates for reported arrests are shown. The data is obtained from 2018-2022 NIBRS. Our sample is restricted to agencies that report all 60 months and aggregated to a state-by-month panel.

Appendix Figure 3.5: Synthetic Control Estimate: Forcing Same Weights

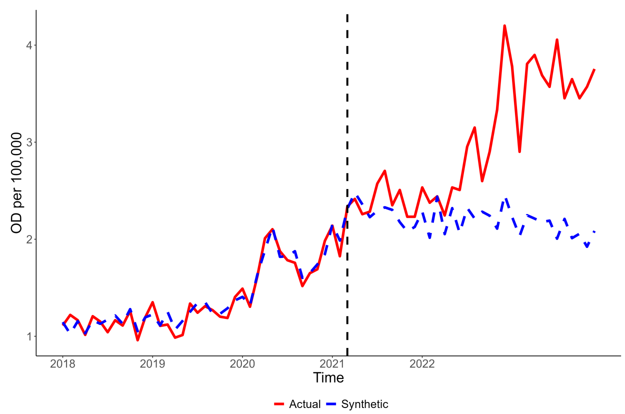
(a) Spencer Outcome Using Joshi et al. Weights: Oregon



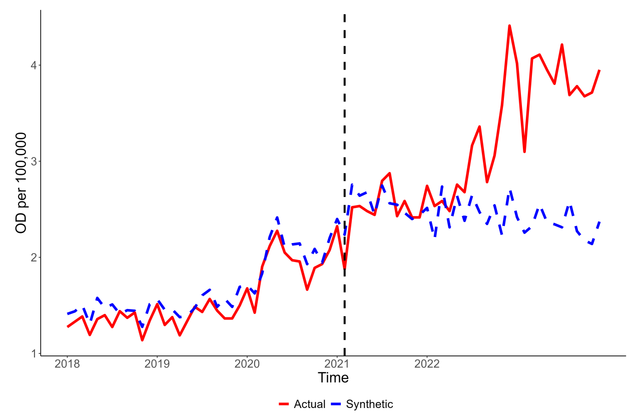
(b) Joshi et al. Outcome Using Spencer Weights: Oregon



(c) Spencer Outcome Using Joshi et al. Weights: Washington

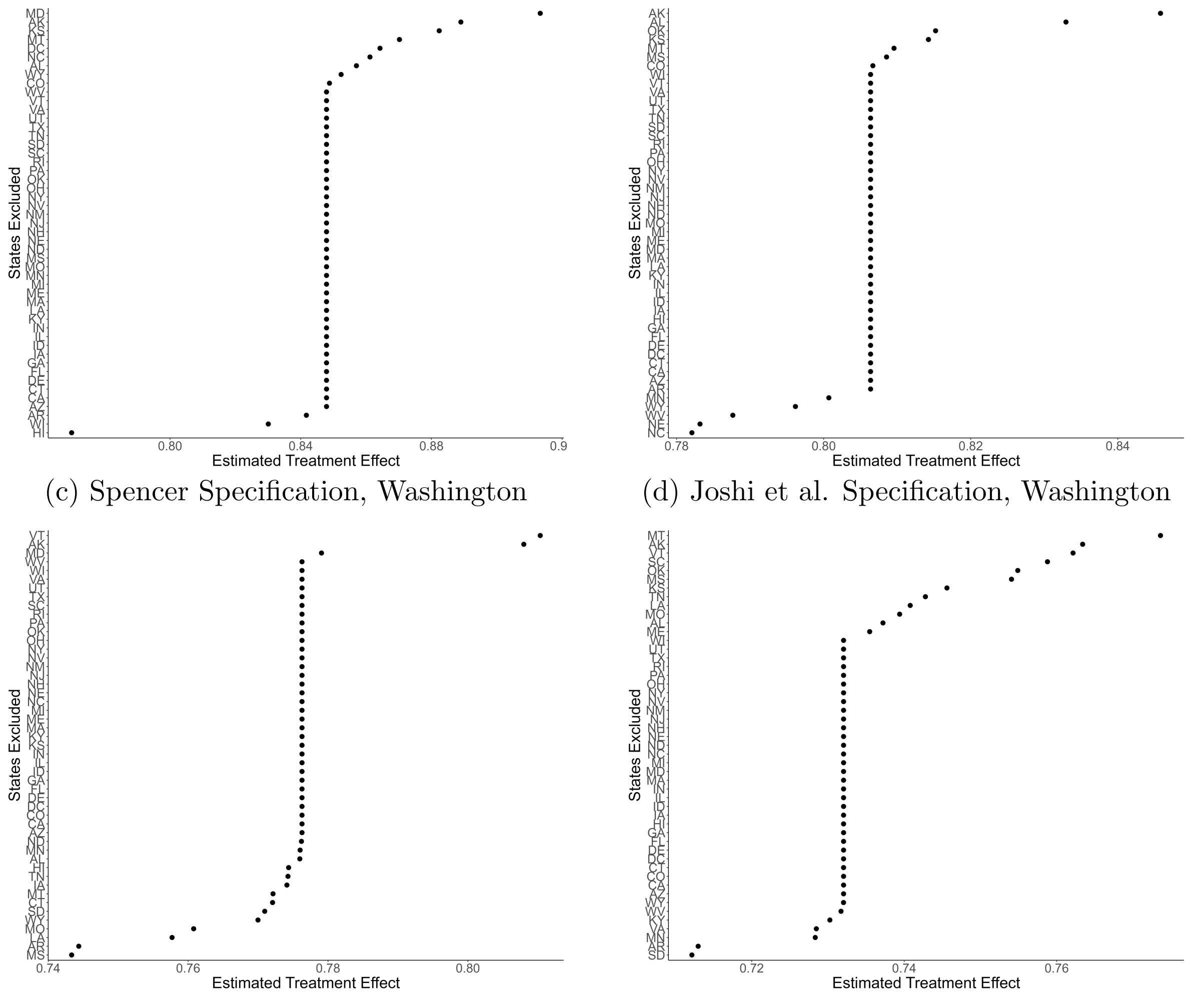


(d) Joshi et al. Outcome Using Spencer Weights: Washington



Notes: These figures are generated using synthetic control weights for the other outcome.

Appendix Figure 3.6: Synthetic Control Estimate: Dropping One Control Units



(a) Spencer Specification, Oregon

(b) Joshi et al. Specification, Oregon

(c) Spencer Specification, Washington

(d) Joshi et al. Specification, Washington

Notes: These figures provide the estimated ATTs when we drop one state from our donor pool.

4.3.2 Chapter 3. Appendix Tables

Appendix Table 3.1: Synthetic Control Estimate Drugs Seized, NFILS

	Fentanyl (1)	Non Fentanyl (2)	All Drugs (3)	Fentanyl Ratio (4)
	Panel I: Oregon			
ATT	-10.814	-55.083**	-66.846*	17.382**
P-Val	{0.18}	{0.02}	{0.06}	{0.02}
	Panel II: Washington			
ATT	-13.615	-23.151**	-17.668**	15.329*
P-Val	{0.2}	{0.02}	{0.04}	{0.08}

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Synthetic control estimates are shown. All estimates use biannual data from 2010 to 2020 to find the synthetic weights. The p-values, calculated via a permutation test, are provided inside the curly brackets.

Appendix Table 3.2: Synthetic Control Estimates with Prediction & Confidence Intervals, Replication

	Oregon		Washington	
	Panel I: Prediction Intervals			
ATT	0.406**	0.264	0.107*	0.098
	[0.13,0.68]	[-0.2,1.01]	[-0.01,0.23]	[-0.08,0.27]
	Panel II: Chernozhukov et al. (2018) T-Test Procedure			
ATT	0.429*	0.331	0.175	0.192
	[0.03,0.83]	[-0.01,0.67]	[-0.01,0.36]	[-0.1,0.48]
Specification	Spencer	Joshi	Spencer	Joshi

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: In panel I, we report the 90 percent prediction intervals (Cattaneo et al. 2021). In panel II, we report the 90 percent confidence intervals calculated using the K=3-fold cross-validation technique (Chernozhukov et al. 2024). These intervals are reported inside the square brackets.

Appendix Table 3.3: Robustness: Other Estimator

	Aug Synth		SDiD		Pop Weighted TWFE		Unweighted TWFE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Oregon							
ATT Overall	0.85*	0.806**	0.691*	0.81**	0.811***	0.838***	0.848	0.883*
P-Val	{0.09}	{0.024}	{0.058}	{0.04}	{0}	{0}	{0.108}	{0.094}
	Panel B: Washington							
ATT Overall	0.775***	0.733***	0.799**	0.831**	0.801***	0.836***	0.853*	0.888*
P-Val	{0.001}	{0}	{0.029}	{0.035}	{0}	{0}	{0.095}	{0.093}
Specification	Spencer	Joshi	Spencer	Joshi	Spencer	Joshi	Spencer	Joshi

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Columns (1) and (2) report results from augmented synthetic control; columns (3) and (4) report results from synthetic difference-in-differences method; columns (5) and (6) report results from generalized synthetic control method; and columns (7) and (8) report results from two-way fixed effects estimator.

Appendix Table 3.4: Robustness: Forcing Same Weights & Estimating Difference-in-Differences

	Oregon				Washington			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATT Overall	0.771*	0.777*	1.089***	1.079***	0.761*	0.792*	0.795*	0.832*
P-Val	{0.06}	{0.06}	{0.04}	{0.04}	{0.06}	{0.06}	{0.06}	{0.06}
Outcome	Spencer	Spencer	Joshi	Joshi	Spencer	Spencer	Joshi	Joshi
Weight	Joshi	Joshi	Spencer	Spencer	Joshi	Joshi	Spencer	Spencer
Treatment	Jan	Feb	Jan	Feb	Feb	Mar	Feb	Mar

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Results from two-way fixed effects estimator are shown. Each regression is weighted via synthetic control weights. Permutation-based p-values are reported inside the curly brackets.

Appendix Table 3.5: Synthetic Control Estimates with Prediction & Confidence Intervals, Full Period

	Oregon		Washington	
	Panel I: Prediction Intervals			
ATT	0.875**	0.784**	0.759**	0.711**
	[0.6,1.15]	[0.32,1.61]	[0.62,0.9]	[0.54,0.88]
	Panel II: Chernozhukov et al. (2018) T-Test Procedure			
ATT	0.86**	0.878**	0.768***	0.816**
	[0.49,1.23]	[0.5,1.26]	[0.7,0.83]	[0.55,1.09]
Specification	Spencer	Joshi	Spencer	Joshi

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: In panel I, we report the 90 percent prediction intervals (Cattaneo et al. 2021). In panel II, we report the 90 percent confidence intervals calculated using the K=3-fold cross-validation technique (Chernozhukov et al. 2024). These intervals are reported inside the square brackets.

Appendix Table 3.6: Synthetic Control Estimate Excluding COVID-19

	Oregon		Washington	
	(1)	(2)	(3)	(4)
ATT	0.878**	0.889**	0.809**	0.954**
P-Val	{0.02}	{0.02}	{0.02}	{0.02}
Data	Spencer	Joshi	Spencer	Joshi

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Synthetic control estimates are shown. All estimates use January 2018 to February 2020 to find the synthetic weights. Columns 1 and 3 focus on all unintentional drug overdoses, which is the outcome used by Spencer (2023). Columns 2 and 4 focus on all drug mortality, which is the outcome used by Joshi et al. (2023). The p-values, calculated via a permutation test, are provided inside the curly brackets.

Appendix Table 3.7: Placebo Test: West Coast State

	Spencer (2023)			Joshi et al. (2023)		
	CA	ID	NV	CA	ID	NV
ATT	0.135	-0.089	0.418	0.176	-0.132	0.255
P-Val	{0.408}	{0.796}	{0.367}	{0.327}	{0.776}	{0.429}

* P-val < 0.1; ** P-val < 0.05; *** P-val < 0.01

Notes: Synthetic control estimates are shown. All estimates use biannual data from 2010 to 2020 to find the synthetic weights. The p-values, calculated via a permutation test, are provided inside the curly brackets.

References

A. B. C. News. (2024). *Oregon officials declare state of emergency to address fentanyl crisis in Portland.* Retrieved 2024-04-18, from https://abcnews.go.com/US/oregon-officials-declare-state-emergency-address-fentanyl-crisis/story?id=106826829

Abadie, A. (2021). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature*, *59*(2), 391–425.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505. (Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/jasa.2009.ap08746)

Abrams, D., Fang, H., & Goonetilleke, P. (2023). Do cops know who to stop? Assessing optimizing models of police behavior with a natural experiment. *National Bureau of Economic Research Working Paper*.

Adams, S., Blackburn, M. L., & Cotti, C. D. (2012). Minimum wages and alcohol-related traffic fatalities among teens. *Review of Economics and Statistics*, *94*(3), 828–840.

Agan, A., & Starr, S. (2018). Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment*. *The Quarterly Journal of Economics*, *133*(1), 191–235.

Anderson, D. M., Hansen, B., & Rees, D. I. (2013). Medical marijuana laws, traffic fatalities, and alcohol consumption. *The Journal of Law and Economics*, *56*(2), 333–369.

Anderson, D. M., Liang, Y., & Sabia, J. J. (2024). Mandatory seatbelt laws and traffic fatalities: A reassessment. *Journal of Applied Econometrics*, *39*(3), 513–521.

Anderson, D. M., & Rees, D. I. (2015). Per se drugged driving laws and traffic fatalities. *International Review of Law and Economics*, *42*, 122–134.

Anderson, M. L., & Davis, L. W. (2023). Uber and traffic fatalities. *Review of Economics and Statistics*, 1–30.

Anthony Johnson, Haven Wheelock, & Janie Gullickson. (2019, August). *Drug Addiction Treatment and Recovery Act.* Retrieved 2024-04-17, from https://egov.sos.state.or.us/elec/web_irr_search.record_detail?p_reference=20200044..LSCYYYDRUG

Anwar, S., & Fang, H. (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review*, *96*(1), 127–151.

Argys, L. M., Mroz, T. A., & Pitts, M. M. (2019). Driven from work: Graduated driver license programs and teen labor market outcomes. *Atlanta Fed. Working Paper*.

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

Banerjee, A., Duflo, E., Keniston, D., & Singh, N. (2019). The efficient deployment of police resources: Theory and new evidence from a randomized drunk driving crackdown in india. *National Bureau of Economic Research Working Paper*.

Banzhaf, H. S. (2021). The value of statistical life: A meta-analysis of meta-analyses. *NBER Working Paper 29185*.

Beck, K. H., & Moser, M. L. (2004). Exposure to the sobriety "checkpoint strikeforce" campaign in Maryland: Impact on driver perceptions of vulnerability and behavior. *Traffic Injury Prevention*, *5*(2), 101–106.

Beck, K. H., & Moser, M. L. (2006). Does the type of exposure to a roadside sobriety checkpoint influence driver perceptions regarding drunk driving? *American Journal of Health Behavior*, *30*(3), 268–277.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, *76*(2), 169–217. (Publisher: The University of Chicago Press)

Becker, G. S., Murphy, K. M., & Grossman, M. (2004). *The Economic Theory of Illegal Goods: The Case of Drugs* [Working Paper]. National Bureau of Economic Research. Retrieved 2023-08-09, from https://www.nber.org/papers/w10976

Beer Institute. (2021). *2021 Brewers Almanac.* Retrieved from https://www.beerinstitute.org/multimedia/brewers-almanac/

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The Augmented Synthetic Control Method. *Journal of the American Statistical Association*, *116*(536), 1789–1803. (Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2021.1929245)

Bergen, G., Pitan, A., Qu, S., Shults, R. A., Chattopadhyay, S. K., Elder, R. W., ... others (2014). Publicized sobriety checkpoint programs: A community guide systematic review. *American Journal of Preventive Medicine*, *46*, 529-539.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275.

Billings, S. B., & Schnepel, K. T. (2022). Hanging out with the usual suspects: Neighborhood peer effects and recidivism. *Journal of Human Resources*, *57*(5), 1758–1788.

Byles, H., Sedaghat, N., Rider, N., Rioux, W., Loverock, A., Seo, B., ... Ghosh, S. (2024). Barriers to calling emergency services amongst people who use substances in the event of overdose: A scoping review. *International Journal of Drug Policy*, *132*, 104559.

Caetano, C., Callaway, B., Payne, S., & Rodrigues, H. S. (2024). *Difference in Differences with Time-Varying Covariates.* arXiv. Retrieved 2025-04-17, from http://arxiv.org/abs/2202.02903 (arXiv:2202.02903)

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*, 200-230.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, *110*(512), 1753–1769.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. (Publisher: Nature Publishing Group)

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, *90*(3), 414–427.

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

Campbell, R. A. (2023). What does federal oversight do to policing and public safety? Evidence from seattle. *Working Paper*.

Caputo, A. (2015). *Chicago police sobriety checkpoints target black, latino neighborhoods.* Retrieved from https://www.chicagotribune.com/investigations/ct-dui-checkpoints-chicago-met-20150507-story.html

Carpenter, C. (2004). How do zero tolerance drunk driving laws work? *Journal of Health Economics*, *23*, 61-83.

Carpenter, C., & Dobkin, C. (2009). The effect of alcohol consumption on mortality: Regression discontinuity evidence from the minimum drinking age. *American Economic Journal: Applied Economics*, *1*(1), 164–182.

Carpenter, C., & Dobkin, C. (2011). The minimum legal drinking age and public health. *Journal of Economic Perspectives*, *25*(2), 133–156.

Cattaneo, M. D., Feng, Y., & Titiunik, R. (2021). Prediction intervals for synthetic control methods. *Journal of the American Statistical Association*, *116*(536), 1865–1880.

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, *134*, 1405-1454.

Center for Disease Control and Prevention, National Center for Health Statistics. (2022). *U.S. Overdose Deaths In 2021 Increased Half as Much as in 2020 - But Are Still Up 15%.* Retrieved 2024-04-18, from https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2022/202205.htm

Chaduvula, R. (2019). *ND bill aims to outlaw DUI checkpoints. What do cops think about that?* The Dickinson Press. Retrieved from https://www.thedickinsonpress.com/news/nd-bill-aims-to-outlaw-dui-checkpoints-what-do-cops-think-about-that

Chaisemartin, C. D., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*, 2964-2996.

Chalfin, A., & Gonçalves, F. M. (2023). Professional motivations in the public sector: Evidence from police officers. *National Bureau of Economic Research Working Paper*.

Chalfin, A., Hansen, B., Weisburst, E. K., & Williams Jr, M. C. (2022). Police force size and civilian race. *American Economic Review: Insights*, *4*(2), 139–158.

Chalfin, A., & McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature*, *55*(1), 5–48.

Chang, A. C., & Li, P. (2015). Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not". *Finance and Economics Discussion Series*, *2015*(83), 1–26.

Chernozhukov, V., Wuthrich, K., & Zhu, Y. (2024). *A t-test for synthetic controls.* arXiv. Retrieved 2024-03-16, from http://arxiv.org/abs/1812.10820 (arXiv:1812.10820 [econ])

Ciccarone, D. (2019). The Triple Wave Epidemic: Supply and Demand Drivers of the US Opioid Overdose Crisis. *The International journal on drug policy*, *71*, 183–188.

Clarke, D., Pailanir, D., Athey, S., & Imbens, G. (2023). Synthetic difference in differences estimation. *arXiv preprint arXiv:2301.11859*.

Cohen, A., & Einav, L. (2003). The effects of mandatory seat belt laws on driving behavior and traffic fatalities. *Review of Economics and Statistics*, *85*(4), 828–843.

Davis, C. S., Joshi, S., Rivera, B. D., & Cerdá, M. (2023). Changes in arrests following decriminalization of low-level drug possession in Oregon and Washington. *International Journal of Drug Policy*, *119*, 104155.

DeAngelo, G., & Hansen, B. (2014). Life and death in the fast lane: Police enforcement and traffic fatalities. *American Economic Journal: Economic Policy*, *6*(2), 231–257.

Dee, T. S. (1999). State alcohol policies, teen drinking and traffic fatalities. *Journal of Public Economics*, *72*(2), 289–315.

Dee, T. S. (2001). Does setting limits save lives? The case of 0.08 bac laws. *Journal of Policy Analysis and Management*, *20*(1), 111–128.

Dube, O., MacArthur, S. J., & Shah, A. K. (2023). A cognitive view of policing. *National Bureau of Economic Research Working Paper*.

Dumont, E., Fortin, B., Jacquemet, N., & Shearer, B. (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics*, *27*(6), 1436–1450.

Eisenberg, D. (2003). Evaluating the effectiveness of policies related to drunk driving. *Journal of Policy Analysis and Management*, *22*(2), 249–274.

Elder, R. W., Shults, R. A., Sleet, D. A., Nichols, J. L., Zaza, S., & Thompson, R. S. (2002). Effectiveness of sobriety checkpoints for reducing alcohol-involved crashes. *Traffic Injury Prevention*, *3*(4), 266–274.

Evans, W. N., Lieber, E., & Power, P. (2018). *How the Reformulation of OxyContin Ignited the Heroin Epidemic* [Working Paper]. National Bureau of Economic Research. Retrieved 2023-12-05, from https://www.nber.org/papers/w24475

Evans, W. N., Neville, D., & Graham, J. D. (1991). General deterrence of drunk driving: Evaluation of recent american policies. *Risk Analysis*, *11*(2), 279–289.

Evans, W. N., & Owens, E. G. (2007). Cops and crime. *Journal of Public Economics*, *91*(1-2), 181–201.

Fagan, J., & Geller, A. (2020). Profiling and consent: Stops, searches, and seizures after soto. *Va. J. Soc. Pol'y & L.*, *27*, 16.

Feigenberg, B., & Miller, C. (2023). Class disparities and discrimination in traffic stops and searches. *Working Paper*.

Fell, J. C., Lacey, J. H., & Voas, R. B. (2004). Sobriety checkpoints: Evidence of effectiveness is strong, but use is limited. *Traffic Injury Prevention*, *5*(3), 220–227.

Fell, J. C., Langston, E. A., & Tippetts, A. S. (2005). Evaluation of four state impaired driving enforcement demonstration programs: Georgia, Tennessee, Pennsylvania and Louisiana. In *Annual proceedings/association for the advancement of automotive medicine* (Vol. 49, p. 311).

Fell, J. C., & Scherer, M. (2017). Administrative license suspension: Does length of suspension matter? *Traffic Injury Prevention*, *18*(6), 577–584.

Ferman, B., Pinto, C., & Possebom, V. (2020). Cherry Picking with Synthetic Controls. *Journal of Policy Analysis and Management*, *39*(2), 510–532. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.22206)

Ferrazares, T. (2024). Monitoring police with body-worn cameras: Evidence from chicago. *Journal of Urban Economics*, *141*, 103539.

Freeman, D. G. (2007). Drunk driving legislation and traffic fatalities: New evidence on bac 08 laws. *Contemporary Economic Policy*, *25*(3), 293–308.

Félix, S., Portugal, P., & Tavares, A. S. (2017). *Going after the Addiction, Not the Addicted: The Impact of Drug Decriminalization in Portugal* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2024-07-09, from https://papers.ssrn.com/abstract=3010673

Gaitán, C. (2023, December). Oregon's Measure 110 drug treatment hotline has served only 577, gets new operator. *Oregon Live*.

Gallup. (2024). *Confidence in institutions.* Retrieved from https://news.gallup.com/poll/1597/confidence-institutions.aspx

Garoupa, N., & Klerman, D. (2002). Optimal law enforcement with a rent-seeking government. *American Law and Economics Review*, *4*(1), 116–140.

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

Goncalves, F., & Mello, S. (2021). A few bad apples? Racial bias in policing. *American Economic Review*, *111*(5), 1406–1441.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277.

Greene, J. W. (2003). Batting DUI: A comparative analysis of checkpoints and saturation patrols. *FBI L. Enforcement Bull.*, *72*, 1.

Gutierrez, A. (2021). *Missouri lawmakers consider ban on sobriety checkpoints.* KSHB News. Retrieved from https://www.kshb.com/news/local-news/missouri-lawmakers-consider-ban-on-sobriety-checkpoints

Hagemann, A. (2020). Inference with a single treated cluster. *arXiv preprint arXiv:2010.04076*.

Hansen, B. (2015). Punishment and deterrence: Evidence from drunk driving. *American Economic Review*, *105*, 1581-1617.

Hansen, B., Miller, K., & Weber, C. (2020). Federalism, partial prohibition, and cross-border sales: Evidence from recreational marijuana. *Journal of Public Economics*, *187*, 104159.

Hansen, B., & Waddell, G. R. (2018). Legal access to alcohol and criminality. *Journal of Health Economics*, *57*, 277–289.

Hausman, C., & Rapson, D. S. (2018). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, *10*, 533–552.

Henderson, K., Campbell, C., & Renauer, B. (2023). Impacts of Successive Drug Legislation Shifts: Qualitative Observations from Oregon Law Enforcement [Interim Report: Year One]. *Criminology and Criminal Justice Faculty Publications and Presentations*.

Hillary, F. G., & Medaglia, J. D. (2020). What the replication crisis means for intervention science. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, *154*, 3–5.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, and Organization*, *7*(special_issue), 24–52.

Holzer, H. J., Raphael, S., & Stoll, M. A. (2006). Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers. *The Journal of Law and Economics*, *49*(2), 451–480. (Publisher: The University of Chicago Press)

Hong, F., Hossain, T., List, J. A., & Tanaka, M. (2018). Testing the theory of multitasking: Evidence from a natural field experiment in chinese factories. *International Economic Review*, *59*(2), 511–536.

Hyndman, R. J. (2021). *Detecting time series outliers.* Retrieved from https://www.r-bloggers.com/2021/08/detecting-time-series-outliers/

Johnson, R. M., Reiley, D. H., & Muñoz, J. C. (2015). "the war for the fare": How driver compensation affects bus system performance. *Economic Inquiry*, *53*(3), 1401–1419.

Join LAPD. (2023). *LAPD career ladder.* Retrieved from https://www.joinlapd.com/career-ladder

Jones, L. E., & Morin, C. B. (2022). Sobriety checkpoint laws, fatal car crashes and arrests. *Ohio State Legal Studies Research Paper*, *724*.

Joshi, S., Rivera, B. D., Cerdá, M., Guy, G. P., Jr, Strahan, A., Wheelock, H., & Davis, C. S. (2023). One-Year Association of Drug Possession Law Change With Fatal Drug Overdose in Oregon and Washington. *JAMA Psychiatry*.

Kaplan, J. (2021a). *Jacob kaplan's concatenated files: Uniform crime reporting program data: Law enforcement officers killed and assaulted (leoka) 1960-2020.* Retrieved from https://www.openicpsr.org/openicpsr/project/102180/version/V11/view

Kaplan, J. (2021b). *Jacob kaplan's concatenated files: Uniform crime reporting (ucr) program data: Arrests by age, sex, and race, 1974-2020.* Retrieved from https://www.openicpsr.org/openicpsr/project/102263/version/V14/view

Kelly, A., Lindo, J. M., & Packham, A. (2020). The power of the IUD: Effects of expanding access to contraception through Title X clinics. *Journal of Public Economics*, *192*, 104288.

Kenkel, D. S. (1993). Drinking, driving, and deterrence: The effectiveness and social costs of alternative policies. *The Journal of Law and Economics*, *36*(2), 877–913.

Kenney, K. (2018). *Sobriety checkpoints don't net many drunk driving arrests but police, madd say they're crucial.* ABC Action News. Retrieved from "https://www.abcactionnews.com/conquering-addiction/sobriety-checkpoints-don-t-net-many-drunk-driving-arrests-but-police-madd-say-they-re-crucial#of%20the%2091,278%20vehicles

Kim, T. (2022). Promotion incentives, career decisions, and police performance. *SSRN Working Paper*.

Knutsson, D., & Tyrefors, B. (2022). The quality and efficiency of public and private firms: Evidence from ambulance services. *The Quarterly Journal of Economics*, *137*(4), 2213–2262.

Lacey, J. H., Jones, R. K., & Smith, R. G. (1999). *Evaluation of checkpoint Tennessee: Tennessee's statewide sobriety checkpoint program.* https://www.ojp.gov/ncjrs/virtual-library/abstracts/evaluation-checkpoint-tennessee-tennessees-statewide-sobriety. US Department of Transportation.

Lacombe, D. (2016). *DUI checkpoints target minority areas.* Retrieved from https://lccrsf.org/news/dui-checkpoints-target-minority-areas/

Leadbeater, B. J., Foran, K., & Grove-White, A. (2008). How much can you drink before driving? The influence of riding with impaired adults and peers on the driving behaviors of urban and rural youth. *Addiction*, *103*(4), 629–637.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355.

Levitt, S. D. (2004). Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. *Journal of Economic Perspectives*, *18*(1), 163–190.

Levitt, S. D., & Porter, J. (2001). How dangerous are drinking drivers? *Journal of Political Economy*, *109*(6), 1198–1237.

Lochner, L., & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review*, *94*(1), 155–189.

Long, W. (2019). How does oversight affect police? Evidence from the police misconduct reform. *Journal of Economic Behavior & Organization*, *168*, 94–118.

Lovenheim, M. F., & Steefel, D. P. (2011). Do blue laws save lives? The effect of sunday alcohol sales bans on fatal vehicle accidents. *Journal of Policy Analysis and Management*, *30*, 798-820.

Luh, E. (2022). Not so black and white: Uncovering racial bias from systematically misreported trooper reports. *Available at SSRN 3357063*.

Luh, E., Pyle, B., & Reeves, J. (2023). Agency incentives and disparate revenue collection: Evidence from chicago parking tickets. *Working Paper*.

MacDonald, J., Fagan, J., & Geller, A. (2016). The effects of local police surges on crime and arrests in new york city. *PLoS one*, *11*(6), e0157223.

Maclean, J. C., Mallatt, J., Ruhm, C. J., & Simon, K. (2020). *Economic Studies on the Opioid Crisis: A Review* [Working Paper]. National Bureau of Economic Research. Retrieved 2023-08-31, from https://www.nber.org/papers/w28067

Makofske, M. (2023). Pretextual traffic stops and racial disparities in their use. *Working Paper*.

Makowsky, M. D., & Stratmann, T. (2009). Political economy at any speed: What determines traffic citations? *American Economic Review*, *99*(1), 509–527.

Makowsky, M. D., & Stratmann, T. (2014). Politics, unemployment, and the enforcement of immigration law. *Public Choice*, *160*, 131–153.

Makowsky, M. D., Stratmann, T., & Tabarrok, A. (2019). To serve and collect: The fiscal and racial determinants of law enforcement. *The Journal of Legal Studies*, *48*(1), 189–216.

Manka, D. (2021). *Washington SB5476 | 2021-2022 | Regular Session.* Retrieved 2024-03-25, from https://legiscan.com/WA/text/SB5476/id/2403634

Matsuzawa, K. (2022). The deterrent effect of targeted and salient police enforcement: Evidence from dui checkpoint bans. *Available at SSRN 4310253*.

McCarthy, J. (2022). *Americans remain steadfast on policing reform needs in 2022.* Retrieved from https://news.gallup.com/poll/393119/americans-remain-steadfast-policing-reform-needs-2022.aspx

Mello, S. (2019). More COPS, less crime. *Journal of Public Economics*, *172*, 174–200.

Mello, S. (2021). Fines and financial wellbeing. *Working paper*.

Mello, S., Ross, M., Stephen, R., & Johnson, H. (2023). Diversity training and employee behavior: Evidence from the police. *Working Paper*.

Midgette, G., Kilmer, B., Nicosia, N., & Heaton, P. (2021). A natural experiment to test the effect of sanction certainty and celerity on substance-impaired driving: North Dakota's 24/7 sobriety program. *Journal of Quantitative Criminology*, *37*, 647–670.

Miller, T. R., Galbraith, M. S., & Lawrence, B. A. (1998). Costs and benefits of a community sobriety checkpoint program. *Journal of Studies on Alcohol*, *59*, 462-468.

Mothers Against Drunk Driving. (2022). *Sobriety checkpoints.* Retrieved from https://madd.org/wp-content/uploads/2022/04/Sobriety-Checkpoints.pdf

Mueller-Smith, M. (2015). *THE CRIMINAL AND LABOR MARKET IMPACTS OF INCARCERATION.*

Naddeo, J., & Pulvino, R. (2023). The effects of reducing pretextual stops: Evidence from saint paul minnesota. *Working Paper*.

National Archive of Criminal Justice Data. (n.d.). *Resource guide uniform crime reporting program.* Retrieved from https://www.icpsr.umich.edu/web/pages/NACJD/guides/ucr.html

National Highway Traffic Safety Administration. (2001). *Determine why there are fewer young alcohol-impaired drivers.* Retrieved from https://one.nhtsa.gov/people/injury/research/FewerYoungDrivers/index.htm

Netherland, J., Kral, A. H., Ompad, D. C., Davis, C. S., Bluthenthal, R. N., Dasgupta, N., ... Wheelock, H. (2022). Principles and Metrics for Evaluating Oregon's Innovative Drug Decriminalization Measure. *Journal of Urban Health*, *99*(2), 328–331.

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777–787.

Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, *61*(4), 631–653.

Oregon Health Justice Recovery Alliance. (2021). *Oregon Health Justice Recovery Alliance.* Retrieved 2024-04-17, from https://web.archive.org/web/20210609033613/https://healthjusticerecovery.org/

Packham, A. (2022). Syringe exchange programs and harm reduction: New evidence in the wake of the opioid epidemic. *Journal of Public Economics*, *215*, 104733.

Parker, S., Ross, M., & Ross, S. (2024). Driving change: Evaluating connecticut's collaborative approach to reducing racial disparities in policing. *National Bureau of Economic Research Working Paper*.

Payton, M., Dawson, B., Stanchis, K., & Jones, T. (2018). *DUI checkpoints: Worth the time and money?* NBC San Diego. Retrieved from https://www.nbcsandiego.com/news/local/dui-checkpoints-worth-the-time-and-money/142194/

Peek-Asa, C. (1999). The effect of random alcohol screening in reducing motor vehicle crash injuries. *American Journal of Preventive Medicine*, *16*(1), 57–67.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., ... others (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, *4*(7), 736–745.

Pitts, W. J., Heller, D., Smiley-McDonald, H., Weimer, B., Grabenauer, M., Bollinger, K., … Pressley, D. (2023). Understanding research methods, limitations, and applications of drug data collected by the National Forensic Laboratory Information System (NFLIS-Drug). *Journal of Forensic Sciences*, *68*(4), 1335–1342. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1556-4029.15269)

Reeves, J. (2024). Multitasking, expectations, and police officer behavior. *Working Paper*.

Representative Jason Kropf, & Senator Kate Lieber. (2024). *House Bill 4002.* Retrieved 2024-04-18, from https://olis.oregonlegislature.gov/liz/2024R1/Measures/Overview/HB4002

Roach, K., Baumgartner, F. R., Christiani, L., Epp, D. A., & Shoub, K. (2022). At the intersection: Race, gender, and discretion in police traffic stop outcomes. *Journal of Race, Ethnicity, and Politics*, *7*(2), 239–261.

Romero, D. (2016). *When it comes to DUI crackdowns, westside residents get a pass.* LA Weekly. Retrieved from https://www.laweekly.com/when-it-comes-to-dui-crackdowns-westside-residents-get-a-pass/

Ruhm, C. J. (1996). Alcohol policies and highway vehicle fatalities. *Journal of Health Economics*, *15*, 435-454.

Rushin, S., & Edwards, G. (2021). An empirical assessment of pretextual stops and racial profiling. *Stan. L. Rev.*, *73*, 637.

Russoniello, K., Vakharia, S. P., Netherland, J., Naidoo, T., Wheelock, H., Hurst, T., & Rouhani, S. (2023). Decriminalization of drug possession in Oregon: Analysis and early lessons. *Drug Science, Policy and Law*, *9*, 20503245231167407. (Publisher: SAGE Publications)

Sabia, J. J., Pitts, M. M., & Argys, L. M. (2019). Are minimum wages a silent killer? New evidence on drunk driving fatalities. *Review of Economics and Statistics*, *101*(1), 192–199.

Sanem, J. R., Erickson, D. J., Rutledge, P. C., Lenk, K. M., Nelson, T. F., Jones-Webb, R., & Toomey, T. L. (2015). Association between alcohol-impaired driving enforcement-related strategies and alcohol-impaired driving. *Accident Analysis & Prevention*, *78*, 104–109.

Shi, L. (2008). Does oversight reduce policing? Evidence from the cincinnati police department after the april 2001 riot. *Journal of Public Economics*.

Shover, C. L., Falasinnu, T. O., Dwyer, C. L., Santos, N. B., Cunningham, N. J., Freedman, R. B., … Humphreys, K. (2020). Steep increases in fentanyl-related mortality west of the Mississippi River: Recent evidence from county and state surveillance. *Drug and Alcohol Dependence*, *216*, 108314.

Smiley-McDonald, H. M., Attaway, P. R., Wenger, L. D., Greenwell, K., Lambdin, B. H., & Kral, A. H. (2023). "All carrots and no stick": Perceived impacts, changes in practices, and attitudes among law enforcement following drug decriminalization in Oregon State, USA. *International Journal of Drug Policy*, *118*, 104100.

Smith, M. R., Makarios, M., & Alpert, G. P. (2006). Differential suspicion: Theory specification and gender effects in the traffic stop context. *Justice Quarterly*, *23*(02), 271–295.

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*, *50*(2), 301–316.

Spencer, N. (2023). Does drug decriminalization increase unintentional drug overdose deaths?: Early evidence from Oregon Measure 110. *Journal of Health Economics*, *91*, 102798.

State of Washington. (2021). *State v. Blake.* Retrieved 2024-03-25, from https://www.co.pacific.wa.us/courts/state-v-blake.html

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*, 175-199.

Surveillance, Epidemiology, and End Results Program. (2022). *U.s. county population data - 1969-2020.* Retrieved from https://seer.cancer.gov/popdata/

Tebes, J., & Fagan, J. (2022). Stopped by the police: The end of "stop-and-frisk" on neighborhood crime and high school dropout rates. *Working Paper*.

The Centers for Disease Control & Prevention. (2015a). *Motor vehicle safety primary enforcement of seat belt laws.* Retrieved from https://www.cdc.gov/motorvehiclesafety/calculator/factsheet/seatbelt.html

The Centers for Disease Control & Prevention. (2015b). *Sobriety checkpoints.* Retrieved from https://www.cdc.gov/motorvehiclesafety/calculator/factsheet/checkpoints.html

The Centers for Disease Control & Prevention. (2020). *Impaired driving: Get the facts.* Retrieved from https://www.cdc.gov/transportationsafety/impaired_driving/impaired-drv_factsheet.html

University of Kentucky Centers for Poverty Research. (2022). *UKCPR national welfare data, 1980-2020.* Retrieved from http://ukcpr.org/resources/national-welfare-data

Urban Institute. (n.d.). *State and local expenditures.* Retrieved from https://state-local-finance-data.taxpolicycenter.org/pages.cfm

U.S. Department of Transportation. (2020). *Highway statistics series.* Retrieved from https://www.fhwa.dot.gov/policyinformation/statistics/2020/mv1.cfm

Voas, R. B., Tippetts, A. S., & Fell, J. C. (2003). Assessing the effectiveness of minimum legal drinking age and zero tolerance laws in the united states. *Accident Analysis and Prevention*, *35*(4), 579–587.

Vote No on Measure 110. (2020). *News | Vote No on Measure 110!* Retrieved 2024-04-17, from https://web.archive.org/web/20210126193538/https://votenoon110.com/news/

West, K. L., Lindquist, K., & Rodda, L. N. (2021). Fentanyl epidemic hits the U.S. West Coast: Opioid-related deaths in San Francisco from 2009–2019. *International Journal of Drug Policy*, *95*, 103402.

Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Political Analysis*, *25*(1), 57–76.

Zibbell, J. E., Aldridge, A., Grabenauer, M., Heller, D., Clarke, S. D., Pressley, D., & McDonald, H. S. (2023). Associations between opioid overdose deaths and drugs confiscated by law enforcement and submitted to crime laboratories for analysis, United States, 2014–2019: an observational study. *The Lancet Regional Health – Americas*, *25*. (Publisher: Elsevier)

Zoorob, M. J., Park, J. N., Kral, A. H., Lambdin, B. H., & del Pozo, B. (2024). Drug Decriminalization, Fentanyl, and Fatal Overdoses in Oregon. *JAMA Network Open*, *7*(9), e2431612.