ESSAYS ON RACIAL DISPARITIES IN LAW ENFORCEMENT

by

MELISSA WILSON

A DISSERTATION

Presented to the Department of Economics

and the Division of Graduate Studies of the University of Oregon

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

June 2022

DISSERTATION APPROVAL PAGE

Student: Melissa Wilson

Title: ESSAYS ON RACIAL DISPARITIES IN LAW ENFORCEMENT

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Economics degree in the Department of Economics by:

Dr. Benjamin Hansen	Chairperson
Dr. Jonathan Davis	Core Member
Dr. Eric Zou	Core Member
Dr. Kristen Bell	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2022

DISSERTATION ABSTRACT

Melissa Wilson

Doctor of Economics

Department of Economics

June 2022

Title: ESSAYS ON RACIAL DISPARITIES IN LAW ENFORCEMENT

Based on recent news and social media discussions, racial bias in police action is currently at the forefront of public interest in the U.S. Whether police operate outside of what is considered fair under our justice system can be challenging to estimate.

I first analyze the incidence of racial bias in traffic stops by city police departments in 9 cities across the country using the "veil of darkness" hypothesis developed in Grogger and Ridgeway (2012). Utilizing SOPP data, I employ a regression discontinuity design around the start of daylight savings time to make an accurate comparison between daylight and nighttime stops drawn from the same distribution of drivers in order to address seasonality bias in their model. I find little evidence of racial disparities in police stops with no significance for minority drivers. This indicates that daylight times do not affect proportion of stops of minority drivers and racial disparities are not affected by visible lighting.

Stanford Open Policing Project

Next, I test the effect of marijuana decriminalization in Illinois on racial disparities in arrests for marijuana possession in Chicago and provide evidence to support that the disparity is driven by racial prejudice. I use amount in possession to determine if racial discrimination affects police decision-making at varying severity levels differentially. I then conduct an Interrupted Time Series estimation to show that marijuana decriminalization led to a substantial drop in the racial disparity for marijuana-related arrests in Chicago. Additionally, the slope changes slightly positive, driven by arrests of black individuals over the decriminalized amount. This implies a shift in resources to target higher severity drug crime activity, but still disproportionately affects black individuals.

Almost universally, drug crimes carry sanctions which vary across weight. A Beckerian model of crime has sharp predictions that those carrying drugs should attempt to sort in response to these crimes. However, police discretion can also vary how the actual weights are recorded. We investigate these competing factors using administrative records from marijuana possession arrests in Chicago. Using a bunching approach, changes in sanction thresholds, and variation in officer-suspect race matches, we test both how decriminalization and officer-race matching affect bunching.

This dissertation includes unpublished co-authored material with Benjamin Hansen.

ACKNOWLEDGEMENTS

This work would not have been possible without the support and assistance of my Advisory Committee Chair, Dr. Benjamin Hansen, who has dedicated a tremendous amount of time helping toward the completion of this dissertation. I am grateful to all of my committee members who have provided me with guidance throughout this process, Dr. Jonathan Davis, Dr. Eric Zou, and Dr. Kristen Bell. The combined support, encouragement, and efforts of these individuals has led to a positive experience and my personal growth as a researcher throughout the projects herein.

I would additionally like to extend my gratitude to my dear friends and colleagues who supported and assisted me along the way, Dr. Xiang Li, Dr. Zachary Kiefer, and Dr. Brett Garcia. Thank you all for being such wonderful people and dedicated friends and researchers.

CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

I. Introduction

Research in economics [1] has provided a profound volume of literature providing evidence of racial disparities throughout the criminal justice system in the U.S. These differences can stem from a multitude of underlying factors, such as social differences, income level, gender, etc. Therefore, policymakers across the U.S. have long tried to adjust laws to ensure equitable and just practices are incorporated. Unfortunately, research shows this has been largely ineffective. Research in this field aims to determine the origins of these disparities as well as the extent to which they affect the effectiveness of resources in order for proper solutions to be developed and implemented and for resource reallocation when it is beneficial. Discrimination in the U.S. justice system weakens its effectiveness and fails to remain impartial and apply equal rights and punishments to citizens.

I first investigate the issue of racial prejudice by examining police stop data using a regression discontinuity approach. I base my premise on the Grogger and Ridgeway (2006) hypothesis, which they dubbed "the veil of darkness," wherein they posit that police officer cannot see the race of a driver when it is dark outside and, therefore, do not take race into account when pulling drivers over at night. They do not find evidence of a racial disparity due to discrimination. However, there are several shortcomings of this method. The first is that there is no way to be certain that officers do observe driver race during the day, but not at not. If this hypothesis

[1] Anwar and Fang (2006), Donohue and Levitt (2001), Abrams et al. (2012), Mustard (2001); Arnold (2018); Arvanites and Asher (2006); Rehavi and Starr (2014); West (2018); Abrams, Bertrand, and Mullainathan 2012

fails, then the test is invalidated. Assuming that their hypothesis is valid, there is still a second issue. This is an issue with the setup of their model. In order to compare arrest rates between light and darkness at the same time of day, they match time of day across months. For example, they compare arrest rates at 5pm in the summer (light out) to 5pm in the winter (dark out) and determine if the arrest rates varied in a way that implies racial discrimination. By doing so, they introduce seasonality into their model where it did not exist before. Given there is research that shows some minority groups work seasonal jobs, it is likely that the underlying distribution of drivers varies by season in a systematic way, as well.

By using a regression discontinuity approach around daylight savings time, I am able to omit any bias in the results caused by seasonal trends in the distribution of drivers. I can compare the hours just before dark prior to DST starting to those just after it starts so as to isolate the disparity around this discrete one hour jump in sunset (and thus daylight hours). I find a lack of evidence of racial discrimination, as well, though this could be due to one of two things. The first may be that the "veil of darkness" hypothesis does not hold true. However, Pierson et al. (2019) retest the Grogger and Ridgeway (2006) model and find significant, though small results which an extremely large dataset. The alternative reason for this could be that I aggregated my data too much, thereby reducing the size of my dataset to a much smaller one compared with Pierson et al. (2019). This is issue can be dealt with by revisiting the data and re-estimating the model with a higher frequency arrest rate to regain some identifying variation, which I intend to do.

Next, I choose to look at different facet of law enforcement, arrests, to estimate the racial disparities here, as well. I have obtained through a FOIA request a dataset of arrests from the Chicago Police Department that allows me to take a new approach to estimate the disparity in arrest rates. I apply my own hypothesis that officer leniency, when applied systematically by race, leads to biased hit rates. This is because there will be a zero reported if leniency is shown and no citation or arrest is made, even if contraband is found. This likely occurs more frequently for less severe offenses; whereas, officers are less likely to show leniency for any reason for more severe offenses. The purpose of this approach is to determine if the relaxation of legality in lower severity crimes leads to a decrease in the racial disparity. If it does, analyzing the effects for different legal severities of crimes can help inform law makers what this change is driven by and help to target particular types of arrests that will have the largest impact with future policy decisions.

To lend credibility to this hypothesis, I run a Kolmogorov-Smirnov test to determine if the density functions arrest rates for marijuana possession differ for white and black arrests more at lower severities than at higher severities.[2] The results of this are consistent with my hypothesis. I then run a linear regression to estimate the per capita racial disparity at lower vs higher severities. The results of this test imply that racial prejudice accounts for twice as much of the disparity as statistical discrimination can account for. Last, I run an interrupted time series to estimate the effect of marijuana decriminalization in Illinois in 2016 differentially for higher and

[2] My data has grams of marijuana which I use to calculate severity; cutoffs are endogenously defined by IL state laws.

lower quantities of marijuana. The results of this show that decriminalization can lead to a lower disparity in arrest rates due to a large decrease in arrests at lower quantities. However, it may lead to a positive in the slope due mostly to the increase in the arrest rate disparity in the higher quantities over time, following the change. I posit the the positive slope change is due to an increase of police resources allocated to arresting individuals who commit more severe offenses.

Finally, along with my coauthor, Benjamin Hansen, we investigate the incidence of bunching around the various legal thresholds in each of the legality regimes for marijuana possession using the aforementioned Chicago PD dataset. We do this for black and white individuals to determine whether there is evidence of differential behavior following a similar approach to Tuttle (2019). Additionally, we estimate how these same bunching outcomes are affected when officer and subject races match or differ.

II. Racial Disparities in Police Stops: A Regression Discontinuity Approach

1 Introduction

After decades of research, it is generally accepted that there are racial disparities in the criminal justice system in the U.S. These disparities might occur in many facets of the justice system, including, but not limited to police stops, sentencing, bail, and parole. These differences can stem from a multitude of underlying factors, such as social differences, income level, gender, etc. It is important for researchers to determine the origins of these disparities in order for effective solutions to be developed and implemented. The U.S. justice system consistently fights to remain impartial and apply equal rights and punishments to citizens, but underlying discrimination can threaten this goal.

If the only difference causing racial disparities within different aspects of the justice system is the individual's race, then the only plausible explanation is underlying racial prejudices. This is because race itself is not a characteristic that should factor into arrests, sentencing, parole, etc., but is often an observable characteristic that officers and judges might unintentionally or otherwise take into account at the time a decision is made. Thus, we must first determine if there is an alternative cause for these racial disparities and narrow down our causal factors to only racial differences. Once we know the incidence of racial disparities due only to race, we can isolate the instances where this occurs and resolve the disparities by targeting those particular points in the system where racial prejudice affects law enforcement decisions.

Mustard (2001) investigates the differences in sentencing that occur based on race to determine both the size and the cause of these disparities. He finds that blacks, males, and offenders with low levels of education and income receive notably longer sentences. Additionally, those disparities are primarily due to departures from sentencing guidelines, rather than differential sentencing within the guidelines. The paper provides evidence that large differences in the length of sentence exist on the basis of race, gender, education, income, and citizenship–in spite of explicit statements in the guidelines that these characteristics should not affect the sentence length. These racial disparities are likely caused by racial discrimination of the sentencing judges since the authors control for a wide range of alternative explanatory variables. Legal sentencing is likely not the only place where racial discrimination leads to disparities, as further research shows it exists in law enforcement, bail setting, and elsewhere.

For example, Norris et al. (1992) analyze data gathered during routine relief and home beat patrolling with police officers. Their data confirm that black people are more likely to be stopped by the police than whites, though, on average, they are stopped for the same types of offenses. Their analysis also suggests black people are stopped on more speculative grounds than white people. This indicates racial profiling may occur when the home beat officer makes a decision to approach a potential suspect. The authors note that the data suggest that, in routine patrolling by relief and home beat officers, prejudice does not significantly lead to differential or discriminatory police action once a stop is underway. Arnold et al. (2018) find evidence that there is substantial bias against black defendants, and rule out statistical discrimination as the sole explanation for the racial disparities in bail. Additionally, other more recent papers find that blacks are more likely to be searched for contraband (Antonovics and Knight 2009), experience more police force (Fryer 2016), have higher chances of being charged with a more serious offense (Rehavi and Starr 2014) or convicted (Anwar, Bayer, and Hjalmarrson 2012), and to be incarcerated at higher rates (Abrams, Bertrand, and Mullainathan 2012).

The research thus far is aimed at determining instances within the justice system where racial prejudice may occur. It is also important to find the original stage(s) at which racial discrimination enters into the criminal justice system. This will allow us to address racial disparities where they initially enter into the system, whether they start from differences in shooting rates, seizures, search rates, or police stops. Police are the first line of law enforcement in the criminal justice system. If racial disparities begin with law enforcement, any racial prejudice that enters into the system at a later stage, such as bail, sentencing, and parole, will be magnified due to the individuals receiving these prejudiced outcomes being minorities at detrimentally higher rates. Thus, we need to determine at what point we first see racial prejudices enter the system and the size of these prejudices.

Millions of people are stopped by police in a given year for various traffic violations. The reasons for these stops are often left up to the discretion of individual traffic cops with limited oversight and accountability. Traffic stops are one of the numerous ways in which the public believes racial bias by law enforcement occurs without consequence. Racial profiling is an issue that the U.S. public are concerned with, due to the difficulty in limiting it and the perpetuation of racial tension by officers. Knowing the extent of racial profiling by traffic cops can help government officials and administrators to target racial sensitivity training to them in an efficient and effective manner. Additionally, it may give insight into the extent of racial profiling across law enforcement positions and other police divisions.

The racial disparity in traffic stop incidence has been measured by Grogger and Ridgeway (2006) using a technique they developed called the "veil of darkness." This method allows us to compare stop distributions when officers can see the race of drivers versus the distribution after dark, when it is difficult or impossible to know the race of a driver prior to stopping them. Ultimately, this gives us the ability to test the difference in these distributions and determine if they change when race is visible prior to a traffic stop. If there is a significant difference, then there is evidence in police bias prior to pulling drivers over. The authors of this paper use variation in daylight times throughout the year, as well as a small window during the intertwilight period to compare the daytime and nighttime stops. They conclude that there is little evidence of racial profiling based on 7,607 stops in 2003.

One issue with the approach used in their paper is that there are no quantitative measures of racial disparities, only qualitative, meaning their approach is centered around finding the sign of the effect of daylight, rather than the magnitude of the effect. Additionally, one assumption used in this paper is that using variation in daylight times from summer to winter, they control for the difference in time of employment, which varies by race. I would argue that this does not account for seasonality of jobs worked by minorities. Timing of jobs differs by race in the time of day *and* the time of year. This is not addressed by the Grogger and Ridgeway (2006) approach. Another limitation is the scope of the data, which is quite small–only one city for about 6 months.

Pierson et al. (2019, working paper) use a new dataset, the Stanford Open Policing project, in which data from dozens of city police and state trooper departments are being compiled into one location, easily accessed by the public. This allows them to look into multiple questions about police stops, including the "veil of darkness" approach with data from 100 million traffic stops across 29 municipal police departments, limited to those cities which provide the necessary variable for the test. They do so using a quantitative approach that identifies the magnitude as well, thereby addressing two of the issues presented by Grogger and Ridgeway (2006). They receive a strongly significant estimate, due in part to the size of the dataset, that is indicative of racial profiling. This result may still be biased.

In this paper, I develop a method to address the final aforementioned issue, differences in driving times due to employment times throughout the year. It is necessary to be sure the data are being drawn from the same underlying distribution of drivers when comparing daylight to nighttime stops. If there is a higher fraction of minority drivers at night, we would expect the fraction of stops of minorities to be higher at night when compared to daytime stops. Thus, the results will be biased toward no racial profiling. In the case of time-of-year differences, we may see, for example, that the distribution of drivers in winter has a smaller fraction of minority drivers. Additionally, the time of sunset is earlier in the evening. Therefore, the estimate may be biased toward racial profiling, as there are fewer minorities being stopped in the darkness-heavy set of the sample where minorities are not on the road at rush hour. This means the result in Pierson et al. (2019) may be still biased toward racial profiling.

The data used in this paper are extracted from the Stanford Open Policing project and include data from city police, all of which include the necessary variables to run our analysis. Those individual observations which do not include all of the variables necessary are then dropped from the set, leaving me with just over 6.3 million individual observations across 10 cities–San Francisco, CA, San Diego, CA, Bakersfield, CA, Aurora, CO, San Antonio, TX, Oklahoma City, OK, New Orleans, LA, Owensboro, KY, Philadelphia, PA, and Hartford, CT. The dataset includes observations from January 1, 2012 through December 31, 2016 at all times of the day, making a much larger panel of data than are used in Gogger and Ridgeway (2006).

To account for changes in driver distributions throughout the year, I utilize a regression discontinuity approach based on daylight savings changes. Suppose sunset begins at 7:00pm prior to the start of daylight savings time. Then the daylight savings time change takes effect. The following day, sunset occurs at 8:00pm. Thus, from 7-8pm the day before, it was dark out, but 7-8pm the day after it is light out. Using this discrete jump in time of sunset, we can compare the data pulled from the same distribution of drivers between one day and the next, but can still implement the veil-of-darkness technique. This allows us to find an estimate for the incidence of racial prejudice in police stops that is unbiased–or at least less biased than those of Grogger and Ridgeway (2006) or Pierson et al. (2019).

The results I find provide little evidence of racial prejudice in police stops. The main estimated change in stops, -1.8 percent, is not statistically significant for black drivers. Additionally, this result is not robust to all functional forms and, in fact, changes signs for two specifications. For Hispanic drivers, the main result, -5.2 percent, is also not statistically significant at any level and this magnitude, sign, and lack of significance is robust to multiple sensitivity checks. Therefore, I conclude that the results in this analysis do not provide evidence of racial prejudice in the stop decision by police officers and, if anything, provides evidence of the disparity being the opposite of what is expected.

In the next section, I discuss the analysis of police stop bias conducted in the literature thus far. In Section 3, I discuss the data used for this research, and in Section 4, I detail the analytical approach used here. In Section 5, I discuss the empirical results the data yield. I then conduct a sensitivity analysis and discuss shortfalls in the estimation approach used herein. Section 7 concludes the paper and discusses the implications and future avenues for research to build on the information presented.

2 Literature Review

Becker (1993) gives a lecture describing how he conducts his research as an analysis that assumes individuals maximize welfare as "they conceive it," regardless of whether they are selfish, altruistic, loyal, spiteful, etc. Treating these preferences as a constant over time, individuals try to anticipate the consequences of their behavior. To understand discrimination against minorities, he widens preferences to accommodate prejudice and hatred of particular groups. This allows them to examine behavior while allowing for preferences based on personal biases in a theoretical framework. Using a theoretical framework that builds off of this and takes the applications a step further, Grogger and Ridgeway (2006) develop an approach utilizing what they call the "veil of darkness." The authors hypothesize that, under the "veil of darkness," i.e. after dark, it is difficult or impossible for traffic cops to determine the race of a driver prior to pulling them over. This means that the decision to stop a driver cannot be influenced by the driver's race after dark. They are able to apply this to their research as a sort of control for racial prejudice in stop decisions. They compare the likelihood of nighttime stops to those during the daytime in order to show whether there is a significant difference in stops of black drivers during the daytime versus nighttime. A significant difference would indicate that visibility of a driver's race changes the stop decision of police officers, providing evidence of racial bias in police stops. The results of their method, based on a dataset from Oakland, CA in 2004, provide little evidence on police stops on a larger scale, as it is a small sample from only one city. The standard errors of their estimates are all relatively large. This is likely due to instances of biased estimates along with having a dataset of only one city in one year.

Further research has looked into post-stop behavior to determine if racial disparities are present here. Anwar and Fang (2006) use a model developed by Knowles et al. (2001) that allows motorists to differ by multiple characteristics observed by the officer that may or may not be recorded in the data. The authors alter this model that allows officers to use information they gather about motorists during traffic stops when they make their search decisions as well as allowing police behavior to vary by their racial group. They fail to reject the hypothesis that troopers of different races do not exhibit relative racial prejudice. However, Antonovics and Knight (2009) investigate the reasons behind racial differences in the rate at which the vehicles of black, Hispanic, and white motorists are searched during traffic stops. They use the idea that, in absence of preferential bias, search decisions should be independent of the race of the police officer that stops a driver. They argue that there is evidence of racial discrimination if searches are more likely to occur when the race of the driver differs from the race of the police officer. They use a dataset containing race of drivers and officer that stopped them from the Boston Police Department spanning 2 years to show that the search likelihood differs based on whether the race of the officer and driver differ and whether these patterns differ by race. Their results provide evidence that the search decision at traffic stops likely is influenced by preferential bias by race. Thus, there is evidence indicating that racial prejudice might influence individual officers to perform stops on a driver of a different race. Depending on the racial distribution of officers and drivers, this may balance out when there are some white officers stopping black drivers and some black officers stopping white drivers. This could explain the difference between their results and Anwar and Fang (2006). This might be an important distinction to keep in mind for further research into police stops as it could be the same mechanism is at work when police stop drivers to begin with. This would require data on the police officer's race as well as the stopped driver's race on an individual level.

In addition to searches during traffic stops, there is research measuring the incidence of racial disparities in frisks at traffic stops. Ryan (2015) measures the individual's probability of receiving a frisk. By exploiting a traffic stop-level dataset from the Pittsburgh Police Department, Ryan estimates the marginal effects of assorted driver characteristics. He shows black drivers are more likely to receive a frisk, as well as identifying several related factors. The results indicate that the interaction of the driver's gender, the time of day, and the existence of passengers with the race of the driver all impact the probability of being stopped. These results alone do not indicate that police are stopping black drivers with the intent to frisk them, or that there is an intention to frisk black drivers. It could be that the police officers tend to feel more threatened when there is a black male driver or when more passengers are present, for example. This may not be intentional, but is still indicative of at least a subconscious racial prejudice when the decision to frisk is made. It also indicates that determining whether the drivers are stopped because they are black or not is a crucial part of finding a solution to this occurrence. Adding to this research, West (2018) shows that officers tend to cite drivers of a different race from their own more frequently. This indicates that police officers show higher tolerance of law breaking towards drivers of their own race, regardless of age, gender, vehicle value, or characteristics of the local community. The author uses data on automobile crash investigations. He then shows that the match of officer and driver is exogenous of race or prejudice because officers are dispatched to the scene of the accident by a remote dispatcher that knows little about the driver or officer who meet at the scene of the accident. Thus, the results are a causal interpretation of how officer behavior differs given the racial combination of officer and driver.

Goel and Shroff (2016) examine New York City's stop-and-frisk program. They show that risk assessment tools can help police officers make considerably better real-time stop decisions. The authors then show that risk assessment tools can also be used to audit past actions. They argue that a significant fraction of New York City police stops were conducted on the basis of relatively weak evidence, in possible violation of constitutional protections. This paper supports the idea that statistical tools can help improve and evaluate decisions throughout the criminal justice system, and, therefore, can reduce the incidence of racial profiling in stops. At the least, they could be used to hold officers accountable when a low level of risk is assigned, yet a stop is still made. This tool could be useful to implement where racial prejudices exist within the criminal justice system, including traffic stops, house stops, sentencing, bail decisions, etc. If racial disparities exist due to prejudice, then it may be worth the cost to implement risk assessment tools. This is both so there is a decrease in the disparity of frisks, stops, etc. and so the laws that officers are employed to uphold are more likely to be upheld because punishment can be enforced more easily when they are not when using these tools.

This research indicates that the answer to the question, “Where does racial prejudice first present itself in the criminal justice system?” is as of yet unclear. Grogger and Ridgeway (2006) do not find evidence of racial disparities, yet many papers following this show disparities in personal interactions, such as frisks and searches. Pierson et al. (2019) assess racial disparities in police interactions with the public, including in police stops, using a dataset they compiled with nearly 100 million traffic stops recorded by 21 state patrol and 29 municipal police departments. They first measure potential disparities in police stops by examining whether black drivers are less likely to be stopped after dark during the “veil of darkness.” The authors find evidence of bias against black drivers both in highway patrol and in municipal police stops. This contradicts the results found in Grogger and Ridgeway (2006), but is not surprising because they use a much larger dataset spanning numerous cities over several years. It is important to note that their main outcome shows a -0.012 effect of darkness on the change in probability of being stopped for a black driver after dark compared to during daylight times. This is significant, but is a relatively small effect. It is important to determine if this effect is truly occurring in the data or if there is a bias due to the seasonal work schedules of minorities or other seasonal factors before taking this as an indication of what types of policy to implement and what officers should be trained for.

3 Data

For traffic stop data, I used the Stanford Open Policing Project data. The Stanford Open Policing Project (SOPP) is collecting and standardizing data on vehicle and pedestrian stops from law enforcement departments across the country. They have collected over 200 million records from dozens of state and local police departments across the country. The SOPP is composed of an interdisciplinary team of researchers and journalists at Stanford University. I have accessed data from a total of 10 U.S. cities, each of which contains the necessary information to run this analysis. The cities included are San Francisco, CA, San Diego, CA, Bakersfield, CA, Aurora, CO, San Antonio, TX, Oklahoma City, OK, New Orleans, LA, Owensboro, KY, Philadelphia, PA, and Hartford, CT. The dataset includes observations from January 1, 2012 through December 31, 2016 at all times of the day, combining to a much larger panel of data than are used in Gogger and Ridgeway (2006).

I aggregate the data to the city-day level for the outcome variable in the model, proportion of stops for which drivers are black or Hispanic. This is because I need observations of the same unit before and after the day DST begins. Thus, our model relies on aggregated observations, rather than at the individual level. After aggregating, we must also exclude those observations that do not occur in a window around sunset so as to only look at those observations that are affected by Daylight Savings Time within the bandwidth, rather than observations that occur, say at 12:00pm, when it is daylight all the time and does not provide us with any information about racial disparities. Gogger and Ridgeway (2006) use the window of

4pm to 9pm and show that this is a reasonable window of time to use. Thus, I will use this same window of time. Once I limit the aggregated sample to this window of time, I am left with approximately 10,543 observations.

Each city's population for each year is needed. to estimate the stop rate. I access this data through the U.S. Census Bureau. Additionally, I need data on Daylight Savings Time start dates for each year, which I have collected from the NASA database. Although I employ a fixed effect to account for the demographic makeup of each city in each year, it is still informative to see this visually, so I have included them here. This data was collected from the DATA USA database. These all span the years and cities in the dataset I have used from the SOPP database.

Figure 1 shows the percentage of black residents in each city in the dataset. Here, it is easy to see that New Orleans and Philadelphia are predominantly black, whereas San Francisco and San Antonio have much smaller black populations relative to white, Hispanic, etc. It is unlikely that San Francisco and San Antonio have the same population demographics, so it is helpful to look at the fraction of the population that is Hispanic, as well. This is shown in Figure 2. San Francisco is approximately one-third white black and Hispanic, whereas San Antonio has a much larger fraction of the population that is Hispanic. Note that the shading key is slightly different for Figure 2 since there are smaller fractions of Hispanic populations black and the key is split into 5 shades for both figures.

These figures give a better understanding of the underlying distribution of demographics within each city of the dataset. These show that each city has a different proportion of minority populations. Therefore, there is a diverse set of cities with varying characteristics within the dataset. I do not want city characteristics to drive the results, so I use a city-year fixed effect in order to estimate the model. More model specifications are discussed in detail in the following section.

The full dataset includes race, gender, age, city, longitude and latitude, time of stop to the second, date, and population, which are all used for my estimates. The data also have officer I.D.'s, division and department names, reason for stop, whether a search was conducted or citation issued, along with vehicle and plate information, all of which will be useful for further research on traffic stops.

In Figure 3, the stop rate frequency distributions are shown for the full dataset, the restricted sample of black drivers, and the restricted sample of Hispanic drivers, respectively. Here we can see that there is a large incidence of days where the stop rate of minorities is close to or equal to zero, meaning those drivers were often white. Keep in mind that the scale is not uniform for all three plots due to the frequency of each rate for each demographic. Since I do not include other races, such as Asian or Pacific/Islander, this is a reasonable incidence of zero stop rates for all drivers, as well in panel a. There is also a larger incidence of zero or close-to-zero rates for Hispanic drivers than black drivers. This is reasonable as there are nearly twice as many observations in the full dataset where the driver is black than that of Hispanics.

There are 2.2 million observations of black drivers and 1.2 million observations of Hispanic drivers. There are over 2.4 millon observations of white drivers with only 0.5 million observations of Asian, Pacific Island, or other race drivers. Of the black driver observations, 1.6 million are male and of the Hispanic drivers 0.8 million are male. I run the regressions on each of these demographics separately in the following sections.

4 Analysis

To consider the effect of the "veil of darkness" by exploiting the discontinuity in time of sunset caused by daylight savings time, I use a regression discontinuity design. The baseline model is

$$log(proportionstops)^{cd}_{race} = \beta_0 + \beta_1 day^{cd} + \beta_2 DST^{cd} + \beta_3 (day \times DST)^{cd} + city \times year + dayofweek \quad (1)$$

where $race = black$, $Hispanic$ and day is the running variable such that 0 is the cutoff for the day daylight savings time begins for city c on day d. Thus, for the baseline model negative running values indicate days prior to DST and positive running values indicate days during DST. To determine if there is an effect on police stops after DST begins, I need to look at the coefficient on DST, β_2. I control for the running variable using a varied slope on either side of the cutoff as per the standard for regression discontinuity. Doleac and Sanders (2015) notes that DST always begins on a Sunday, which has different driving patterns and crime patterns than other days of the week. To control for these differences, I include day-of-week fixed effects. As I previously mentioned, city-year fixed effects are included to account for those city and year specific characteristics that might affect the outcome variable, such as the demographic makeup of each city in each year. Additionally, all estimates use standard errors clustered by city and year.

The model is expanded to include the squared, cubed, and quartic terms because the RD model function *rdplot* selects it. I conduct a robustness check to show alternative functional forms do not give different results in section 6, so I keep the global fourth-order polynomial form. Although the window used in this analysis is typically restricted to 4pm to 9pm, it is important to recall daylight savings shifts sunset discretely by one hour. This hour that has a discontinuous change in daylight is where the majority of the explanatory power will come from. Though, sunset does change time from day to day throughout the period of time in the bandwidth, so there is some explanatory power coming from this. I do not want to use this variation in daylight from day to day over the course of the year to explain the change in stops, so I attempt using several bandwidths to determine if this has an effect on the outcome variable, which I will discuss more in the Robustness section. For the main regressions, the bandwidth includes up to 100 days. Thus, the majority of the identification comes from spring and summer drivers so that the outcome is largely unaffected by seasonality.

The baseline model is then run again when restricting the data to include only male drivers from each racial demographic group. This is because Ryan (2015) and other previous research suggests that racial discrimination is more prevalent against males in the criminal justice system. To determine if there is a difference in the size of the effect due to gender, I run the model for male drivers of each demographic as part of my heterogeneity analysis.

5 Results

5.1 Main Results

The baseline model is a log-linear regression discontinuity design where the outcome is the $log(proportionstops)^{cd}_{race}$ for particular city in a given day. This is run for $race = black$ and $race = Hispanic$ to begin with. The standard window form 4pm to 6pm local time is used unless otherwise noted. The results for the model where the outcome is the logged proportion of stops for black drivers are shown in Table 1. I cannot conclude that Daylight Savings Time leads to any statistically significant change in the proportion of stops of black drivers. The outcome of importance here is the coefficient on $During\ DST$. There is significance on the interaction of day^3 and day^4 with daylight savings time, so there is some change in the curvature, though the magnitude of these estimates is quite small.

For $During\ DST$, the estimate I find is the opposite sign of the estimate that Pierson et al. (2019) find. My results show an approximately -1.8 percentage change in the proportion of stops of black drivers and is not significant at any level. This estimate would imply, if anything, a decrease in the proportion of black drivers being stopped when DST begins and sunset has moved back by one hour. This effect is the opposite of what would occur if police officers systematically stopped black drivers when they can see the driver's race. In the tables, I multiplies the estimates by 100 for ease of reading because the numbers are small and the estimated effect gives the $100 \times \beta_n$ change in the proportion of stops, so this also makes them more logical.

The baseline model is still used for Table 2, but the regression discontinuity design is altered slightly. The design used here is the "donut" RD, meaning there is a small range of the running variable around the cutoff day that is dropped from the sample. Barreca et al. (2016) demonstrate that the RD design's smoothness assumption is inappropriate when there is non-random heaping. In order to account for this possibility, I exclude the day before to the day after the start of DST in column 1 and two days before to two days after the start of DST in column 2. The reason this might be necessary is due to the fact that people's sleep schedules may be intentionally altered leading up to DST and sleep is definitely affected once DST begins as an hour of the night is skipped. Janszky and Ljung (2008) note that changing clocks "can disrupt chronobiologic rhythms and influence the duration and quality of sleep" for several days, and also hypothesize negative physical effects as a result of the policy. They also note that most of these costs are due to the switch from standard time to DST. This might cause other changes to driving behavior that affect the distribution of drivers on the days before and after DST begins or that may affect likelihood of being pulled over depending on the time of day drivers work. The magnitude of our main estimate increases, but the sign is unchanged in column 1. However, the sign changes to positive when we exclude 2 days before and after DST begins. If DST causes other behavior changes that bias my estimate toward a negative outcome, then column 2 provides evidence of a racial disparity, but a small one of less than 2 percent. Additionally, the estimate remains statistically insignificant with less precision than the previous estimates. In the second row of Table 2, the estimate for Hispanic drivers is still negative, at -6.7 and -3.3 percent, respectively, but are not statistically significant.

Comparing these to the point estimate for Hispanic drivers in Table 4 column 1, which shows the estimates for the baseline model for multiple demographics of drivers between 4pm and 9pm. The magnitude of the estimate for Hispanic drivers is similar, here, yet slightly larger at -6.7 percent. This estimate is also not significant. These results do not indicate that there is a disparity in stops for Hispanic drivers due to the change in daylight hours, just as is seen for black drivers in this dataset. It is imperative to keep in mind that this tells us that the incidence of daylight does not lead to a change in the proportion of minority drivers stopped. It is plausible that racial disparities exist in traffic stops that are not captured by this change in daylight, but are consistent throughout the day and throughout the year. Other factors, such as neighborhood, vehicle characteristics, or particular roads or highways, may be used to determine if a driver is likely to be a minority, which may affect officer behavior, though I do not explore them in this paper.

5.2 Heterogeneity

In this section I explore the outcomes of my heterogeneity analysis and discuss how these affect the conclusion of this paper. The first heterogeneity test I run is reported in Table 3, where observations are separated by the region of the U.S. they fall into, then the baseline model of the fourth degree is run on the sample for several demographic groups: black drivers, Hispanic drivers, male black drivers and male Hispanic drivers. The North is defined as any city above the 36th parallel, whereas the South is any city below the 36th parallel. I use this cutoff because it is a historical indicator as the cutoff for the cultural northern and southern states. Here it is notable that the magnitude of the estimated effect on black drivers in northern cities is quite large and negative, whereas the south is small and positive. This means our baseline estimate is likely driven more by the effect of northern cities rather than southern cities. None of these outcomes are significant, however. Additionally, the positive outcome for black male drivers is driven by northern and western cities. This estimate tells us that there is more likely to be a disparity for black males than black female drivers in these regions, which is what the literature supports. The estimate for black males in southern cities is negative, which is not expected. One possible reason for this may be that the majority of the southern cities in the data are in the western U.S. and the overall effect for black drivers in the western region is large and negative. Referring to the outcomes for Hispanic male drivers, there is a positive effect in the northern and southern cities for Hispanic male drivers, indicating a disparity for this demographic, but neither estimate is significant. For Hispanic drivers as a whole, there is an estimated negative effect for northern and southern regions, but they are not significant.

Table 3 also shows the outcome when the U.S. is split east to west. The western region includes all cities in the Pacific and Mountain time zones. The east then includes all Central and Eastern time zone cities. I intended to split by time zone, but mountain time does not have enough observations to properly identify the regression discontinuity since only Aurora, CO falls into this time zone. There are too many days in the sample where no minority drivers are stopped in the time frame used. Therefore, in this analysis, I use east and west U.S. regions, instead. The results for columns 3 and 4 show that the positive effect for black male drivers comes mostly from western cities. For black driver as a whole, there is a small positive effect in the eastern cities. One possible explanation for this is there is more historical tension between white officers and African Americans in eastern cities. However, the estimate is not significant.

This heterogeneity analysis shows that there may be some variation in the effect of daylight savings time on proportion of stops based on the region of the city. This might be caused by the fact that sunset is at a different time in the northern states, meaning the jump in time of sunset occurs at a different time within the window. It is also notable that the effect on Hispanic drivers is less pronounced in the east versus the west, but are both negative; whereas, the effect in eastern cities for black drivers is positive. This could be partly due to the regions that are more heavily black populated are eastern cities and those that are more heavily Hispanic fall further to the west, which could cause heterogeneity in the attitude of officers toward specific groups of minorities.

6 Sensitivity

The next step of my analysis is the sensitivity analysis and robustness checks. The outcome of the first robustness check is reported in Table 4. Here, I varied the time window used to see if the results are robust to a different time of day. Column 1 shows the baseline model of the fourth order estimates for each demographic. Column 2 shows the results when we include only observations between 5pm and 7pm. The results change slightly in magnitude, but the signs remain the same for black and Hispanic drivers. For the male groups, the estimates both become negative on the same magnitude of the overall group estimates. In columns 3 and 4, the estimated effects keep the same signs for the overall groups, but change in magnitude. The effect for Hispanic males becomes smaller in magnitude and the sign for Black males becomes negative and very large, yet all estimates are still insignificant. In column 5, I use a time window of 8pm to 10pm. Here, all of the signs are negative except for black drivers. This becomes positive, but is very small in magnitude at less than 0.3 percent. The cities in our sample are not affected by daylight changes between 8pm and 10pm, so this effect is likely due to some other factor, potentially that of daytime variation in work hours for minorities. Black workers might tend to work later hours, whereas Hispanic workers work earlier hours, since seasonal jobs tend to be outside work. Time of day variation in drivers is an effect Grogger and Ridgeway (2006) try to avoid by choosing the time window of 4pm to 9pm, so the change in the 8pm-10pm estimates do not concern me.

After trying different time windows, I run the model again for each demographic, while varying the bandwidth used arbitrarily. The results from this analysis are reported in Table 5. Column 1 shows the results with a bandwidth of 60. Here the magnitudes of some estimates vary while keeping the same signs. The estimate for black males, however, becomes negative. Again, we see a similar occurrence in column 2 where the results for a bandwidth of 40 are shown; though, the sign for Hispanic male drivers also becomes negative in this column. For a bandwidth of 20, the magnitude of the estimated effect on black drivers increases by nearly 5 percentage points, but, for Hispanic drivers, the estimated effect increases by 9 percentage points. When the bandwidth is limited to 10 days, every estimated effect sign becomes positive and range from 2.4 percent to 6.7 percent. These are still insignificant, but indicate a possible racial disparity in this small bandwidth. The estimates in this table largely imply that there is, if anything, a decrease in the proportion of black drivers stopped when it is daylight, rather than an increase. The positive signs of the point estimates for column 3 is likely driven by the fact that there is less variation in the proportion of minority drivers stopped in the few days or so surrounding DST. For example, in Figure 9 panel a, we can see that there are more lower observations past 40 days out.

In the second and third rows of Table 5, it is notable that the sign for Hispanic drivers becomes negative, though not significantly so, as the bandwidth decreases. It is possible the estimates for Hispanic drivers are not as well identified, likely due to the fact that there are fewer observations of Hispanic drivers, so even small changes in the number of Hispanic drivers stopped will lead to a large percentage change in the proportion of Hispanic drivers stopped. This might cause the estimated sign and magnitude to change with only a change in bandwidth.

Finally, the last sensitivity check I run is varying degrees of polynomial in the baseline model. As I noted in the analysis section, the RD model function uses the fourth-order. However, it is important to determine if the estimated effect and its sign are affected by the functional form, or if this outcome is robust to such changes. Looking at the 3^{rd} order column, there appear to be sign changes for the black drivers and Hispanic male driver groups, though these only change by about 4 and 2 percentage points, respectively, and neither of these estimates are statistically significant. The same is true for the 1^{st} degree polynomial estimates. The sign for Hispanic male drivers is negative for the 2^{nd} degree, as well. Although the outcomes are not completely insensitive to different degrees of polynomials, there does not appear the be a better functional form that contradicts the outcomes of the fourth order, nor do any of them provide more evidence of a racial disparity due to DST. Thus, I argue that the conclusions made from these results are reasonably acceptable.

7 Conclusion

It is generally accepted that there are racial disparities within the criminal justice system. These disparities are magnified at each step where they occur, from arrest, to charges, to setting bail, to sentencing, and even to probationary release.

It is important to determine where these disparities first occur so our government officials and law enforcement can target these areas for racial sensitivity training and enforce punishment for targeting individuals based on race. This paper analyzes police stop data to determine if there is evidence of police disparities in the decision to stop drivers. Based on the "veil of darkness" hypothesis and employing a regression discontinuity approach, I find little evidence of racial disparities in stops due to the start of daylight savings time.

To identify the occurrence of racial disparities, using the Grogger and Ridgeway (2006) hypothesis, I would expect to see an increase in the percentage of stops of minority drivers during the evening time after daylight savings time begins. This is because the time of sunset is pushed back by one hour the next day, thereby allowing me to take advantage of a discrete jump in daylight hours from the day before to the day after DST begins. My estimates show that there is not a significant effect on the incidence of stops for minority drivers and, if we chose to take the estimates as significant, we still see the opposite effect occur for black and Hispanic drivers. Alternatively, there is a positive effect for male drivers, which indicates a racial disparity. It is important to keep in mind, however, that this estimate for Hispanic drivers and its sign are not robust to different functional forms of the baseline model and is insignificant for black and Hispanic males.

These results point to a lack of evidence of racial disparity, but I take an important assumption as given that may not be true. I use the "veil of darkness", which assumes that officers observe a driver's race more often than not before stopping them during daylight hours. It is important to ask whether officers do in fact see the driver's race prior to a stop during the day or if this base assumption itself may be invalid. To test this assumption, it may be necessary to find a different identification strategy and require the use of other observable characteristics, such as neighborhoods, highway routes, or vehicle make/model/condition, which might attributed to particular demographics by police officers. If there is a constant higher rate of stops for minority drivers than there should be, given the number of drivers of each race, this would mean the incidence of daylight does not affect incidence of racial disparities in police stops. This would render the "veil of darkness" approach invalid, motivates further research to test the validity of the Grogger and Ridgeway (2006) hypothesis. The alternative is that there is not an abnormally high constant stop rate for minorities, which indicates no racial disparity in police stops. This would motivate further research regarding disparities in search rates and citations issued, rather than other tests for stop rate disparities. This requires data on driving rates for each demographic in each city, which I do not currently have access to, but plan to acquire or build in order to determine the direction of future research.

Determining if other observable characteristics are used to target minorities in traffic stops requires further research into the behavior of traffic cops and whether certain other driver characteristics can be found to explain police officers systematically stopping drivers or not. This question is one I plan to explore in future research and can help determine if the results here and in Grogger and Ridgeway (2006) or Pierson et al. (2019) and other traffic stop research give an accurate picture of the incidence of racial disparities in police stops. It is possible that discrimination does not occur when the stop decision is made, in which case, policy implications are that officials should target training of police officers to interact with minority groups, rather than on how they decide which drivers to stop initially. Whether racial discrimination starts with the decision to stop or the personal interaction of an officer with a driver helps determine what policies would make society better off and minority drivers safer.

References

Abrams, David S. and Bertrand, Marianne and Mullainathan, Sendhil. 2012, "Do Judges Vary in Their Treatment of Race?," *The Journal of Legal Studies*, Vol. 41, No. 2, pp. 347-383.

Antonivics, Kate and Brian G. Knight. 2009, "A New Look at Racial Profiling: Evidence from the Boston Police Department," *The Review of Economics and Statistics*, Vol. 91, No. 1, pp. 163-177.

Anwar, Shamena and Hemming Fang. 2006, "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *The American Economic Review*, Vol. 96, No. 1, pp. 127-151.

Anwar, Shamena, Patrick Bayer, Randi Hjalmarsson. 2012, "The Impact of Jury Race in Criminal Trials," *The Quarterly Journal of Economics*, Vol. 127, No. 2, pp. 1017-1055.

Arnold, Davis, Will Dobbie, Crystal Yang. 2018, "Racial Bias in Bail Decisions," *The Quarterly Journal of Economics*, Vol. 133, No. 4, pp. 1885-1932.

Barreca, Alan, Jason Lindo, and Glen Waddell. 2016, "Heaping-Induced Bias in Regression-Discontinuity Designs," *Economic Inquiry*, Vol. 54, No. 1, pp. 268-293.

Becker, Gary S. 1993, "Nobel Lecture: The Economic Way of Looking at Behavior," *Journal of Political Economy*, Vol. 101, No. 3, pp. 385-409.

Doleac, Jennifer and Nicholas Sanders. 2015, “Under the Cover of Darkness: How Ambient Light Influences Criminal Activity,” *The Review of Economics and Statistics*, Vol. 97, No. 5, pp. 1093-1103.

Goel, Sharad, Justin M. Rao, and Ravi Shroff. 2016, “Personalized Risk Assessments in the Criminal Justice System,” *American Economic Review: Papers and Proceedings*, Vol. 106, No. 5, pp. 119-123.

Grogger, Jeffrey and Greg Ridgeway. 2006, “Testing for Racial Profiling in Traffic Stops from Behind a Veil of Darkness,” *Journal of American Statistical Association*, Vol. 101, No. 475, pp. 878-887.

Janszky, I., and R. Ljung. 2008, “Shifts to and from Daylight Saving Time and Incidence of Myocardial Infarction,” *New England Journal of Medicine*, Vol. 359, pp. 1966-1968.

Knowles, John, Nicola Persico, and Petra Todd. 2001, “Racial Bias in Motor Vehicle Searches: Theory and Evidence,” *Journal of Political Economy*, Vol. 109, No. 1, pp. 203-229.

Marit, Rehavi and Sonja Starr. 2014, “Racial Disparity in Criminal Sentences,” *Journal of Political Economics*, Vol. 122, No. 6, pp. 1320-54.

Mustard, David B. 2001, “Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts,” *Journal of Law & Economics*, Vol. 44, No. 1, pp. 285-314.

Norris, Clive, Nigel Fielding, Charles Kemp, and Jane Fielding. 1992, “ Black and blue: an analysis of the influence of race on being stopped by the police,” *Wiley on behalf of The London School of Economics and Political Science*, Vol. 43, No. 2, pp. 207-204.

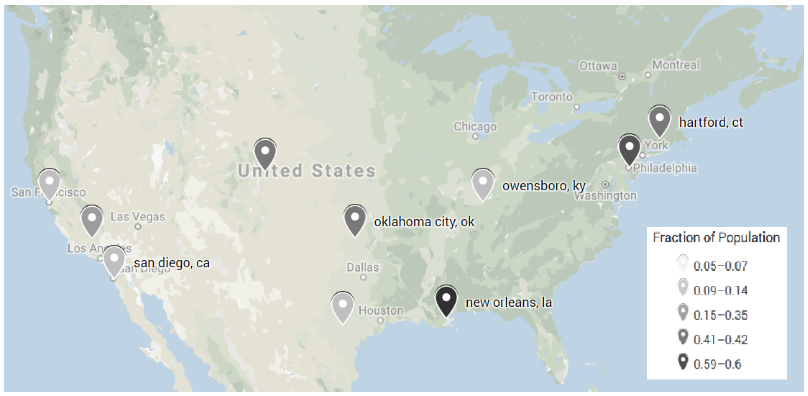
Pierson, Emma, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Pheobe Barghouty, Cheryl Phillips, Ravi Shroff, and Sharad Goel. 2019, “A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States,” *Stanford Computational Policy Lab*, Working paper.

Ryan, Matt E. 2015, “Frisky Business: Race, Gender and Political Activity During Traffic Stops,” *European Journal of Law and Economics*, Vol. 41, pp. 65-83.

West, Jeremy. 2018, “Racial Bias in Police Investigations,” *Revise and resubmit at The Review of Economics and Statistics.*

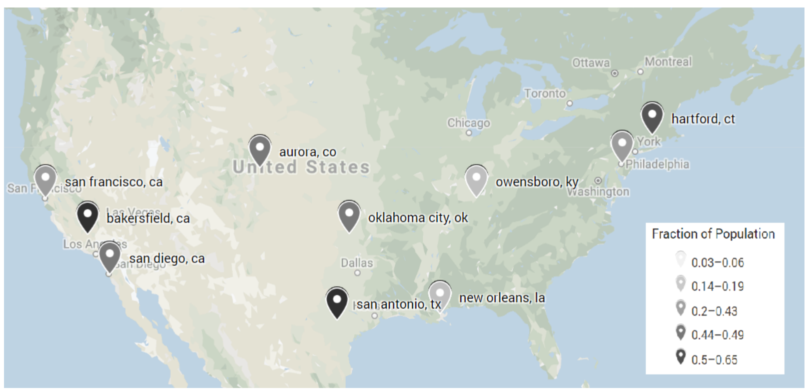
Figures

Figure 1
Fraction of Population that is Black in 2017



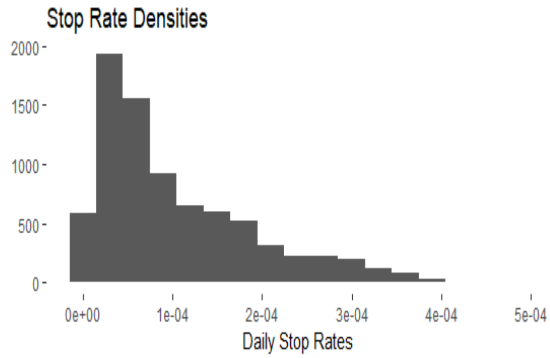
Notes: This map shows the percentage of the population that is black for each city in the sample using forecasted population data from the US Census based on the American Community Survey.

Figure 2
Fraction of Population that is Hispanic in 2017

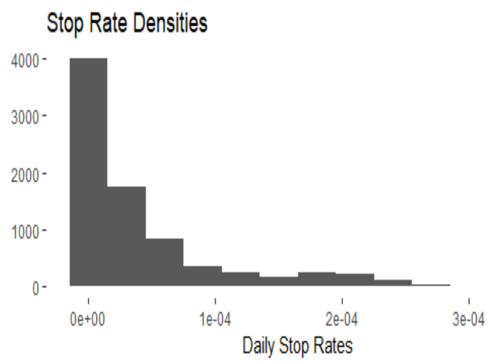


Notes: This map shows the percentage of the population that is Hispanic for each city in the sample using forecasted population data from the US Census based on the American Community Survey.

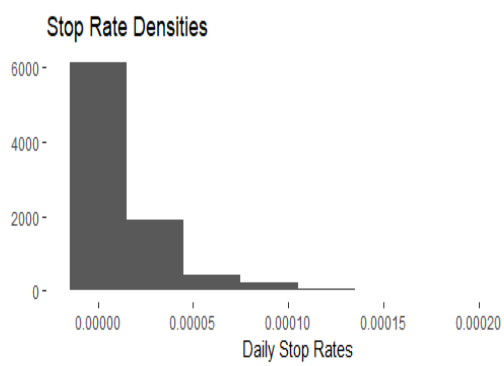
Figure 3
Stop Rate Distributions



(a) All Drivers



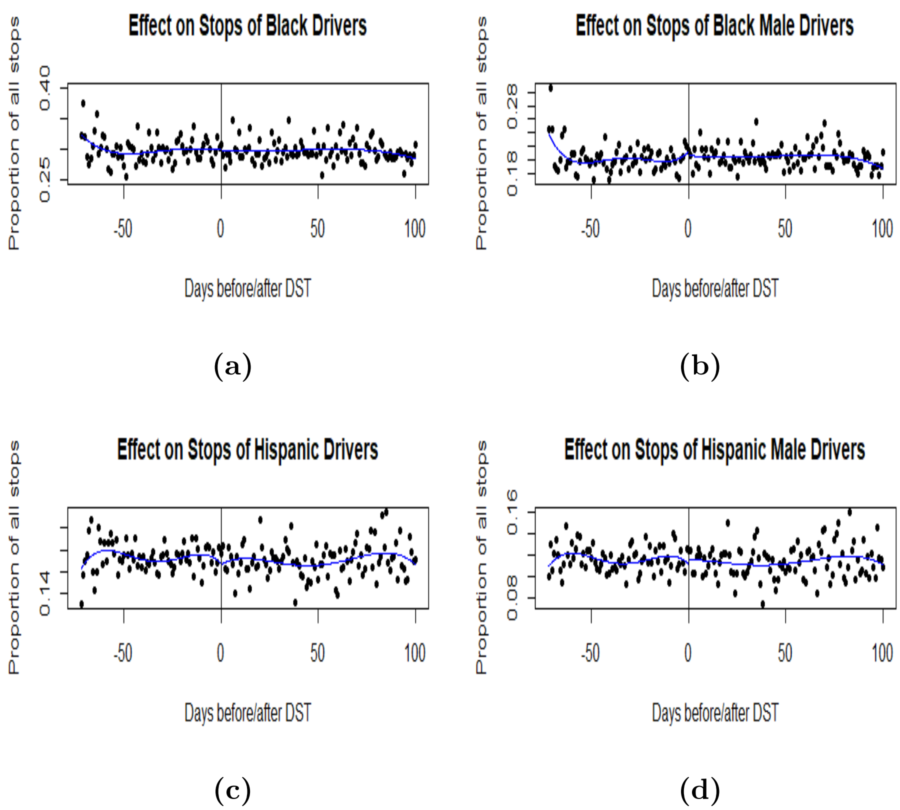
(b) Black Drivers



(c) Hispanic Drivers

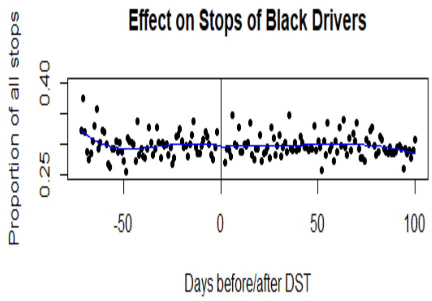
Notes: This figure shows the stop rate distributions for the SOPP dataset. Panel A shows the overall stop rates per entire population on each day. Panel B restricts the data to only black drivers and B restricts the data to only Hispanic drivers.

Figure 4
Effect on Stops of Drivers 4pm to 9pm

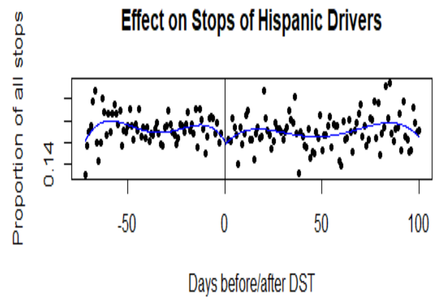


Notes: This figure shows the effect of DST on stops of drivers of different demographics in the window of time between 4:00pm and 9:00pm. The running variable cutoff 0 is the day DST begins. The data for this figure comes from the Stanford Open Policing Project and includes all 10 cities.

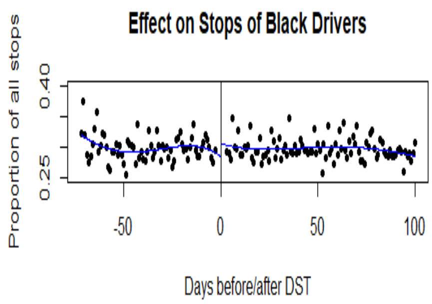
Figure 5
Effect on Stops of Drivers Using "Donut" RD



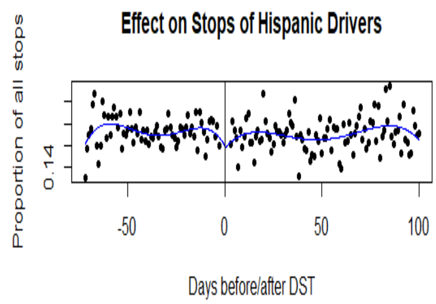
(a) Excluding day before to day after DST



(b) Excluding day before to day after DST



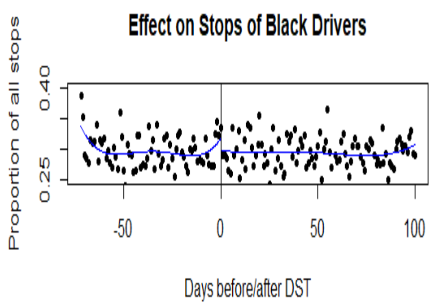
(c) Excluding two days before and after DST



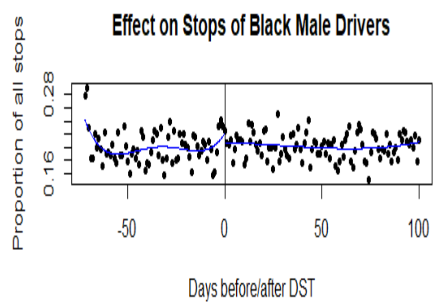
(d) Excluding two days before and after DST

Notes: This figure shows the effect of DST on stops of drivers of different demographics in the window of time between 4:00pm and 9:00pm. The "donut" RD drops the day(s) before and after the cutoff. Panels a and b exclude -1 to 1, while panels c and d exclude -2 to 2. The running variable cutoff 0 is the day DST begins, though we do not observe the stops on those day. The data for this figure comes from the Stanford Open Policing Project and includes all 10 cities. model.

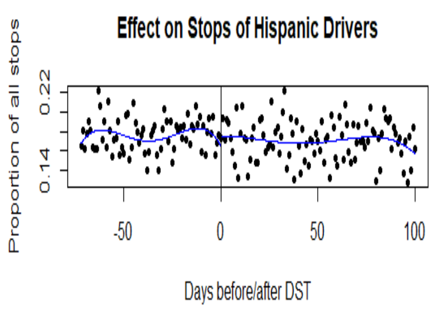
Figure 6
Effect on Stops of Drivers 5pm to 7pm



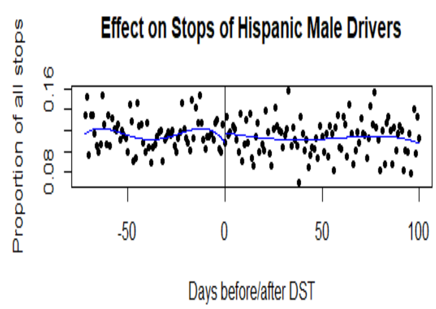
(a)



(b)



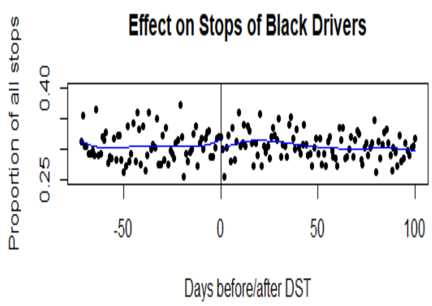
(c)



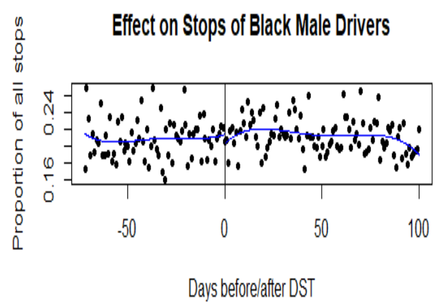
(d)

Notes: This figure shows the effect of DST on stops of drivers of different demographics in the window of time between 5:00pm and 7:00pm. The running variable cutoff 0 is the day DST begins. The data for this figure comes from the Stanford Open Policing Project and includes all 10 cities.

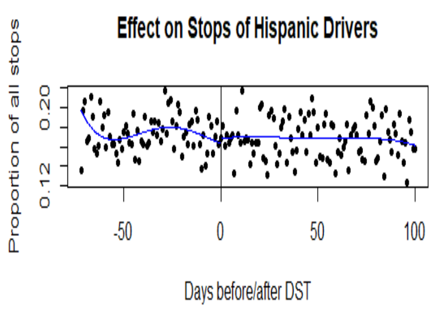
Figure 7
Effect on Stops of Drivers 6pm to 8pm



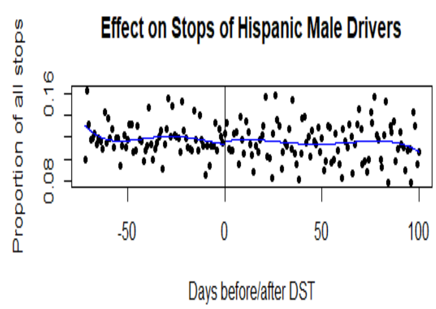
(a)



(b)



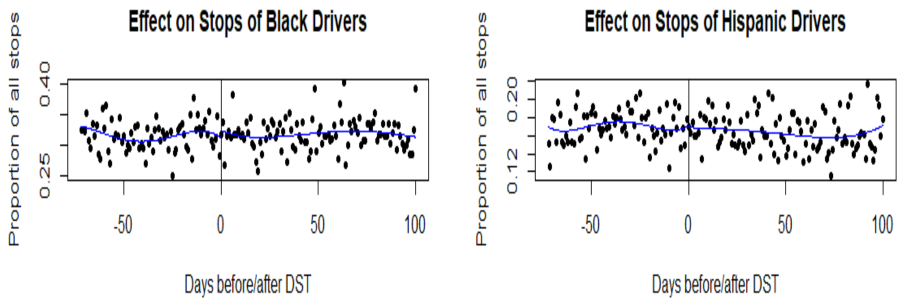
(c)



(d)

Notes: This figure shows the effect of DST on stops of drivers of different demographics in the window of time between 6:00pm and 8:00pm. The running variable cutoff 0 is the day DST begins. The data for this figure comes from the Stanford Open Policing Project and includes all 10 cities.

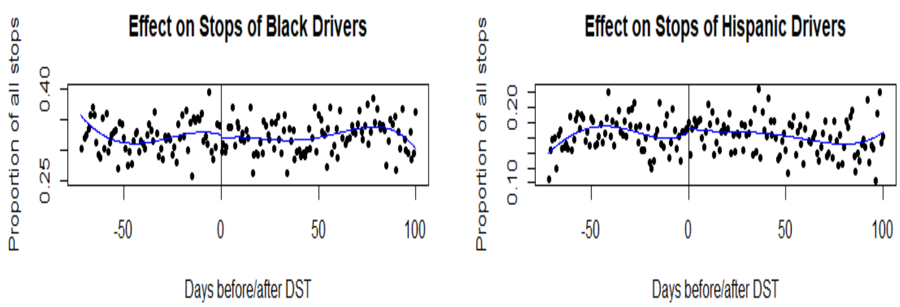
Figure 8
Effect on Stops of Drivers 7pm to 9pm



(a) **(b)**

Notes: This figure shows the effect of DST on stops of black and Hispanic drivers, respectively, in the window of time between 7:00pm and 9:00pm. The running variable cutoff 0 is the day DST begins. The data for this figure comes from the Stanford Open Policing Project and includes all 10 cities.

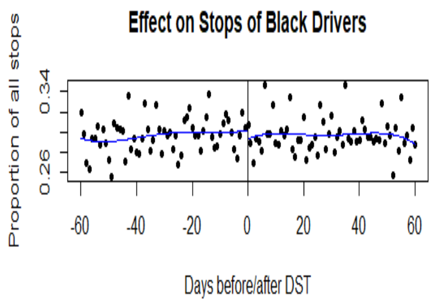
Figure 9
Effect on Stops of Drivers 8pm to 10pm



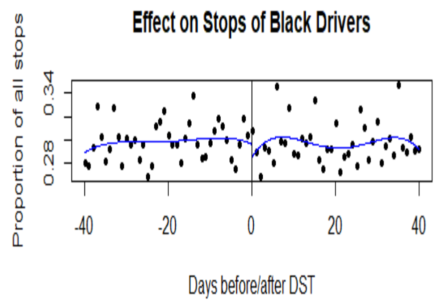
(a) **(b)**

Notes: This figure shows the effect of DST on stops of black and Hispanic drivers, respectively, in the window of time between 8:00pm and 10:00pm. The running variable cutoff 0 is the day DST begins. The data for this figure comes from the Stanford Open Policing Project and includes all 10 cities.

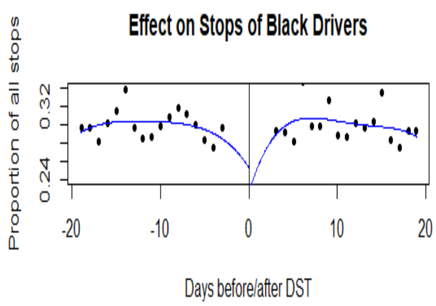
Figure 10
Effect on Stops of Black Drivers Varying Bandwidths



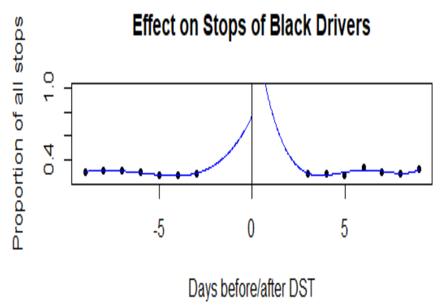
(a) bandwidth = 60



(b) bandwidth = 40



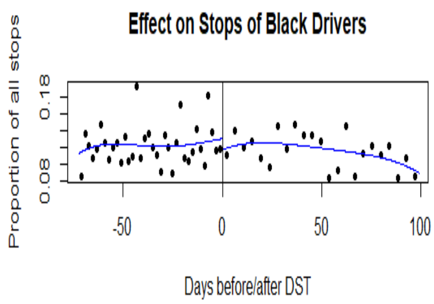
(c) bandwidth = 20



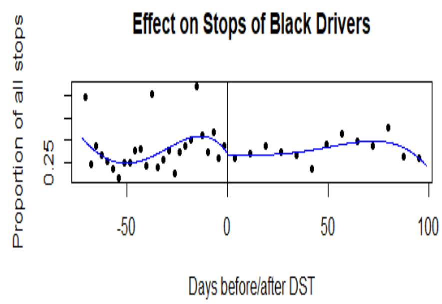
(d) bandwidth = 10

Notes: This figure shows the effect of DST on stops of black drivers in the window of time between 4:00pm and 9:00pm. The running variable cutoff 0 is the day DST begins. The bandwidth is varied here restricted to 60, 40, 20, and 10 days, respectively. The data for this figure comes from the Stanford Open Policing Project and includes all 10 cities.

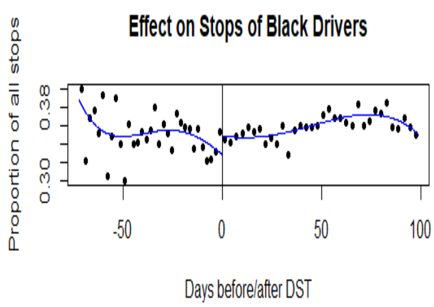
Figure 11
Effect on Stops of Black Drivers By Time Zone



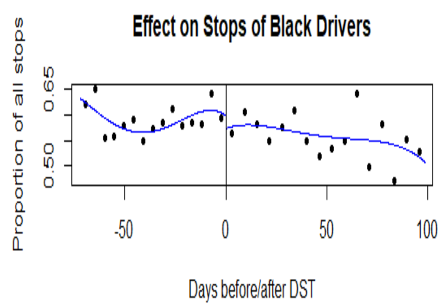
(a) Pacific Time



(b) Mountain Time



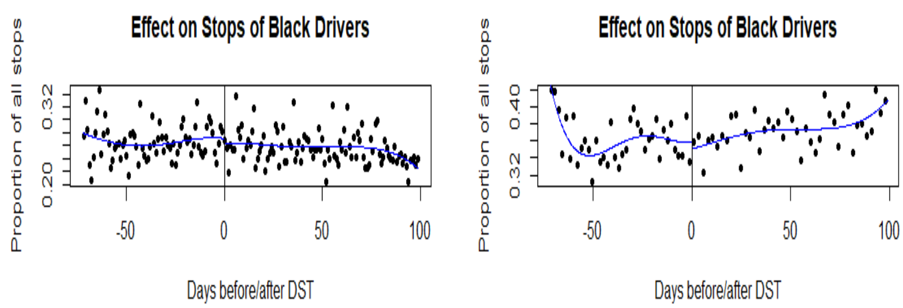
(c) Central Time



(d) Eastern Time

Notes: This figure shows the effect of DST on stops of black drivers in the window of time between 4:00pm and 9:00pm grouped by time zone. The Pacific, Mountain, Central, and Eastern Time Zones are each shown separately. The running variable cutoff 0 is the day DST begins. The data for this figure comes from the Stanford Open Policing Project and includes all 10 cities, each in their respective time zone model.

Figure 12
Effect on Stops of Black Drivers Split at 36th Parallel



(a) Above 36th Parallel **(b) Below 36th Parallel**

Notes: This figure shows the effect of DST on stops of black drivers in the window of time between 4:00pm and 9:00pm grouped by region. The Northern and Southern US regions are each shown separately. The north and south regions are determined by the cities' respective direction from the 36th parallel, which is used here as it is an indicator of a cultural barrier between the northern and southern states. The running variable cutoff 0 is the day DST begins. The data for this figure comes from the Stanford Open Policing Project and includes all 10 cities, each in their respective time zone model.

Tables

Table 1
Effect of DST on Stops of Black Drivers

	Estimate	Clustered SE	t	P
day	0.806	0.752	1.07	0.284
day^2	0.053	0.033	1.60	0.109
day^3	0.001**	0.001	2.25	0.024
day^4	0.000***	0.000	2.83	0.005
During DST	-1.826	4.955	-0.37	0.713
day $\times$ During DST	-0.971	0.784	-1.24	0.216
day^2 $\times$ During DST	-0.051*	0.030	-1.68	0.093
day^3 $\times$ During DST	-0.001*	0.001	-1.95	0.051
day^4 $\times$ During DST	-0.000***	0.000	-3.20	0.001

Estimates and errors are all multiplied by 100
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table reports the effect of DST on stops of black drivers in the window of time between 4:00pm and 9:00pm. This fits a line with differing slopes prior to and following the start of DST. The coefficient on tells us the effect we care about. The data for this table comes from the Stanford Open Policing Project and includes all 10 cities. model.

Table 2
Effect of DST on Stops of Drivers using "Donut" Regression Discontinuity

	(1) Excluding *day* $\leq \lvert 1 \rvert$	(2) Excluding *day* $\leq \lvert 2 \rvert$
Black Drivers		
During DST	-4.893	1.900
	(7.772)	(8.639)
Hispanic Drivers		
During DST	-6.732	-3.336
	(7.286)	(7.527)

Estimates and errors are all multiplied by 100
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table reports the effect of DST on stops of black drivers in the window of time between 4:00pm and 9:00pm. This fits a line with differing slopes prior to and following the start of DST. The coefficient on *During DST* tells us the effect of daylight savings time on the proportion of stops, so it is shown for each demographic using 2 different "donut" regression discontinuity designs–the first excludes the day before and after DST, while the second excludes 2 days before and after DST. The data for this table comes from the Stanford Open Policing Project and includes all 10 cities. model.

Table 3
Effect of DST on Stops of Drivers By Region

	(1) North	(2) South	(3) East	(4) West
Black Drivers				
During DST	-3.813	0.330	1.615	-11.78
	(7.160)	(3.350)	(3.532)	(15.47)
Hispanic Drivers				
During DST	-4.570	-6.073	-1.849	-2.065
	(7.662)	(11.94)	(8.517)	(5.117)
Black Male Drivers				
During DST	2.644	-3.225	-2.252	7.953
	(5.641)	(7.634)	(5.661)	(11.07)
Hispanic Male Drivers				
During DST	0.334	2.074	4.771	2.342
	(7.193)	(7.793)	(7.108)	(1.543)

Estimates and errors are all multiplied by 100
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table reports the effect of DST on stops of black drivers in the window of time between 4:00pm and 9:00pm. This fits a line with differing slopes prior to and following the start of DST. The coefficient on *During DST* tells us the effect of daylight savings time on the proportion of stops, so it is shown for each demographic in each US Region. The US is split such that North and South are on either side of the 36th parallel. The east and west are split such that west includes the Pacific and Mountain Time Zones, while the other cities make up the east. The sample is not split by 4 smaller regions, but 2 regions at time (North/South and East/West). The data for this table comes from the Stanford Open Policing Project and includes all 10 cities. model.

Table 4
Effect of DST on Stops of Drivers Varying Time Windows

	(1) 4pm-9pm	(2) 5pm-7pm	(3) 6pm-8pm	(4) 7pm-9pm	(5) 8pm-10pm
Black Drivers					
During DST	-1.826	-3.661	-5.869	-3.474	0.255
	(4.955)	(7.792)	(5.397)	(3.741)	(5.412)
Hispanic Drivers					
During DST	-5.204	-3.550	-0.781	-4.492	-2.028
	(5.065)	(7.020)	(3.816)	(3.919)	(10.43)
Black Male Drivers					
During DST	0.395	-3.587	-9.482	-7.897	-1.249
	(6.058)	(7.427)	(7.116)	(5.807)	(2.751)
Hispanic Male Drivers					
During DST	1.000	-3.388	5.381	-2.619	-4.222
	(4.050)	(7.619)	(1.000)	(2.666)	(6.886)

Estimates and errors are all multiplied by 100
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: This table reports the effect of DST on stops of black drivers in the window of time between 4:00pm and 9:00pm. This fits a line with differing slopes prior to and following the start of DST. The coefficient on *During DST* tells us the effect of daylight savings time on the proportion of stops, so it is shown for each demographic group during each time window used. The data for this table comes from the Stanford Open Policing Project and includes all 10 cities. model.

Table 5
Effect of DST on Stops of Drivers with Varying Bandwidths

	(1)	(2)	(3)	(4)
	60	40	20	10
Black Drivers				
During DST	-6.365	-3.165	-7.715	6.690
	(5.480)	(5.661)	(11.15)	(11.80)
Hispanic Drivers				
During DST	-6.294	-8.738	-17.90	2.388
	(6.604)	(9.412)	(18.63)	(22.06)
Black Male Drivers				
During DST	-1.228	-3.090	-2.419	4.593
	(6.137)	(6.590)	(18.02)	(14.381)
Hispanic Male Drivers				
During DST	0.622	-6.187	-4.828	3.601
	(3.206)	(5.144)	(8.282)	(5.751)

Estimates and errors are all multiplied by 100
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table reports the effect of DST on stops of black drivers in the window of time between 4:00pm and 9:00pm. This fits a line with differing slopes prior to and following the start of DST. The coefficient on *During DST* tells us the effect of daylight savings time on the proportion of stops, so it is shown for each demographic using 4 arbitrary alternative bandwidths. The data for this table comes from the Stanford Open Policing Project and includes all 10 cities.

Table 6
Effect of DST on Stops of Drivers with Varying Polynomial Orders

	Polynomial Order			
	(1)	(2)	(3)	(4)
	1^{st}	2^{nd}	3^{rd}	4^{th}
Black Drivers				
During DST	0.455	-1.366	2.399	-1.826
	(1.959)	(1.488)	(4.508)	(4.955)
Hispanic Drivers				
During DST	-6.027**	-6.261	-3.878	-5.204
	(2.659)	(4.204)	(3.656)	(5.065)
Black Male Drivers				
During DST	4.017	1.147	7.259	0.395
	(2.541)	(3.102)	(4.717)	(6.058)
Hispanic Male Drivers				
During DST	-4.380	-4.408	-0.771	1.000
	(2.536)	(4.208)	(5.300)	(4.050)

Estimates and errors are all multiplied by 100
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table reports the effect of DST on stops of black drivers in the window of time between 4:00pm and 9:00pm. This fits a line with differing slopes prior to and following the start of DST. The coefficient on *During DST* tells us the effect of daylight savings time on the proportion of stops, so it is shown for each demographic using 4 different polynomial orders. The data for this table comes from the Stanford Open Policing Project and includes all 10 cities. model.

III. Drug Laws, Police Leniency, and Racial Disparities in Arrest Rates

1 Introduction

Decades of research[3] has provided a profound volume of literature showing evidence of racial disparities in the criminal justice system in the U.S., including in traffic stops, arrests, sentencing, bail setting, and parole decisions. These differences can stem from a multitude of underlying factors, such as social differences, income level, gender, etc. Inequitable treatment by law enforcement, the justice system as a whole, can lead to further issues, such as the poverty cycle and social unrest. Therefore, policymakers across the U.S. have long tried to adjust laws to ensure equitable and just practices are incorporated. Unfortunately, research shows this has been largely ineffective. Thus, it as an issue that still must be dealt with. Researchers must first determine the origins of these disparities in order for effective solutions to be developed and implemented. Discrimination in the U.S. justice system weakens its effectiveness and fails to remain impartial and apply equal rights and punishments to citizens.

If citizens believe police officers, judges, or other law enforcement will not be impartial, then they may be less likely to report a crime. Lack of trust in the police might lead to ineffective law enforcement; for example, if a minority individual fears there is a break in at their neighbor's house. Uncertain that police will not arrest them or think they were in some way involved might lead the individual to delay or refrain from calling

[3] Anwar and Fang (2006), Donohue and Levitt (2001), Abrams et al. (2012), Mustard (2001); Arnold (2018); Arvanites and Asher (2006); Rehavi and Starr (2014); West (2018)

the police altogether. There is not much research available describing how trust in law enforcement affects reporting, but it is logical to assume fear of being wrongly arrested, and therefore lack of trust in the justice system, might deter someone from calling the police when they witness a crime. Additionally, racially prejudiced decisions could mean government resources are wasted on detaining, holding, and rehabilitating individuals who may not necessarily deserve as much time in jail. Alternatively, those who might otherwise deserve to be in jail might go free due to an officer using their own discretion in a particular arrest and allowing a individual to go free when they may actually have broken the law only because they are white. There is the direct ineffectiveness here due to the individual not facing any legal consequences, but it also sends a signal that even getting caught might not lead to them being arrested; they may then be more likely to commit crimes in the future. The inability to remain impartial then results in many circumstances where the justice system breaks down.

It is the goal of the criminal justice system to impose laws and policies that equally benefit and punish people of any and all races, rather than disproportionately punishing or favoring individuals of a certain race. In fact, Justice Lewis Powell Jr. said "Equal justice under law is not merely a caption on the facade of the Supreme Court building, it is perhaps the most inspiring ideal of our society. It is one of the ends for which our entire legal system exists... it is fundamental that justice should be the same, in substance and availability, without regard to economic status."[4] Government resources are wasted when their tools are applied inappropriately and ineffectively.[5] Racial prejudice could lead government resources to be wasted on detaining, holding, and rehabilitating individuals who may not

[4]Supreme Court Associate Justice 1971 to 1987.

[5]There is not much research available describing how trust in law enforcement affects reporting, but it is logical to assume fear of being wrongly arrested, and therefore lack of trust in the justice system, might deter someone from calling the police when they witness a crime.

need it; or others go free due to an officer using their own discretion in a particular arrest when they may actually have broken the law only because they are white. There is the direct ineffectiveness due to the individual not facing any legal consequences, but also sends a signal that even getting caught might not leading to fewer crime disincentives. A lack of impartiality then leads to the justice system failing on a fundamental level. Therefore, it is necessary to accurately determine if the evolution of laws is pushing toward or away from equity. This paper specifically focuses on the equity of law enforcement and aims to determine the mechanisms through which arrests occur at disproportionately higher rates for minorities. Information excluded from police reports can skew the data leading to inaccurate estimated outcomes for stops and arrests because it fundamentally involves latent outcomes. Specifically, if officers choose not to report something systematically for individuals of a particular race, we would not know if these drivers did, in fact, break certain laws–I will give a more specific example related to this issue in the next section. Therefore, I look at arrest outcomes from a different angle. Taking that contraband is found at a stop as given, I determine that there are differing racial disparities in the rates at which arrests occur at different levels of severity.

Mustard (2001) shows that sentencing, especially for drug-trafficking offenses, disproportionately harms blacks and Hispanics. The sentencing disparity is even primarily driven by departures from the sentencing guidelines. The implication, here, is that judges make the decision to give harsher sentences to minorities, regardless of not making these decisions for white or wealthier individuals. It is unclear if these are conscious or subconscious harsher sentences. Regardless, it is clear that white people may be more likely to have less harsh sentences. If this same sort of disparity enters into the previous phase of law enforcement, then it may present in such a way that police officers are more likely to arrest minorities to whites for the same offenses. Additionally, it is reasonable to assume that these disparities are more likely to occur for lower level offenses. To address this, I use data on stops and arrests which include measures and types of contraband that allow me to see the severity of the contraband found. I then compare the disparity in arrest rates for minorities at multiple levels of severity to determine whether disparities are higher for lower levels of contraband, such as possession of a small amount of marijuana. If arrests occur at disproportionately higher at lower contraband levels, then this indicates that harsher drug laws that criminalize marijuana are more likely to lead to racial disparities in arrests. In this case, it is more equitable to relax marijuana laws so as to decrease the costs of racial prejudices within the justice system. In this paper, I look at racial disparity in arrest rates over several severity levels of contraband found on an arrestee to determine one way marijuana law relaxation can lead to positive social effects in the criminal justice system.

This paper contributes to a broad spectrum of literature examining the prevalence of racial discrimination in the criminal justice system and to more recent literature examining the consequences of marijuana legalization laws. Knowles, Persico, and Todd (2001) develop the KPT model to estimate discrimination that is expanded upon later by other researchers.[6] Bjerk (2004) develops a model to include a noisy signal of guilt and tests it to find it is difficult to determine if there is racial prejudice or if statistical discrimination can explain the arrest disparity. Antonovics and Knight (2009) find that police officers in Boston are less likely to conduct a search if the race of the officer matches the race of the driver and vice-versa. Alesina and La Ferrara (2014) find that death sentences of minority defendants convicted of killing white victims are more likely to be reversed on appeal. Additionally, Abrams et al. (2012) show that there is between-judge variation in

[6] Anwar and Fang (2006). Donohue and Levitt (2001) looked at the effect on arrest rates based no the demographics of officers.

incarceration rates. Ryan (2015) shows that interaction of the gender of the driver, the time of day of the traffic stop, and the existence of passengers in the stopped vehicle with the race of the driver all impact the probability of receiving a frisk. Alternatively, Anwar and Fang (2015) fail to reject the hypothesis that troopers of different races do not exhibit relative racial prejudice, though their model is prone to type-II errors, so this result is somewhat inconclusive. There is also a large literature examining racial discrimination in other settings, such as the labor market[7], the provision of healthcare[8], and the sharing, housing, sports, and credit markets.[9]

There are several reasons we might care about the potential positive effects of marijuana legalization. First, it can help determine positive social effects through increased police trust and equitable treatment of different races. It can also help determine a means of reducing behavioral differences in justice system, so laws can specifically target reducing racial disparities in arrest rates. Additionally, a drop in arrests for misdemeanor drugs might allow officers to focus on more serious drug activity and allowing resources to be spent more wisely on more harmful and costly crimes. Finally, it helps provide evidence for marijuana legalization, and subsequently for other minor legal offenses that might disproportionately favor individuals who are white or wealthy.

There is a large volume of literature suggesting that a portion of the racial disparities in the criminal justice system come from discrimination.[10] However, there is often little

[7]Lang et al. (2005) and Goldin and Rouse (2000)

[8]Anwar and Fang (2012)

[9]Edelman et al. (2017), Schafer (1979) and Edelman et. al (2017), Hamilton (1997), Dougal et al. (2019), Storey (2004), Bocian et al. (2008), and Dhakal (2019)

[10]Abrams et al. (2012), Mustard (2001): sentencing; Arnold (2018): stops; Arvanites and Asher (2006): imprisonment; Rehavi and Starr (2014): mandatory minimums; West (2018): traffic citations

evidence this is caused by taste-based (rather than statistical) racial discrimination. This research aims to show the magnitude of the racial disparity due to taste-based discrimination in arrests for contraband using leniency. I will test for the unobservable leniency exercised by police officers toward white individuals in this paper.

Consider the following example: a police officer stops a white male college student and notices he has a couple grams of pot in his center console. Since he is a young white man, the officer decides to pretend he did not notice. The stop is reported, but the marijuana possession is never written down. Therefore, this stop does not end up in the data. If a young black man is stopped in the same situation and racial bias factors into the officer's decision on how to proceed, he likely will take further action and arrest the man for possession of marijuana. If both drivers had heroin on them, the officer would likely not show leniency in either case as it is a schedule one drug.

Hit rate models will be biased toward no disparity if officers are systematically reporting a hit equal to 0 where it should be 1 for individuals with certain characteristics. It will appear the disparity is caused by statistical discrimination, i.e. black people are more likely to have drugs, but really is due to taste-based discrimination, i.e. the officer chose not to report the white driver's contraband. Therefore, another method is needed to try to tease out the effect of an officer's leniency that does not appear in the data, rather than simply looking at hit rates. By looking at different contraband severity–where low severity would be a small amount of a non-dangerous drug, e.g. 1g of marijuana, and high severity would be larger quantities, e.g. 5oz of marijuana–I can compare the arrest racial disparity at lower levels of contraband to that at higher levels, giving a better idea of how much racial prejudice affects the arrest rate. If the disparity is smaller at higher levels, controlling for individual and stop characteristics, this indicates police discretion explains the disparity at lower levels. This implies taste leads to less leniency for minorities at lower offense levels. I discuss this further in the Methodology section.

The data used in this paper are individual arrests from Chicago spanning 2016-2019 and totaling approximately 400,000 observations. The data includes the time and location of each stop, contraband found and its corresponding severity, race, gender, vehicle description, and other individual characteristics, e.g. age, sex, hair and eye color, etc. Using this detailed data, I can see whether a driver is arrested for a gram of marijuana or a gram of heroin, and whether the driver is white. If there is a larger racial disparity of arrests for lower levels of drug contraband, then this implies police discretion might be a determining factor in the arrests for smaller drug crimes, allowing racial prejudices to seep into the decision-making process. Thus, I employ a linear regression model to first find the difference in arrests rates, then find the difference in this disparity at each contraband level and determine if there is evidence of unequal treatment at lower levels.

In the next section, I provide a thorough explanation of Laws on marijuana possession within Illinois followed by a discuss the analysis of arrest rate disparities conducted in the literature thus far. In Section 3, I discuss the data used for this research, and in Section 4, I detail the analytical approach used here. In Section 5, I discuss the empirical results the data yield. I then conduct a sensitivity analysis and discuss shortfalls in the estimation approach used herein. Section 7 concludes the paper and discusses the implications and future avenues for research to add to and improve upon the information presented.

2 Background

2.1 Illinois Marijuana Laws

Throughout the period of the sample data, there have been multiple changes to Illinois state laws regarding possession and consumption of marijuana. Prior to July 29, 2016, The Cannabis Control Act–that was originally passed in 1978–outlined the laws surrounding marijuana use. Up to this point, possession or use of less than 2.5 grams of marijuana was a Class C misdemeanor with a fine of $100-$200. Possessing 10g-30g was a Class A misdemeanor. Possession of 30 grams or more was a felony, where classes increase with amounts. In order to improve law enforcement effectiveness for more violent and severe crimes, the penalties and classification of the aforementioned crimes were reduced. The most notable change was the decriminalization of possession of smaller amounts of cannabis.

On July 29, 2016, Public Act 99-0697 was enacted, thereby amending the Cannabis Control Act. This act states that possession of up to 10 grams of cannabis is a civil law violation punishable by a fine between $100-$200 and removes the class C misdemeanor status of possessing less than 2.5 grams of cannabis, as well as the class B misdemeanor status of possessing 2.5g-10g of cannabis. This act makes possession of 10g-30g of cannabis a class B misdemeanor and possession of 30g-100g of cannabis a class A misdemeanor (down from a felony). Possession of 100 grams or more remains a felony that increases in class as the amount in possession increases. A class C misdemeanor could result in up to 30 days in jail, while a class B misdemeanor could result in up to 6 months. A felony charge for marijuana possession carries a mandatory minimum of one year in jail and could result in up to 6 years. Thus, these changes to the Cannabis Control Act mean that individuals found in possession of up to 10 grams of marijuana will not face the possibility of jail time, but be sentenced to jail time if they have more than 10 grams, dependent upon the judge assigned to the case and various other legal factors.

Moving forward, the Illinois legislature began discussing the possibility of legalizing recreational sale and use of marijuana over the following two years before introducing to the state senate Public Act 101-0027. After the state legislature passed this bill, governor Pritzker signed the bill on June 25, 2019, to officially take effect on January 1, 2020. Thus, the Cannabis Control Act of 1978 became partially obsolete. In this act, under Article 10, personal use of cannabis by individuals 21 year of age or older is legal. This is, of course, provided that they do not carry more than the legal limit of 30 grams of cannabis. Carrying over the legal limit results in the same charges outlined in Public Act 99-0697. Additionally, anyone under 21 years of age found in possession of marijuana faces a fine for a civil law violation.

In addition to changes in marijuana laws, there was another major change that affected the arrest rates in Chicago. To reduce the amount of discretion officers could use in stops related to stop and frisk. After this change in stop and frisk policies, officers are required to report reasoning for each stop and all officers are now required to receive training directed at ensuring each stop is made due to "reasonable suspicion of criminal conduct." l. On May 30, 2015, senate bill 1304 passed the Illinois Congress and it was signed into law officially on August 12, 2015. This was as part of Public Act 099-0352.

As part of their requirements to report reasoning for each stop, the data for these stops must now be collected and later released to the public. This is must explain all of "the reasons that led to a protective pat down and whether it was with consent." This potentially decreases arrests due to stop and frisk, which directly impacts marijuana arrests as a portion of these are a result stop and frisks.

2.2 Literature Review

Abrams et al. (2012) measure the between-judge variation in the difference in incarceration rates and sentence lengths between black and white defendants in felony cases from Cook County, Illinois. using a Monte Carlo simulation in order to construct the a counterfactual, in which race has no influence in sentencing. They find significant between-judge variation in incarceration rates, although not in sentence lengths. These results mirror Arvanites and Asher (2006) who show that there is a positive and significant effect of race on imprisonment rates, though the indirect effect is greater than the direct effect. However, Abrams et al. (2012) contrasts with Mustard (2001), who does find evidence that judges make the decision to give harsher sentences to minorities, regardless of not making these decisions for white or wealthier individuals. Additionally, Rehavi and Starr (2014) find that black individuals are 1.75 times more likely to face a charge carrying a mandatory minimum than white individuals. Though these are not necessarily harsher sentencing decisions by the judge, these charges automatically come with a minimum prison sentence and could exacerbate sentencing differentials. The data in this paper span a multitude of legal regimes under which larger amounts of marijuana possession carry a mandatory minimum until legal restrictions on marijuana are relaxed. The cutoff for mandatory minimums sends a signal of severity and can affect the distribution of arrests. Here, we would expect that racial prejudices do not play a role in the decision to arrest when marijuana is found.

There is less research into the disparity in bail setting. However, using estimates from Miami and Philadelphia, Arnold et al. (2018), tests for racial bias in bail decisions. They show that bail judges are racially biased against black defendants, with more racial bias among inexperienced and part-time judges. In addition, they find that both black and white judges are biased against black defendants; therefore, they argue that this is consistent with bail judges making racially biased prediction errors, rather than being racially prejudiced, meaning they find evidence of a statistical bias.

To deal with the difficulty of teasing out taste-based racial bias in traffic citations, West (2018) examines automobile crash investigations for which officer dispatch is exogenous to driver race. He finds that state police issue more traffic citations to individuals of a different race from their own, indicating a preference for discriminatory leniency towards same-race individuals. Additionally, Donohue and Levitt (2020) use panel data from 122 large U.S. cities to show that increases in the number of minority police are associated with increases in arrests of white individuals, but have little impact on arrests of minorities. Similarly, more white police officers is associated with an increase in the number of arrests of minorities but do not systematically affect the number of white arrests. They estimate that maximizing same-race policing would lead arrests to decrease by over 15 percent.

Grogger and Ridgeway (2006) examine the disparity in police stops and do not find evidence of racial bias, though Pierson et al. (2019) re-examines this model using a larger, broader dataset across dozens of U.S. cities and find that there is a disparity in stop rates. Feigenberg and Miller (2020) use unique Texas administrative data to isolate variation in search behavior across sate troopers. They find that search rates are unrelated to the proportion of searches that yield contraband. According to their results, troopers appear unable to distinguish between those who are more or less likely to carry contraband.

Knowles et al. (2001) do not find evidence of racial prejudice against black drivers. They note that, if police have utility only for searches yielding large drug finds, then their analysis would suggest bias against white drivers. However, they provide no argument as to why they believe police have a higher utility for larger drug finds. In their model, they show that, if troopers are not racially prejudiced, all motorists must carry contraband with equal probability regardless of their race and other characteristics, in equilibrium. However, this implies that a motorist's characteristics other than race do not provide information about the presence of contraband when a trooper decides whether to search. However, trooper guidelines require officers to base their search decisions on the information the motorist presents to the trooper at the time of the stop, such as their demeanor and the contents of their vehicle that are in plain view. (Riksheim and Chermak 1993). Anwar and Fang (2005) allows the possibility that police behavior might vary by racial group by relaxing the assumptions made by KPT by using race-matchin, using Florida data that provides race of the trooper in addition to driver's race. When replicating the KPT test, the result immediately implies that the troopers show racial prejudice against black and Hispanic motorists. However, this conclusion is only valid if their model of motorist and trooper behavior is true. Anwar and Fang's (2005) results show that, without strong assumptions, they cannot confidently prove the presence of relative, racial prejudice within the modified KPT framework. Cox and Cunningham (2017) estimate the effect of a federal grant program as part of the 1988 Anti-Drug Abuse Act to combat illicit drugs and provide evidence that federal involvement in narcotic control and trafficking leads to an increase in drug arrests; disproportionally affecting black individuals. This implies that targeting drug use leads to a larger racial disparities and relaxation of these policies may lead to a smaller racial disparity in arrests, and therefore a smaller disparity in drug-related crimes that lead to prison sentences. This could reduce the disparity at numerous levels of the criminal justice system and could instead allow funding, such as the ADAA grant, to fund reduction of other more violent crimes or to fund drug abuse mitigation through recovery for addicts.

There is some evidence of racial bias at each stage of the criminal justice system, though the magnitudes of this are still disputed, as is the underlying cause for the disparity, i.e. whether the estimates accurately estimate taste-based racial bias or statistical bias. The aim of this paper is to assess how much of the racial disparity in arrests is due to racial bias and what portion of this is due to statistical vs taste-based preferential treatment by race.

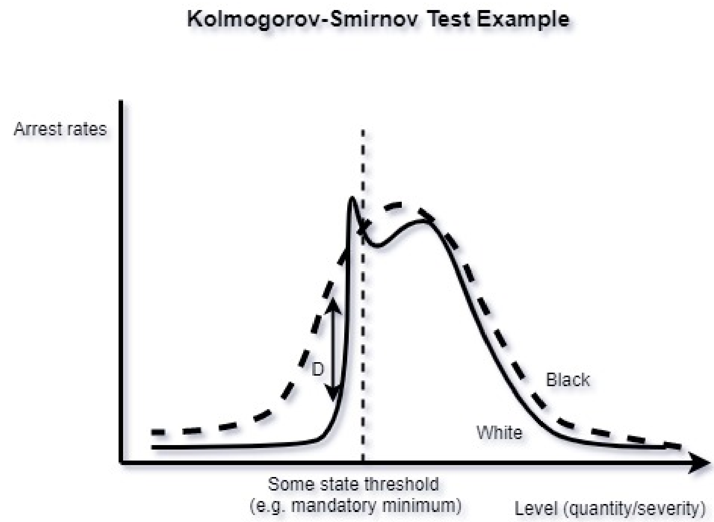
2.3 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) test is a non-parametric and distribution assumption-free test. It requires no assumptions to be made about the distribution of the data. The KS test can be used to compare a sample with a reference probability distribution, or to compare two samples. These two aspects of the test make it versatile. In order to determine if the two data samples I am interested in, arrest rates of black vs white people, the KS test can be used. The null hypothesis is that the samples do indeed come from the same distribution; if we reject the null, then we accept the alternative hypothesis that the samples do not come from the same distribution.[11] To clarify this test, I provide a hypothetical illustration in Figure 13.

The estimated D is the maximum difference between the two distributions. This tells the biggest difference, but the p-value from this test gives more information. With a

[11] Motivation for when to use the KS test is detailed by Kawwa (2020) and Lopes et al. (2007)

Figure 13



Notes: This figure provides an example diagram of the estimation provided by the Kolmogorov-Sirnov Test.

small enough p-value to find significance, we can say that the distributions are statistically significantly different. However, this simple test does not provide information on what ways the distributions differ. For this, further analysis is needed through plots and regression analysis. For this paper, I use the KS test to determine whether the distribution of white arrest rates and black arrest rates over quantity differ over the whole distribution, then whether they differ specifically in the lower and higher quantity regions. I will expand upon this further in the methodology section 7.

3 Data

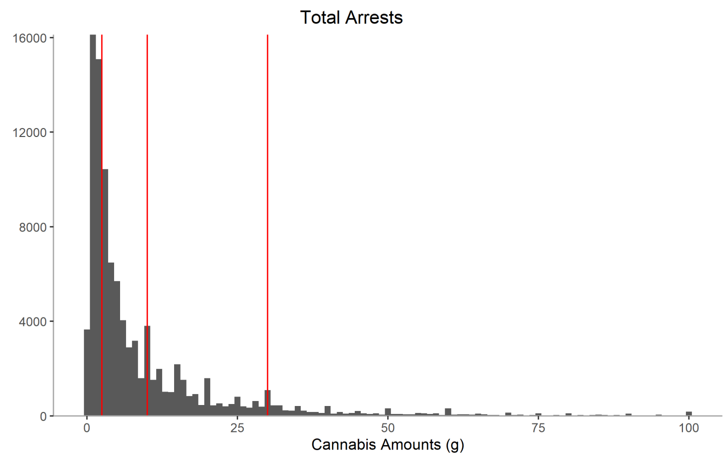
The data used in this paper are contraband arrests from Chicago, IL from January 1, 2012 to October 19, 2020. This data was procured through a FOIA request that has been ongoing for much of 2020 and was released to me in October 2020 by the Data

Fulfillment and Analysis Section in the Strategic Data Analytics Division of the Chicago Police Department. There are nearly 220,000 contraband arrests, almost half of which are specifically marijuana arrests, which is the primary subset I focus on in the following analysis. The data includes the time and location of each stop, amount of contraband found, individual's name, race, gender, and other individual characteristics, e.g. age, sex, hair and eye color. I also received these same characteristic descriptions for the arresting officer. Approximately 79% of the individuals arrested for marijuana possession are black with less 5% being white and 7% Hispanic. Of the overall dataset, approximately 78% are black and 6% are white, with approximately 15% Hispanic.

The incidence of arrests at each weight are shown in Figure 25b. The vast majority of all marijuana arrests occur at less than 100 grams. Any possession of marijuana 30g or more is a felony under Illinois laws with 30g-100g being downgraded to a Class A misdemeanor when legalization occurs. The laws for possession of marijuana up to 30g change over the sample period, with 2.5g or less being the least severe bracket. Between 10g-30g is still relatively low severity due to its decriminalization status as of 2016 and makes up about one-third of the marijuana arrests. Any possession under 30g is considered legal as of 2020.

The time series plots in Figure 15 show the arrest rates across the data period for each race. Panel 15a shows the arrests for entire dataset, including arrests for non-cannabis related controlled substance possession arrests. Panel 15b shows the arrest rates for just individuals carrying marijuana. The shaded regions indicate changes in the policies for legal marijuana possession and use in Illinois. There is an overall downward trend for all contraband arrests with a few noticeable jumps, or trend changes that seem to correspond with policy shifts, and of course a Covid-19 shock at the very end of the data period. Look-

Figure 14



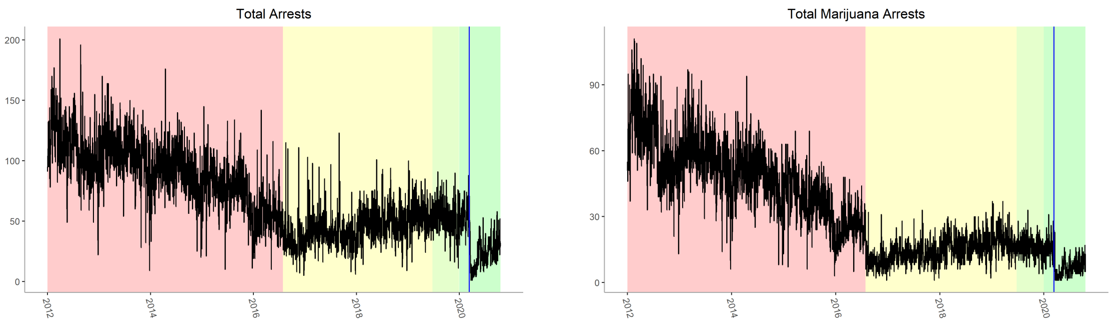
Notes: This graph shows the number of arrests for marijuana possession in the dataset using proprietary data from the Chicago Police Department. Each vertical red bar denotes a legal threshold

ing at 15b, it is notable that there appears to be a drop when decriminalization occurs at the start of the yellow region, but the more dramatic shift here is the change in the overall trend. It seems to flatten out, which is possibly due to a more stable enforcement policy, though it there are other possible explanations. Looking at the start of the green section, it is difficult to see a drop in arrest rates due to legalization, which may be because decriminalization already occurred for overlapping amount brackets that legalization applies to. Less than 10 grams was already a civil law violation requiring no arrests and 10-30 grams was a class B misdemeanor. Therefore, up to 10 grams already did not lead to any arrests during the yellow region, prior to legalization.

These trends are broken up by race in Figure 27a. It clear here just how large the difference in number of arrests is over the sample period. Additionally, although all races do follow a similar trend, we can see that the overall trends are largely driven by the

Figure 15
Time Series of Arrests

(a) **(b)**



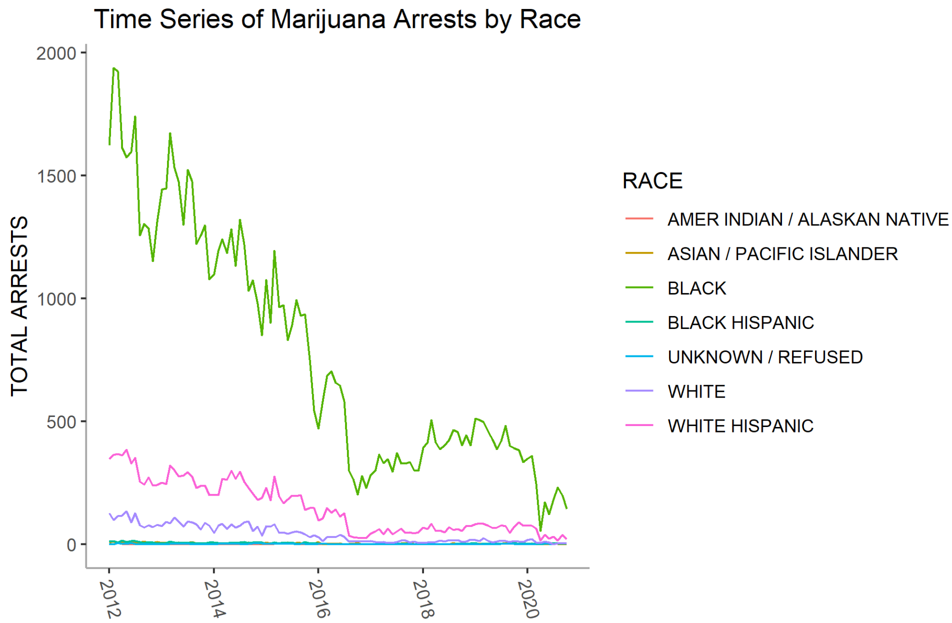
Notes: These time series show the daily number of drug possession arrests in the dataset using proprietary data from the Chicago Police Department. Panel (a) depicts the time series of all arrests and Panel (b) depicts the time series for marijuana arrests.

changes in arrests of black people.

The makeup of the marijuana arrest data are explained by Figure 26a and Figure 26b, for all races then for white and black individuals only, respectively. Here, you can see that the incidence of marijuana arrests is higher for black individuals for smaller amounts (in grams) of marijuana in their possession, but this may not be true for larger amounts. The vertical red lines show the cutoff amounts based on IL state laws for marijuana. Looking at the histogram, it is possible that bunching behavior is occurring at the lower bounds of each legal bracket, especially for black individuals.

Moving into the analysis, the density of arrests will be important, as I need to test my hypothesis to provide motivation for the next step of my analysis. Figure 18 shows

Figure 16

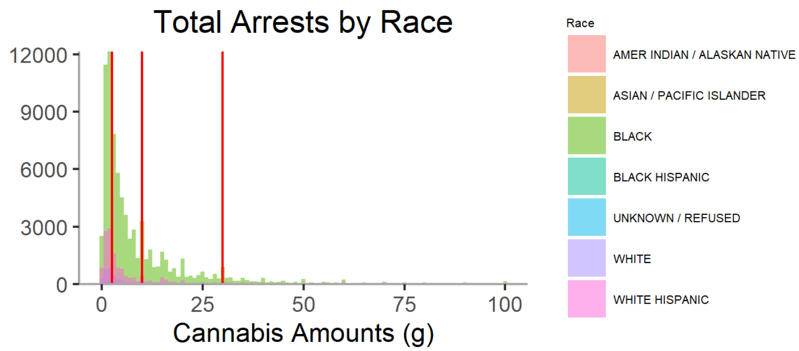


Notes: This graph shows the time series for monthly marijuana possession arrests in the dataset by race, using proprietary data from the Chicago Police Department.

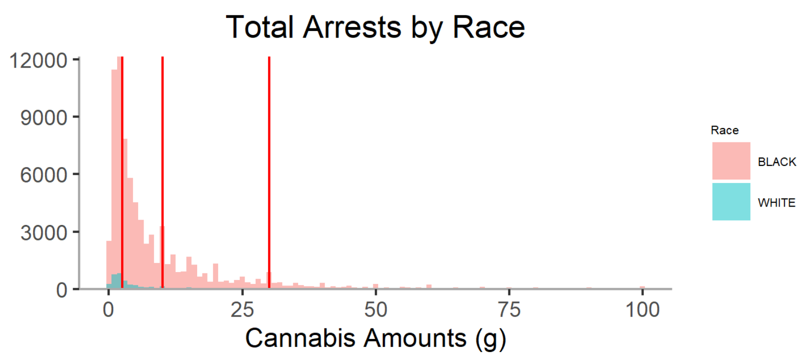
the density plot of marijuana arrests for white and black individuals. Before testing the density, it seems plausible that my hypothesis is supported by the data since, at the lower severity end, there is a noticeable racial disparity; however, at higher severities, the disparity actually flips–though I postulate that this disparity will not be statistically significant. This preliminary figure supports the idea that other papers estimating the racial disparity across all severities are picking up the average across the whole distribution and, therefore, their estimates are likely biased toward zero. Further support will be provided in the next section where the KS test is performed.

Figure 17
Histogram of Marijuana Arrests by Race

(a)



(b)



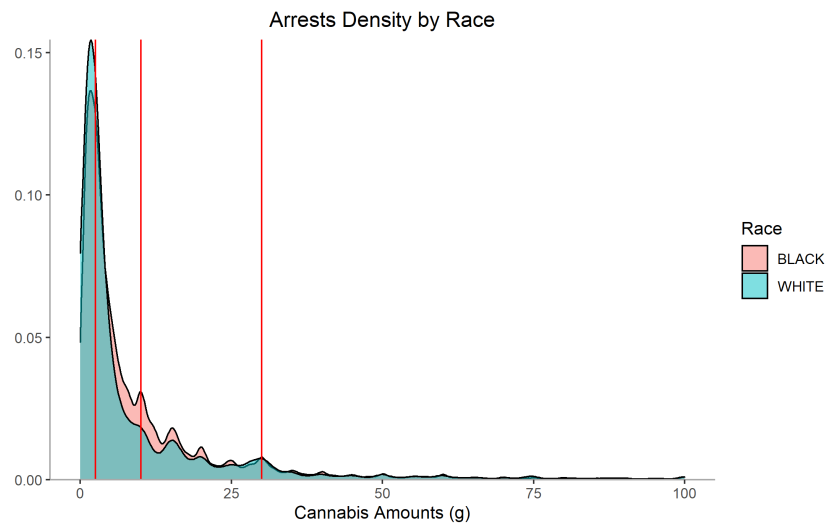
Notes: These graphs show the number of marijuana possession arrests in the dataset using proprietary data from the Chicago Police Department, by race. The vertical red bars denote the legal thresholds.

4 Methodology

4.1 Kolmogorov-Smirnov Test

To test whether the difference in arrest rate density is significantly different between white and black individuals, I first use the Kolmogorov-Smirnov (KS) test. This tests if the CDFs of each set of data are the same. I hypothesize that the distribution for arrests at lower contraband levels shows higher densities for blacks, but higher levels are statistically

Figure 18



Notes: The figure shows the distribution for marijuana possession arrests in the dataset for black and white individuals, using proprietary data from the Chicago Police Department. The vertical red bars denote the legal thresholds.

similar. In order to test the lower and higher parts of the CDF separately, I must run the KS test twice. The cutoffs I use are the levels at which marijuana possession sentencing carries a higher citation and penalties or a mandatory minimum, which is legally mandated by the state of Illinois. Where previous research often finds small or no racial bias using a simple hit rate model, this distribution comparison might find more evidence of racial bias because it shows if there are differences in behavior of officers at different levels of severity. If this is occurring, then previous papers understate racial bias by concluding arrest rates are higher for blacks, so the disparity is explained by the higher propensity to commit a crime based on race. However, their results are missing some criminal behavior on the part of white individuals. The distance, D, on the left tail is the region where white people are actually committing crimes that would lead to arrest, but that are not reported.

The Two-Sample KS test shows that the two samples follow differing distributions for smaller, but not larger, amounts. The I compare are the those from below 10 grams, then below 30 grams, then 30 - 100 grams, 100-300 and 300 - 500, since, for IL, these are the legal cutoffs. Therefore, this is where policy changes are likely to lead to more heterogeneity in arrest rates. Additionally, I choose to then exclude outliers in the higher weight range. It is prudent to do this because there are a small number of extremely large amounts, e.g. over 200,000 grams, and are likely to be associated with drug cartel or large gang busts, rather than average individuals. Large busts like this occur very infrequently and it is reasonable to assume individuals of these gangs are more likely to share race due to the segregated nature of gangs as described by Hagedorn (2006). Thus, there is more likely to be a one time jump in the distribution by race and can throw off the estimates for larger amounts that otherwise follow a similar distribution. As is shown in the previous plots, I choose to use 100 grams as the higher cutoff because it is a natural legal threshold, above which possession is classified as a class 4 felony.

One shortcoming of this portion of the analysis is that there is potentially a systematic difference im how black vs white people choose the amounts of marijuana which they carry. However, according to Keyes et al. (2017), many epidemiological studies of race/ethnicity and marijuana use have found that among adolescents, white adolescents are more or equally likely to use marijuana than their black peers. While some studies of adults indicate that Black Americans surveyed in 2012–2013 had higher rates of cannabis use disorders than whites, other surveys of the same sampling found that white adults had higher rates. Additionally, according to Feigenberg and Miller (2020), during routine traffic stops, police are more likely to search black and Hispanic motorists for contraband than white motorists, yet searches of black and Hispanic motorists are equally likely to yield contraband. Chung et al. (2006) notes that their results show that rates of use of marijuana among black youths tend to be lower than among their white counterparts and note that this is confirms a “well-known” result. In my data, the average age of back individuals is 26 with white being 27. The 90th percentile for black arrestees is only 39 (41 for white), so the data consists of more young adults than older. This suggests that there is no reason to expect use of marijuana to differ by race and quantity to the extent that this test shows occurs in arrest rates in this systematic way.

4.2 Linear Model

Differences in the distribution also mean that the outcomes of standard hit rate models, implying either statistical or taste-based discrimination, are not trustworthy. It misses crucial information about the change in the mean over different levels of severity. I re-examine the link between subjective perceptions and objective measures of race discrimination by estimating the mean over several severity brackets in the distribution of arrest rates between white and black arrestees in attempt to decompose the race arrest rate disparity into the part attributed to different driver characteristics and the part attributable to differential treatment at points other than the conditional mean. To do so, I employ a linear regression, then take the difference between two severity ranges of the arrest rate densities between white and black individuals.[12] This gives a more precise estimate of the disparity at different severities that is conditioned on driver and stop characteristics. I start with my hypothesis.

Hypothesis: Officers are more likely to show leniency for less severe offenses than for more severe ones.

[12]Tuttle (2019)

Hit rate models tend to be some variation of this baseline model

$$hit_i = \beta_0 + \beta_1 Black_i$$

The problem is, this model's estimates depend on police reporting of "hits."

$$hit = \begin{cases} 0, & \text{when no contraband is found} \\ 1, & \text{when contraband is found} \end{cases}$$

If an officer shows leniency to an individual and chooses to not arrest them for illegal contraband, they must report a 0; otherwise, they are obligated to take legal action as is dictated by their precinct, local laws, and federal laws. This means there appear to be more 0s than what occurs in reality because a 1 was not reported when a hit occurred.

The KS test is consistent with the idea that standard hit rate model outcomes are biased because of latent variables they do not account for. Therefore, I employ a linear regression approach then decompose the arrest rate racial disparity into the part attributed to different statistical discrimination and the part attributed to differential treatment. The baseline model for race group r in quantity group q in month t is

$$arrestrates_{rqt} = \beta_0 + \gamma_1(\text{Race} \times \text{Quantity})_{rq} + \epsilon_{rqt} \tag{2.1}$$

This allows me to estimate equation 1, while using this approach. I find the disparity in the lower and higher severities, then take the difference in those disparities:

$$\gamma_1^{B1} - \gamma_1^{W1} = \beta_1^1 \tag{a}$$

$$\gamma_1^{B4} - \gamma_1^{W4} = \beta_1^4 \tag{b}$$

$$\beta_1^1 - \underbrace{\beta_1^4}_{\text{Statistical}} = \underbrace{\widetilde{\beta_1}}_{\text{Taste-based}} \tag{c}$$

where q=1 is the lowest severity. I then run the equation (1) again and add a time fixed effect to the model.

$$arrestrates_{rqt} = \beta_0 + \gamma_1(\text{Race} \times \text{Quantity})_{rq} + \gamma_2\alpha_t + \epsilon_{rqt} \tag{2.2}$$

This ultimately provides a more precise estimate of taste-based racial discrimination by differencing out the statistical discrimination in the racial disparity.

4.3 Interrupted Time Series

Given that changes in arrest laws may change the size of the disparity in arrest rates, I estimate the effects law changes in Illinois using an Interrupted Time Series.[13] This helps to determine if relaxation of marijuana laws affects the disparity in arrest rates, as well. [14] Knowing that quantity does have an effect on the disparity, I then estimate the interrupted time series taking quantity into account. This is important because policies targeting the arrest rate disparity in marijuana may need to be adjusted to depend on where this disparity is coming from, i.e. the quantities of possession that drive it.

[13]I follow the methodology for ITS outlined by Bernal et al. (2017)
[14]Recall that the regions are color coded in Figure 15b

To test the effect on the arrest rate disparity of relaxing marijuana laws, I conduct an interrupted time series (ITS) around decriminalization. I begin with the following model that assumes both the level and slope of the arrest rate are affected by decriminalization.

$$arrestrates_{rt} = \delta_0 + \delta_1 \text{Day} + \delta_2 \text{Decrim} + \delta_3 (\text{Day} \times \text{Decrim}_t) + \upsilon_{rt} \tag{ITS 1}$$

where r=race and t=day. It is possible that this policy shift leads to a response model wherein marijuana decriminalization leads to a change in the level of the arrest rate, but not a change in the slope over time. Therefore, I then run the following variation on ITS 1:

$$arrestrates_{rt} = \delta_0 + \delta_1 \text{Day} + \delta_2 \text{Decrim} + \upsilon_{rt} \tag{ITS 2}$$

Additionally, it is feasible for decriminalization to lead to a change in the slope of the arrest rate, but not a change in the level. To estimate this case, I run the following:

$$arrestrates_{rt} = \delta_0 + \delta_1 \text{Day} + \delta_3 (\text{Day} \times \text{Decrim}_t) + \upsilon_{rt} \tag{ITS 3}$$

After running these three models for the overall dataset, as well the low vs high quantity groups, I determine the most likely appropriate response model based on the estimates. Based on the visible trend in time series of arrests in Figure 27a, I posit that ITS 1 will be the most appropriate model to describe the policy change effect. In the final time series estimation, I use the best model estimated for each quantity group and run the model for each race, independently, to determine how the arrest rate changes for each race due to decriminalization. The results from these regressions are reported in Tables 9 through 12. I will discuss these estimates in the results section 7.

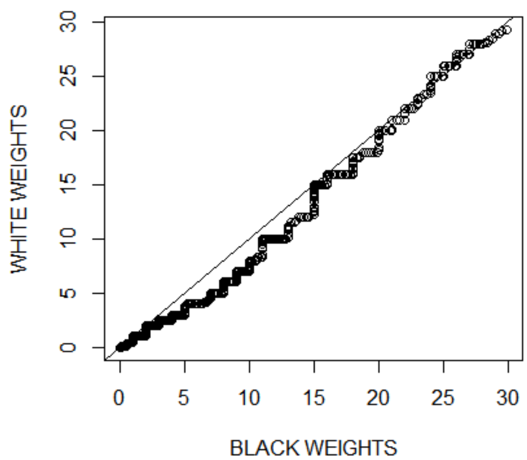
5 Results

5.1 Kolmogorov-Smirnov Test Results

As per the Methodology section, the first step is to run the KS test. When comparing the distributions for arrests of white individuals to black individuals, we get a positive D estimate of approximately 0.135. This estimate can be found in Table 7 and is the maximum distance between the two curves in the lower severity range, so it gives little to no information regarding the disparity and is, therefore, not important apart from being non-zero. The more important estimate is the p-value of 2.2×10^{-16}, essentially 0. This means that the distributions are significantly different between 0 and 30 grams. When looking at the higher severity, between 30 and 100 grams, D is 0.047, a much smaller maximum difference. But again, I care about the p-value to determine if the more sever distributions are statistically different. The p-value for this range is 0.49 and therefore, which tells us the distributions are similar enough to be statistically identical. These outcomes are in line with my hypothesis and motivate the next section of this paper.

As a follow up to this test, I also create a quantile-quantile plot, which can be used to eyeball whether two datasets follow the same distribution. If they do, the plot creates relatively straight line, diagonally, across the plot box. This plot is shown in Figure 19 and provides further evidence that the distributions differ on the lower end of the distribution, but less so on the upper end. There is too much curvature to say these follow the same distribution. In the figure, “WEIGHTS” refers to the approximate weight of marijuana in their possession.

Figure 19
Quantile-Quantile Plot of Distribution of Marijuana Arrests



Notes: The above figure shows the Quantile-Quantile plot for marijuana possession arrests for the black and white distribution of arrests in the dataset, using proprietary data from the Chicago Police Department. It describes the deviation from following the same distribution.

5.2 Linear Regression Results

As is detailed in the previous section, I model the affect of race on arrest rates by running a linear regression where arrest rates is the outcome and Race × Quantity is the main regressor. This provides the estimated difference in arrests rates due to being in race group r at quantity level q. I then calculate the difference in this disparity at each contraband level using equations (a) through (c) and determine if there is evidence of unequal treatment at lower levels.

The results for equation (1) are reported in Table 8. Here, the omitted dummy variable is that of Black individuals in Quantity region 1. Therefore, this is our estimate of the less than 2.5 grams quantity black arrest rate, and thus γ_1^{B1} is essentially zero. It is approximately 31 arrests/100,000 population. This is significant at 99.999% level. The estimate for γ_1^{W1} is -28.9, or there are nearly 29 fewer marijuana arrests per 100,000 population of white people with less than 2.5 grams in their possession. The estimated γ_1^{B4} is -21.2 and γ_1^{W4} is -30.2. Given the base level is γ_1^{B1}, β_1^1 is simply 28.9 (I round up to 30), or the difference between γ_1^{B1} and γ_1^{W1}. Taking the difference between γ_1^{B4} and γ_1^{W4} (in absolute value, since both deviate from γ_1^{B1} in the negative direction), we find $\beta_1^4 = 9$. Now it is straight forward to calculate the differential in the disparity:

$$\beta_1^1 - \underbrace{\beta_1^4}_{\text{Statistical}} = \underbrace{\widetilde{\beta_1}}_{\text{Taste-based}} \tag{c}$$

$$30 - 9 = 21$$

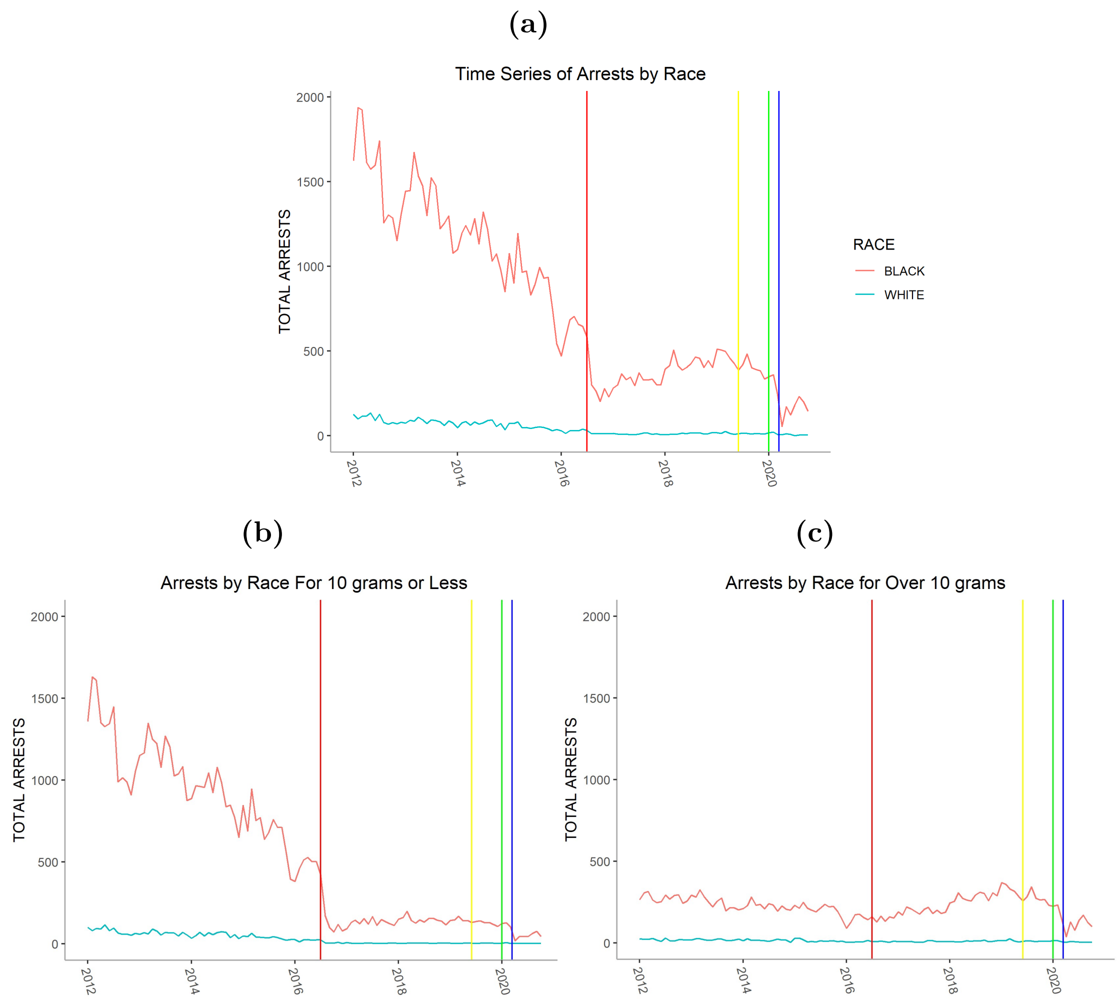
Thus, $\widetilde{\beta_1} = 21$ and is consistent with my hypothesis because it shows that the disparity in arrest rates by race is more than two times higher for smaller quantities. The estimates for these values from equation (2) are almost exactly the same, except the intercept is 46. This, however, does not affect the differential of the disparity estimates, only the level of the disparity. Therefore, $\widetilde{\beta_1}$ is still 21.

5.3 Interrupted Time Series Results

Figure 20 includes four panels, detailing the time series of arrests by race to visualize the disparity in overall arrests.

Looking at these figures, it is clear that, prior to decriminalization, there was a downward trend in overall black arrests for marijuana possession; this was largely driven by decreases

Figure 20
Time Series of Arrests by Race
with Regime Changes



Notes: These graphs show the time series of monthly marijuana possession arrests in the dataset using proprietary data from the Chicago Police Department. The vertical bars denote the start of end of each legal regime.

in arrest rates at lower quantities. At the time of decriminalization there is a stark drop in arrest rates that leads to a new level of arrests, but also to a different slope of arrest rates over the following years. There are a couple other noticeable drops ion monthly arrest rates between 2012-2016, but they both appear to be temporary and leave the overall trend relatively unaffected. In the overall time series in Figure 15, there is a larger downward slope going from mid 2015 to the end of 2015. I believe this is due to the change in stop and frisk policy that occurred over the course of several months, but it appears to by only lead to a temporary drop in marijuana arrests.[15]

For panel 20c, there is seemingly no drop in the monthly arrest rate at the time of decriminalization; if there is a level change in the disparity, it is quite small, though may be slightly more pronounced at the day frequency when I can more clearly estimate the level before and after the policy change. It does appear to be followed by an increase in the black arrest rates for higher quantities of marijuana and, therefore, a change in the slope of the disparity. Since the plot of arrest rates does not give the statistical significance of level and trend changes, I estimate the ITS under each possible scenario: (1) both trend and level are affected by the policy, (2) only the level is changed, (3) only the slope is changed. I then use the estimates of these models to determine which best describes the effect of the policy change in the arrest rate disparity and run this model for each race group.

The results presented in Table 9 through 11 show that the effect of decriminalization likely, even if to a small degree, had an effect on the level and slope of each group. Therefore, I use ITS 1 to estimate the results for Table 12. The first thing to take away from these results is that the decriminalization policy appears to have no affect on white arrest rates for possession of less than 10 grams of marijuana. For black arrest rates, the policy change leads to a decrease in the level by approximately 0.63 per 100000 population per day. Therefore, the disparity here sees a large decrease (by 0.63 arrests per 100000 pop per day). There is also a resulting positive change in the slope, increasing the trend by about

[15]I describe this policy shift in 7

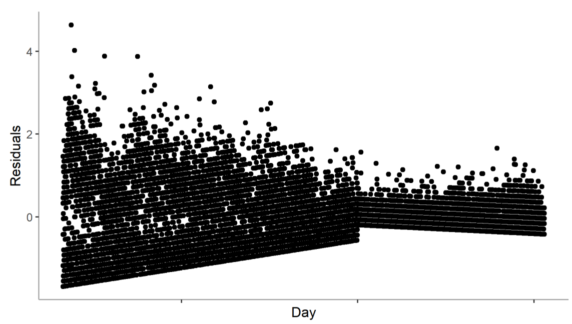
10 percent. This makes sense because the number of arrests are quite low after the policy change, but means that decriminalization in Chicago did result in a flatter downward trend. For high quantities, decriminalization is associated with a decrease in black arrests by 0.055 per 100000 population and for white arrests a decrease in arrests by 0.02 per 100000 population per day. Therefore, the disparity decreases by 0.03 arrests per 100000 population per day. These changes are much smaller, which is what we would expect because the laws changed such that up to 10 grams results in a fine, but no arrest, while over 10 grams is still a crime in this period.

A more interesting result at higher quantities is that the trend increase enough after decriminalization to become positive. I posit that this affect is due to resource reallocation. Police resources are no longer dedicated as much to arresting and stopping individuals who much have small amounts of marijuana on their person. Instead, they can use their time to focus on more severe offenses. If they care about getting drugs off of the street, they will then focus more time on arresting individuals with large amounts of marijuana, rather than small. This leads to more time focused on stopping gang and cartel related drug crimes, rather than average people with a joint on them for later. Additionally, there is less money allocated toward sentencing and jailing of these individuals that can be used for higher severity offenses. I would like to point out that this trend shift is more pronounced for black arrest rates than white arrest rates and may indicate that this higher quantity increase in enforcement may disproportionately affect minorities. Looking at Figure 20a, it appears this upward trend in black arrests could have kept increasing to the pre-policy change levels. However, this does not occur. It is possible that the passing of legalization in IL changed the overall trend enough to keep this from occurring. I plan to analyze the policy change here in future research and determine if this is the case.

5.4 Robustness and Heterogeneity

The main issue with an Interrupted Time Series approach is the presence of partial or serial autocorrelation. It is often present in the model when one or more important variables are omitted. I plotted the residuals for my data, it does not appear that this is too much of a concern based on the plotted residuals for the black arrest rate models; however there is a bit more of a concern with regards to ITS 1 for white arrests as is seen in Figure 22. Unfortunately, with this data, I do not have a way of dealing with this issue at this time. It could lead to some bias in the estimated outcome of the ITS models.

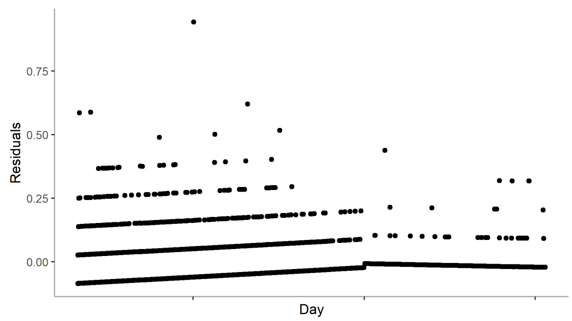
Figure 21
Plotted Residual for Black Arrests



Notes: This figure depicts the plotted residuals for marijuana possession arrests of black individuals in the data, using proprietary data from the Chicago Police Department.

Figure 23a shows the incidence of arrests of black people for possession of marijuana in Chicago. The black colored region is the rest of cook county, outside of CPD jurisdiction. Alternatively, Figure 23b shows the incidence of arrests of white people for possession of marijuana in Chicago. The scales for these figures differ, so this is meant to show where the majority of arrests are occurring for each race, though the highest number of arrests in a particular district are much higher for black individuals, at almost 10,000. That of white individuals is 288. It is notable that the zip codes with higher frequencies occur mostly

Figure 22
Plotted Residual for White Arrests



Notes: This figure depicts the plotted residuals for marijuana possession arrests of white individuals in the data, using proprietary data from the Chicago Police Department.

to the south east portion of Chicago for black arrests, with those of white occurring more toward the northern districts of the city. Also, note that many of the blackened districts are outside of CPD jurisdiction, though not all of them. Within their jurisdiction, some districts do have very few or zero arrests, especially for white arrests.

Figure 24 shows the arrests per population of each zip code for possession of marijuana in Chicago.[16]This map specifically details the arrests within the city limits of Chicago, rather than the whole Chicago Metro area. The district in gray is the Chicago O'Hare International airport, so it has a population of 0. Panel 24a shows all of the marijuana arrests per total population of the district. Panel 24b shows the marijuana arrests of all white people per population within each zip code district. Panels 24c and 24d show the same information for black and Hispanic peoples, respectively.[17] The large gray district is gray because, although there are arrests in it, there is a 0 population as it is the Chicago O'Hare

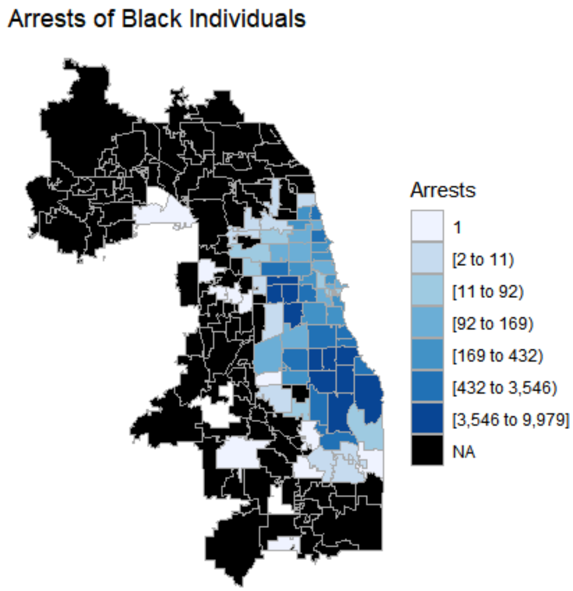
[16]The population by Zip code comes from 2018 census provided by the U.S. Census Bureau.
[17]Here, Hispanic is the sum of black and white Hispanic individuals.

International Airport.

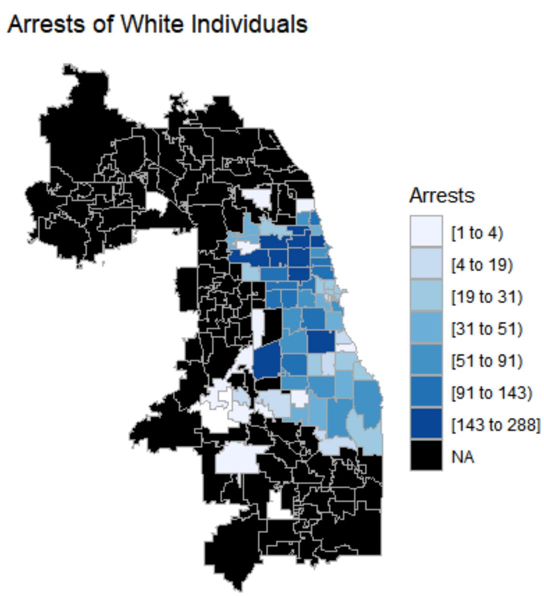
Looking at this figure, it is strikingly noticeable that much of the denser zip codes. Keep in mind that this shows the arrests of each race per total population of the district, not the population by race. It is also clear that there are certain districts that tend to have higher arrest rates for marijuana possession than others across all races. White arrest rates are the most evenly dispersed of all three races. Keep in mind that Chicago is approximately 30% white, 30% black, and 30% Hispanic, according to the U.S. Census Bureau.

Figure 23
Total Arrests by Zip Code

(a)



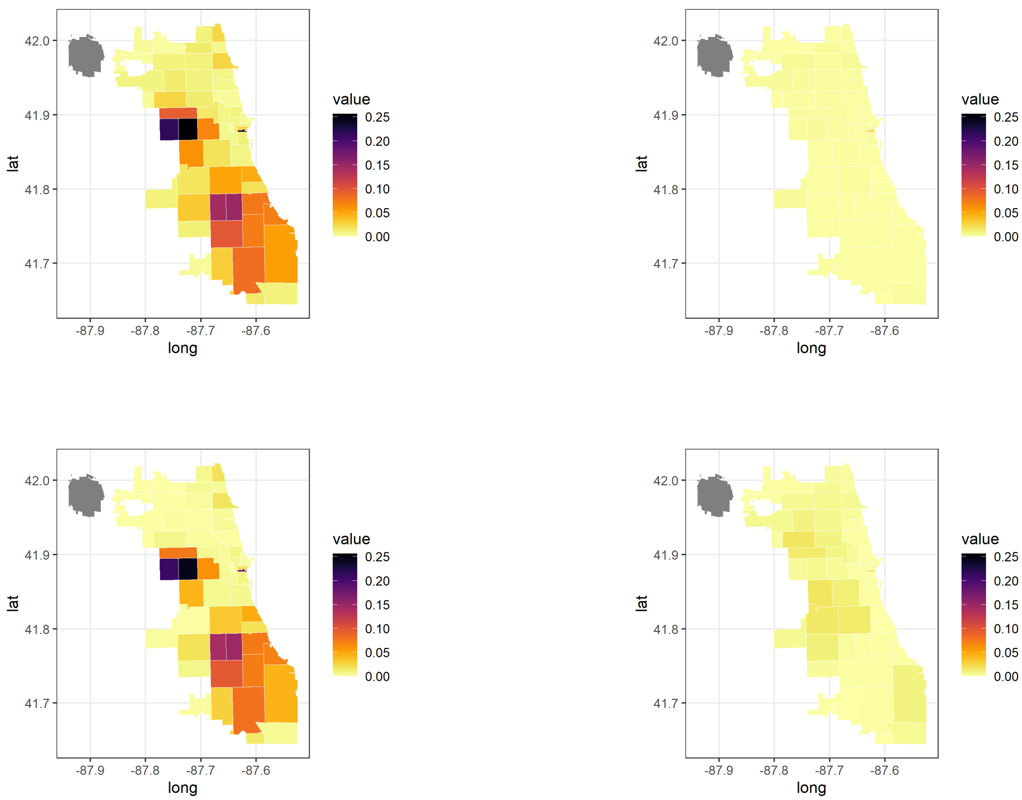
(b)



Notes: Theses figures show the number of marijuana possession arrests in the data for each zip code region for black and white individuals, respectively, using proprietary data from the Chicago Police Department

Figure 24
Arrest Rates by Zip Code

(a) All Arrests per Zip Pop **(b) White Arrests per Zip Pop**



(c) Black Arrests per Zip Pop **(d) Hispanic Arrests per Zip Pop**

Notes: Theses figures show the rates of marijuana possession arrests in the data for each zip code region as a fraction of the overall zip code population, using proprietary data from the Chicago Police Department. (Rates are calculated as racial group population divided by zip code total population.)

6 Conclusion

For decades, researchers, social activists, and the media have brought attention to racial disparities across a multitude of platforms, from wage disparities to the ability to take out a mortgage. In recent years, this attention has focused more closely on how the police interact with minority citizens. There has been a surge in protests of police brutality and prejudice against minorities. Though it is important the public be active in keeping others safe and playing a part in ensuring equity to hold policy makers and law enforcement accountable, it is also necessary to first have all of the information to make a positive change. Much of the literature on racial disparities provide evidence of racial discrimination in the criminal justice system[18]. Others find little or no evidence of racial discrimination[19]. Though there are mixed results when trying to determine if there is racial discrimination, there is still a racial disparity present in stop rates, arrest rates, sentencing and parole, etc. that has not been fully explained by the characteristics of the individual or crime committed. Determining how much of the racial disparity is due to discrimination is essential in order to properly allocate government resources to law enforcement and to reduce costs to society of unequal punishment for crimes.

This paper uses density function comparison techniques to illustrate how the arrest rate disparity is determined by the prevalence of racial discrimination. It uses data from the Chicago police department on the severity of contraband found leading to arrest, which has not been used in prior research. Using severity allows me to run a heterogeneity analysis to see the affect of leniency at different levels of severity. Officers are less likely to show leniency to any driver for a more serious crime, but they may show leniency for less serious

[18]Mustard 2001, Arvanites and Asher 2006, Abrams et al. 2012, Alesina and La Ferrara 2014, West 2018, Pierson et al. 2019

[19]Grogger and Ridgeway 2006, Anwar and Fang 2012

crimes. Thus, if discrimination were to enter into the arrest decision, we would see this at lower levels, but not higher levels of severity. Therefore, the racial disparity in arrests should smaller at higher levels. The results of the model estimates is consistent with this type of behavior causing a significant portion of the racial disparity in arrest rates.

Based on this outcome, I use an Interrupting Time Series model to estimate the effects of marijuana decriminalization, that occurred in Illinois in 2016. These results imply that decriminalization decreases the arrest rate disparity by 0.63 arrests per 100,000 population for lower quantities with only a very small effect on higher quantity arrests (as is expected)in the city of Chicago; however, it leads to an increase in the trend of the disparity over time in the higher quantities that may only have failed to continue past 2019 due to legalization, a question I would like to examine further in future research. The effect on the disparity in the lower range is larger than estimates in previous research, but it is important to recall that this is partially due to it being only for arrests of less than 10 grams of marijuana and partially could be due to this data being for the city of Chicago, where arrests of black people tend to be around 75% of overall arrests. The estimate for all arrests is very similar to the estimates seen in other research and implies an approximately 0.36 per 100,000 population decline in the arrest rate disparity.[20].

The policy implications of the results presented in this paper are that the government should allow laws to evolve in a way the more equitably penalizes citizens for their crimes. Relaxing laws for crimes that cause little or no harm to others, such as marijuana use, can decrease the amount of racial discrimination within the justice system. This is directly beneficial to minority populations who no longer need to fear unequal treatment

[20]These can be seen in 12

for these crimes. Additionally, it means the government can save on costs to imposing such laws and to imprisonment and detainment of people who break them. This allows police to spend more money and time on enforcing laws for more serious crimes, such as trafficking schedule one drugs. The result is a more efficient and effective police force and a more fair justice system.

The results of these tests provide a more accurate picture of the leniency officers apply based on driver's race, and provide evidence for the prevalence of taste-based discrimination. Using this idea to analyze the effects of

The policy implications of this would be that relaxation of minor crime laws can decrease the racial disparity in arrests and allow police department funding to be reallocated toward more severe crime enforcement and increased effectiveness.

References

[1] Abrams, David S. and Bertrand, Marianne and Mullainathan, Sendhil. 2012, "Do Judges Vary in Their Treatment of Race?," *The Journal of Legal Studies*, Vol. 41, No. 2, pp. 347-383.

[2] Alesina, Alberto, and Eliana La Ferrara. 2014. "A Test of Racial Bias in Capital Sentencing," *American Economic Review*, Vol. 104, No.11, pp.3397-3433.

[3] Antonovics, Kate L. and Brian G. Knight. 2004. "A New Look at Racial Profiling: Evidence from the Boston Police Department," *The Review of Economics and Statistics, MIT Press*, Vol. 91, No. 1, pp 163-177.

[4] Anwar, Shamena and Hemming Fang. 2006, "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *The American Economic Review*, Vol. 96, No. 1, pp. 127-151.

[5] Anwar, Shamena and Hemming Fang. 2012. "Testing for the Role of Prejudice in Emergency Departments Using Bounceback Rates," *The B.E. Journal of Economic Analysis & Policy*, Vol. 12, No. 3, pp.1935-1682.

[6] Arnold, Davis, Will Dobbie, Crystal Yang. 2018, "Racial Bias in Bail Decisions," *The Quarterly Journal of Economics*, Vol. 133, No. 4, pp. 1885-1932.

[7] Arrow, Kenneth. 1973, "The Theory of Discrimination," *Princeton University Press*, 3-33.

[8] Arvanites, Thomas M. and Martin A. Asher. 2006, "State and County Incarceration Rates: The Direct and Indirect Effects of Race and Inequality," *The American Journal of Economics and Sociology*, Vol. 57, No. 2, pp. 207-222.

[9] Becker, Gary S. 1993, "Nobel Lecture: The Economic Way of Looking at Behavior," *Journal of Political Economy*, Vol. 101, No. 3, pp. 385-409

[10] Bjerk, David. 2004, "Racial Profiling, Statistical Discrimination, and the Effect of a Colorblind Policy on the Crime Rate," *Journal of Public Economic Theory*, Vol.9, No. 3, pp. 521–546.

[11] Bocian, Debbie Gruenstein, Keith S. Ernst, and Wei Li. 2008. "Race, ethnicity and subprime home loan pricing," *Journal of Economics and Business*, Vol. 60, No. 1–2, pp.110-124.

[12] Bunting, W.C., Lynda Garcia, Ezekiel Edwards. 2013, "The War on Marijuana in Black and White," *American Civil Liberties Union.*

[13] Bushway, Shawn D., and Jonah B. Gelbach. 2011. "Testing for Racial Discrimination in Bail Setting Using Nonparametric Estimation of a Parametric Model." Working Paper.

[14] Cox, Robynn and Jamein Cunningham. 2017, "Financing the War on Drugs: The Impact of Law Enforcement Grants on Racial Disparities in Drug Arrests," CESR-Schaeffer Working Paper No. 2017-005.

[15] Chung, Hwan, Brian P. Flaherty, and Joseph L. Schafer. 2006, "Latent class logistic regression: application to marijuana use and attitudes among high school seniors," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 169, No. 4.

[16] Dhakal, Chandra K., Cesar L. Escalante and Charles Dodson. 2019. "Heterogeneity of farm loan packaging term decisions: a finite mixture approach," Applied Economics Letters, Vol. 26, No. 18, pp.1528-1532.

[17] Dominitz, Jeff. 2003, "How Do the Laws of Probability Constrain Legislative and Judicial Efforts to Stop Racial Profiling?" *American Law and Economic Review*, Vol. 5 No. 2, pp. 412-432.

[18] Donohue, John J. and Levitt, Steven D. 2001, "The Impact of Race on Policing and Arrests." *Journal of Law and Economics*, Vol. 44, pp. 367-394.

[19] Dougal, Casey, Pengjie Gao, William J. Mayew, Christopher A. Parsons. 2019. "What's in a (school) name? Racial discrimination in higher education bond markets," *Journal of Financial Economics*, Vol. 134, No. 3, pp.570-590.

[20] Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment," *American Economic Journal: Applied Economics*, Vol. 9, No. 2: pp.1-22.

[21] Feihenberg, Benjamin and Conrad Miller. 2020. "Racial Disparities in Motor Vehicle Searches Cannot Be Justified by Efficiency." *NBER Working Paper*. No. 27761.

[22] Garcia, Jaume, Pedro J. Hernandez, and Angel Lopez-Nicolas. 2001. "How wide is the gap? An investigation of gender wage differences using quantile regression," *Empirical Economics*, Vol. 26, No. 1, pp.149–167.

[23] Goel, Sharad, Justin M. Rao, and Ravi Shroff. 2016, "Personalized Risk Assessments in the Criminal Justice System," *American Economic Review: Papers and Proceedings*, Vol. 106, No. 5, pp. 119-123.

[24] Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians." *American Economic Review*, Vol. 90, No. 4, pp.715-741.

[25] Grogger, Jeffrey and Greg Ridgeway. 2006, "Testing for Racial Profiling in Traffic Stops from Behind a Veil of Darkness," *Journal of American Statistical Association*, Vol. 101, No. 475, pp. 878-887.

[26] Hagedorn, John M. 2006. "Race not space: A revisionist history of gangs in Chicago," *The Journal of African American History*, Vol. 91, No. 2, pp.194-208.

[27] Hamilton, Barton Hughes. 1997. "Racial discrimination and professional basketball salaries in the 1990s," *Applied Economics*, Vol. 29, No. 3, pp.287-296.

[28] Hernandez-Murillo, Ruben and Knowles, John. 2004, "Racial Profiling or Racist Policing?: Testing in Aggregated Data," *International Economic Review*, Vol. 45, No. 3, pp. 959-989.

[29] Kawwa, Nadim. 2020, "When to Use the Kolmogorov-Smirnov Test: Theory, Application, and Interpretation," *Towards Data Science*.

[30] Keyes, Katherine M., Melanie Wall, Tianshu Feng, Magdalena Cerda, and Deborah Hasin. 2017, "Race/ethnicity and marijuana use in the United States: Diminishing differences in the prevalence of use, 2006 to 2015," *Drug and Alcohol Dependency*, Vol. 179, pp. 379–386.

[31] Knowles, John, Nicola Persico, and Petra Todd. 2001, "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, Vol. 109, No. 1, pp. 203-229.

[32] Lang, Kevin, Michael Manove, and William T. Dickens. 2005. "Racial Discrimination in Labor Markets with Posted Wage Offers." *American Economic Review*, Vol. 95, No. 4, pp.1327-1340.

[33] Lopes, Raul H.C., Ivan Reid, Peter Hobson. 2007, "The two-dimensional Kolmogorov-Smirnov test," *XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research in Amsterdam, the Netherlands.*

[34] Marit, Rehavi and Sonja Starr. 2014, "Racial Disparity in Federal Criminal Sentences," *Journal of Political Economics*, Vol. 122, No. 6, pp. 1320-54.

[35] Martin, Pedro, Pedro Pereira. 2004. "Does education reduce wage inequality? Quantile regression evidence from 16 countries," *Labour Economics*, Vol. 11, No. 3. pp.355-371.

[36] Mueller, R. 1998. "Public-Private Sector Wage Differentials in Canada: Evidence from Quantile Regression," *Economics Letters*, Vol. 60, pp.229-235.

[37] Mustard, David B. 2001, "Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts," *Journal of Law & Economics*, Vol. 44, No. 1, pp. 285-314.

[38] Norris, Clive, Nigel Fielding, Charles Kemp, and Jane Fielding. 1992, " Black and blue: an analysis of the influence of race on being stopped by the police," *Wiley on behalf of The London School of Economics and Political Science*, Vol. 43, No. 2, pp. 207-204.

[39] Pierson, Emma, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff and Sharad Goel. 2019, "A large-scale analysis of racial disparities in police stops across the United States," *Stanford Computational Policy Lab Working Paper.*

[40] Riksheim, E.C. and S.M. Chermak. 1993, "Causes of Police Behavior Revisited." *Journal of Criminal Justice*, Vol. 21, pp. 353-382.

[41] Ryan, Matt E. 2015, “Frisky Business: Race, Gender and Political Activity During Traffic Stops,” *European Journal of Law and Economics*, Vol. 41, pp. 65-83.

[42] Sabia, Joseph J. and Thanh Tam Nguyen. 2016, “The Effect of Medical Marijuana Laws on Labor Market Outcomes,” *Discussion Paper No. 9831, IZA Institute for the Study of Labor.*

[43] Schafer, Robert. 1979. “Racial discrimination in the Boston housing market,” *Journal of Urban Economics*, Vol. 6, No. 2. pp.176-196.

[44] Storey, D.J. 2004. “Racial and Gender Discrimination in the Micro Firms Credit Market?: Evidence from Trinidad and Tobago,” *Small Business Economics*, Vol. 23, No. 5, pp.401–422.

[45] Tuttle, Cody. 2019. “Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums,” University of Maryland Job Market Paper.

[46] West, Jeremy. 2018, “Racial Bias in Police Investigations,” Revise and resubmit at The Review of Economics and Statistics.

[47] Young, Ian T. 1977, “Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources,” *The Journal of Histochemistry and Cytochemistry*, Vol. 25, No. 7, pp. 935-941.

Tables

Table 7
Results of
Kolmogorov-Smirnov Test

	D	p-value
Less than 10g	0.12282	2.2×10^{-16}
Less Than 30g	0.135	1.487×10^{-5}
10g to 30g	0.0755	0.1713
30g to 100g	0.047	0.4888
100g to 300g	0.118	0.0297
300g to 500g	0.12674	0.4643

Notes: This table reports the estimated D (maximum difference) and p-value for the Kolmogorov-Smirnov tests. The p-value is what we care about, here. The data for this table comes from the Chicago Police Department and is not available to the public. It includes all marijuana arrests in Chicago from Jan 2012-Oct 2020.

Table 8
Results of Linear Models

	(1)	(2)
Intercept	**30.979*****	**45.8403*****
	(1.469)	(8.468)
Black with 2.5-10g	**6.287****	**6.287***
	(2.077)	(3.095)
Black with 10-30g	$-$**12.313*****	$-$**12.313*****
	(2.077)	(2.240)
Black with >= 30g	$-$**21.199*****	$-$**21.199*****
	(2.077)	(2.375)
White with < 2.5g	$-$**28.882*****	$-$**29.600*****
	(2.125)	(2.271)
White with 2.5-10g	$-$**29.398*****	$-$**30.616*****
	(2.142)	(2.290)
White with 10-30g	$-$**30.273*****	$-$**30.745*****
	(2.097)	(2.311)
White with > 30g	$-$**30.227*****	$-$**30.433*****
	(2.092)	(2.317)
R^2	.456	.623
Adj. R^2	.452	.563
N	820	
Statistic		10.777
P Value		.000
DF Resid.		707.000

$^{***}p$¡0.001; $^{**}p$¡0.01; $^{*}p$¡0.05

Notes: Regression output of equation 1, 2, and 3. This estimation uses Chicago PD data aggregated to the appropriate level for arrest rates in each model.

Table 9
Results of Interrupted Time Series

	(ITS 1)	(ITS 2)	(ITS 3)
Intercept	0.59989***	0.68101***	0.50351***
	(0.01754)	(0.01654)	(0.01445)
Day	−0.00039***	−0.00030***	−0.00048***
	(0.00002)	(0.00002)	(0.00002)
Decrim	−0.29498***	−0.11017***	
	(0.03069)	(0.02750)	
Day × Decrim	0.00058***		0.00039***
	(0.00004)		(0.00004)
R^2	0.12209	0.11041	0.11591
Adj. R^2	0.12189	0.11028	0.11577
Num. obs.	13113	13113	13113

$^{***}p$¡0.001; $^{**}p$¡0.01; $^{*}p$¡0.05

Notes: Interrupted time series output of ITS 1, 2, and 3. This estimation uses Chicago PD data aggregated to the appropriate level for arrest rates in each model. Changes in the arrest rate can be interpreted as a change in the number of arrests per 100,000 population.

Table 10
Results of Interrupted Time Series Quantity 10g or Less

	(ITS 1)	(ITS 2)	(ITS 3)
Intercept	0.77867***	0.85226***	0.64579***
	(0.02644)	(0.02524)	(0.02263)
Day	−0.00061***	−0.00052***	−0.00072***
	(0.00003)	(0.00002)	(0.00002)
Decrim	−0.48080***	−0.24573***	
	(0.05029)	(0.04299)	
Day × Decrim	0.00066***		0.00029***
	(0.00007)		(0.00006)
R^2	0.23227	0.22367	0.22230
Adj. R^2	0.23194	0.22345	0.22208
Num. obs.	7043	7043	7043

***p¡0.001; **p¡0.01; *p¡0.05; #p¡0.1

Notes: Interrupted time series output of ITS 1, 2, and 3 for possession of 10 grams or less of marijuana. This estimation uses Chicago PD data aggregated to the appropriate level for arrest rates in each model. Changes in the arrest rate can be interpreted as a change in the number of arrests per 100,000 population.

Table 11
Results of Interrupted Time Series more than Quantity 10g

	(ITS 1)	(ITS 2)	(ITS 3)
Intercept	0.36051^{***}	0.42768^{***}	0.34243^{***}
	(0.01115)	(0.01045)	(0.00874)
Day	-0.00009^{***}	-0.00002	-0.00011^{***}
	(0.00001)	(0.00001)	(0.00001)
Decrim	-0.04695^{**}	0.05535^{**}	
	(0.01796)	(0.01700)	
Day × Decrim	0.00039^{***}		0.00036^{***}
	(0.00003)		(0.00002)
R^2	0.03991	0.0241	0.03883
Adj. R^2	0.03944	0.0208	0.03851
Num. obs.	6070	6070	6070

$^{***}p$¡0.001; $^{**}p$¡0.01; $^{*}p$¡0.05

Notes: Interrupted time series output of ITS 1, 2, and 3 for possession of more than 10 grams of marijuana. This estimation uses Chicago PD data aggregated to the appropriate level for arrest rates in each model. Changes in the arrest rate can be interpreted as a change in the number of arrests per 100,000 population.

Table 12
Results of ITS using Best Model for Each Race-Quantity Group

	Black	White	Black Low	White Low	Black High	White High
(Intercept)	0.68827***	0.13409***	0.93712***	0.13372***	0.38826***	0.13463***
	(0.01895)	(0.00413)	(0.01893)	(0.00531)	(0.01175)	(0.00519)
Day	−0.00067***	−0.00004***	−0.00119***	−0.00005***	−0.00015***	−0.00000
	(0.00002)	(0.00000)	(0.00002)	(0.00001)	(0.00001)	(0.00000)
Decrim	−0.36748***	−0.01558	−0.62828***	−0.00411	−0.05466**	−0.02083*
	(0.03135)	(0.01024)	(0.03260)	(0.01926)	(0.01872)	(0.00901)
Interact	0.00088***	0.00005***	0.00124***	0.00004	0.00050***	0.00003*
	(0.00004)	(0.00002)	(0.00005)	(0.00003)	(0.00003)	(0.00001)
R^2	0.28361	0.05399	0.72571	0.06691	0.06918	0.00897
Adj. R^2	0.28340	0.05302	0.72555	0.06546	0.06863	0.00600
Num. obs.	10170	2943	5105	1938	5065	1005

***p¡0.001; **p¡0.01; *p¡0.05

Notes: Interrupted time series output of ITS 1 for each race-quantity group(all, 10g or less, more than 10 grams, for each all, black, and white race groups). ITS 1 appears the best model for each quantity group, so I use it to estimate the results in this table. This estimation uses Chicago PD data aggregated to the appropriate level for arrest rates in each model. Changes in the arrest rate can be interpreted as a change in the number of arrests per 100,000 population.

IV. Bunching Behavior in Chicago Marijuana Arrests

The methodology described in this chapter was contributed to by Melissa Wilson and Benjamin Hansen contributed substantially to this work by participating in the development of the methodology herein. Benjamin Hansen was helpful in verifying the approach and contributed to writing of the Introduction and Conclusion sections. I was the primary contributor in operationalizing the use of the approach in this context, and did all other writing.

1 Introduction

Drug crimes continue to be one of the most common arrests and convictions in the United States. In 2019, this amounts to 1,155,610 arrests, a number greater than all violent crimes combined together ([1]). Moreover, the vast majority of the arrests (over 85 percent) are drug possession, rather than their manufacture or distribution. Almost universally, how people are charged and ultimately sanctioned depends on the weight of the controlled substance they are found to be carrying at the time of arrest.

Becker [5] offers strong predictions about how sanctions based on weights should affect the distribution of weight found. Under perfect information and with no police discretion, we would see individuals in possession of weights just barely under the legal thresholds in question. If sanctions change, this would in turn alter the degree of bunching and sorting we see around the threshold. Indeed, in the context of speeding, Traxler et al. [16] find evidence of bunching in driver behavior using automated speed cameras and changes in speed limits on the autobahn.

However assuming that officers exhibit no discretion is strong assumption. The prior empirical literature find many factors influence police officers decisions to search or cite individuals. Makowsky and Stratmann [12] finds that officers strategically respond to the likelihood that someone will contest thereby discounting their reported speed. Makowsky and Stratmann [13] finds officers give out more tickets to generate revenue during budgetary shortfalls. West [18] produces evidence that minority drivers are more likely to receive citations when a white police officer responds to a traffic incident. Lastly, Goncalves and Mello [9] find 42 percent of police officers exhibit some form of bias based upon speed discounting. This evidence supports the hypothesis that police officers exhibit taste based discrimination driven by in-group biases.

Possession of drugs is a crime with vastly different penalties than speeding. As such, we might expect that police discretion could be different for harsher crimes versus more minor crimes like speeding, which are normally civil offenses (except for extremely high speeds). In fact, Agan et al. [2] find that prosecution for crimes can increase future recidivism. Analyzing the role of officer discretion, Tuttle [17] finds that when weight threshold for mandatory minimum sanctions increased for crack from 50g to 280g, that the fraction of cases sentenced at 280g jumped.

Bunching estimators have been used in a wide variety of prior contexts. They have been quite notably common in the taxation literature. Saez [15] finds evidence of kinking at the first threshold of the earned income tax credit, particularly for self employed people. Likewise, Bastani and Selin [4] study Swedish earners bunching behavior to estimate the taxable income elasticity, and Ruh and Staubli [14] find strong bunching at a disability income threshold. Finally Ghanem and Zhang [8] and Liu and Kong [11] find evidence of bunching in reporting pollution levels in China, providing evidence that official pollution data may be misreported there. However, in many of these contexts, Blomquist et al. [6] show generally strong assumptions are needed to identify elasticities using bunching approaches. With that in mind, our goal is not to estimate an elasticity, but rather to estimate the presence of bunching behavior, similar to other papers of crime.

In that context, we study the role of both officer discretion and offender sorting using a similar bunching approach for marijuana possession. While marijuana is increasingly accessible for legal consumption and possession, many people continue to be arrested annually around the country. During the period we study, Chicago transitioned from a regime where marijuana possession was illegal, to one where it was decriminalized, to one where it was eventually legalized for recreational use. Thus, our estimates inform about a variety of legal structures civilians commonly encounter across the United States even today. We use administrative data from the Chicago Police Department. Wilson [19] also uses this data to study the absolute changes in arrest counts that resulted with decriminalization. Indeed, Wilson [19] provides evidence officers do exhibit discretion in who they stop, search and report, as the number of arrests fall at a time when penalties and sanctions are increasing. Our approach focuses on a different margin, instead measuring how the bunching behavior shifted around relevant weight thresholds before and after decriminalization. This approach is similar to Tuttle [17] taking advantage of changes in the weight thresholds used for sanctions, the race of the arrested suspect, and the race of the arresting officer. In particular, when we study the effects of race matching on bunching, this informs about officers using discretion to minimize the sanctions potential offenders could see.

The rest of this chapter is organized in the following manner. In the next section we discuss the marijuana arrest data from the Chicago Police Department and the policy changes we study. Following that we review the bunching methodology we use. In Section 4 we discuss the results. Finally we conclude and discuss the policy implications.

2 Data

We use drug possession arrests from the Chicago police department January 1, 2012 to October 19, 2020. This data was procured through a FOIA request from the Chicago Police Department. We use the same dataset utilized and described by [19]. There are nearly 220,000 contraband arrests, approximately half of which are specifically marijuana arrests. The data includes the time and location of each stop, amount of contraband found, individual's name, race, gender, and other individual characteristics, e.g. age, sex, hair and eye color. Additionally, we have these same characteristic descriptions for the arresting officer. Approximately 79% of the individuals arrested for marijuana possession are black with less than 5% being white and 7% Hispanic. Of the overall dataset, approximately 78% are black and 6% are white, with approximately 15% Hispanic.

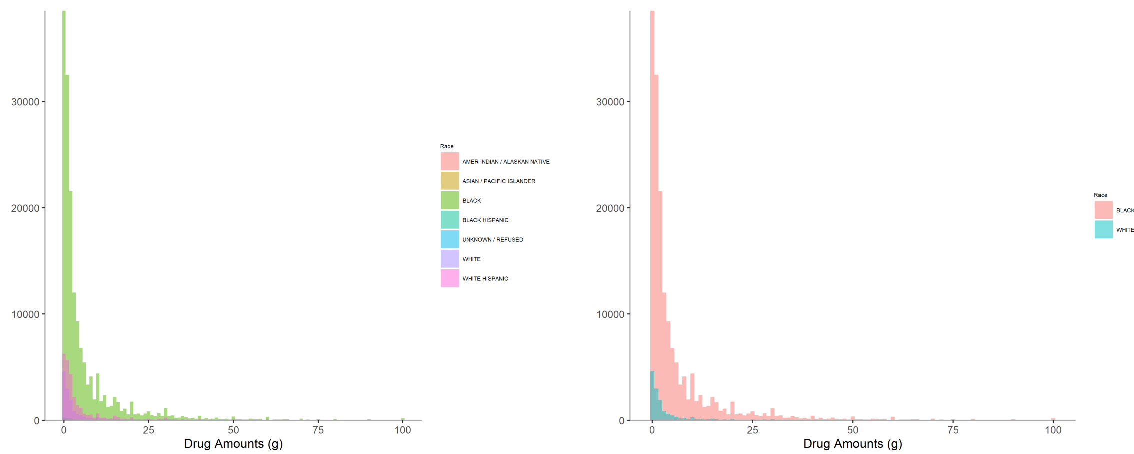
The incidence of arrests at each weight are shown in Figure 25b.[21] The vast majority of all arrests occur at less than 100 grams. Any possession of marijuana 30 grams or more is a felony under Illinois laws with 30g-100g being downgraded to a Class A misdemeanor when legalization occurs. The laws for possession of marijuana up to 30 grams change over the sample period, with 2.5g or less being the least severe bracket. Between 10g-30g is still relatively low severity due to its decriminalization status as of 2016 and makes up

[21] These are reconstructed figures of [19] Figure 5.

about one-third of the marijuana arrests. Any possession under 30 grams is considered legal as of 2020. Possession of other controlled substances remains a felony over the entire period of the data. Figure 25b depicts the histogram above with arrests of black and white individuals.

Figure 25
Total Drug Arrests by Amount (g)

(a) All Races **(b) Black and White**



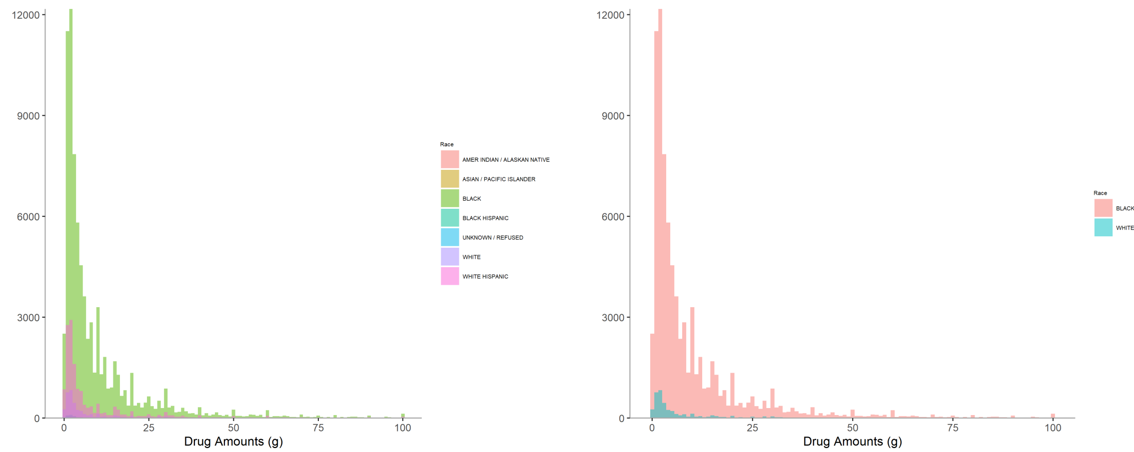
Notes: These graphs show the number of arrests that occur at each weight in the dataset using proprietary data from the Chicago Police Department, by reported racial categories.

The makeup of the marijuana arrest data are explained by Figure 26a and Figure 26b, for all races then for white and black individuals only, respectively. Here, you can see the vast majority of arrests in the dataset occur on the lower end of the distribution. Looking at the histogram, it is possible that bunching behavior is occurring at the lower bounds of each legal bracket, especially for black individuals.

Looking at only these histograms, there appears to be possible bunching around legal thresholds. Bunching could occur due to behavior of officers or due to behavior of the arrestees. If officers bunch in a way to maximize penalties, they would be could round

Figure 26
Total Marijuana Arrests by Amount (g)

(a) All Races **(b) Black and White**



Notes: These graphs show the number of arrests that occur at each weight in the dataset for marijuana possession arrests using proprietary data from the Chicago Police Department, by reported racial categories.

upward to just above legal thresholds, or downward just below legal thresholds, if it is to minimize. The direction of this bunching behavior depends on the driving forces for the officers, e.g. if making more drug arrests is a goal, rounding to just above the decriminalized threshold would occur. If discrimination played a role, we would expect to see differentiation in bunching behavior by race of the subject. If the arrestees exhibit bunching behavior, then we would expect bunching to occur uniformly across races just below legal thresholds, e.g. possessing just under the misdemeanor threshold in the civil violation realm after decriminalization takes place. Bunching behavior is difficult to distinguish in a the figures, so we will estimate the incidence of bunching in the following sections. Given that there is evidence of bunching in criminal behavior as well as officer behavior in the literature, it plausible that one might assume bunching occurs in drug arrests, as well, especially when considering the histograms of arrests from the data presented in this paper [22]. Therefore, it is important to conduct a bunching estimation analysis in order to determine if this behavior is exhibited here.

These time series of arrests by race are shown in Figure 27a (for marijuana arrests in Figure 27b and other controlled substances in Figure 27c). It is clear here just how large the difference in number of arrests is over the sample period. Additionally, although all races do follow a similar trend, it appears that the overall trends are largely driven by the changes in arrests of black people.

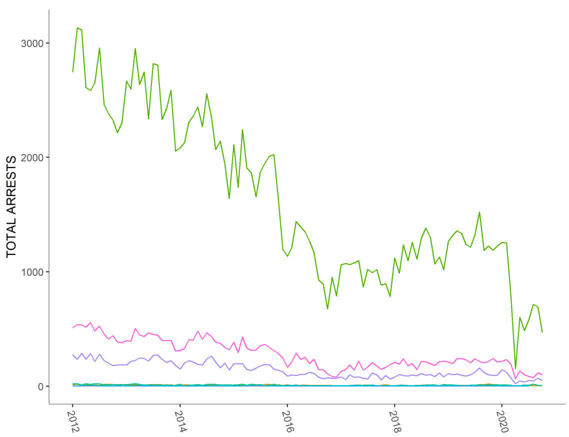
2.1 Bunching Plots

We next take a look at bunching plots for each legal regime at each threshold. The legal regimes throughout the dataset include, illegality (start of data in Jan 2012 through July 28, 2016), decriminalization (July 29, 2016 through December 31, 2019), and legality (January 1, 2020 onward). Decriminalization is the policy shift we are focusing on here because there are possible changes to police behavior around the start of the legality regime due to the start of Covid-19. Decriminalization reduced the severity of possession crimes within several ranges of weights. We focus our analysis on two major thresholds, 30 grams and 100 grams. In the following bunching plots, there is some possible indication of bunching on various thresholds under different legal regimes. However, to be certain of whether this is occurring, we need to estimate the excess mass, which we discuss further in the following sections.

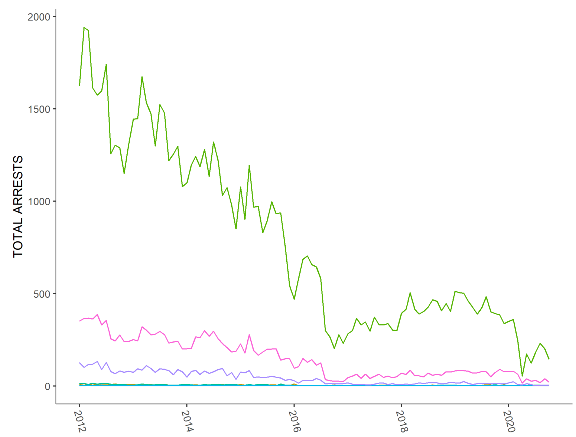
[22] [17], [10]

Figure 27
Time Series of Arrests by Race

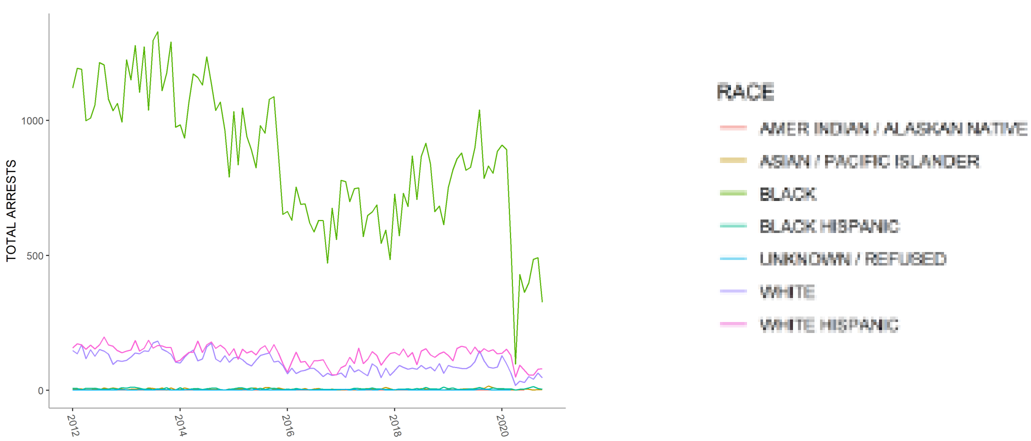
(a) All Drug Arrests



(b) Marijuana Arrests

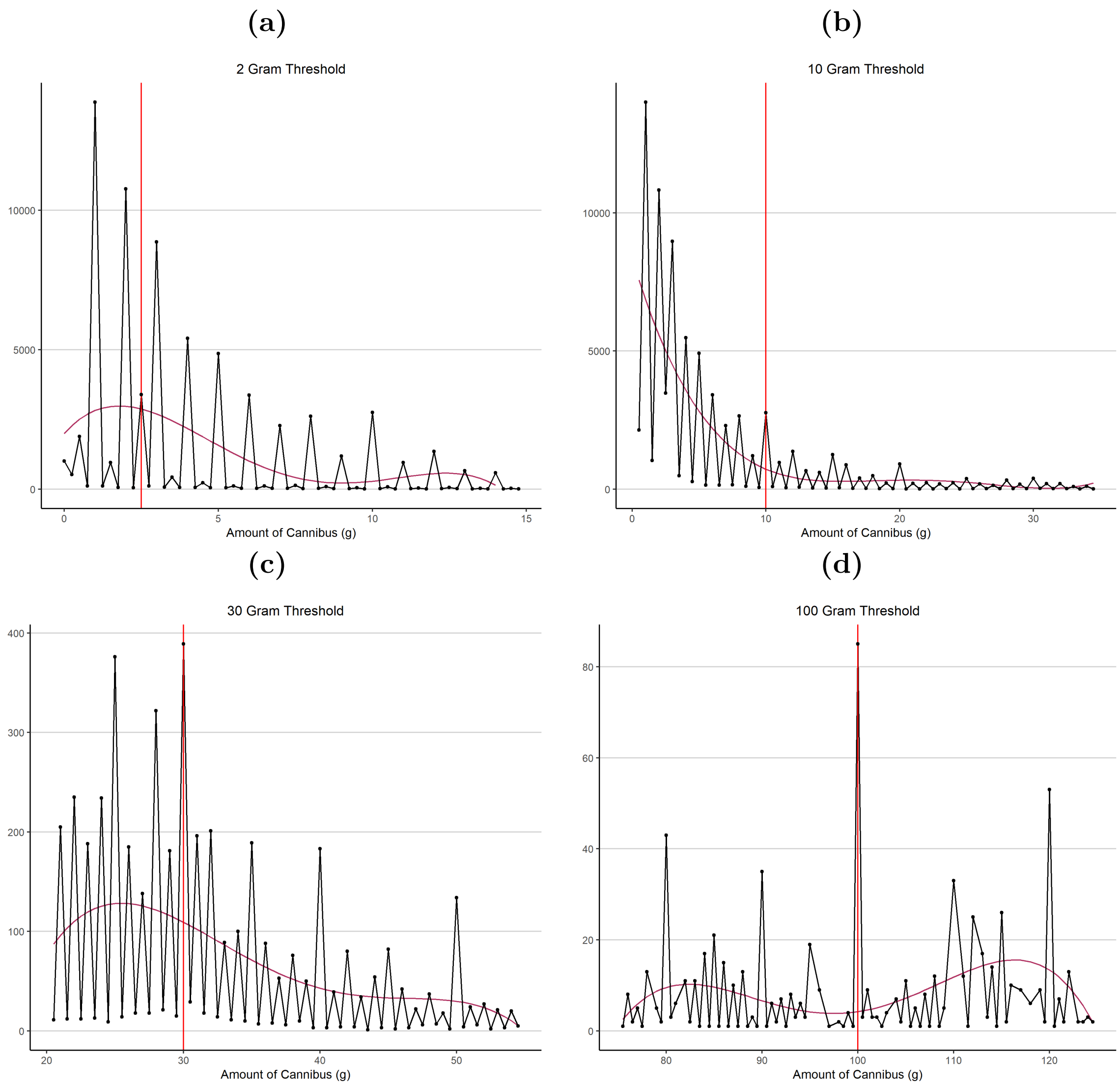


(c) Other Drug Arrests



Notes: These time series graphs shows the number of arrests that occur each month in the dataset using proprietary data from the Chicago Police Department, by reported racial categories. The legend is provided in the bottom right panel.

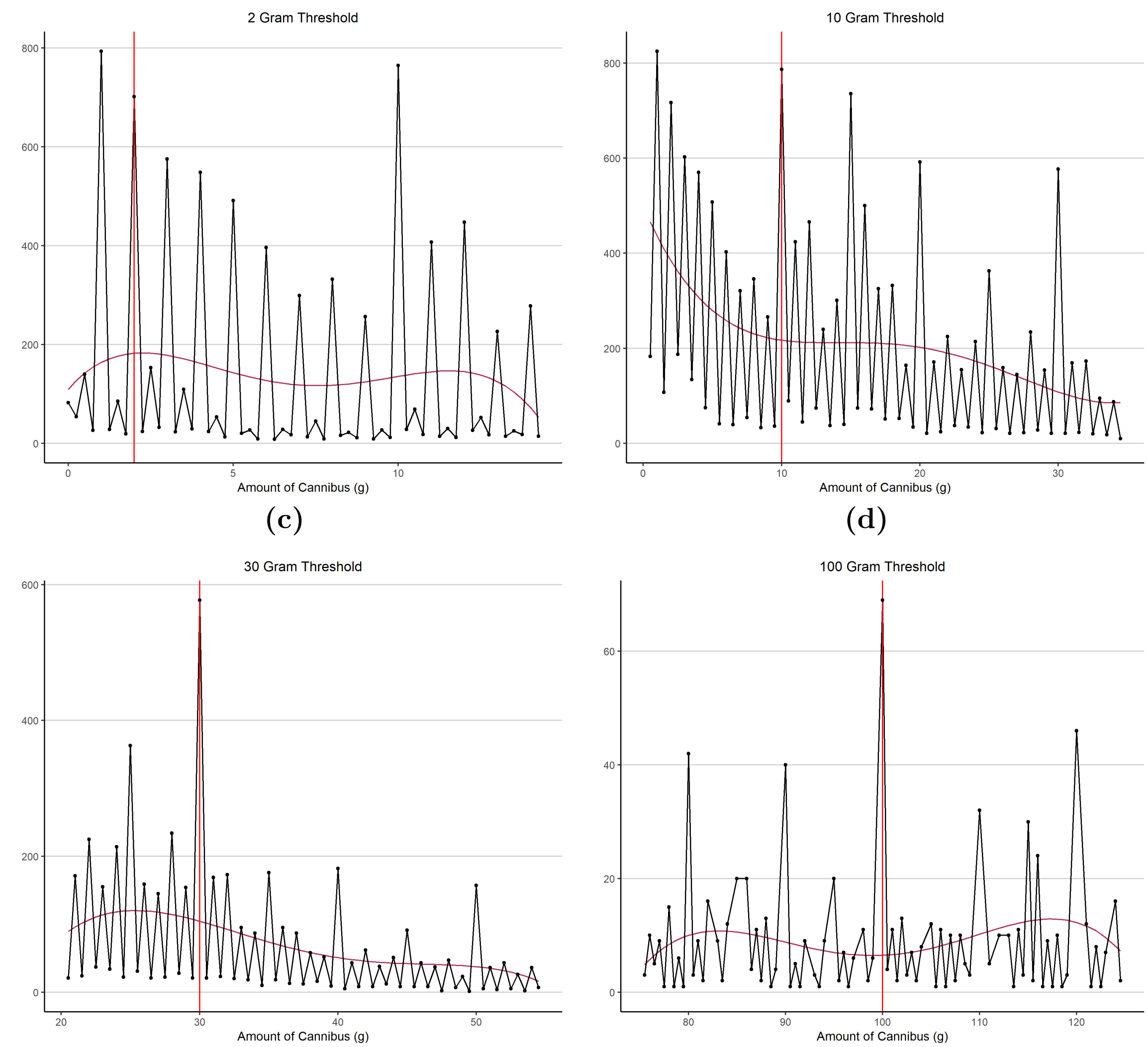
Figure 28
Bunching Plot for Each Threshold in Illegal Regime



Notes: These graphs show the bunching plots for marijuana possession arrests in the dataset using proprietary data from the Chicago Police Department. Each panel provides the plot for the specified legal regime at the specified legal threshold

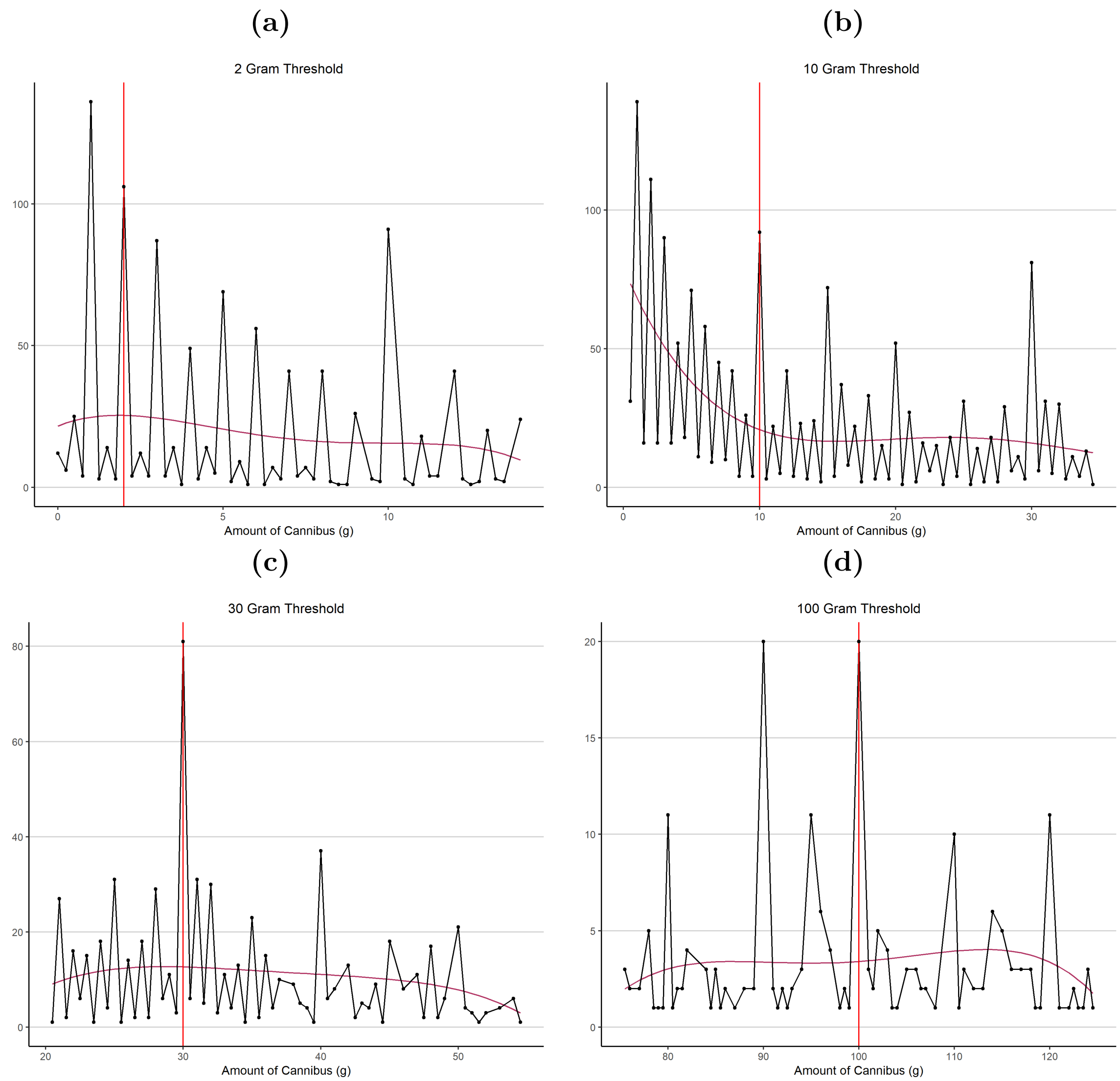
Figure 29
Bunching Plot for Each Threshold in Decriminalized Regime

(a) **(b)**



Notes: These graphs show the bunching plots for marijuana possession arrests in the dataset using proprietary data from the Chicago Police Department. Each panel provides the plot for the specified legal regime at the specified legal threshold

Figure 30
Bunching Plot for Each Threshold in Legalized Regime



Notes: These graphs show the bunching plots for marijuana possession arrests in the dataset using proprietary data from the Chicago Police Department. Each panel provides the plot for the specified legal regime at the specified legal threshold

3 Methodology

As noted in previous sections, decriminalization reduced the severity of possession crimes within several ranges of weights. We focus our analysis on two major legal thresholds, 30 grams and 100 grams. We compare arrests of subjects reported to be carrying within a range of 30 grams in the pre-decriminalization distribution of arrests to the arrests within this same range in the post-decriminalization distribution of arrests. We focus on arrests around the threshold of 30 grams of marijuana in possession and, therefore, define a case as "bunched" at 30 grams as any case in the range of 30 grams specified. We first estimate the mass in the range of above 30 grams up to and including 35 grams, then estimate the mass right at 30 grams. Following this, we vary the bandwidths to determine if the outcomes are robust. Prior to July 2016, 30 grams was the misdemeanor-felony threshold prior to decriminalization, separating crimes with a possibility of facing time in jail and crimes accompanied by a mandatory minimum. After decriminalization went into effect, it was the threshold separating possession of Misdemeanor Class B from Misdemeanor Class C. The ideal counterfactual is the post-decriminalization distribution with the pre-decriminalization thresholds. We assume the pre-decriminalization distribution is a good-enough counterfactual in this sense for each section of the drug quantity distribution.[23] Additionally, we restrict the sample for the main estimation to arrests with weights in possession to the bandwidth of above 15 grams up to and including 45 grams. We do this to avoid bias in our estimates given that 10 grams is a legal threshold at which decriminalization changed the severity of offense. If there is bunching around 10 grams either before or after decriminalization, it will lead to the estimated bunching at 30 grams to be biased either up or downward, depending on when the bunching occurred. Since

[23] [17]

we are focused on the 30 gram legal threshold here, we want to avoid the range in which bunching could occur around 10 grams for the main analysis, so we use a 15 gram bandwidth on either side of the 30 gram legal threshold. Based off of the model utilized by [17], we estimate the following baseline linear probability model of bunching incidence around the legal threshold.

$$(\text{Arrested30gRange})_{it} = \beta_0 + \beta_1 \text{Decrim}\ _{it} + \beta_2 X_{it} + T(t) + \epsilon_{it} \tag{3.1}$$

where $(\text{Arrested30gRange})_{it}$ is equal to one when the subject i is reported at time t to have been carrying an amount within the range of 30 grams specified and zero if any other amount was in possession. β_1 is the change in a subject's probability of being arrested for possession of an amount in the range of 30 grams. $\text{Decrim}\ _{it}$ is a dummy variable equal to 1 when the arrest occurs in the post-decriminalization period. X_{it} is the vector of arrest level covariates (race, age, gender, zip code FE, etc.) and $T(t)$ is the time trend.

The baseline model is then extended to estimate the heterogeneity in bunching by race. We would expect to see heterogeneity in bunching if there is a difference in carrying behavior by subjects along racial lines, or if there is a difference in behavior of officer reporting along racial lines. According to [7], there is little evidence that usage rates of marijuana differ by race, implying there is no reason to suspect that individual possession behavior does in fact differ.

$$
\begin{aligned}
(\text{Arrested30gRange})_{it} = \; & \delta_0 + \delta_1 \text{Decrim} + \delta_2(\text{Decrim} \times \text{White})_{it} + \\
& \delta_3(\text{Decrim} \times \text{Black})_{it} + \delta_4(\text{Decrim} \times \text{Hispanic})_{it} + \\
& \gamma X_{it} + T(t) + \epsilon_{it}
\end{aligned} \tag{3.2}
$$

where δ_1 is the change in a white subject's probability of being arrested for possession reported in the range of 30g. δ_2 is the change in a black subject's probability of being arrested for possession reported in the range of 30g.

We further extend this model beyond [17] because we have data with both subject and arresting officer's race and, therefore, are able to estimate the following extension of Model 3.1.

$$
\begin{aligned}
(\text{Arrested30gRange})_{it} = \; & \delta_0 + \delta_1 \text{Decrim} + \delta_2(\text{Decrim} \times \text{White})_{it} + \\
& \delta_3(\text{Decrim} \times \text{Black})_{it} + \delta_4(\text{Decrim} \times \text{Hispanic})_{it} + \\
& \delta_5(\text{Match})_{it} + \delta_6(\text{Match} \times \text{White})_{it} + \delta_7(\text{Match} \times \text{Black})_{it} + \\
& \delta_8(\text{Match} \times \text{Hispanic})_{it} + \gamma X_{it} + T(t) + \epsilon_{it}
\end{aligned} \tag{3.3.1}
$$

where δ_3 is the change in probability of being arrested for possession reported in the range of 30 grams if the subject and arresting officer's races are the same. $(\text{Match} \times \text{White})_{it}$ and $(\text{Match} \times \text{Black})_{it}$ is the change in probability of being arrested for possession reported in the range of 30 grams if the subject and arresting officer's races are both white and black, respectively. This is a pivotal extension because it allows us to examine whether officer behavior changes when interacting with subjects of their own race versus a differing race. This type of race-matching technique has been used in the racial discrimination literature to determine if officers are more likely to cite, search, or arrest individuals only due to their race.[24] These models are then run using varying bandwidths to check for robustness of these estimates. Bunching is most likely to occur closer to the legal threshold, but we should expect wider bandwidths to have a larger magnitude since they encompass a larger portion of the overall distribution of arrests.

In addition to the analysis around the 30 gram threshold, we also conduct a similar analysis around the 100 gram threshold. This is because 100 grams was a legal threshold between two different classes of felony prior to decriminalization, but became the misdemeanor-felony threshold following decriminalization. This threshold could potentially lead to altered behavior after the policy change takes place. To avoid overlap in the bandwidths for these two separate thresholds, and therefore any bias from the effects of decriminalization might cause at the 30 gram threshold, we restrict the sample for the 100 gram threshold analysis to 56 grams to 155 grams, with 55 grams on either side of the threshold included in the bandwidth of the distribution of weights. We begin with the same baseline Model 3.1 with the alternative dependent variable of the probability of an arrest occurring within a range of 100 grams. We then expand upon it just as in Model 3.2 and Model 3.3.1. Finally, the analysis is repeated using the full distribution of weights in possession to determine whether the results are sensitive to the bandwidth of the distribution we use here. The following section discusses the results of these analyses.

[24] Antonovics and Knight [3], West [18]

$$
\begin{aligned}
(\text{Arrested100gRange})_{it} = \; & \delta_0 + \delta_1\text{Decrim} + \delta_2(\text{Decrim} \times \text{White})_{it} + \\
& \delta_3(\text{Decrim} \times \text{Black})_{it} + \delta_4(\text{Decrim} \times \text{Hispanic})_{it} + \\
& \delta_5(\text{Match})_{it} + \delta_6(\text{Match} \times \text{White})_{it} + \delta_7(\text{Match} \times \text{Black})_{it} + \\
& \delta_8(\text{Match} \times \text{Hispanic})_{it} + \gamma X_{it} + T(t) + \epsilon_{it}
\end{aligned}
\tag{3.3.2}
$$

4 Results of Linear Probability Model

4.1 Linear Probability Model Estimates

The Methodology section details the results of Models 3.2 through 3.3.2 for each of the various bandwidths around the 30 gram and around 100 gram thresholds. We then check for robustness and sensitivity to changes in the bandwidth. In these models, we include 3 racial categories–white, black, and Hispanic. The reference category is any subject in the white racial category for each bandwidth of weights around the legal threshold, as well for each bandwidth of the distribution of weights in possession.

Tables 13 through 16 show the results of each model specification estimated for each of the various bandwidths around the 30 gram legal threshold. As is shown in Table 13, there does not appear to be any evidence of bunching in the above 30 up to and including 35 gram range. Table 14 shows the results for Models 3.1 through 3.3.1 at exactly 30 grams. There appears to be a positive effect of decriminalization on the probability of being arrested for 30 grams of possession in the baseline estimate, as well as the estimates for Model 3.2. However, this effect seems to disappear when we take matching race into account. The estimates for Model 3.3.1 show a strongly significant positive effect of $Match \times Hispanic$ on the probability for being arrested for a reported 30 grams of marijuana in possession. This implies that there is an the probability that an individual will be arrested for possession of 30 grams, rather than any other amount between 15g-45g, increases by 0.061 if the arresting officer and subject are both Hispanic. This result could be driven by Hispanic officers rounding down to reduce the severity of punishment for criminal activity the Hispanic subject will face due to them sharing a similar racial group. It is worth noting that this effect is not robust to increasing the bandwidth to 27-30 grams as seen in Table 16 column 3. The effect seen here disappears.

Table 15 shows estimated for the range of above 30 grams up to and including 32 grams, a slightly condensed bandwidth compared to Table 13. The results are smaller in magnitude–as we would expect given that this is a smaller bandwidth–and are not statistically significant, as well. The results for each of the other variations in bandwidth around the 30 gram threshold are reported in Table **??**. This table shows the Model 3.3.1 estimates for the bandwidths of 30g-40g, 30g-42g, and 27g-30g.[25] Here, there appears to be a negative effect of decriminalization of the probability of individuals being arrested in the 27 to 30 gram range. The estimated effect of -0.045 is significant at the 10 percent level. Additionally, there is an estimated effect of -0.046, significant at the 5 percent level, on probability of being arrested in the 27 to 30 gram range if the arresting officer and subject share a racial group. There also appears to be an additional small positive effect of Decriminalization on arrests of black and Hispanic individuals in this same range.[26] Finally, there is a small positive effect on black individuals if the arresting officer is also black for this same range, as well. These results imply that black and Hispanic individuals have a

[25]The lower bound is not included in the bandwidth, while the upper bound is included.
[26]Significant at the 5 percent level.

higher probability of being arrested for 27-30 grams of possession after decriminalization. It is possible that these effects stem from the fact that this is no longer a felony offense and that makes these individuals more likely to carry larger amounts. Given that the severity of punishment has decreased, there is less of a disincentive for carrying about 30 grams. This would lead to an increase in possession arrests in this range, as it is only a misdemeanor offense. The effect is similar in magnitude for both black and Hispanic subjects.

Following the analysis around the 30 gram legal threshold, we investigate the 100 gram threshold. Table 17 shows the estimates for the baseline model and both extensions thereof for the above 100 grams up to and including 105 grams range. There are no significant effects of any independent variable of interest here. At 100 grams, depicted in Table 18, there is a positive effect of $Match \times Black$ on the probability of being arrested. This is similar to the effect seen at 30 grams on Hispanic subjects arrested by Hispanic officers. These same results are seen in Table 20 column 3, for the 97-100 gram range. There appears to be a negative effect of decriminalization and of subject's race matching arresting officer's in the 100 to 115 gram range, shown in Table 20. This is the broadest range above the 100 gram legal threshold and is the only bandwidth that shows significance for these estimates. The largest significant estimate is in this same range and is the coefficient on $Decrim \times Black$. It implies a 0.071 increase in probability of being arrested in the 100g-115g range after decriminalization for black individuals. This coefficient is not significant in either of the other smaller bandwidths above 100 grams.

Overall, the results in this analysis provide little evidence to suggest that there is bunching behavior at the major legal threshold or that the behavior varies by race. There is some significance of estimates; however, we are cautious of claiming causal relationships as each model has an extremely low R^2, all being less than 1 percent. This adds to the overall literature by providing evidence that reductions in legal severity for marijuana possession will likely not lead to perverse incentives for bunching or unexpected changes in officer behavior or criminal activity.

4.2 Robustness

To determine that the results found here are not overly sensitive to changes in the bandwidth of the distribution of weight in possession, the results for Model 3.1 and each extension over each range in the baseline model are run over the full distribution of weights. These estimates are reported in Table 21. There appears to be a positive and statistically significant, though small in magnitude, effect on the probability of specific racial groups ending up within this range following decriminalization. This result implies that there is a slightly larger mass between 30 and 35 grams after 30 grams changes from the threshold between no mandatory minimum and a mandatory minimum to being solidly in the middle of the misdemeanor realm of legal severity. The results from Models 3.2 and 3.3.1 imply that any statistically significant effect on *Decrim* actually stems from the effects on black individuals and Hispanic individuals which have an increase in probability of being arrested for possession within this range of 0.008 and 0.009, respectively after decriminalization takes effect. These estimates are significant at the 1 percent level and are nearly the same for both extensions of the baseline model. There does not appear to be any effect when races match for any group within this range. Theses results appear to be quite similar across all bunching bandwidths for the full distribution. Although they are significant, the estimates here are all quite small in magnitude at less the 0.01 increase in probability across all specifications.

In Table 22, we see similarly small, yet significant estimates on certain independent variables. There appears to be significance on the estimated coefficients for $Decrim \times B$ and $Decrim \times H$ in the above 100 grams ranges. The increase in probability here is in the order of 0.001 to 0.004, so these are again quite small in magnitude. It is likely these nearly zero estimates are strongly significant due to the precision of the estimates increasing when using a much larger dataset with over 200,000 observations, rather than a subset of the overall data falling within the distribution bandwidths used above. Overall, these outcomes do not appear to be very sensitive to changes in the bandwidth of the distribution of weights in possession and the estimates do not negate the results found in the previous subsection.

5 Conclusion

We study the role of discretion and deterrence in drug arrests. We expand on the prior research by Tuttle [17], West [18], and Wilson [19] by using both a change in sanctions, and the role of police and suspect race matching.

Notably, with decriminalization, sanctions were decreased so that possession of weights over 30gs was no longer a felony. Using administrative data from the Chicago police department on marijuana arrests, we find limited evidence that bunching changed with decriminalization at the 30 gram threshold. The lack of change in bunching following decriminalization (which led to a substantial change in potential punishments at legal thresholds), suggests deterrence was not a substantial driver of bunching behavior.

However, we do find some evidence of bunching at the 30 gram threshold, particular when both officers and offenders are Hispanic. This suggests in-group association partially drives officer discretion, similar to evidence from West [18]. At the 100 gram threshold, we find evidence of bunching for race matching with black officers and suspects.

These findings suggest officers may be engaged in some level of discounting to report weights at exactly the legal cutoffs for lower sanctions. While this shares similarities with some prior work on discounting with speeding based upon in-group matching, it also stands out. In speed discounting, bunching was far more prevalent for white speeders matched with white police officers. In contrast, we find bunching behavior is more in race matching for minority police officers and suspects.

This does not necessarily imply white officers do not also engage in other types of discretion when involved with white suspects. Indeed, the largest margin of discretion may be reporting nothing at all, or where and when police search people for marijuana possession, as white suspects charged with possession are vastly under represented in the set of those who are arrested, relative to surveys of with recent drug use.

References

[1] Drug Arrests drug policy facts. https://www.drugpolicyfacts.org/chapter/crime_arrests. Accessed: 2022-5-13.

[2] Amanda Y Agan, Jennifer L Doleac, and Anna Harvey. Misdemeanor prosecution. Technical report, National Bureau of Economic Research, 2021.

[3] Kate Antonovics and Brian G Knight. A new look at racial profiling: Evidence from the boston police department. *The Review of Economics and Statistics*, 91(1):163–177, 2009.

[4] Spencer Bastani and Håkan Selin. Bunching and non-bunching at kink points of the swedish tax schedule. *Journal of Public Economics*, 109:36–49, 2014.

[5] Gary S Becker. Crime and punishment: An economic approach. In *The economic dimensions of crime*, pages 13–68. Springer, 1968.

[6] Sören Blomquist, Whitney K Newey, Anil Kumar, and Che-Yuan Liang. On bunching and identification of the taxable income elasticity. *Journal of Political Economy*, 129 (8):2320–2343, 2021.

[7] Benjamin Feigenberg and Conrad Miller. Racial profiling without efficiency gains: Evidence from motor vehicle searches. 2020.

[8] Dalia Ghanem and Junjie Zhang. 'effortless perfection:'do chinese cities manipulate air pollution data? *Journal of Environmental Economics and Management*, 68(2): 203–225, 2014.

[9] Felipe Goncalves and Steven Mello. A few bad apples? racial bias in policing. *American Economic Review*, 111(5):1406–41, 2021.

[10] Louis-Pierre Lepage. Are criminals strategic? offender responses to drug sentencing cutoffs. *Labour Economics*, 66:101906, 2020.

[11] Chenhao Liu and Dongmin Kong. Does political incentive shape governments' disclosure of air pollution information? *China Economic Review*, 69:101659, 2021.

[12] Michael D Makowsky and Thomas Stratmann. Political economy at any speed: what determines traffic citations? *American Economic Review*, 99(1):509–27, 2009.

[13] Michael D Makowsky and Thomas Stratmann. More tickets, fewer accidents: How cash-strapped towns make for safer roads. *The Journal of Law and Economics*, 54(4): 863–888, 2011.

[14] Philippe Ruh and Stefan Staubli. Financial incentives and earnings of disability insurance recipients: Evidence from a notch design. *American Economic Journal: Economic Policy*, 11(2):269–300, 2019.

[15] Emmanuel Saez. Do taxpayers bunch at kink points? *American economic Journal: economic policy*, 2(3):180–212, 2010.

[16] Christian Traxler, Franz G Westermaier, and Ansgar Wohlschlegel. Bunching on the autobahn? speeding responses to a 'notched'penalty scheme. *Journal of Public Economics*, 157:78–94, 2018.

[17] Cody Tuttle. Racial disparities in federal sentencing: Evidence from drug mandatory minimums. *Available at SSRN*, 2019.

[18] Jeremy West. Racial bias in police investigations. *Retrieved from University of California, Santa Cruz website: https://people. ucsc. edu/˜ jwest1/articles/West_RacialBiasPolice. pdf*, 2018.

[19] Melissa Wilson. Drug laws, police leniency, and racial disparities in arrest rates. *Job Market Paper*, 2022.

6 Tables

Table 13
Bunching at 30-35 grams

	In 30g to 35g Range		
	(1)	(2)	(3)
Decrim	-0.010	-0.022	-0.018
	(0.010)	(0.023)	(0.023)
Match			0.022
			(0.022)
Decrim×B		0.016	0.012
		(0.022)	(0.022)
Decrim×H		-0.005	-0.009
		(0.024)	(0.025)
Decrim×W			
Match×B			-0.035
			(0.023)
Match×H			-0.022
			(0.028)
Match×W			
Obs		18,966	
Reference Group Mean		0.1255	
Note:	*p<0.1; **p<0.05; ***p<0.01		

Linear Probability Model output of models 3.1 through 3.3 for the bandwidth specified, inclusive of the upper bound and not inclusive of the lower bound. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The 30 gram threshold is estimated with a subset of the data including arrests for amounts of 15-45g in order to avoid any affects that may occur at the 10g threshold. The reference group is white individuals prior to decriminalization.

Table 14
Bunching at 30 grams

	At 30g		
	(1)	(2)	(3)
Decrim	0.024***	0.033*	0.030
	(0.008)	(0.018)	(0.018)
Match			-0.025
			(0.017)
Decrim×B		-0.008	-0.004
		(0.017)	(0.017)
Decrim×H		-0.014	-0.018
		(0.019)	(0.020)
Decrim×W			
Match×B			0.028
			(0.018)
Match×H			0.061***
			(0.022)
Match×W			
Obs		18,966	
Reference Group Mean		0.0657	
Note:			*p<0.1; **p<0.05; ***p<0.01

Linear Probability Model output of models 3.1 through 3.3 for the given amount. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The 30 gram threshold is estimated with a subset of the data including arrests for amounts of 15-45g in order to avoid any affects that may occur at the 10g threshold.

Table 15
Bunching at 30-32 grams

	In 30g to 32g Range		
	(1)	(2)	(3)
Decrim	-0.011	-0.006	-0.006
	(0.007)	(0.017)	(0.017)
Match			0.003
			(0.016)
Decrim×B		-0.004	-0.005
		(0.016)	(0.016)
Decrim×H		-0.011	-0.010
		(0.018)	(0.018)
Decrim×W			
Match×B			-0.011
			(0.017)
Match×H			-0.011
			(0.021)
Match×W			
Obs		18,966	
Reference Group Mean		0.0540	
Note:			*p<0.1; **p<0.05; ***p<0.01

Linear Probability Model output of models 3.1 through 3.3 for the bandwidth specified, inclusive of the upper bound and not inclusive of the lower bound. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The 30 gram threshold is estimated with a subset of the data including arrests for amounts of 15-45g in order to avoid any affects that may occur at the 10g threshold.

Table 16
Bunching Above and Below 30 grams

	30g to 40g	30g to 42g	27g to 30g
	(1)	(2)	(3)
Decrim	-0.036	-0.041	-0.045*
	(0.029)	(0.030)	(0.025)
Match	0.017	0.010	-0.046**
	(0.027)	(0.028)	(0.023)
Decrim×B	0.029	0.026	0.061**
	(0.027)	(0.028)	(0.024)
Decrim×H	0.023	0.019	0.057**
	(0.031)	(0.032)	(0.027)
Decrim×W			
Match×B	-0.027	-0.019	0.043*
	(0.028)	(0.029)	(0.025)
Match×H	-0.015	-0.004	0.029
	(0.035)	(0.036)	(0.031)
Match×W			
Obs	18,966	18,966	18,966
Reference Group Mean	0.1956	0.2161	0.1810

Note: *p<0.1; **p<0.05; ***p<0.01

Linear Probability Model output of model 3.3 for the bandwidths specified, inclusive of the upper bound and not inclusive of the lower bound. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The 30 gram threshold is estimated with a subset of the data including arrests for amounts of 15-45g in order to avoid any affects that may occur at the 10g threshold.

Table 17
Bunching at 100-105 grams

	In 100g to 105g Range		
	(1)	(2)	(3)
Decrim	0.010	0.007	0.008
	(0.008)	(0.016)	(0.017)
Match			0.002
			(0.015)
Decrim×B		0.002	0.001
		(0.015)	(0.016)
Decrim×H		0.007	0.006
		(0.017)	(0.017)
Decrim×W			
Match×B			-0.008
			(0.016)
Match×H			
Match×W			
Obs.		6,953	
Reference Group Mean		0.0252	
Note:	*p<0.1; **p<0.05; ***p<0.01		

Linear Probability Model output of models 3.1 through 3.3 for the bandwidth specified, inclusive of the upper bound and not inclusive of the lower bound. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The 100 gram threshold is estimated with a subset of the data including arrests for amounts of 46-155g.

Table 18
Bunching at 100 grams

	At 100g		
	(1)	(2)	(3)
Decrim	-0.010	0.012	0.006
	(0.010)	(0.019)	(0.020)
Match			-0.027
			(0.018)
Decrim×B		-0.027	-0.021
		(0.018)	(0.018)
Decrim×H		-0.011	-0.005
		(0.020)	(0.021)
Decrim×W			
Match×B			0.042***
			(0.020)
Match×H			
Match×W			
Obs.		6,953	
Reference Group Mean		0.0283	
Note:	*p<0.1; **p<0.05; ***p<0.01		

Linear Probability Model output of models 3.1 through 3.3 for the given amount. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The 100 gram threshold is estimated with a subset of the data including arrests for amounts of 46-155g.

Table 19
Bunching at 100-102 grams

	100g to 102g		
	(1)	(2)	(3)
Decrim	0.006	0.004	0.004
	(0.006)	(0.011)	(0.011)
Match			-0.003
			(0.010)
Decrim×B		0.003	0.004
		(0.010)	(0.011)
Decrim×H		-0.002	-0.001
		(0.012)	(0.012)
Decrim×W			
Match×B			0.002
			(0.011)
Match×H			
Match×W			
Obs.		6,953	
Reference Group Mean		0.0094	
Note:	*p<0.1; **p<0.05; ***p<0.01		

Linear Probability Model output of models 3.1 through 3.3 for the bandwidth specified, inclusive of the upper bound and not inclusive of the lower bound. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The 100 gram threshold is estimated with a subset of the data including arrests for amounts of 46-155g.

Table 20
Bunching Above and Below 100 grams

	100g to 110g	100g to 115g	97g to 100g
	(1)	(2)	(3)
Decrim	-0.012	-0.052***	-0.015
	(0.024)	(0.030)	(0.022)
Match	0.001	-0.047***	-0.032
	(0.022)	(0.027)	(0.020)
Decrim×B	0.035	0.071***	0.002
	(0.022)	(0.028)	(0.021)
Decrim×H	0.031	0.044	0.017
	(0.025)	(0.031)	(0.023)
Decrim×W			
Match×B	0.00004	0.041	0.045***
	(0.024)	(0.030)	(0.022)
Match×H			
Match×W			
Obs.	6,953	6,953	6,953
Reference Group Mean	0.0692	0.1195	0.0440
Note:			*p<0.1; **p<0.05; ***p<0.01

Linear Probability Model output of model 3.3 for the bandwidths specified, inclusive of the upper bound and not inclusive of the lower bound. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The 100 gram threshold is estimated with a subset of the data including arrests for amounts of 46-155g.

Table 21
Bunching for Full Distribution at 30 gram Threshold

	Bunching Bandwidth									
	30g-35g			At 30g			30g-32g	30g-40g	30g-42g	27g-30g
	(1)	(2)	(3)	(1)	(2)	(3)	(3)	(3)	(3)	(3)
Decrim	0.006***	-0.002	-0.002	0.007***	0.003	0.003	-0.001	-0.004	-0.004	-0.003
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)	(0.002)
Match			0.003			-0.001	0.001	0.003	0.003	-0.001
			(0.002)			(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Decrim×B		0.009***	0.008***		0.004***	0.004***	0.003**	0.015***	0.016***	0.013***
		(0.002)	(0.002)		(0.001)	(0.002)	(0.001)	(0.003)	(0.003)	(0.002)
Decrim×H		0.010***	0.009***		0.004**	0.003*	0.004***	0.018***	0.019***	0.014***
		(0.002)	(0.002)		(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Decrim×W										
Match×B			-0.003			0.002	-0.001	-0.002	-0.002	0.002
			(0.002)			(0.001)	(0.001)	(0.003)	(0.003)	(0.002)
Match×H			-0.001			0.007***	-0.001	-0.0004	0.001	0.001
			(0.003)			(0.002)	(0.002)	(0.004)	(0.004)	(0.003)
Match×W										
Obs.					18,966					
Reference Group Mean		0.0081			0.0042		0.0035	0.0126	0.0139	0.0117

Note: * p<0.1; ** p<0.05; *** p<0.01

Linear Probability Model output of the model and bandwidths specified, inclusive of the upper bound and not inclusive of the lower bound. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The analysis here includes the full bandwidth of amounts in the dataset.

Table 22
Bunching for Full Distribution at 100 gram Threshold

	Bunching Bandwidth:									
	100g-105g			At 100g			100g-102g	100g-110g	100g-115g	97g-100g
	(1)	(2)	(3)	(1)	(2)	(3)	(3)	(3)	(3)	(3)
Decrim	0.001***	0.0002	0.0003	0.0004	0.0004	0.0003	0.0001	-0.0004	-0.002	-0.0004
	(0.0003)	(0.001)	(0.001)	(0.0003)	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)
Match			0.0002			-0.001	-0.00004	0.0004	-0.001	-0.001
			(0.001)			(0.001)	(0.0003)	(0.001)	(0.001)	(0.001)
Decrim×B		0.001	0.001		-0.0003	-0.0001	0.0004	0.002***	0.004***	0.001
		(0.001)	(0.001)		(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)
Decrim×H		0.001***	0.001***		0.001	0.001	0.0004	0.003***	0.004***	0.002***
		(0.001)	(0.001)		(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)
Decrim×W										
Match×B			-0.0003			0.001***	0.00003	-0.0003	0.001	0.001***
			(0.001)			(0.001)	(0.0004)	(0.001)	(0.001)	(0.001)
Match×H										
Match×W										
Obs	201,482									
Reference Group Mean	0.0008			0.0008			0.0003	0.0021	0.0036	0.0013

Note: *p<0.1; **p<0.05; ***p<0.01

Linear Probability Model output of the model and bandwidths specified, inclusive of the upper bound and not inclusive of the lower bound. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The analysis here includes the full bandwidth of amounts in the dataset.

Table 23
Bunching Estimates Controlling for Rounding/Heaping

	30-35g	At 30g	100-105g	At 100g
Decrim	-0.016	0.022	0.008	0.006
	(0.023)	(0.016)	(0.017)	(0.020)
Match	0.026	-0.027*	0.002	-0.027
	(0.021)	(0.015)	(0.015)	(0.018)
Decrim × B	0.011	-0.002	0.001	-0.021
	(0.022)	(0.015)	(0.016)	(0.018)
Decrim × H	-0.011	-0.009	0.006	-0.005
	(0.024)	(0.017)	(0.017)	(0.021)
Decrim × W				
Match × B	-0.038*	0.025	-0.008	0.042**
	(0.023)	(0.016)	(0.016)	(0.020)
Match × H				
Match × W				
Obs.	19,164	19,164	6,953	6,953
Reference Group Mean	0.1255	0.0657	0.0252	0.0283
Note:				*p<0.1; **p<0.05; ***p<0.01

Linear Probability Model output of model 3.3 with a control for numbers being rounded added into the covariate matrix for the bandwidths specified, inclusive of the upper bound and not inclusive of the lower bound. This estimation uses Chicago PD data at the individual arrest level. The dummy variable and interaction dummies for white subjects are left out of the estimation to avoid perfect multicolinearity. The 30 gram threshold is estimated with a subset of the data including arrests for amounts of 15-45g in order to avoid any affects that may occur at the 10g threshold. The 100 gram threshold is estimated with a subset of the data including arrests for amounts of 46-155g.