

The Effects of Reducing Pretextual Stops: Evidence from Saint Paul Minnesota

J.J. NADDEO* RORY PULVINO*

October 10, 2023

Abstract

Traffic stops make up the vast majority of law enforcement-initiated contact with the public in the United States. As such, it is an important gateway into the criminal legal system. Elected prosecutors—from across the political spectrum—are currently grappling with the growing consensus that the cost of bringing marginal individuals into the criminal legal system likely outweighs the benefits. An important question is how policymakers can reduce intake into the system while maintaining—or even improving—public safety. This article investigates the effectiveness of a policy by the Ramsey County Attorney’s Office (RCAO) in conjunction with the Saint Paul Police Department (SPPD) to decrease the number of traffic stops for minor equipment violations. Using an interrupted time series (ITS) framework, we find that the policy met its goal of virtually eliminating traffic stops for vehicle violations, a common source of police-citizen interactions. Furthermore, we show that the policy reduced stops for Black motorists more relative to White motorists. Finally, we utilize ITS to also document changes in short- and long-term crime trends, the number of gun seizures, the number of victims of gun violence, traffic incidents, and response times to 911 calls. The results suggest that the policy was effective in reducing contact between law enforcement and the public—in a way that reduced racial disparities—without adversely affecting long-term public safety.

Acknowledgments: Both authors would like to acknowledge the Ramsey County Attorney’s Office, the Saint Paul Police Department, and the Ramsey County Emergency Communications Center for their willingness to provide data and invaluable context. Our analysis would not have been possible without the help of Mark Haase, Michael Sullivan, Alyssa Arcand, and Assistant Chief Jack Serier. We also thank Don Braman, Jacob Kaplan, Felix Owusu, Tom Scott, Jess Sorensen, and Kevin Wilson for helpful comments that greatly improved this paper. J.J. Naddeo thanks Georgetown University and Lynn and Anisya Fritz for generous financial support while drafting this paper. Rory Pulvino thanks George Washington University and the St. Paul Foundation for their support while drafting this paper.

* Justice Innovation Lab

1 Introduction

The purpose of law enforcement is to maintain and improve public safety. To do so, they are required to engage with residents of their respective jurisdictions and make difficult decisions that have lasting impacts on the lives of both the accused and victims. Survey data from the Bureau of Justice Statistics (BJS) show that over 20% of all US residents 16 years and older report having some contact with law enforcement per year, and more than half of this contact was reported to be initiated by police. Although there are substantial differences in policing tactics across the nearly 18,000 law enforcement agencies, almost all rely on traffic stops as a means of deterring and detecting criminal activity. For example, the BJS survey also shows that roughly 90% of police-initiated contact was made via traffic stops of motorists ([Tapp and Davis, 2022](#)).

The traffic code in the US is expansive, making it nearly impossible for even the most careful driver to complete a trip without violating it. For example, there are traffic codes that prohibit “cruising”, profanity, hanging anything from the rear view mirror, and driving in the left lane of a two lane highway, to name a few.¹ Moreover, each jurisdiction has its own idiosyncratic laws and enforcement patterns—causing behavior that is unpunished in one jurisdiction, to draw punishment only a few miles away in another jurisdiction. Therefore, as is common in the criminal legal system (CLS) in the US, discretion of individual decision makers plays a significant role in who is pulled into the system—and for what. Unfortunately, where there is discretion, there exists the possibility of bias, in the criminal legal system, this bias tends to manifest along racial lines ([Doleac, 2022](#)).

The Fourth Amendment protects against unreasonable searches and seizures by the government. Generally, this protection requires that police have probable cause that an

¹See City of Mahnomen, Minnesota Code of Ordinances, Title VII, § 70.12 Cruising Prohibited for example of law prohibiting driving around the same area. See Rockville City Code of Ordinances, § 13-53 Profanity; violation of section declared misdemeanor, for example, of prohibition of use of profanity while driving. Numerous states have laws that prohibit drivers from hanging objects from their rear-view mirrors; an example can be found in NY Veh & Traf L § 375. See NJ Rev Stat § 39:4-82 for an example of a law prohibiting driving in the left lane unless passing or turning left.

individual is involved in a crime before stopping the person. While traffic stops fall under these protections, the US Supreme Court in *Whren v. United States*, 517 U.S. 806 (1996) held that traffic stops made for violations of the traffic code, but where the officer made the stop to investigate a hunch that, by itself, would not amount to probable cause, are constitutional. The Court explained that, as long as the officer had probable cause for a legal violation, including minor traffic violations, any ulterior motives for the stop did not amount to a Fourth Amendment violation. As detailed in [Rushin and Edwards \(2021\)](#) the scholarly response was “overwhelmingly critical”, as most scholars recognized that—paired with the aforementioned fact that almost all motorists violate the traffic code—*Whren* opened almost all drivers up to traffic stops technically made for minor violations, but with the intent to search for more serious crimes. These stops are colloquially referred to as pretextual stops, as the minor violation is a *pretext* to conduct further investigations.

In the *Whren* decision the court discussed the difficulty of identifying pretextual stops by saying that in order to identify stops whereby the vehicle violation was used as a pretext would require “speculating about the hypothetical reaction of a hypothetical constable an exercise that might be called virtual subjectivity.” THe court was highlighting that identifying these stops requires understanding the subjective intentions of an officer, requiring one to pin down officer’s true motivation for each traffic stop. Further complicating things is the lack of standardized training around pretextual stops. Some departments—including the federal government—have trained officers to use minor violations as a pretext to search for drugs.² However, some—including the police department examined in this paper—do not train officers to use minor vehicle violations to justify stops motivated by the suspicion of greater criminal activity. Therefore, it is unclear how often—or in what way—pretextual stops were used as an investigatory tool. However, even without specific training to carry out pretextual stops, allowing—or even encouraging—enforcement of minor vehicle violations provides fertile ground for their use.

²See ?

This paper focuses on evaluating a policy enacted by the Ramsey County Attorney's Office (RCAO) in conjunction with the Saint Paul Police Department (SPPD), whose aim was to eliminate the use of traffic stops for minor vehicle equipment violations.³ The assumption motivating this policy is that although not all traffic stops for minor vehicle equipment violations are pretextual stops, most pretextual stops are traffic stops for minor vehicle equipment violations. This article utilizes the policy announcement in an interrupted time series (ITS) framework to not only estimate the effectiveness of policies in reducing traffic stops for minor vehicle equipment violations (vehicle violations henceforth for brevity), but also any corresponding impact on crime, gun seizures, and traffic incidents. Finally, we use a difference-in-differences framework to show the impact on racial disproportionalities in contact with the criminal legal system. To preview our results, we find that the policy was remarkably effective in immediately reducing traffic stops for vehicle equipment violations with an approximately 80% decrease compared to the pre-policy mean of 487 stops for vehicle equipment violations per month, with no evidence of a rebound in a nearly two-year follow-up. Though all minor vehicle equipment violations are not pretextual stops, such stops were part of historical pretextual stop training and are a likely source of any current pretextual stops, thus for much of the rest of the paper we discuss the stops at issue as pretextual.

To our knowledge, this paper represents the first empirical examination of the effect of a policy aimed at eliminating pretextual traffic stops by curtailing the most common tool for making such stops—minor vehicle equipment violations. However, there exists a thick literature in the area of pretextual stops that inspired much of the methodology adopted in this paper. To begin, our interest in evaluating this type of policy is inspired by existing literature that theorizes about the potential costs and benefits of pretextual traffic stops.

(Wu and Lum, 2020; Haywood, 2023)

One of the main motives for reducing pretextual stops is reducing the disproportionate burden they place on communities of color. This disproportionality is not just an ethical

³See the full policy memo [here](#). This policy is unique in that it was implemented with the support of the district attorney's office and law enforcement.

issue, but also decreases trust between communities of color and law enforcement—thus decreasing law enforcement’s ability to protect and serve these communities (Tyler et al., 2014; Blanks, 2016; Skogan, 2016). Furthermore, racial disproportionalities in police contact likely contribute to the substantial racial disparities observed in downstream outcomes in the CLS. Finally, routine traffic stops are commonly portrayed as one of the most dangerous encounters police have with the public (Woods, 2019). Therefore, reducing the number of traffic stops may not only reduce racial disparities, but also increase the safety of police officers.

Our analysis of racial heterogeneity in policy effects mirrors the analysis conducted in Rushin and Edwards (2021), who look into the racial impact of a 2012 decision by the Washington Supreme Court that eased restrictions on pretextual stops. The authors find that the decision caused an increase in the number of traffic stops for Black drivers relative to White drivers and this relative increase was concentrated during the daytime. In this paper, we estimate the racial heterogeneity in impact the policy had on pretextual traffic stops. Using a similar difference-in-differences framework, we find that Black motorists experience a significantly larger drop in pretextual traffic stops. We then conduct a triple difference, as in Rushin and Edwards (2021), and show that this racial heterogeneity is largest during the daytime, when the motorist race is most salient. We remain mute on the direct mechanisms that generate these empirical patterns, but note that they suggest that pretextual stops do fall disproportionately on motorists of color when their race is most easily identified. These results highlight the potential benefit of such policies.

Proponents of pretextual stops argue that they are an invaluable tool to deter and detect criminal activity (Kirkpatrick et al., 2022; Lauer, 2022). Increasing stops in a vacuum unambiguously increases the probability of apprehension. In the classic Becker model, individuals decide whether or not to commit certain crimes by weighing the expected costs against the benefits (Becker, 1968). Therefore, *ceteris paribus* decreasing traffic stops decreases the cost of committing certain crimes, such as carrying illegal firearms or other contraband. With that

in mind, it is worth noting that this policy does not prevent officers from making the same number of stops as before the policy, rather it focuses officers on making stops where they can articulate a public safety reason for the stop. To test whether eliminating a significant source of pretextual stops resulted in an increase in criminal activity, we leverage data from multiple sources such as 911 calls for service from civilians, reported incidents by SPPD, and traffic accident data. To preview our results, we find short-term increases in 911 calls (2.9%, not statistically significant) and reported criminal incidents (20% statistically significant at the 99% level).⁴ However, these general trends mask important heterogeneity in the types of criminal activity possibly affected by the policy. Using two different data sources, we find that reported shots fired dramatically increased (58-67%) in the first two months after the policy. Although this short-term increase is concerning, our models also estimate that the long-term impact of the policy is a 4-5% monthly decrease in crimes involving firearms.⁵

Finally, we were granted access to time series data on how many guns were seized by the SPPD. This is important because, in theory, pretextual stops may represent an important tool used to find illegal guns and remove them from the public. However, we do not find evidence supporting this hypothesis, as our models do not estimate a discernible effect on the short-term number of monthly gun seizures (-6%, not statistically significant), or the long-term trends (1.6%, not statistically significant). This is an important finding to contrast with the finding of a short-term increase in weapon-related calls for service and incidents involving a discharge of a firearm recorded by SPPD. Given that there was no significant change in firearm seizures and how few seizures there were during traffic stops before the policy went into effect, the short-term increase in shots fired after the policy with a continual decline requires more investigation to understand how it might be connected to the policy.

The paper is organized as follows: Section 3 begins with an in-depth description of the policy that will provide the reader with institutional background helpful to understanding

⁴It should be noted that in both cases, the short run increase disappears after 12-18 months.

⁵This is also supported by more sophisticated time series modeling that shows long run reductions in firearm related crimes.

the context of the policy. Section 4 then describes each primary data source and some of the cleaning performed to obtain our final analytic datasets. Section 5 then describes the empirical frameworks used to obtain the results found in Section 6. Section 7 concludes and offers some potential high-level takeaways for policy makers. Throughout the paper we refer to supporting information found in the Appendices A-E.

2 Related Literature

This article contributes to multiple strands of literature that span numerous disciplines from law and economics to criminology. The first literature focuses on court decisions that codified the legal right of law enforcement to make pretextual traffic stops. This literature has tracked state and federal cases and provides rich contextual details and pertinent qualitative observations. While a full review of the law literature is beyond the scope of this article, we highlight a few papers here that informed our work.

To begin, [Harris \(1997\)](#) is an essay that presents an immediate response to the 1996 *Whren* decision. This article outlines the potential harms of increasing the ability of law enforcement to make pretextual stops. Specifically, the author presents four stories that illustrate how law enforcement conducts pretextual traffic stops disproportionately more for minority drivers. The essay ends with two policy recommendations to ameliorate the effect of the *Whren* decision. First, a call is made to police departments to enact agency-wide standards that define well-defined parameters for when traffic stops should be made. Second, the author hypothesizes that the collection of data on traffic stops will allow for more large-scale studies of traffic stops and the issues they raise, and “allow for a more rigorous analysis than I have presented here”. The current study is in fact the direct result of both recommendations, as we rigorously study—using large-scale data—a police department that standardized the reasons officers should make traffic stops. Our results also support the qualitative claim in [Harris \(1997\)](#), that minority drivers are more exposed to pretextual

traffic stops relative to White drivers.

Similarly, Haywood (2023) provides a more recent outline of the history of pretextual policing and “urges policymakers and advocates to identify and pursue reforms to limit pretextual policing without jeopardizing true public safety”. Haywood also calls attention to Virginia—the first state to pass legislation aimed at curtailing pretextual traffic stops. The current article provides a clear framework to investigate the impact of this statewide policy on racial disparities and public safety.

The second literature that our paper contributes to is the racial disproportionality in contact with law enforcement. As mentioned above, scholars have long posited that decisions like *Whren* would widen racial disproportionalites. Recent work has quantified these fears by showing that Black drivers are more likely to be stopped and searched, yet less likely to have searches turn up illegal contraband (Meehan and Ponder, 2002; Gelman et al., 2007; Antonovics and Knight, 2009; Horrace and Rohlin, 2016; Baumgartner et al., 2017; Chohlas-Wood et al., 2018; Pierson et al., 2020; Vito et al., 2020). There is also exceptional work that documents racial bias in police use of force (Hoekstra and Sloan, 2022) and traffic citations (Anbarci and Lee, 2014; West, 2018; Goncalves and Mello, 2021).⁶

The current paper’s analysis of the racial heterogeneity of policy impact is inspired by Rushin and Edwards (2021). Rushin and Edwards (2021) empirically analyzed the racial impact of a 2012 decision made by the Washington Supreme Court that eased restrictions on pretextual stops. The authors find that the decision caused an increase in the number of traffic stops for Black drivers relative to White drivers. Moreover, the authors find that this relative increase was more pronounced in the daytime when race was more salient. The authors conclude that increases in pretextual stops fall disproportionately onto drivers of color, a finding supported by the results in this paper.

The third literature we make contributions to aims to understand the relationship be-

⁶It should be noted that there are also studies that find no evidence of racial bias in Knowles et al. (2001); Grogger and Ridgeway (2006); Weisburst (2023), likely due to heterogeneity across settings, methods, and outcomes tested.

tween traffic enforcement and criminal activity. There are a few papers that directly test this using small sample experiments. Weiss and Freels (1996) analyzes an experiment in Dayton, Ohio aimed at identifying the effect of increased traffic enforcement on traffic accidents, robberies, auto thefts, and arrests. The authors found that treatment in the form of proactive policing (increased traffic stops) did not have any noticeable impact on any aforementioned outcome except “special arrests” (arrests made for DUI, drugs, or weapons offenses).⁷

Josi et al. (2000) presents results from an experiment that also increased traffic enforcement and recorded its impact on crime levels. The study paired four experimental districts with four “similar” control districts. The officers working in the experimental districts were instructed to substantially increase the number of traffic stops made, while no such directive was given to the control groups. After the experiment, the authors analyzed the data and found statistically significant declines in crimes involving larceny, weapons, and loitering in the experimental areas that were not present in the control areas. The result that more traffic stops reduced crimes involving weapons is echoed by the current paper, however, only in the two months immediately following the policy.⁸

3 Description of Policy

The Ramsey County Attorney’s Office crafted their policy to reduce the number of traffic stops made for minor vehicle violations. Their motivation was that these stops offer little value to public safety and place a disproportionate burden on communities of color.⁹ A tragic example of the terrible outcomes that can be triggered by pretextual traffic stops is the shooting of Philando Castile by a St. Anthony Police Department in 2016. This shooting

⁷It should be noted that the study had some concerning limitations driven mainly by the fact that only two sites were compared and there was no measurement of the treatment (traffic stops) before the policy was administered.

⁸Additionally, the treatment in Josi et al. (2000) involved “aggressive enforcement” with increased traffic contacts “by a minimum of 100%” in experimental target areas. This brings into question the external validity of this study.

⁹<https://www.ramseycounty.us/sites/default/files/County%20Attorney/Non-Public-Safety%20Traffic%20Stop%20Press%20Release.pdf>

occurred in the Twin Cities and began driving local press investigations of racial bias in traffic stops.¹⁰ Those investigations, as well as academic researchers, find that pretextual stops are a significant source of racial bias in policing and there is little evidence that such stops are an effective crime prevention strategy.¹¹ In addition to concerns over racial disproportionalities, stops for vehicle violations—such as expired tags—also disproportionately affect lower income individuals. Figure 1a illustrates this relationship in our data, whereby 91% of stops made for vehicle violations are made in block groups in the bottom half of the median income distribution.¹² Finally, prosecutors have long questioned the value of pretextual traffic stops given that most cases arising from such stops are for low-level crimes such as vehicle violations or small amounts of drugs. This intuition is supported by recent literature that suggests pulling the marginal individual into the system adversely effects public safety (Agan et al., 2023; Naddeo, 2022). In sum, pretextual stops tend to disproportionately involve minority drivers, are a potential source of danger to both officers and citizens, and rarely result in arrest and prevention of serious crime. This evidence spurred the Ramsey County Attorney’s Office to begin discussing with the St. Paul Police Department a policy to end the use of non-safety related pretextual traffic stops.¹³

On September 8, 2021, the Ramsey County Attorney’s Office (RCAO) announced that the office would not prosecute most felonies resulting from pretextual traffic stops.¹⁴. RCAO’s policy states that RCAO will not pursue felony charges arising out of a non-public-safety stop where there is not a danger to public safety or the vehicle is not stopped due to a dangerous condition. RCAO announced this policy—in consultation with St. Paul Police

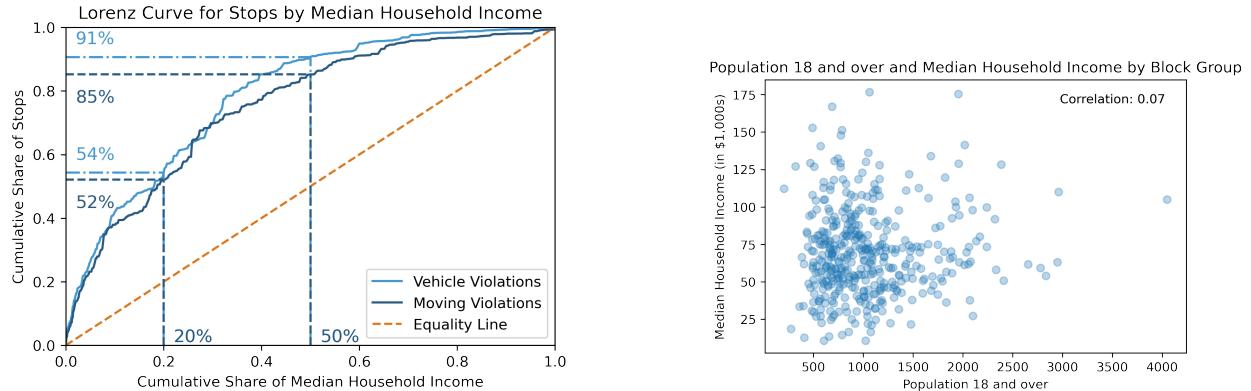
¹⁰See Gottfried and Magan (2021) accessed [here](#).

¹¹See Policing Project at New York University School of Law (2018); Pierson et al. (2020); Rushin and Edwards (2021).

¹²See Tapp and Davis (2022) for nationwide trends. Figure 1b shows that this concentration of stops in lower income block groups is not simply due to those block groups having more residents in the jurisdiction of interest. However, this pattern could also be driven by differences in traffic patterns by the income of the neighborhoods. We leave a deeper investigation for future work.

¹³<https://www.ramseycounty.us/sites/default/files/County%20Attorney/Non-Public-Safety%20Traffic%20Stop%20Press%20Release.pdf>; <https://www.twincities.com/2021/06/12/st-paul-data-shows-black-drivers-nearly-4-times-more-likely-to-be-pulled-over-than-white-drivers/>

¹⁴See policy memo [here](#).



(a) Lorenz curve illustrating what share of stops are made in block groups at X% of the median income distribution

(b) Relationship between median income of Census block group and total population ≥ 18 years old

Figure 1: Data from the American Community Survey 5-Year Data (2015-2019)

Department (by far the largest police department in Ramsey County)—in a coordinated effort to move policing to target other crimes with greater benefits to public safety.¹⁵ Prior to implementation of the policy, vehicle violation stops made up around one third of all stops for SPPD—the second most frequent reason for a stop behind moving violations. These stops are mostly for non-public-safety reasons and thus would likely not be prosecuted under the policy. By decreasing the incentive for officers to make such stops, the police might increase enforcement of more serious crimes, including traffic stops for dangerous driving.

4 Data

For all time series data used in this paper, we group data at the monthly level and examine changes from before to after the policy for various features of interest. Since the policy analyzed was announced on September 8, 2021 the policy only partially affected the month of September 2021. To address this issue, the date associated with each data point and the policy date were shifted back seven days. This aligns the policy start date with the beginning of the month—September 1, 2021—and all monthly data is grouped into “months” that are

¹⁵<https://www.ramseycounty.us/sites/default/files/County%20Attorney/Charging%20Policy%20Regarding%20Non-Public-Safety%20Traffic%20Stops%209.8.21.pdf>

either wholly before or after the policy. This also necessitates removing a small number of observations that are shifted to the month before each analysis period begins. For example in the analysis of stops, some stops shifted into December 2016 and were dropped. Similarly, data in the partial last month in the data, 7 days in June 2023, are removed. These date shifts remove 940 rows as outside the date range January 2017 through May 2023.

4.1 Vehicle Stops Calls For Service

The Ramsey County Emergency Communications Center (ECC) provided calls for service for vehicle stops from January 2016 through September 2022 for the St. Paul Police Department (SPPD). For every call, the data includes the date and time the call started and ended, the latitude and longitude of the incident, the reason for the call, and demographic text comments. The demographic text comments are entered by the call taker when speaking with the officer and are entered as codes for the race and sex of the driver, the reason specified by the officer for the stop and binary information as to whether the person or the car was searched. Each stop is identified with a unique alphanumeric id. Multiple rows of data may share an id where multiple officers called in the same stop or where multiple individuals were in the car and included in the officer's demographics report.

The original data set that we received contained 182,984 stops by SPPD. 29,299 stops that occurred before January 2017 were removed, since the police did not report the reason for the stop in the demographic text comments until December 2016. 6,102 stops with a non-unique id were removed. These duplicate id rows were removed as a precaution, since there were multiple reasons each duplicate might exist. 940 stops were removed after the date shift described above. Finally, 219 stops were removed where the demographic codes used were improperly entered, e.g., where a numeric code for race was expected, but a letter was provided instead. After these steps, the dataset contained 146,424 number of stops—80% of the original dataset. As shown in Table 1 there is substantial heterogeneity in the number of traffic stops made in a month, with some months pre-policy seeing as few as 72

and 175 stops and as many as 742 and 2,837 for vehicle and moving violations, respectively. As in most jurisdictions, Black motorists are disproportionately represented in the stops data (38%), compared to their share of the total population (12%). Another salient pattern that will be returned to repeatedly throughout this paper is the massive decrease in the number of stops for vehicle violations after the policy, with the average dropping from 480 stops per month to 49.

Table 1: Traffic Stops

Demographic	Pre-Policy		Post-Policy	
	Average (Std. Dev.)	Min. - Max.	Average (Std. Dev.)	Min. - Max.
Race				
Asian	233 (105.9)	27 - 540	125 (58.5)	43 - 256
Black	805 (240.7)	178 - 1219	497 (227.5)	175 - 1000
Hispanic	131 (45.0)	19 - 214	107 (64.3)	23 - 275
Native American	11 (4.1)	3 - 19	6 (3.1)	1 - 12
Other	87 (54.1)	6 - 209	33 (14.2)	10 - 67
White	868 (342.3)	85 - 1600	467 (184.9)	158 - 798
Gender				
Female	748 (293.8)	86 - 1393	447 (200.4)	137 - 859
Male	1387 (458.3)	233 - 2206	785 (338.6)	283 - 1494
Non-binary	2 (1.6)	1 - 5	3 (1.9)	1 - 7
Reason For Stop				
Vehicle Violation	480 (154.6)	72 - 742	49 (19.6)	17 - 103
Moving Violation	1537 (638.0)	175 - 2837	1103 (502.9)	359 - 2159
Investigative	102 (22.9)	158 - 170	80 (21.4)	41 - 122
Citizen Report	7 (4.6)	1 - 22	3 (2.5)	1 - 11
No Reason Given	10 (10.4)	1 - 47	3 (1.8)	1 - 7

4.2 911 Calls For Service

The ECC also provided data for all 911 calls initiated by civilians for service from January 2016 through April 2023. This data is used to estimate any changes in crime levels in areas

policed by SPPD. Similar to the data for vehicle stops, every call includes the date and time the call started and ended, the latitude and longitude of the incident, a call type, the disposition of the call, and the agency who responded to the call.

The original dataset contained over 5 million observations from all of Ramsey County. To measure the impact on crime, the data is limited to 911 calls which make up about 2.5 million observations. Of those, about 800,000 calls were removed at the direction of ECC that did not have an id. As described earlier, we shift all dates of calls to seven days earlier to group calls into whole months. We then filter the data to calls from January 1, 2016 to April 30, 2023. After these steps, the dataset contained 1,723,059 911 calls from all of Ramsey County.

In order to focus the analysis on areas policed by SPPD calls were labeled based on the police department that historically policed the surrounding area rather than relying on the agency that responded to the call. To determine what agency historically responded to calls in the area calls were geo-coded to 2010 census block groups, then each census block group was assigned the police department that most frequently responded to calls in that block group in 2020. Finally calls were assigned to police departments based on the census block group from which the call originated. When comparing this labeling to the agency that responded to the call, this results in 1,450 additional calls being labeled as policed by SPPD—less than one percent of all SPPD calls.

After labeling calls by police department and limiting the data to SPPD, there are initially 1,141,987 calls. 619,038 calls were removed at the direction of ECC because they either did not have a call disposition or the call disposition was a “hang-up” or “canceled” call. All of these calls were instances where the call taker was unable to determine if it was a valid 911 emergency call. After removing these rows, the final dataset analyzed contains 522,949 calls.

Each analyzed call includes a call type—a categorical description of the reason for the call. The call “problem” is set by the ECC call taker based upon their initial discussion

with the caller. When the emergency responders actually arrive at the call, they may find that the actual nature of the emergency is different from what the ECC call taker initially thought. The data may not contain the “final” determination of the call problem.¹⁶ In Section C, our main estimates are broken out by call type. Calls for shots fired and calls for weapons are combined because they both include firearm offenses. However, we note that calls for weapons do not all concern firearms, as they are also reports of other weapons such as knives. The calls for weapons comprise 30% of the calls in the shots fired and weapons group in Section C. The 911 hang-up calls included in the table are those calls where the call taker believes there was an emergency, but the caller hung up before the call taker could assess the nature of the emergency.

4.3 St. Paul Police Department Crime Data

The city of St. Paul, Minnesota, maintains an open data portal that includes a criminal incident dataset.¹⁷ The dataset includes publicly reported criminal incidents recorded by SPPD since 2014. We use these data to assess the impact of the policy on reported crime collected by SPPD. We downloaded the data in May 2023, which after our date shift described above means the data extends through April 2023. These data—along with the 911 call data described in Section 4.2—allow us to measure crime with two independent sources. While these two datasets do overlap, neither completely encapsulates the other. The public SPPD crime data includes crimes that police observed or were reported in person, conversely the 911 calls includes incidents where the police arrive but no one is present to report the crime or, based on their assessment, there is not a crime to report. Furthermore, there are instances where the police respond to a 911 call, but after responding the call, categorize the incident as a different type of crime than what was reported by the 911 call taker.

The SPPD incident data include a unique identifier for each incident, the date of the

¹⁶This provides one major difference between the 911 calls and crime incident data provided by SPPD. As the latter is classified by officers after responding to the incident.

¹⁷See [here](#).

incident, and a description of the crime involved. Our historical dataset begins mid-August 2014, thus incidents in the first partial month are dropped. Incidents with duplicate incident ids—3,565 (< 1%) observations total—were removed as a precaution since it is unclear why duplicate ids would exist e.g. multiple victims, multiple crime types, or errant duplication.

Based on the description of the incidents, there are some non-criminal incidents included in the dataset - community events that police held and proactive policing - these are removed for the analysis. Incidents without a crime description are removed because they cannot be properly classified. More than half of incidents from March 2022 did not have a crime description, but did have a crime type, we have replaced those missing crime descriptions with the crime type.

4.4 St. Paul Police Department Firearm Seizures

SPPD provided internally tracked gun seizures since January 2020. The data provide the date the gun was seized and a unique identifier for the gun seizure. From January 2021 onward, the data also include a binary indicator for whether the gun was seized during a traffic stop. The data provided to us extended through October 2022.

5 Empirical Framework

5.1 Interrupted Time Series

This paper aims to measure the effect of a policy on various outcomes. Measuring the true causal effect requires calculating the difference between the realized outcome and a counterfactual outcome in a world where the policy was not enacted. We are never able to actually observe the latter, and therefore we are left trying to come up with our best guess. In this study, this calculation is further complicated by the fact that there is no untreated control group. We are thus left estimating the counterfactual world by modeling the trajectory of outcomes before the policy to make forecasts about what the world would

have looked like post-policy. Finally, we can come up with the effect of the policy by comparing these forecasts with the observed outcomes post-policy. Operationally, we use an interrupted time series (ITS) framework to make these forecasts and compare them to realized outcomes.

ITS requires a number of design choices and assumptions. To begin—and for simplicity—we use a linear model to estimate the effect of the policy. This choice assumes that the outcome followed a linear time trend and that the policy impacts the outcome through a level shift and change in linear trend.¹⁸

As a result of our linearity assumption, we identify the level shift and trend change by estimating the following model via OLS:

$$Y_t = \alpha + \beta_1 T_t + \beta_2 D_t + \beta_3 P_t + \theta W_t + \epsilon_t , \quad (1)$$

where T , D , and P represent a variable that represents the time since the observations began, a binary indicator of whether the observation was collected before ($D = 0$) or after ($D = 1$) the policy was implemented and the time since the policy was enacted, respectively. Therefore, β_2 represents an estimate of the change in level after the policy was implemented and β_3 represents the change in the linear trend after the policy. Neither are inherently causal estimates without assuming that if a policy was never implemented, the outcomes could be reliably predicted using the linear model $Y_t = \alpha + \beta T_t + u_t$ and that nothing different happens in the post-period except the implementation of the policy itself. After discussions with SPPD, concerns were raised that the unseasonably warm weather immediately following the policy may have contributed to changes in our outcome variables. Therefore, we added a measure of temperature W , which is the average of the average daily high and low temperature for each month t .¹⁹

¹⁸In Appendix A we relax these assumptions and perform more rigorous time series modeling to produce higher quality forecasts to show they do not change our qualitative findings.

¹⁹The inclusion of this control did not substantially change any of our estimates. This is likely because there is very little variation in monthly temperature after removing calendar month fixed effects (e.g. the average temperature in September is fairly stable over time).

Theoretical violations of this assumption are plentiful, making claims of causality nearly impossible. Instead, we treat our estimates as the best attempt to quantify the effect of the policy given the constraints of our data. We also augment our model to get closer to plausibly satisfying both of the aforementioned assumptions. First, we remove seasonality by replacing the raw outcome, Y_t (e.g. stops, calls, gun seizures) with the outcome after removing the monthly average. This ensures that we are not mistakenly inferring policy effects that are simply regular seasonal variations. Second, we conduct a more robust time series analysis that takes into account the temporal correlation in the outcomes. This allows us to make more accurate predictions of the counterfactual.

5.2 Difference in Differences

To better understand the racial heterogeneity in the effect of the policy on pretextual stops, we adopt a difference in differences (DiD) framework. Here, we aim to measure the difference before and after the policy in the difference between mean monthly stops of Black and White motorists for vehicle violations. The intuition is that if the policy was effective at reducing pretextual stops, then a change in the difference between Black and White motorists may indicate which group experienced more pretextual stops before the policy was implemented. Formally, to identify the DiD estimator, we estimate the following equation:

$$S_{rt} = \alpha + B_{rt} + D_t + \beta B_{rt} D_t + \tau + \epsilon_{rt}, \quad (2)$$

where S_{rt} represents the total stops made for vehicle violations by SPPD in month t of motorists of race r , B_{rt} represents a dummy for Black motorists, D_t is a binary indicator that is 1 post policy and 0 otherwise, and τ are month and COVID fixed effects. Therefore, β is the DiD estimator for Black motorists, which represents a difference in the policy effects on Black motorists relative to White motorists, after accounting for pre-policy differences between Black and White motorists.

There are a few limitations of our setting that may violate some of the underlying assumptions of this framework, which we would like to highlight here. First, in order for DiD to properly identify the effect of the racial heterogeneity of the policy effect, we rely on a parallel-trends assumption. This assumption states that without the policy, stops of Black and White motorists would have remained on a similar—parallel—trends as observed before the policy (i.e., with no convergence or divergence). For example, a violation of this assumption would be if law enforcement’s behavior was already decreasing their differential use of pretextual stops by race before the policy. A potentially valid example of this would be if law enforcement’s patrolling strategy (or motorists driving behavior/patterns) were affected by the Black Lives Matter movement after the unlawful killing of George Floyd (an unarmed Black civilian), on May 25th, 2020, by police in Minneapolis, Minnesota, which forms the “Twin Cities” with Saint Paul, Minnesota (the jurisdiction of interest). Appendix B provides empirical evidence that, in fact, the parallel trends assumption is reasonable.

To further motivate our results, we look at the heterogeneity of this DiD estimator in a triple difference framework. This framework allows us to conduct a “veil of darkness” test as developed by [Grogger and Ridgeway \(2006\)](#). Here we estimate the following equation:

$$S_{rt} = \alpha + B_{rt} + D_t + L_{rt} + \theta L_{rt} \cdot D_t + \gamma L_{rt} \cdot B_t + \beta^D B_{rt} \cdot D_t + \beta^L L_{rt} \cdot B_{rt} \cdot D_t + \tau + \epsilon_{rt}, \quad (3)$$

where all the variables are the same as in equation 2 except now we include interactions with L_{rt} which is an indicator for stops made before nautical twilight at night and after nautical twilight in the morning—or in plain English—traffic stops made when it is light out.²⁰ Therefore, β^L represents the difference in the DiD estimator β from equation 2 between stops made during the daytime versus those made at night. The theory is that if the differential impact of the policy is larger during the daytime it suggests that the prominence of the race of the driver was a significant driver of pretextual stops pre-policy.

²⁰Formally, nautical twilight is defined when the center of the Sun is between 6° and 12° below the horizon. Given the latitude and longitude and exact time of each stop we calculate the Sun’s position for every stop in our data. Then using the sun’s position we label “light” as the sun being above 12° below the horizon.

6 Results

6.1 Traffic Stops

As detailed in Section 3 the stated goal of the policy was to reduce the number of traffic stops for non-public-safety reasons. Consequently, we begin our analysis by investigating the impact of the policy on the volume and composition of traffic stops. Figure 2 shows the raw time series of stops and searches with the policy date shown. From this raw time series, there is a salient decrease at the time of the policy; however, it should be noted that this decrease seems to be present only in the short run. Figure 2 also illustrates how the composition of the reasons for stops changed significantly with the policy, as the share of stops due to vehicle violations decreased dramatically from around 20-30% of stops before the policy to less than 5% post-policy. Unlike the trend in total stops, there is no noticeable rebound. However, without controlling in some way for seasonality and preexisting time trends, we caution the reader about drawing conclusions based solely on Figure 1. For example, simple comparisons of pre- and post-policy averages are contaminated with global time trends (e.g. more drivers over time causing an increase in stops) as well as seasonal changes (e.g. more drivers on the road in the summer compared to the winter).

To better control for things like time trends, seasonality, and COVID, we model stops using an interrupted time series framework as described in Section 5. First, we regress monthly stops on a vector of indicators for the calendar month (ie, January, February, etc.) and a COVID dummy²¹ and keep the residuals for each month from this model. These residuals represent the deviation from the calendar month-COVID average for each observation or, in another way, represent the variation in stops after removing the monthly seasonality and COVID shock. We then take this variable and use it to estimate equation 1. Figure 3 visualizes a linear ITS model that allows for a level shift and trend shift due to

²¹We define the COVID dummy as an indicator that takes on the value of 1 for the period from March 2020—when a state of emergency was declared—to May 2021, when it was rescinded, and 0 otherwise.

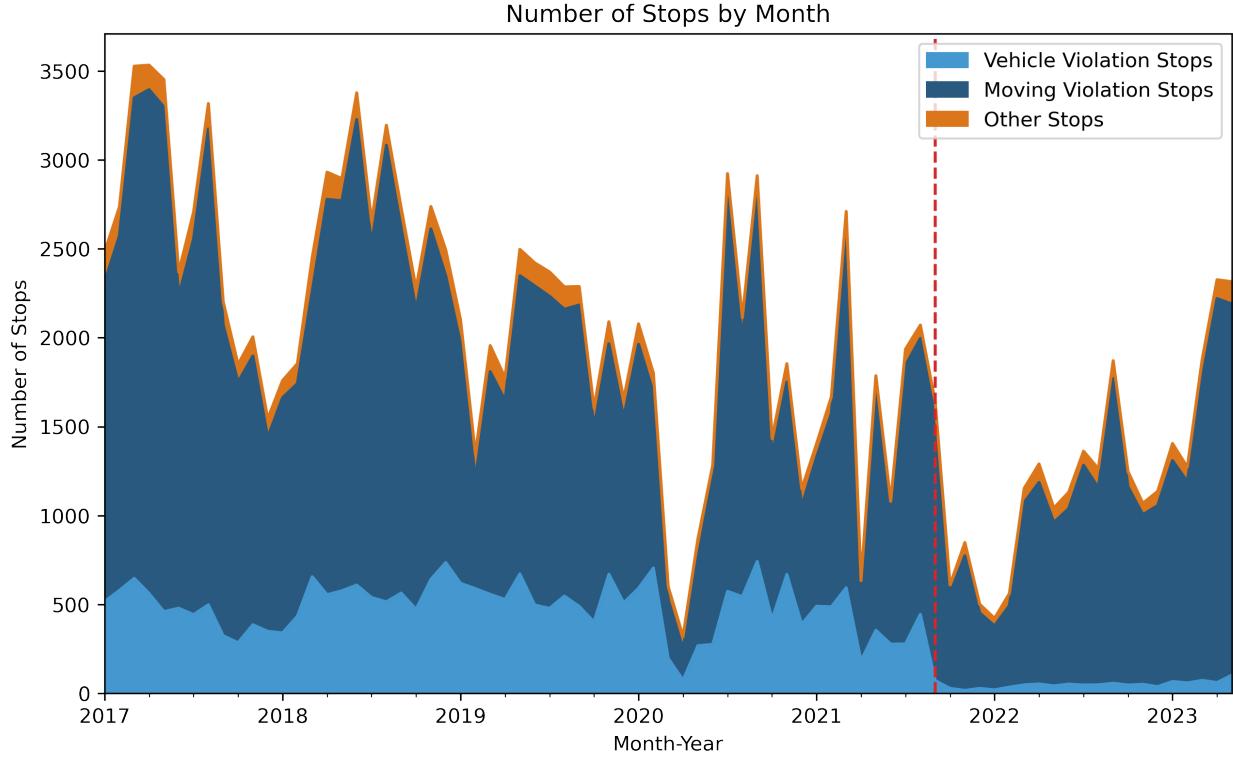


Figure 2: Monthly total traffic stops by category. The policy date is highlighted with a red dashed vertical line at September 2021.

the policy. In addition, we include the full regression output of the model in Table 2.

Table 2: Interrupted Time Series Results

	Vehicle Violations (1)	Moving Violations (2)	Calls (3)	Crimes (4)	Gun Seizures (5)	Crashes (6)	Resp. Time (7)
Post	-352.94** (48.59)	-578.09** (198.42)	170.52 (130.25)	423.18** (128.49)	-2.62 (4.83)	2.72 (16.29)	0.13 (0.09)
Time Since Policy	5.03 (2.76)	65.96** (12.12)	-27.55** (7.31)	-36.98** (10.89)	-0.68 (0.44)	-0.07 (1.43)	-0.04** (0.01)
Time	-2.95** (1.05)	-15.04** (4.06)	8.38** (2.00)	-0.24 (1.03)	0.34 (0.35)	-0.58 (0.37)	0.01** (0.00)
Temperature	-0.66 (0.58)	0.33 (2.13)	0.14 (1.28)	-0.03 (1.30)	0.01 (0.05)	-0.02 (0.19)	-0.00 (0.00)
Pre-policy mean	487	1537	5998	2065	43	257	6
Obs	77	77	87	103	34	74	87
Adj. R ²	0.70	0.39	0.25	0.13	-0.03	0.06	0.54
F-stat	86.28	17.86	18.44	3.21	0.75	1.77	55.29

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

All dependent variables are demeaned by month and COVID shock.

As is clear in both Figure 3 and Column (1) of Table 2 there is a clear downward level shift due to the policy. As shown in column (1) of Table 2 the linear ITS model estimates a

353 stop per month decrease in stops due to the policy. This amounts to an approximately 73% decrease compared to the pre-policy mean of 487 stops per month. Therefore, it seems that the policy all but ceased traffic stops made for vehicle violations.

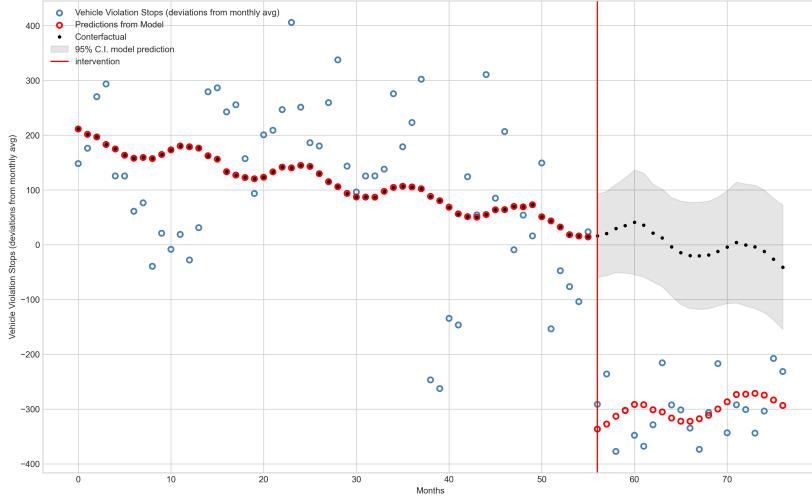


Figure 3: Visualization of ITS analysis of traffic stops for vehicle violations

As evidenced by this quote from the press release announcing the RCAO policy, it is clear that the motivation behind the policy was to reduce the burden of law enforcement contact borne by motorists of color relative to White motorists.

“Drivers of color, and those who are under-resourced who may not be able to afford to make needed repairs, are disproportionately subject to such stops, eroding trust and confidence in the justice system, and among law enforcement and the communities they serve.”

To empirically test the assumption that drivers of color were disproportionately subject to pretextual stops, we estimate equation 2, which tests the effect of the policy on monthly stops of Black drivers relative to the effect on White drivers. The assumption is that the policy acted to remove all stops for vehicle violations that were pretextual in nature, and therefore if we see a larger drop in the number of stops attributed to the policy for a racial group, it implies that they were experiencing more pretextual stops before the policy. Column 1 in

Table 3 shows that after estimating the equation 2 using OLS, we find $\hat{\beta} = -57$ (robust se = 12.4). Therefore, it seems that Black drivers were subject to more pretextual stops before the policy relative to White drivers.

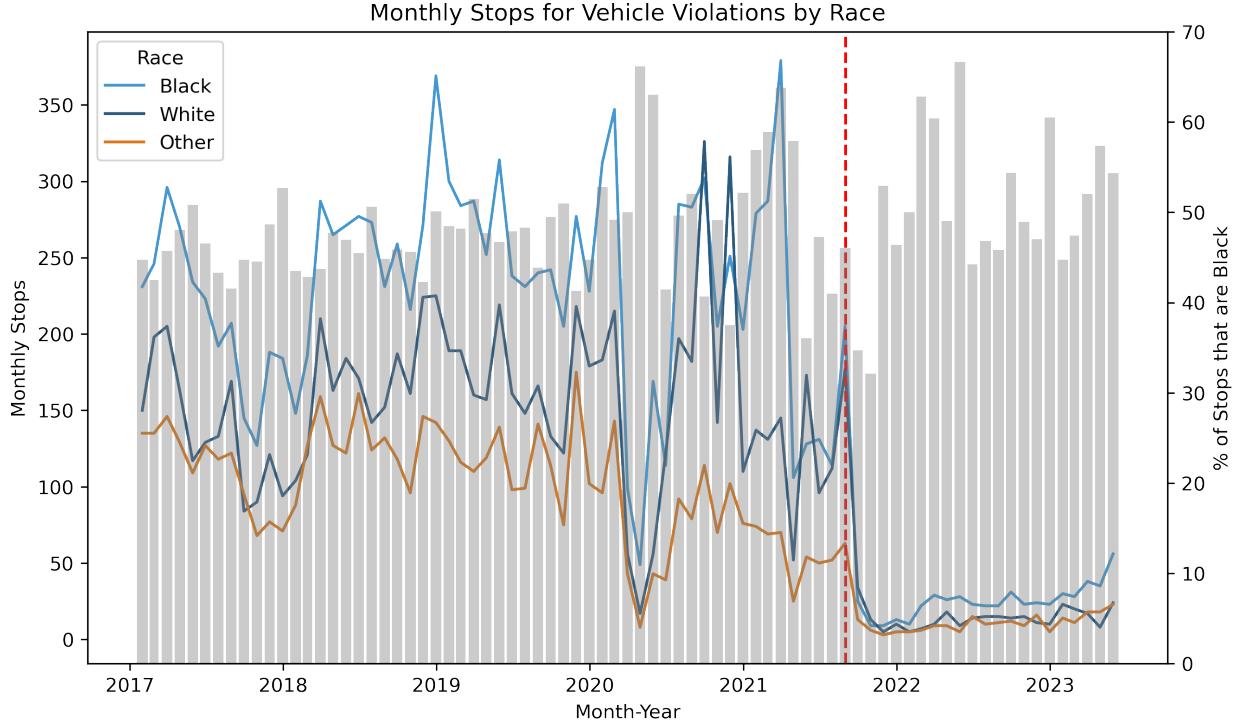


Figure 4: Monthly total traffic stops by race with histogram showing the percentage of all stops made for vehicle violations of Black drivers. The policy date is highlighted with a red dashed vertical line at September 2021.

Column 2 of Table 3 presents estimates of a subset of parameters in equation 3, that is, it shows that $\hat{\beta}^L = -47.4$ (robust se = 11.6) the differential impact of the policy identified in column 1 is driven primarily by an outsized decrease in pretextual stops for Black motorists when it is light out. This result suggests that the racial disproportionality in pretextual stops was larger during the daytime—when it was likely easier for officers to discern the race of drivers—compared to the night. As noted in Horrace and Rohlin (2016), simply using the sun as the only light source revealing a driver’s race can be inaccurate in urban areas well lit by artificial streetlights. As Saint Paul is the 2nd most populous city in the state of Minnesota, it is likely to have a dense network of streetlights that may reduce the “veil

of darkness”. Column 3 presents estimates of a subset of parameters obtained by looking at how β^L and β^D were interpreted vary in areas with different levels of artificial lighting (measured by light intensity measured by satellites, described in Appendix E). To measure heterogeneity by artificial light intensity, we group stops into four equally sized bins that represent the quartiles of nightlight intensity in our sample. We then estimate equation 3 but include another interaction with dummies for each quartile of nightlight intensity. Column (3) of Table 3 presents a subset of the estimates that shows how each of the top three quartiles (and a bin for observations missing location data) compares to the lowest quartile. As is evident from Table 3, there appears to be no clear difference in our estimates across areas with different lighting at night. Discussions with an officer who patrolled for SPPD revealed that there is dense tree cover on most streets in Saint Paul. This may explain why artificial light measured from above (satellites) does not translate to increased visibility on the ground.

Although the number of vehicle violation stops never “rebounded” after the policy, there has been a marked increase in traffic stops due to moving violations (see Figure 2). One worry may be that officers simply began to record stops from vehicle violations to moving violations. Figure 5 shows that stops for moving violations increased equally for Black and White motorists after the policy, making it unlikely that officers were simply substituting vehicle violation stops (which showed large racial disparities) with moving violation stops.

6.2 Crime

Theoretically pretextual stops may be effective means to suppress crime by providing specific and general deterrence. For example, pretextual stops allow more frequent contact with law enforcement and motorists. This increased contact may lead to an increase in the perception of police presence and provides the ability to detect and remove contraband, such as illegal weapons. To investigate whether the policy—which, as shown in Section 6.1 to almost eliminate pretextual traffic stops—correlated with an increase in criminal activity, we utilize

Table 3: Difference in Differences Results

	DiD (1)	Triple DiD (2)	Quadruple DiD (3)
Post Policy x Black	-57.00** (12.38)	-4.80 (4.80)	-0.14 (1.17)
Light Out x Post Policy x Black		-47.39** (11.55)	-6.32 (3.59)
25 th -50 th pctile NL x Light Out x Post Policy x Black			-8.60 (4.67)
50 th -75 th pctile NL x Light Out x Post Policy x Black			-8.83 (4.66)
>75 th pctile NL x Light Out x Post Policy x Black			-2.83 (5.19)
Missing light x Light Out x Post Policy x Black			3.38 (3.84)
25 th -50 th pctile NL x Dark x Post Policy x Black			-1.29 (1.55)
50 th -75 th pctile NL x Dark x Post Policy x Black			-2.07 (1.74)
>75 th pctile NL x Dark x Post Policy x Black			-0.36 (1.92)
Missing light x Dark x Post Policy x Black			-0.23 (1.37)
Mean pre-policy White	139	139	139
Mean pre-policy Black	206	206	206
Obs	154	308	1500
Adj. R ²	0.71	0.79	0.71
F-stat	63.28	57.14	58.88

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

All dependent variables are demeaned by month and COVID shock.

two datasets.

First, we obtained calls for service made by citizens (i.e. 911 calls made by civilians and not police officers). Column (2) of Table 2 shows a statistically insignificant short-term increase in calls for service, which represents only 3% increase from the baseline number of calls received before the policy. Additionally, this short-term increase is offset by a statistically significant negative effect on the long-term trend in 911 calls.

To corroborate these findings, we also use publicly available SPPD data on reported crimes. Column (3) of Table 2 shows similar results to our results using calls for service. In both, there are short-term shifts up and long-term trend shifts down in crime and calls for

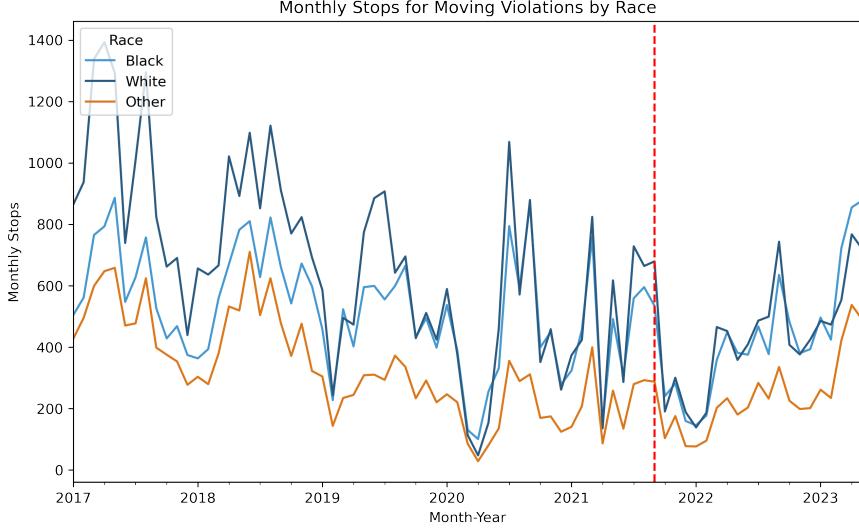


Figure 5: Monthly total traffic stops for moving violations by race. The policy date is highlighted with a red dashed vertical line at September 2021.

service. In summary, it appears that the elimination of pretextual stops was not correlated with large or sustained increases in general criminal activity.

However, this overall effect may mask significant heterogeneity in the effect the policy had on certain types of crime. Tables 7-9 and Tables 10-12 contain estimates produced by iteratively running the same ITS model for each crime category for calls for service and SPPD incidents, respectively. To summarize, we find statistically and economically significant short-term increases in calls for domestic violence / disturbances (33% of pre-policy mean), calls for shots fired or weapons (58%), incidents reported by SPPD of firearm discharges (78%), theft (22%), property damage (26%), and auto theft (75%). Combined with all these short-term increases, we also find decreases in the long-term trends of: 2%, 4%, 5%, 2%, 0.7%, and 5%, respectively. Finally, we found a statistically significant short-term decrease in calls for property crime of 10%, with a further 0.5% (not statistically significant) decrease in the long-term trend. We also examine the dynamics of these effects in Appendix A and find that all the positive effects on crime are concentrated in the 2-3 months post policy, with long run reductions in each crime category.

As alluded to earlier, a clear cost of eliminating the probability of being pulled over

for a vehicle violation is a dramatic decrease in probability one will be caught carrying a firearm/weapon. Therefore, we focus a discussion on the significant increases in calls for service regarding shots fired and weapons (58%) and SPPD incidents regarding discharges of a firearm (78%). In both cases the short-term increase is eroded by a significant decrease in the long-term trend of both calls for shorts fired and weapons (4% per month) and SPPD incidents regarding the discharge of a firearm (5% per month). Finally, we also use data on reported gunshot victims from SPPD. These data record the race, age, and description of how the victim was shot/the outcome of the shooting (e.g. homicide, self-inflicted, suicide). Table 4 summarizes the results of running our analysis over these data—aggregated to the monthly level. Similar to results for shots fired/weapon and discharge of weapon there appears to be a significant short-term increase in the reported number of gun shot victims that is quickly diminished by a negative long-term trend.

Table 4: Interrupted Time Series Results: Gun Shot Victims

	Total Victims	Homicide	Non-Homicide/Suicide	Black	White	Other Race	Age (0, 20]	Age (20, 26]	Age (26, 35]	Age (35, 100]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Post	6.43 (2.84)	-0.03 (0.99)	6.45 (2.59)	3.32 (1.99)	1.59 (1.15)	1.52 (0.97)	0.01 (1.24)	2.11 (1.09)	3.46 (1.56)	0.84 (1.04)
Time Since Policy	-0.54 (0.19)	-0.08 (0.07)	-0.45 (0.16)	-0.37 (0.14)	-0.09 (0.08)	-0.08 (0.06)	-0.08 (0.09)	-0.02 (0.07)	-0.13 (0.11)	-0.26 (0.11)
Time	0.13 (0.04)	0.05 (0.02)	0.05 (0.04)	0.10 (0.03)	0.03 (0.02)	0.01 (0.02)	0.03 (0.03)	0.00 (0.02)	0.02 (0.02)	0.08 (0.02)
Temperature	0.01 (0.02)	-0.00 (0.01)	0.02 (0.02)	0.01 (0.02)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.01)	-0.00 (0.01)
Pre-policy mean	16	2	12	11	2	2	4	4	4	3
Obs	66	66	66	66	66	66	66	66	66	66
Adj. R ²	0.41	0.14	0.26	0.23	0.20	0.09	0.01	0.03	0.17	0.38
F-stat	8.35	4.50	3.40	5.53	4.10	1.76	0.97	1.35	2.26	9.73

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

These results possibly indicate that reallocating officer time away from vehicle violations towards other productive work that reduces or deters firearm-related crimes may take time. This result is further discussed below. Given the trade-off in decreasing traffic stops potentially leading to a decrease in gun seizures, we next examine changes in the number of guns seized by SPPD.

6.3 Gun Seizures

As shown in Section 6.2 there is suggestive evidence that firearm-related incidents increased in the short term after the policy, an effect that decayed (and eventually reversed) over time. Theoretically, the elimination of pretextual stops may also have longer-term impacts on crime by limiting the ability of law enforcement to remove illegal weapons from the street that are used for multiple crimes in the future. To investigate this possibility, we applied the same ITS model to data on monthly gun seizures made by SPPD. Column (4) of Table 2 shows a small imprecise estimate of a decrease in both the long-term trend and the short-term level of monthly gun seizures by SPPD. We note that of all the outcomes we test gun seizures suffers the most from missing data as SPPD only started gathering data beginning in 2020, limiting our analysis to a rather short time series. Furthermore, the number of guns seized per month is relatively small, with quite a bit of variance. For this reason, we caution drawing any firm conclusions from these results but view them as suggestive that monthly gun seizures did not substantially decrease after the elimination of traffic stops for vehicle violations.

6.4 911 Call Response Time

An open question for future research is what police do with the time that would otherwise have been spent making traffic stops. One possibility is that this excess time allows officers to respond to calls for service faster. To test this, we utilized 911 calls for service data and modeled the policies impact on call response time using the same ITS framework.²² Column 7 in Table 2 shows that there is no appreciable short-term change to the median monthly response time and a statistically significant but negligible decrease in long-term trend. Therefore, we do not find strong evidence that the policy allowed police to reduce

²²Response time was measured as the elapsed time from when a call was assigned to an officer until the officer was on the scene.

the time it takes to respond to calls for service.²³

6.5 Traffic Incidents

Finally, we apply the same ITS model to data on traffic incidents to investigate whether reduced traffic stops for vehicle violations decreased the safety of drivers. As column (6) of Table 2 shows, there was a statistically and economically insignificant change in the number of reported traffic accidents. Table 5 breaks the effect of the policy on the total number of incidents into the effect on the number of crashes resulting in death or injury, as well as whether the incidents were caused by alcohol, distracted driving, or speeding. As can be seen from the large standard errors, we lack statistical precision to make any strong claims about how the policy affected these different types of traffic incidents.

Table 5: Interrupted Time Series Results: Traffic Incidents By Type

	Number of Injuries	Number of Fatalities	Alcohol Related	Distraction Related	Speed Related
	(1)	(2)	(3)	(4)	(5)
Post	-6.76 (5.40)	0.35 (0.56)	-0.30 (2.75)	2.66 (1.39)	2.38 (2.13)
Time Since Policy	0.04 (0.43)	-0.04 (0.04)	-0.07 (0.22)	-0.23 (0.11)	-0.04 (0.16)
Time	-0.17 (0.11)	0.01 (0.01)	-0.03 (0.04)	-0.12** (0.03)	-0.04 (0.04)
Temperature	-0.02 (0.06)	0.00 (0.01)	-0.01 (0.03)	0.01 (0.02)	0.00 (0.02)
Pre-policy mean	71	1	18	10	12
Obs	74	74	74	74	74
Adj. R ²	0.18	-0.01	-0.01	0.37	-0.03
F-stat	5.68	0.88	0.85	15.08	0.40

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

7 Discussion

Pretextual traffic stops are seen by some as an invaluable tool for law enforcement and by others as a significant contributor to the racial harm done by the criminal legal system in

²³This may be because the median response time was already an impressive 6 minutes, leaving little room for improvement.

the US. In this paper, our results support both statements to varying degrees. We find that pretextual stops are disproportionately used against Black drivers, and suggestive evidence that this is driven in part by the salience of the drivers' race. We also found that eliminating these stops did not result in massive increases in general criminal activity, traffic incidents, or significant decreases in the number of illegal firearms seized by law enforcement. However, we did find evidence of a significant short-term increase in criminal activity for a few categories of crime. However, our estimates also show that this effect was short-lived and dissipated within 2 months after the policy, with long-term reductions in most crime categories.

Several theories could explain a short-term increase in shots fired and discharges followed by a steady decline in such incidents. Our current hypothesis is that the pattern may be the result of a transition from one equilibrium where pretextual stops featured prominently in law enforcement's toolbox, to one in which they reallocated their time and efforts to other—equally productive—policing tactics. Although we are unable to test this theory, we did speak with SPPD analysts and command staff who stated that there was no office-wide change policy regarding what officers should do with time otherwise spent making traffic stops for vehicle violations.²⁴ However, we concede—as is the case in almost all ITS studies—the results may also be explained by other variables that are correlated with the timing of the policy. For example, in conversations with SPPD it was mentioned that shortly after the policy on October 10th 2021 there was a mass shooting incident that may have generated the short term increase in calls/incidents involving firearms.²⁵ However, after dropping all observations in our data linked to this event we still found nearly identical point estimates. This is because we already attempt to collapse multiple calls for the same incident into one observation.

Accordingly, our normative takeaway from this policy is that it was successful in eliminating a source of racial disproportionality in contact with the criminal legal system, with little evidence of a long-term adverse impact on public safety. We hope that these results

²⁴The lack of an office-wide policy does not rule out officers re-optimizing given an increase "free" time.

²⁵See [here](#).

inspire policy makers throughout the nation to consider similar strategies regarding pretextual traffic stops. Additionally, we hope that other jurisdictions develop these and other policies collaboratively—taking input from both prosecutors and law enforcement—as well as community members. We also hope—albeit selfishly—that when these policies are implemented offices consider empirical testing using rigorous methods that allow for causal claims to be made about their effectiveness. Moving forward, more empirical work should also focus on testing the effectiveness of policies surrounding how additional officer time is redeployed after reducing the time spent enforcing vehicle violations.

References

- Agan, Amanda, Jennifer L Doleac, and Anna Harvey**, “Misdemeanor Prosecution,” *The Quarterly Journal of Economics*, 1 2023.
- Anbarci, Nejat and Jungmin Lee**, “Detecting racial bias in speed discounting: Evidence from speeding tickets in Boston,” *International Review of Law and Economics*, 2014, 38, 11–24.
- Antonovics, Kate and Brian G. Knight**, “A new Look at racial profiling: Evidence from the Boston Police Department,” *Review of Economics and Statistics*, 2009, 91 (1), 163–177.
- Baumgartner, Frank R., Derek A. Epp, Kelsey Shoub, and Bayard Love**, “Targeting young men of color for search and arrest during traffic stops: evidence from North Carolina, 2002–2013,” *Politics, Groups, and Identities*, 2017, 5 (1), 107–131.
- Becker, Gary S**, “Crime and Punishment : An Economic Approach Published by : The University of Chicago Press,” *Journal of Political Economy*, 1968, 76 (2), 169–217.
- Blanks, Jonathan**, “Thin Blue Lies: How Pretextual Stops Undermine Police Legitimacy,” *Case Western Reserve Law Review*, 2016, 66 (4), 931–946.
- Chohlas-Wood, Alex, Sharad Goel, Amy Shoemaker, and Ravi Shroff**, “An analysis of the Metropolitan Nashville Police Department’s traffic stop practices,” *Technical report, Stanford Computational Policy Lab.*, 2018, pp. 1–10.
- Doleac, Jennifer L**, “Racial Bias in the Criminal Justice System,” 2022.
- Gelman, Andrew, Jeffrey Fagan, and Alex Kiss**, “An analysis of the New York City police department’s “stop-and- frisk” policy in the context of claims of racial bias,” *Journal of the American Statistical Association*, 2007, 102 (479), 813–823.
- Goncalves, Felipe and Steven Mello**, “A few bad apples? racial bias in policing,” *American Economic Review*, 2021, 111 (5), 1406–1441.
- Gottfried, Mara H. and Christopher Magan**, “Black drivers nearly 4 times more likely to be pulled over than white drivers,” 12 2021.
- Grogger, Jeffrey and Greg Ridgeway**, “Testing for racial profiling in traffic stops from behind a veil of darkness,” *Journal of the American Statistical Association*, 2006, 101 (475), 878–887.
- Harris, David A.**, ““Driving while black” and all other traffic offenses: The supreme court and pretextual traffic stops,” *Journal of Criminal Law and Criminology*, 1997, 87 (2), 544.
- Haywood, Bradley R**, “Ending Race-Based Pretextual Stops: Strategies For Eliminating America’s Most Egregious Police Practice,” *Richmond Public Interest Law Review Volume*, 2023, 26 (1), 47–84.
- Hoekstra, Mark and Carly Will Sloan**, “Does Race Matter for Police Use of Force? Evidence from 911 Calls,” *American Economic Review*, 2022, 112 (3), 827–860.
- Horrace, William C. and Shawn M. Rohlin**, “How dark is dark?: Bright lights, big city, racial profiling,” *Review of Economics and Statistics*, 2016, 98 (2), 226–232.
- Josi, Don A, Michael E Donahue, and Robert Magnus**, “Conducting blue light specials or drilling holes in the sky: are increased traffic stops better than routine patrol in taking a bite out of crime?,” *Police Practice*, 2000, 1 (4), 477–507.
- Kirkpatrick, David D., Steve Eder, and Kim Barker**, “Cities Try to Turn the Tide on Police Traffic Stops - The New York Times,” 4 2022.

- Knowles, John, Nicola Persico, and Petra Todd**, “Racial Bias in Motor Vehicle Searches: Theory and Evidence,” *Journal of Political Economy*, 2001, 109 (1), 203–229.
- Lauer, Claudia**, “Police union sues over Philadelphia ban on low-level stops,” 2 2022.
- Meehan, Albert J. and Michael C. Ponder**, “Race and place: The ecology of racial profiling African American motorists,” *Justice Quarterly*, 2002, 19 (3), 399–430.
- Naddeo, J.J.**, “Race , Criminal History , and Prosecutor Case Selection : Evidence from a Southern U.S. Jurisdiction,” 2022.
- Pierson, Emma, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff, and Sharad Goel**, “A large-scale analysis of racial disparities in police stops across the United States,” *Nature Human Behaviour*, 2020, 4 (7), 736–745.
- Policing Project at New York University School of Law**, “An Assessment of Traffic Stops and Policing Strategies in Nashville,” Technical Report 2018.
- Román, Miguel O., Zhuosen Wang, Qingsong Sun, Virginia Kalb, Steven D. Miller, Andrew Molthan, Lori Schultz, Jordan Bell, Eleanor C. Stokes, Bharatendu Pandey, Karen C. Seto, Dorothy Hall, Tomohiro Oda, Robert E. Wolfe, Gary Lin, Navid Golpayegani, Sadashiva Devadiga, Carol Davidson, Sudipta Sarkar, Cid Praderas, Jeffrey Schmaltz, Ryan Boller, Joshua Stevens, Olga M. Ramos González, Elizabeth Padilla, José Alonso, Yasmín Detrés, Roy Armstrong, Ismael Miranda, Yasmín Conte, Nitza Marrero, Kytt MacManus, Thomas Esch, and Edward J. Masuoka**, “NASA’s Black Marble nighttime lights product suite,” *Remote Sensing of Environment*, 2018, 210 (April), 113–143.
- Rushin, Stephen and Griffin Edwards**, “An empirical assessment of pretextual stops and racial profiling,” *Stanford Law Review*, 2021, 73 (3), 637–726.
- Skogan, Wesley G.**, “Stop-and-frisk and trust in police in Chicago,” 2016.
- Smith, Taylor G and others**, “{pmdarima}: ARIMA estimators for {Python}.”
- Tapp, Susannah N and Elizabeth J Davis**, “Contacts between police and the public, 2020,” Technical Report November, Bureau of Justice Statistics 2022.
- Tyler, Tom, Jeffrey Fagan, and Amanda Geller**, “Street Stops and Police Legitimacy: Teachable Moments in Young Urban Men’s Legal Socialization,” *Journal of Empirical Legal Studies*, 2014, 11 (4), 751–785.
- Vito, Anthony G., Vanessa Woodward Griffin, Gennaro F. Vito, and George E. Higgins**, ““Does daylight matter”? An examination of racial bias in traffic stops by police,” *Policing*, 2020, 43 (4), 675–688.
- Weisburst, Emily**, ““Whose help is on the way ?” The importance of individual police officers in law enforcement outcomes,” *Journal of Human Resources*, 2023, 58 (3).
- Weiss, Alexander and Sally Freels**, “The effects of aggressive policing: the Dayton traffic enforcement experiment,” *American Journal of Police*, 1996, 15 (3), 45–64.
- West, Jeremy**, “Racial Bias in Police Investigations,” 2018.
- Woods, Jordan Blair**, “Policing, danger narratives, and routine traffiffic stops,” *Michigan Law Review*, 2019, 117 (4), 635–712.
- Wu, Xiaoyun and Cynthia Lum**, “The practice of proactive traffic stops,” *Policing*, 2020, 43 (2), 229–246.

Supplementary Details

A Robustness of ITS to Forecasting Methods

In Section 5.1 we define a simple ITS framework that utilizes a linear model to identify the effect of the policy. In this section, we test whether this assumption drives the results presented in Section 6. To better understand the consequences of using a linear model, we turn to modeling each time series using an autoregressive integrated moving average, or ARIMA model. These models use differencing as well as temporal lags of the dependent variable and error term to use a time series to forecast future periods.

Operationally, we use the *pmdarima* Python package to determine the ARIMA model that best captures the variation in the monthly time series before the policy Smith and others (n.d.). Columns 5 and 6 in Table 6 summarize the ARIMA model parameters selected that minimize the Akaike information criterion (AIC). After estimating these models, we then apply them to post-policy time series to obtain a forecast for each month post-policy. Figure 6 represents the difference between the forecasts using pre-policy data and the actual observed outcomes. The differences for the total post period and the short-term 3 month window are also summarized in columns 3 and 4 of Table 6, respectively.

The results in Table 6 present similar patterns as found in the main text but suggest a much less dramatic short-term increase in calls for service for firearms / weapons (35%) and incidents involving the discharge of a weapon (40%). Figure 6 shows how the short-term increase dissipates shortly after the shaded 3 month window, while the number of traffic stops for vehicle violations remains virtually eliminated.

B Parallel Trends

A key assumption underlying any difference in differences design is parallel trends. In the setting of this paper, this assumption translates to assuming that stops of Black and White

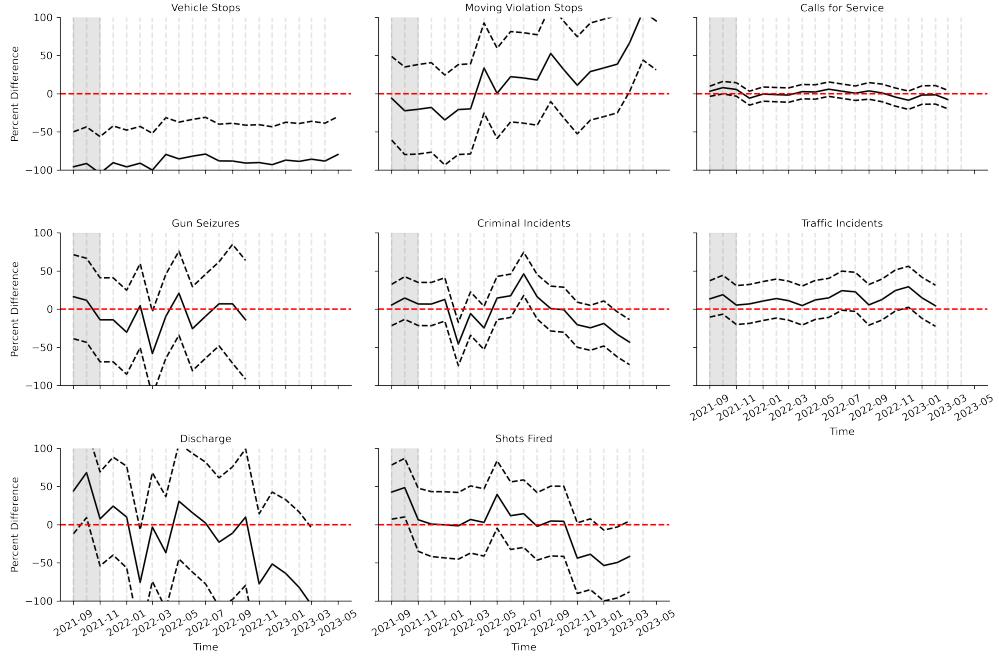


Figure 6: Difference between ARIMA prediction (using pre-policy data) and actual observed outcome as a percent of pre-policy mean. The gray shading represents the short-term 3 month window post-policy.

Table 6: Results from ARIMA Analysis

Outcome	Pre-Policy Mean	Overall	Short-Term	ARIMA Order	Seasonal Order
Vehicle Stops	487 (19)	-435 (7)	-473 (19)	(1, 0, 0)	(0, 0, 1, 12)
Moving Violation Stops	1537 (86)	308 (129)	-249 (79)	(2, 1, 2)	(0, 0, 1, 12)
Calls for Service	5998 (95)	11 (63)	334 (82)	(1, 0, 0)	(1, 1, 0, 12)
Seizures	43 (2)	-3 (2)	2 (4)	(0, 0, 0)	(0, 1, 0, 12)
Crimes	2065 (42)	-82 (113)	185 (59)	(1, 0, 0)	(0, 0, 2, 12)
Traffic Incidents	257 (6)	36 (5)	32 (10)	(0, 1, 2)	(1, 0, 0, 12)
Discharge	101 (6)	-17 (11)	40 (18)	(0, 1, 1)	(1, 0, 1, 12)
Shots Fired	148 (6)	-4 (10)	48 (19)	(2, 0, 0)	(0, 0, 2, 12)

Notes: Short-term refers to the average effect for the 3 months post policy. All means have standard errors in parentheses.

motorists for vehicle violations would have continued to trend as before the policy if the policy was in fact never implemented. As we never observe a world in which the policy did not occur, there is no direct way to empirically test this assumption. Instead, we rely on testing trends in each group prior to the policy to build a case that the parallel-trends assumption is reasonable. To begin, Figure 7 shows the number of monthly stops for vehicle violations by race after removing the monthly averages and COVID shock (to remove seasonality and

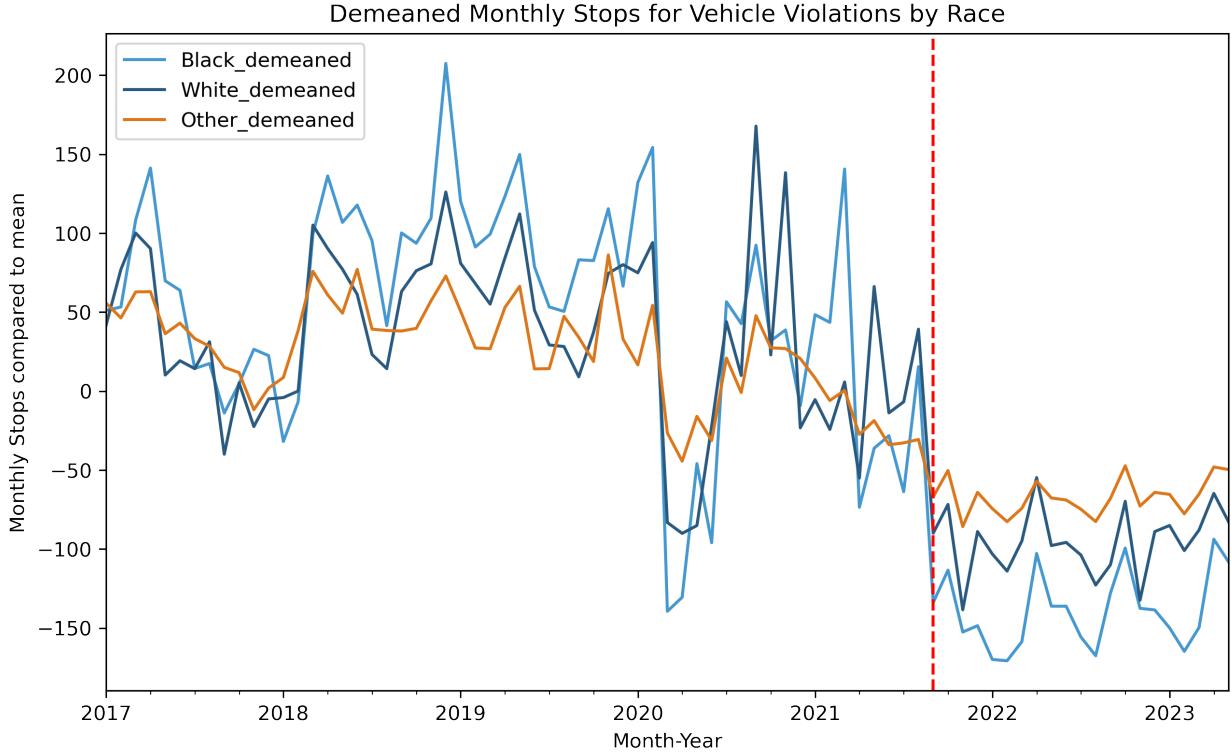


Figure 7: Total stops per month after removing the calendar month mean and COVID mean. The policy date is highlighted with a red dashed vertical line at September 2021.

COVID shock common between races). From this Figure it seems like the time series for Black and White individuals follow similar trends, satisfying the “eye test” that assuming the trend to continue is reasonable. To more rigorously test this, we estimate the following equation:

$$S_{rt} = \alpha + \sum_{r \in \{B,W,H\}} R^r + Time_t + \theta^r \sum_{r \in \{B,W,H\}} R^r \cdot Time_t + \tau + \epsilon_{rt}, \quad (4)$$

over only data before the policy, where S_{rt} represents the total stops made for vehicle violations by SPPD in month t of motorists of race r , R^r represents a dummy for race (with omitted reference category), $Time_t$ is a running variable measuring the time since the first observational period, and τ are month fixed effects. Therefore, θ^B and θ^H represent differential linear time trends for Black and Hispanic motorists, respectively. Therefore, testing if θ^B or θ^H is significantly different from zero represents a test of whether there are differ-

ences in linear trends between races. Our estimates for $\hat{\theta}^B$ and $\hat{\theta}^H$ are -0.4219 (robust se = 0.769, p-value=0.583) and -0.1874 (robust se = 0.491, p-value=0.703), both statistically insignificant from zero.²⁶ To ensure that this result is not simply due to the assumption of a linear time trend, we also tested for differences up to the quartic time trend and found no statistically significant differences between races.

C Results by Type

C.1 911 Calls

Table 7: Interrupted Time Series Results: 911 Calls By Type

	Disorderly Conduct/Suspicious Activity (1)	Property (2)	Support - Person, Alarm, or Agency (3)	911 hangup (4)	Other (5)
Post	215.69 (83.53)	-86.15* (28.83)	-28.22 (26.88)	-88.79 (47.90)	4.64 (16.95)
Time Since Policy	-18.98** (4.35)	-4.12 (2.17)	-2.82 (1.92)	12.03 (5.22)	-0.15 (1.37)
Time	6.45** (1.35)	-2.10** (0.49)	2.20** (0.39)	1.96** (0.29)	0.39 (0.19)
Temperature	0.14 (0.72)	-0.52 (0.33)	0.05 (0.42)	0.28 (0.33)	0.03 (0.18)
Pre-policy mean	1551	882	647	560	367
Obs	87	87	87	87	87
Adj. R ²	0.46	0.72	0.17	0.55	0.05
F-stat	61.74	62.99	10.89	18.40	2.49

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

²⁶It should be noted that while these estimates are statistically indistinguishable they are rather large compared to the point estimate of the time trend for White motorist 0.0177 (robust se = 0.472, p-value=0.970). Therefore, it is a little unclear whether with more statistical precision (from a longer time series or better functional form) we would be able to reject common prior linear trends.

Table 8: Interrupted Time Series Results: 911 Calls By Type

	Domestic Violence/Disturbance (1)	Investigatory (2)	Civil Problem (3)	Other Crime (4)	Crime involving injury/violence (5)
Post	109.03** (15.40)	-18.88 (13.42)	-31.96 (15.32)	-16.30 (10.08)	15.26 (6.10)
Time Since Policy	-4.87** (0.71)	-1.44 (1.29)	1.65 (1.02)	-1.36 (0.75)	-1.57** (0.32)
Time	2.04** (0.31)	0.91** (0.17)	-2.44** (0.24)	-0.48** (0.14)	0.07 (0.11)
Temperature	0.25 (0.21)	-0.02 (0.15)	-0.08 (0.18)	-0.06 (0.11)	0.01 (0.08)
Pre-policy mean	331	290	296	237	187
Obs	87	87	87	87	87
Adj. R ²	0.80	0.21	0.78	0.52	0.02
F-stat	276.88	9.98	84.41	36.15	7.27

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

Table 9: Interrupted Time Series Results: 911 Calls By Type

	Civil Ordinance Issue (1)	Vehicle (2)	Shots Fired/Weapon (3)	Uncategorized (4)	Drugs (5)
Post	3.40 (12.42)	-2.10 (11.69)	85.86** (9.45)	16.34 (7.01)	-7.32 (5.06)
Time Since Policy	0.94 (0.75)	-1.63 (0.93)	-6.08** (0.54)	0.73 (0.68)	0.09 (0.34)
Time	-0.82** (0.19)	0.01 (0.12)	0.00 (0.17)	0.26 (0.16)	-0.08 (0.10)
Temperature	0.04 (0.16)	-0.07 (0.10)	0.05 (0.11)	0.02 (0.12)	-0.00 (0.07)
Pre-policy mean	192	177	148	69	63
Obs	87	87	87	87	87
Adj. R ²	0.24	0.12	0.36	0.23	0.06
F-stat	15.13	5.07	46.82	12.03	4.90

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

C.2 Criminal Incidents

Table 10: Interrupted Time Series Results: Criminal Incidents By Type

	Proactive Policing (1)	Theft (2)	Property Damange (3)	Auto Theft (4)
Post	-459.67 (304.19)	176.76* (57.60)	70.70* (22.09)	173.27** (40.29)
Time Since Policy	-98.79** (15.71)	-17.28** (4.59)	-2.34 (2.25)	-11.10** (2.92)
Time	34.03** (4.10)	0.08 (0.48)	-0.16 (0.18)	0.89** (0.23)
Temperature	-2.89 (4.22)	-0.07 (0.57)	0.11 (0.23)	0.04 (0.29)
Pre-policy mean	2090	814	273	231
Obs	103	103	103	103
Adj. R ²	0.41	0.13	0.12	0.54
F-stat	24.89	3.58	3.63	12.56

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

Table 11: Interrupted Time Series Results: Criminal Incidents By Type

	Narcotics (1)	Burglary (2)	Discharge (3)	Domestic Assault (4)
Post	-35.63 (13.41)	1.61 (16.89)	79.08** (11.05)	-27.54 (10.07)
Time Since Policy	0.36 (1.40)	-0.84 (1.02)	-5.39** (0.70)	-0.11 (0.67)
Time	-0.62** (0.16)	-0.60* (0.20)	0.38 (0.15)	0.12 (0.07)
Temperature	-0.09 (0.17)	-0.04 (0.16)	0.08 (0.12)	-0.03 (0.08)
Pre-policy mean	196	191	101	100
Obs	103	103	103	103
Adj. R ²	0.45	0.18	0.38	0.19
F-stat	24.03	9.55	50.41	9.04

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

Table 12: Interrupted Time Series Results: Criminal Incidents By Type

	Community Event (1)	Violent Crime (2)	Robbery (3)	Agg Assault Dom (4)	Other (5)
Post	-99.23** (19.20)	0.08 (4.89)	-8.90 (5.33)	-6.25 (3.37)	-0.03 (0.04)
Time Since Policy	-1.34 (0.68)	-0.45 (0.41)	0.12 (0.40)	0.05 (0.29)	-0.00 (0.00)
Time	1.51** (0.31)	0.04 (0.06)	-0.30** (0.06)	-0.07 (0.03)	0.00 (0.00)
Temperature	-0.21 (0.40)	-0.00 (0.05)	-0.01 (0.06)	-0.01 (0.04)	-0.00 (0.00)
Pre-policy mean	88	67	58	35	0
Obs	103	103	103	103	103
Adj. R ²	0.18	-0.02	0.46	0.18	-0.03
F-stat	7.42	0.44	21.70	6.02	0.84

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

C.3 Response Time to 911 Calls

Table 13: Interrupted Time Series Results: Response Time to Calls By Type

	Disorderly Conduct/Suspicious Activity (1)	Property (2)	Support - Person, Alarm, or Agency (3)	911 hangup (4)	Other (5)
Post	0.37* (0.11)	-0.41 (0.18)	0.18 (0.15)	0.15 (0.24)	-0.37 (0.22)
Time Since Policy	-0.05** (0.01)	-0.03 (0.01)	-0.03 (0.01)	-0.03 (0.02)	-0.00 (0.02)
Time	0.01** (0.00)	0.01** (0.00)	0.01** (0.00)	0.02** (0.00)	0.02** (0.00)
Temperature	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Pre-policy mean	5.27	6.73	5.23	5.88	5.99
Obs	87	87	87	87	87
Adj. R ²	0.60	0.15	0.14	0.53	0.34
F-stat	83.62	7.22	4.17	27.27	17.37

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

Table 14: Interrupted Time Series Results: Response Time to Calls By Type

	Domestic Violence/Disturbance (1)	Investigatory (2)	Civil Problem (3)	Other Crime (4)	Crime involving injury/violence (5)
Post	0.53* (0.16)	0.38 (0.25)	0.43 (0.20)	0.01 (0.35)	0.02 (0.16)
Time Since Policy	-0.04* (0.01)	-0.04 (0.02)	-0.05 (0.02)	-0.06 (0.03)	-0.02 (0.02)
Time	0.01** (0.00)	0.02** (0.00)	0.02** (0.00)	0.02** (0.00)	0.01** (0.00)
Temperature	0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Pre-policy mean	5.95	6.17	5.32	6.33	4.88
Obs	87	87	87	87	87
Adj. R ²	0.42	0.40	0.50	0.21	0.08
F-stat	28.73	13.06	26.92	7.71	3.74

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

Table 15: Interrupted Time Series Results: Response Time to Calls By Type

	Civil Ordinance Issue (1)	Vehicle (2)	Shots Fired/Weapon (3)	Uncategorized (4)	Drugs (5)
Post	0.400 (0.347)	0.228 (0.312)	0.772** (0.193)	-0.027 (0.032)	1.231 (0.836)
Time Since Policy	-0.051 (0.026)	-0.063 (0.033)	-0.055 (0.022)	0.002 (0.003)	-0.141 (0.066)
Time	0.019** (0.003)	0.013** (0.003)	0.014** (0.002)	-0.000 (0.001)	0.015* (0.006)
Temperature	0.000 (0.003)	-0.002 (0.003)	0.000 (0.002)	0.000 (0.000)	-0.000 (0.006)
Pre-policy mean	5.41	6.94	3.95	0.02	5.8
Obs	87	87	87	87	87
Adj. R ²	0.437	0.131	0.568	-0.036	0.178
F-stat	20.401	5.492	39.791	0.633	2.899

Robust standard errors in parentheses ** $p < 0.01$, * $p < 0.05$

Note: p-values adjusted for multiple hypothesis using the Benjamini-Hochberg procedure

All dependent variables are demeaned by month and COVID shock.

D Employment Level

A potential reason for the drop in stops made could be a decrease in deployable officers around the same time as the policy. To test this we received data on the number of deployable officers from SPPD. Unfortunately, this data was only available starting in September 2020 and therefore we were unable to add it as a control in our main models without seriously shortening our panel. Figure 8 shows the general trend of stops made per officer. As in the main article, there was a precipitous decrease in stops per officer made for vehicle violations. This supports the idea that it was not simply a decrease in total workforce but a change in the portfolio of stops being made by individual officers.

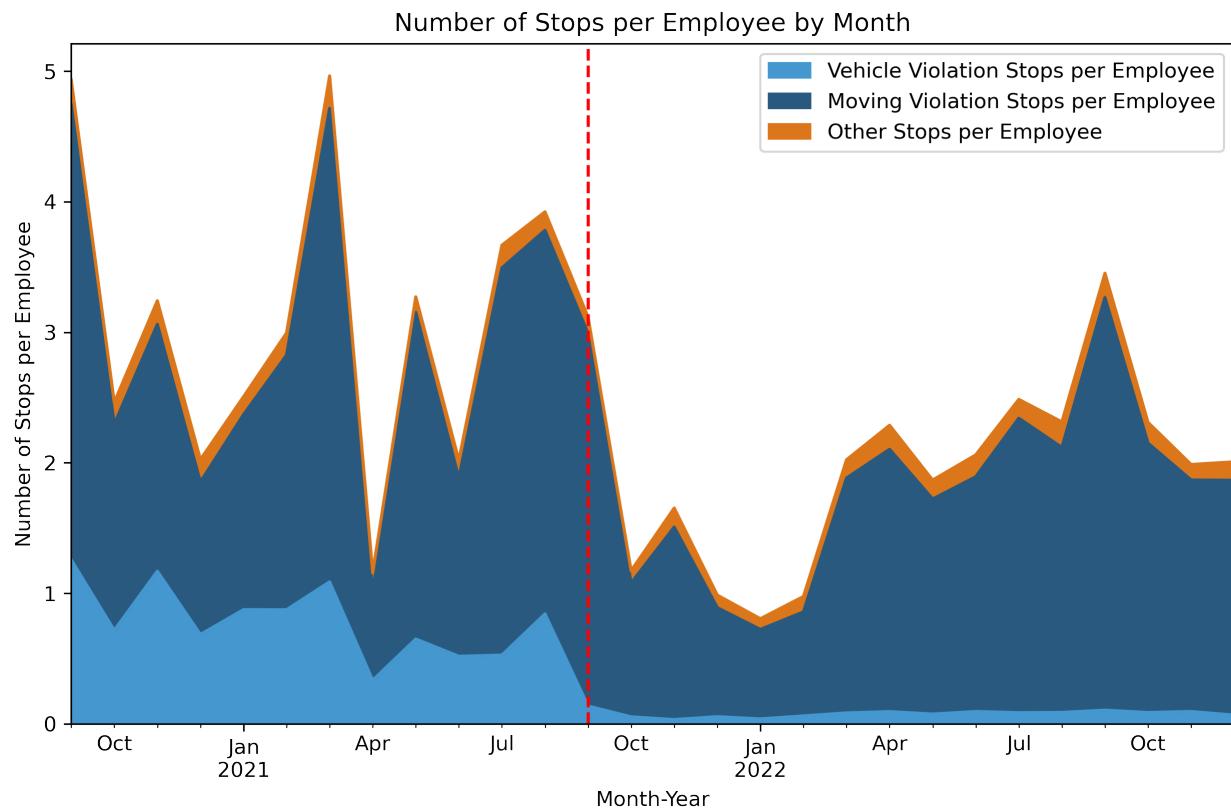


Figure 8

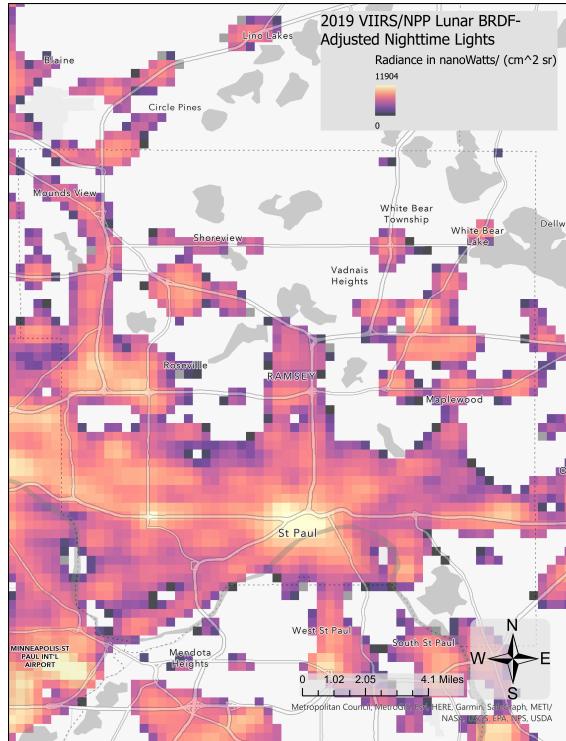
E Nightlight Data

As suggested in [Horrace and Rohlin \(2016\)](#), the veil-of-darkness hypothesis can be impacted in larger urban areas that have a dense network of streetlights. To test whether our results regarding the difference in the racial gap in policy effectiveness between the daytime and nighttime are different in areas with more/less artificial nightlight.

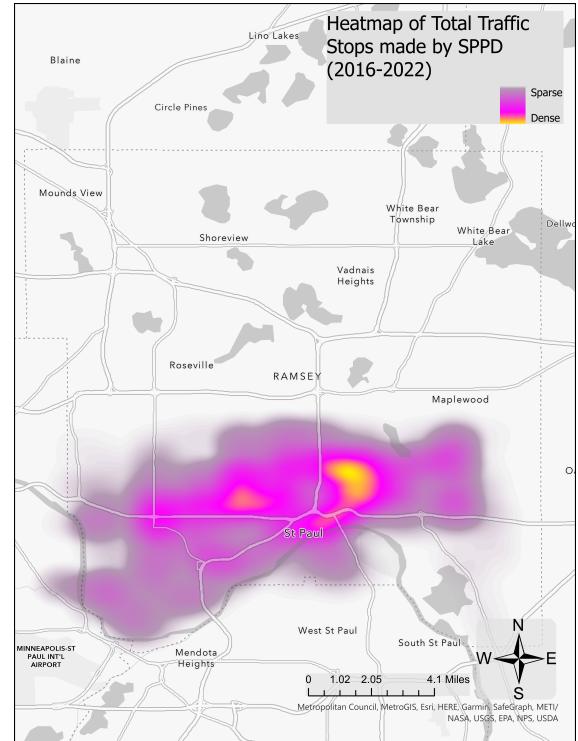
To measure artificial light at night we turn to NASA's Black Marble data, which are derived from measurements taken from the visible Infrared Imaging Radiometer Suite (VI-IRS) instrument, a component of the Suomi National Polar-orbiting Partnership (NPP). As shown in [Román et al. \(2018\)](#) the construction of Black Marble required a massive amount of effort and expertise to correct for environmental factors among other sources of measurement error. For our purposes, we utilize the VNP46A4 product which provides annual average measurements of nighttime light (at approximately 1:30am CST) at the pixel level, with a pixel measuring approximately 0.5 mile by 0.5 mile.

After obtaining annual measurements from 2012-2022 we then overlay each traffic stop location and find the nightlight value for each.²⁷ The nightlight distribution of our sample was remarkably log-normal, and therefore we took the natural log of the nightlight value and classified each stop by the quartile of $\ln(\text{nightlight})$ in which the stop fell.

²⁷Operationally we use a bi-linear interpolation method to create a smooth surface of light from the pixel values, and then pull the value for each (latitude, longitude) that represents a stop.



(a) Average nightlight in 2019



(b) Heatmap of traffic stops made by SPPD

Figure 9: Maps of NTL and traffic stops, showing clear correlation between areas that are well lit and traffic stops.

F Weather

Discussions with SPPD brought up the potential for unseasonably nice weather to affect crime and traffic patterns. To account for this we pulled climate data for Saint Paul from the Minnesota Department of Natural Resources (MN DNR).²⁸ These data allowed us to construct and average monthly temperature by first finding the midpoint for each day between the high and low temperature and then taking the monthly mean.²⁹ We then include this variable as a control in our analysis. Likely because we already isolate to variation within calendar month (using month fixed effects) the addition of this control does not change our results and is not correlated with any of our outcomes of interest. Figure 10 illustrates the deviations from calendar month mean temperature. There does not seem to be any interesting time trends or abnormal temperature changes around the policy implementation.

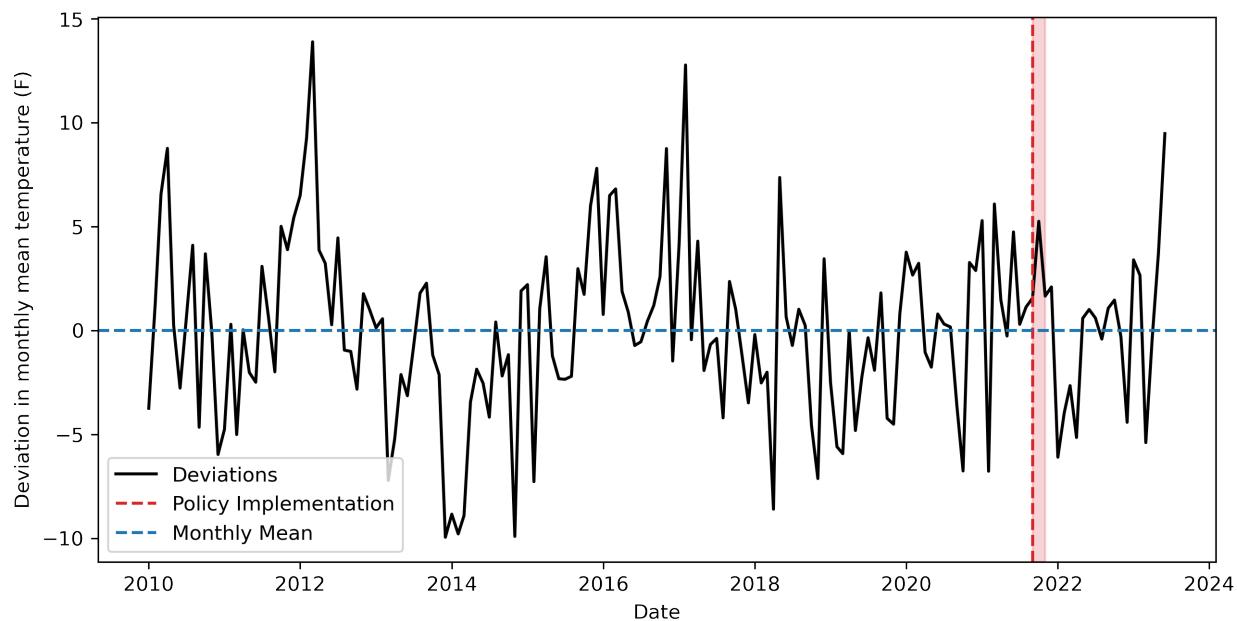


Figure 10: Deviations in monthly mean temperature

²⁸ Accessed [here](#).

²⁹We also calculated the median and included with the mean the standard deviation, neither of which changed the results.