A Few Bad Apples?
Racial Bias in Policing*

Felipe Goncalves   Steven Mello [†]

October 15, 2018

Abstract

We estimate the degree to which individual police officers practice racial discrimination. Using a bunching estimation design and data from the Florida Highway Patrol, we show that minorities are less likely to receive a discount on their speeding tickets than white drivers. Disaggregating this difference to the individual police officer, we find that 40% of officers explain all of the aggregate discrimination. We then apply our officer-level discrimination measures to various policy-relevant questions in the literature. In particular, reassigning officers across locations based on their lenience can effectively reduce the aggregate disparity in treatment.

JEL Classification: J15, K42
Keywords: Discrimination, Racial Bias, Police, Traffic Enforcement.

*We are grateful to Will Dobbie, Ilyana Kuziemko, and Alex Mas for guidance and support throughout this project. We benefited from helpful comments by Leah Boustan, Jessica Brown, Nicholas Buchholz, Janet Currie, Rebecca Diamond, Jennifer Doleac, Nik Engbom, Hank Farber, Sara Heller, Nathaniel Hendren, Daniel Herbst, Bo Honoré, Sierra Kuzava, Michael LaForest, Andrew Langan, Michael Mueller-Smith, Christopher Neilson, Emily Owens, Aurelie Ouss, Jakob Schlockermann, Petra Todd, and participants in various seminars. We thank Beth Allman, Jeffrey Bissainthe, Kiara Guzzo, Wilton Johnson, Timothy Kutta, Stacy Lehmann, and Brenda Paige for assistance with data from various agencies. The Princeton University Industrial Relations Section provided generous financial support. Any errors are our own.

†Industrial Relations Section, Princeton University, Princeton, NJ 08544. Goncalves: fmg@princeton.edu. Mello: smello@princeton.edu

1 Introduction

The disparate treatment of whites and minorities in the criminal justice system is a central policy concern in the United States. Blacks and Hispanics are more likely to be stopped by the police (Coviello and Persico, 2013), convicted of a crime (Anwar *et al.*, 2012), denied bail (Arnold *et al.*, 2018), and issued a lengthy prison sentence (Rehavi and Starr, 2014) relative to observably similar whites. In light of these disparities, a literature has developed to test whether these outcomes can be explained by discrimination on the part of police officers, judges, and other criminal justice agents (Knowles *et al.*, 2001; Anwar and Fang, 2006; Grogger and Ridgeway, 2006; Antonovics and Knight, 2009; Persico, 2009; Abrams *et al.*, 2012; Horrace and Rohlin, 2016; Fryer, 2018; Arnold *et al.*, 2018). The view that discrimination is responsible for these disparate outcomes has gained traction in recent years, particularly within minority communities, following several highly publicized police killings of minorities. A 2013 Gallup poll found that half of black adults agreed that racial differences in incarceration rates are "mostly due to discrimination," while only 19% of white respondents agreed.[1]

While current methods focus on detecting the presence of racial discrimination *on average*, an unresolved challenge is how to identify discrimination at the level of the individual criminal justice agent. Existing approaches largely do not differentiate between discrimination that is widespread versus that which is concentrated among a few agents. However, the optimal policy for mitigating the presence of discrimination depends crucially on how it varies across individuals. Without knowing which agents are discriminatory, it is not possible for institutions to target individuals for discipline or training. More generally, the optimal remedy will depend on the concentration of discrimination across agents. If misbehavior is widespread, a targeted policy of disciplining specific individuals will be ineffectual, and the appropriate response may require a department-wide solution.[2]

[1] See www.gallup.com/poll/175088/gallup-review-black-white-attitudes-toward-police.aspx.

[2] The question of whether misbehavior is systemic or the product of a few bad individuals has also garnered policy interest with regard to federal oversight of local police departments. In January 2017, Attorney General nominee Jeff Sessions stated, "I think there's concern that good police officers and good departments can be sued by the Department of Justice when you just have individuals within a department who have done wrong. These lawsuits undermine the respect for police officers and create an impression that the entire department is not doing their work consistent with fidelity to law and fairness."

In this paper, we study traffic policing by the Florida Highway Patrol and examine whether officers discriminate when enforcing punishments for speeding. We exploit a common institutional feature in traffic policing and use a bunching estimation design to identify discrimination. In many states, the punishment for speeding increases discontinuously with the speed of the driver, exhibiting "jumps" in harshness. A jump may involve not only a higher fine, but also a mandated court appearance or permanent mark on the driver's record. Although officers typically observe a driver's speed via radar before stopping them, they are free to choose what speed to charge. It is thus a common practice for officers to reduce the written speed on a driver's ticket to right below a jump in the fine schedule.[3] Our objective is to identify discrimination in discounting at the level of the individual officer, where we define discrimination as the differential treatment of drivers on the basis of their race when stopped for the same speed.

Several features of our setting are ideal for studying discrimination. When testing for discrimination in many criminal justice outcomes, a central concern is accounting for unobserved differences in criminality across individuals. In the context of speeding tickets, guilt is summarized by the driving speed, which is both one-dimensional and typically observed by the ticketing officer. Further, in many criminal justice contexts, the lenience of an agent is calculated relative to his peers' behavior. In our setting, officers make an explicit decision to reduce a driver's speed, allowing us to see each officer's absolute degree of lenience and observe officers who practice no lenience. Perhaps most importantly, we observe agents making many decisions in very similar contexts, which allows us to construct an accurate measure of discrimination for each officer by comparing his treatment of white and nonwhite drivers.

As shown in Figure 1, the distribution of speeds ticketed by the Florida Highway Patrol between 2005 and 2015 shows substantial excess mass at speeds just below the first fine increase, where speeds are reported relative to the speed limit. Meanwhile, a remarkably small portion of tickets are issued for speeds just above. We take this bunching as evidence that officers systematically manipulate the charged speed, commonly charging speeds just below fine increases after observing a higher speed, perhaps to avoid an onerous punishment for the driver. However, when disaggregated by driver race in Figure 2, we see that minorities

[3] This practice is similar to teachers' bunching up of grades on high-stakes exams (Dee *et al.*, 2016; Diamond and Persson, 2016).

are significantly less likely to be found at the bunch point.

The first task of this paper is to confirm that this disparity is evidence of officer discrimination. Our central challenge is in ruling out that racial differences in treatment are due to differences in criminality. Minorities may be driving faster than whites when stopped, leading officers to treat them less leniently. While our data record the speed that is charged on a ticket, we do not observe the true *stopped* speed of the drivers in our data. To deal with this challenge, we use the fact that one-third of officers practice no lenience. Namely, they exhibit no bunching in their distribution of ticketed speeds.[4] For these officers, we argue that their distribution of ticketed speeds reflects the true distribution of driven speeds among stopped and ticketed drivers. We show that, conditional on location and time, driver characteristics are not predictive of whether the officer he encounters is lenient. Non-lenient officers do not write fewer tickets than lenient officers, and a similar share of their tickets are for speeding offenses. These facts suggest that lenient and non-lenient officers are pulling over similar types of drivers, and thus non-lenient officers can be used to identify the "true" distribution of speeds.

Using a difference-in-differences framework, we then find that white drivers differentially benefit from being stopped by a lenient officer. White drivers stopped by lenient officers are six percentage points more likely to be discounted than minority drivers, off a base of 45%. This gain stems from the fact that minorities are treated less leniently when stopped for speeds ranging from 12 to 25 MPH over the limit.

The central contribution of our paper is to further provide an estimate of the discrimination of *each individual officer*. Specifically, we compute an officer's lenience toward minorities relative to his own treatment of white drivers, differencing out the treatment of each race by non-lenient officers and adjusting for other features of the stop, and treat that difference as the officer's discrimination. Disaggregating to the officer level reveals significant heterogeneity in the degree of discrimination. An officer at the 90th percentile of discrimination is nearly twice as likely to discount a white driver as a minority driver. The modal officer practices no discrimination, and forty percent of officers explain the entirety of the aggregate disparity. Correlating officer-level discrimination to demographics, we find that minority and female officers tend to practice less discrimination than other officers.

[4] The existence of non-lenient officers also leads us to conclude that the bunching of ticketed speeds is not due to drivers strategically driving below the jump in fine.

We then show that a police department could feasibly use our approach to identify discriminatory officers early in their careers. We construct our measure of each officer's discrimination using only his first 100 tickets and show that this early measure is closely correlated with the full-sample estimate of her discrimination. An officer in the top 2% of discrimination in the early measure is on average at the 8th percentile of discrimination in our full-sample estimate, suggesting that a department can quickly identify the worst offending officers.

The remainder of the paper exploits our officer-level measures of lenience and discrimination to understand the mechanisms that lead to the disparity in treatment. To what extent are minorities being discounted less often because they are driving faster? Conversely, how much of the gap in discounting is caused by discrimination? And what policies can be used to reduce any disparity that is due to discrimination?

To answer these questions, we estimate a simple model that identifies both differences in driving speeds, by each race and county, and preferences for discounting, by each officer and race of driver. Model estimates indicate that, within location, forcing all officers to treat minority drivers the same as they treat white drivers removes 83% of the gap in discounting. Only 17% of the gap is due to minorities driving faster. Across locations, a large share of the disparity in treatment is due to the fact that minorities drive in areas where officers are less lenient to all motorists.

Performing the counterfactuals discussed above, we find that policies that target discrimination directly are only mildly effective for reducing the treatment gap. Firing the most discriminatory officers (both for and against minorities) reduces the gap, as does increasing the presence of minority or female officers, but the gains are limited. Perhaps most effective and easily implemented, reassigning officers across counties within their troops so that minorities are exposed to more lenient officers can remove essentially the entire white-minority discounting gap.

While the central focus of the paper is not to differentiate between taste-based discrimination (Becker, 1957) and statistical discrimination (Arrow, 1973; Phelps, 1972), several pieces of evidence suggest that the discrimination we observe is taste-based. First, our setting is not as conducive to statistical discrimination as other criminal justice interactions. In a speeding stop, the officer is aware of the crime committed (i.e., the speed driven) and does not need to use race as a signal of criminality. This knowledge contrasts with cases such as vehicle searches or stop and frisk, where the officer may use demographics to infer whether an individual is carrying contraband. Further, the fact that minority and female officers are less discriminatory on average suggests that the discrimination we observe is a function of preferences rather than statistical inference. We also provide evidence that officers are not statistically discriminating on the basis of whether drivers are deterred from future speeding by getting the full ticket. While we do find evidence that officers discount partly on the basis of whether the individual will contest the ticket in court, this selection cannot explain the racial disparity in discounting. Therefore, for the remainder of the paper, we use *discrimination* and *bias* interchangeably.

This paper contributes to a growing literature on methods of testing for the presence of discrimination in criminal justice and beyond. Popular approaches include audit studies that vary individual race (Bertrand and Mullainathan, 2004; Edelman *et al.*, 2017; Agan and Starr, 2016), studies that vary the observability of race or gender (Goldin and Rouse, 2000; Grogger and Ridgeway, 2006; Donohue, 2014), and studies of settings with rich controls for underlying behavior and context (Fryer, 2018). Another popular approach to testing for bias is the "hit rate test," pioneered by Becker (1957), where discrimination is identified by comparing the success in treatment across two groups where the treator ostensibly cares about a single objective (Knowles *et al.*, 2001; Arnold *et al.*, 2018).

Another popular set of methods for detecting racial bias are *benchmarking* procedures, whereby the behavior of one agent is compared to a proposed control group.[5] Ridgeway and MacDonald (2009) compare the racial makeup of NYPD officers' stop and frisks to those of nearby officers and are able to identify a set of officers with a disproportionately high share of minority stops. To date, Ridgeway and MacDonald (2009) is the only study that aims to identify discrimination of individual criminal justice agents. As they concede, the central limitation of their approach is that they are unable to identify an overall level of discrimination since they use the average officer within a beat as the comparison group for officers who disproportionately stop minorities.

In the paper most closely related to ours, Anbarci and Lee (2014) study the discounting behavior of traffic officers and, using a benchmarking design, find that the racial makeup of discounted tickets is whiter for white officers than for minority officers, suggesting that at least one group is biased in favor of their own race. Our approach broadly falls into the

[5]See Ridgeway and MacDonald (2010) for a review of the benchmarking literature.

benchmarking literature, as we use the set of non-lenient officers as a benchmark for the behavior of other officers. Relative to this existing literature, a strength of our approach is that the non-lenient officers are by construction non-discriminatory. This fact allows us to avoid the common benchmarking challenge that the comparison group may itself be discriminatory, leading to an underestimate of overall discrimination.

This paper also falls into a broad category of recent research using "bunching" estimators to recover behavioral parameters (Kleven, 2016). Predominantly used in the literature on taxation, these studies traditionally attempt to estimate the hypothetical distribution of interest in the absence of bunching by looking at the distribution outside a region around the manipulated area and inferring out-of-sample how the distribution should look at the discontinuity (Chetty *et al.*, 2011; Saez, 2010). Bunching is then estimated to be the difference between the true and hypothetical distribution around the bunch point. In contrast, our approach is similar to Best *et al.* (2015) in that we use panel data and differences across individuals in propensity to bunch to identify the true underlying distribution.

The rest of the paper is organized as follows. Section 2 provides institutional background on the Florida Highway Patrol and describes the data. Section 3 presents a conceptual framework, and Section 4 describes our empirical strategy. Section 5 presents the central findings, and Section 6 considers specification checks and alternative interpretations of our results. Section 7 discusses applications of our officer-level measures of discrimination. In Section 8, we present and estimate a model of officer behavior and perform counterfactuals, and Section 9 concludes.

2 Institutional Background and Data

2.1 Institutions of the Florida Highway Patrol

State-level patrols are the primary enforcers of traffic laws on interstates and many highways. When on patrol, officers are given an assigned zone, within which they combine roving patrol and parked observation patrol. During the course of a traffic stop for speeding, officers have two primary ways to exercise discretion. They can give a written or verbal warning, which leads to no fine or points on the driver's license, or they can reduce the speed charged on the ticket. Florida Highway Patrol (FHP) officers are told explicitly in their training manuals that no enforcement actions during a traffic stop can be based on any demographic characteristics, including race and gender.

In Florida, driving 10 MPH over the limit leads to about a $75 higher fine than 9 MPH over.[6] While drivers receive points on their license for speeding, tickets received for 9 and 10 MPH over the limit carry the same number of license points. While it is also common to find a jump in fine between 19 and 20 MPH over, the data strongly suggest that officers prefer to reduce the ticket to 9 MPH over.

Officers in the FHP are divided into one of 12 troops, almost all of which patrol six to eight counties each. Officer assignments operate on eight-hour shifts and cover an assignment region that roughly corresponds to a county, though the size of a "beat" can vary based on the population density of the region. In practice, because we do not observe the exact beat policed by an officer, we will use the county of the stop as a proxy for the officer's assignment region.

Officers face no revenue incentive to collect tickets, as all fines paid by drivers are collected by the government of the county in which the fine was issued. There is also, to the best of our knowledge, no quota system for a minimum number of tickets officers must write.[7] Officers do, however, potentially have a promotion incentive to write a certain number of tickets, as the number of tickets they write appears on their performance evaluations. We believe these set of institutional factors contribute to an environment in which officers are encouraged to write tickets but also have the freedom to write reduced charges, which is ideal for our research design.

While all speeding beyond 5 MPH over the limit commands a statutory fine, the evidence suggests that drivers are not regularly pulled over for less than 10 MPH over, and the data show very few tickets for 8 MPH over and 10 MPH over. As we will reiterate in Section 4, many officers have almost no tickets issued at 9 MPH over the limit, suggesting that the majority of the bunching of tickets is for higher speeds that have been reduced.

[6] The actual fine schedule depends on the county in Florida, though the jump point is the same across all counties and always includes at least a $50 jump in fine.

[7] We checked for a spike in the number of issued tickets at certain days of the month or days of the week, and found no evidence of an "end of the period" effect.

2.2 Data

From the Florida Court Clerks & Comptrollers, we obtained data on traffic citations issued by the Florida Highway Patrol (FHP) for the years 2005-2015. These data include all information provided on the stopped motorist's driver's license – name, address, race, gender, height, and date of birth, as well as driver's license state and number. The make, model, and year of the stopped automobile is provided, but this information is recorded inconsistently. In the final sample of citations, 69% of tickets list the vehicle make and year. The citing officer is identified by name, rank, troop number, and badge number.[8] While we see the speed charged by the officer, we do not see the original speed recorded by the officer. We also do not see stops and interactions that do not result in a traffic citation.[9]

To supplement the citations data, we obtained officer demographic information from the Florida Department of Law Enforcement (FDLE). These data include officer race, sex, age, education level, and the Florida law enforcement employment history of all law enforcement officers employed in the State of Florida. It further includes every misconduct investigation made by the state against an officer, the type of alleged violation, and the ultimate verdict of the state. From the FHP, we also collected information on all use of force incidents and civilian complaints against officers for the period 2010-2015, which list the name of the officer, the date of the incident, and a description of the incident.

While the citations record the driver race, there appear to be inconsistencies in the recording of Hispanic. For example, Miami-Dade County issues fewer than 1% of their tickets to Hispanic drivers. To address this issue, we match the drivers' names to Census records, which record all names that appear more than 1,000 times and the share of white, black, Hispanic, and other that carry that name. If an individual in our data has a name that is more than 80% Hispanic, we record them as such.

We restrict the sample to citations in which the main offense is speeding; no accident is reported; the cited speed is between zero and 40 above the posted speed; race of the driver

[8]The full data from the FCC contain all traffic citations for 2005-2015, including tickets not given by the highway patrol. We use these tickets to measure an individual's previous driving record. We do not use non-FHP tickets in our measures of bias, because officers are harder to identify in these data. Further, much of the personnel information we collected is unique to the FHP.

[9]The problem of only seeing interactions that lead to enforcement is general in the discrimination literature. For a recent paper that addresses this issue, see West (2018).

is reported as white, black, or Hispanic (or is imputed as such); and the gender, age, and driver's license number are not missing. To link citations and officer information, we first narrowed the list of FDLE personnel to include only officers with an employment spell as a sworn officer with the FHP covering some portion of the 2005-2015 period. We then match the list of candidate officers with the citations data using the officer name. We exclude stops that cannot be matched to an officer. Lastly, we restrict the sample to officers issuing at least 100 citations, with at least 20 given to minorities and 20 to whites.

The final sample includes 1,142,628 citations issued by 1,591 officers, from an initial sample of 2,124,692 speeding citations. The two most binding restrictions are requiring that race be specified (84% of tickets) and requiring that the officer be linkable to the FDLE (77%). In the appendix Section A we include a table that documents the sample reduction from each restriction we make. In all of our analyses, we consider speed relative to the speed limit (or posted speed) rather than absolute speed. We often refer to this quantity as *MPH Over* or simply as "the speed."

Beginning in 2013, about 40% of tickets are geocoded with the latitude and longitude of a stop (135,586 observations). We link the geocoded tickets to a Florida Department of Transportation roadmap shapefile using ArcGIS.[10] The shapefile is at the level of road "segments," which are on average 6.7 miles long and roughly correspond to entire streets within cities and uninterrupted stretches of road on interstates and highways. Tickets are linked to the nearest segment, and we remove tickets that are more than 100 meters from the nearest road (dropping 1.5% of observations). Officers in more rural areas and on interstates are given priority for vehicles with GPS, as they cannot clearly describe the location of their ticket using street intersections. 40% of officers have fewer than 5% of their stops geocoded, and there is some variation across counties in the share of tickets geocoded. Throughout the analysis, we provide results for the restricted sample of tickets with GPS with corresponding fixed effects at the road-segment level. Because we do not have perfect information on officer assignment (and use the county of the stop as a proxy), the road-segment analysis allows us to consider a more granular comparison of drivers.

[10]http://www.fdot.gov/planning/statistics/gis/road.shtm; We use the "Basemap Routes with Measures" shapefile.

2.3 Summary Statistics

Table 1 presents summary statistics for the sample, broken out by driver race. 58% of drivers are white, 18% are black, and about 23% are Hispanic. Drivers are 35% female and about 36 years old on average, with Hispanics less likely to be female and minority drivers typically younger. In-state drivers account for 84% of tickets. The average driver has been cited about 0.34 times in the past year, though minorities have 0.13 more prior tickets. On average, minority drivers are charged with higher speeds than whites: just over 1 MPH higher for blacks and almost 3 MPH higher for Hispanics. Consistent with Figures 1 and 2, drivers of all races have a high probability of being ticketed at 9 MPH over the limit, which is just below the first jump in the fine schedule. However, minority drivers are also less likely to be charged this speed. As we show in Appendix Tables A.1 and A.2, these disparities in speed and ticketing below the jump persist after controlling for all stop characteristics and time and location fixed effects.

A notable feature of the distribution of tickets is the heaping of charged speeds at multiples of five above the bunch point. This heaping occurs because, in many instances, officers do not use a radar gun, and their recording of the speed may be approximate. For 51% of the tickets, the officers do record the "method of arrest," and 17% of these tickets report that the officer used a radar gun. We report in Appendix Figure A.1 the distribution of ticketed speeds for this subsample, and there is no heaping at multiples of five.[11]

In Table 2, we compare the racial distribution of speeding tickets with the racial distribution of residents and drivers in Florida using the 2006-2010 American Community Survey (ACS) 1% samples.[12] These data demonstrate that whites account for about 65% of Florida's population, and 63% of its drivers (an ACS respondent is considered a driver if they indicate that they drive to work), and about 59% of tickets.[13] Blacks represent around 14% of the population and driving population, but 18% of tickets. Similarly, Hispanics are 20% of the population, almost 19% of the driving population, and 24% of tickets. In Columns

[11] In Appendix Table 5, we also show that our main result is not changed when restricting attention to this subsample.

[12] We obtained these data from Integrated Public Use Microdata Series (IPUMS). So that the samples are parallel, we use only citations from 2006-2010 and keep only white, black, or Hispanic individuals aged 16 or over in the ACS. We use sampling weights when computing the shares from the ACS data.

[13] To match to the shares in our data, we restrict attention to ACS respondents who report their race as white, black, or Hispanic.

(4) and (5), we present the racial distribution of black, white, and Hispanic drivers involved in crashes and crashes with injuries over the 2006-2010 period. These shares are computed from records provided by the Florida Division of Motorist Services that contain information on all auto accidents known to police. These data likely correspond more closely to the demographic composition of speeders than the general population of drivers. The racial shares in the crash data correspond very closely to the citations data, with black drivers slightly overrepresented and Hispanic drivers slightly underrepresented among crashes with an injury. Overall, we do not have the impression that minorities are severely overrepresented or underrepresented in the tickets data relative to the population or the distribution of speeding drivers.

3 Conceptual Framework

In the previous section we documented the disparity in ticketing at 9 MPH over between whites and minorities. Here we introduce a simple framework of officer decision-making that can explain the disparity in discounting through two mechanisms – differences in speeding and discrimination – and motivates our empirical strategy in Section 4 and our modeling exercise in Section 8.

Officer j stops motorist i for speeding. His observed speed x is drawn from some discrete distribution $F_r(\cdot)$, which can be a function of the driver's race r. For simplicity, we suppress here the possible dependence of the distribution on other driver characteristics. If the driver's speed is above x_d, the officer has the choice to reduce the charged speed to x_d to reduce the fine the driver will face. Otherwise, the speed is set to x. When deciding whether to reduce the ticket, we suppose the officer weighs a mix of personal concerns such as the inconvenience of attending traffic court; policing objectives such as the blameworthiness of the individual and the potential deterrence effect of ticketing the individual; and bias against certain groups r. Balancing these objectives, the officer has some probability $P_j(x, r(i))$ of discounting the individual, which may be a function of the driver's race r and the driver's speed x.

In this framework, it is natural to define discrimination in the following way: We say that officer j is *discriminatory* if $P_j(x, r(i) = w) > P_j(x, r(i) = m)$ for a given speed x. While we describe the officers' preferences as potentially reflecting bias, we are not yet taking a stand on whether any disparity in treatment is taste-based versus statistical. For example, it is possible that some officers prefer whites because they believe the likelihood of having to go to court later is lower. We discuss statistical discrimination in Section 6 and why we believe the observed discrimination in discounting is taste-based.

The first empirical step we take is to model the likelihood of an individual appearing at the discount point and above, given his observables. In our model, the probability of being charged the discount speed is the summed likelihood of appearing at or above that speed times the likelihood of being discounted:

$$\Pr(X_i = x_d | \, i, j \,) = F_{r(i)}(x_d) + \sum_{k > x_d} F_{r(i)}(k) \cdot P_j(x, r(i)) \tag{1}$$

and the probability of appearing at a point above the discount point,

$$\Pr(X_i = x > x_d | \, i, j \,) = F_r(x) \cdot \big(1 - P_j(x, r(i))\big) \tag{2}$$

is the likelihood of having driven that speed and then *not* being discounted.

4 Empirical Strategy

From Equations (1) and (2), we see that racial differences in the likelihood of appearing at the bunch point and above can arise from either differences in speeds $F_r(x)$ or differences in speed-specific discounting, $P_j(x, r(i))$. Primarily in the latter case will the disparity be of policy interest, as it would be due to discrimination rather than differences in behavior. To determine whether the observed disparity is due to differences in driving speed, we use the fact that one-third of officers in our sample practice no lenience. In other words, these officers have no bunching in their distribution of speeds.

In Figure 3, we motivate this approach by documenting the significant heterogeneity in discounting across officers. Panel A plots the officer-level distribution of lenience, defined as the share of tickets written for 9 MPH or above that are for exactly 9 MPH. A large share of officers appear to exhibit very little lenience, with 30% writing less than 1% of tickets for this bunching speed. Panel B plots the distribution of officer lenience after residualizing county and month-of-stop fixed effects and driver characteristics. The observed disparity suggests that the heterogeneity across officers is not due to differences in location or characteristics of the stopped drivers.

The lower two panels confirm that officers are persistent in their level of lenience across time and location. In Panel C, we plot each officer's residualized lenience in his year with the second-most stops (y-axis) against his residualized lenience in his year with the most stops (x-axis). A strong correlation is evident: an officer who charges 9 MPH relatively more often in one year also does so in other years. In Panel D, we plot lenience in the county where the officer has made the second most stops against lenience in the county where he has made the most stops, confirming that officer lenience is highly correlated over space.

We treat the 33% of officers with fewer than 2% of their tickets issued at 9 MPH over as non-lenient officers, and we use these officers for two purposes.[14] First, we suppose that these officers' ticketing distribution reflects the true distribution of speeds within their location and shift and use them to uncover the true racial difference in speeding. Secondly, we use these officers as a control group in a difference-in-differences style framework to estimate the effect of encountering a lenient officer on the likelihood of being discounted for each racial group.

To do so, we run a linear probability model, where the outcome is an indicator S_{ij}^k of whether a driver is stopped at a given speed k, and the race of the driver is interacted with the lenience of the officer:

$$\begin{aligned} S_{ij}^k &= \beta_0 + \beta_1 \cdot \text{White}_i + \beta_2 \cdot \text{Lenient}_j \\ &\quad + \beta_3 \cdot \text{White}_i \cdot \text{Lenient}_j + X_{ij}\gamma + \epsilon_{ij} \end{aligned} \tag{3}$$

For all regressions, the primary coefficient of interest is β_3, the interaction between white driver and lenient officer. For the bunch point of 9 MPH over the limit, β_3 reflects how much more a white driver benefits from encountering a lenient officer than a minority driver. For all speeds above 9 MPH, the interaction reflects how much less likely minorities are to be discounted by a lenient officer. X_{ij} contains the set of all observable characteristics of the drivers, including gender, age, age squared, number of previous tickets, whether the driver is in-state, the log average income of the driver's home zip code, vehicle age and age squared, and indicators for vehicle make.

[14] An alternative approach is to explicitly test for the presence of bunching officer-by-officer. When we do so using the Frandsen (2017) test, the set of officers identified as non-lenient remain very similar and the regression results do not change. These results are reported in Appendix Table 5.

We also include fixed effects interacted at the level of the stop's year, month, day of the week, shift, county, and whether it was on a highway, which we henceforth refer to as the time and location of the stop. The purpose of the fixed effects is to make the difference-in-differences comparison among drivers stopped in the same beat and shift. As mentioned earlier, county is our best available approximation to an officer's beat. To provide an even more granular comparison, we will also report results for our GPS sample, where we include fixed effects interacted at the year, month, day of the week, shift, and road segment level.

To calculate each officer's individual discrimination coefficient, we take a similar approach and use non-lenient officers as a control for the baseline frequency of tickets at 9 MPH over, but we allow the coefficients for Lenient_j and $\text{White}_i \cdot \text{Lenient}_j$ to vary by individual officer:

$$\begin{aligned} S_{ij}^{9} &= \beta_0 + \beta_1 \cdot \text{White}_i + \beta_2^j \cdot \text{Lenient}_j \\ &\quad + \beta_3^j \cdot \text{White}_i \cdot \text{Lenient}_j + X_{ij}\gamma + \epsilon_{ij} \end{aligned} \tag{4}$$

The coefficients of interest, β_3^j, are identified from each officer's difference in discounting between whites and minorities, differencing out the disparity in ticketing for non-lenient officers. We denote β_3^j as officer j's degree of discrimination. For the purpose of reporting the distribution of discrimination across officers, we treat non-lenient officers as having $\beta_3^j = 0$, since by definition they cannot be discriminatory.

The intuition for our difference-in-differences procedure is shown in the top two images in Figure 4. Here we plot the histogram for non-lenient officers over the histogram for lenient officers, separately by driver race. The gap in histograms between lenient and non-lenient officers above 9 MPH over indicates the speeds at which drivers are reduced to 9 MPH over. The difference in these gaps between white and minority drivers indicate the difference in discounting between races for each speed.

For lenient officers to be a valid control group, it must be the case that, conditional on location and time of the stop, the lenience of the officer is uncorrelated with the error term, $\text{Cov}(\text{Lenient}_j, \epsilon_{ij}) = 0$. This assumption entails two presumptions about the stop. First, we require that officers in the same shift and beat are not systematically different in who they stop; second, officers do not systematically differ in the characteristics of drivers to whom they give a warning, which would lead to differential selection into our data. As mentioned above, we see no information about stops that do not result in a ticket, so one concern is that officers who differ in their lenience toward discounting may also differ in their lenience in the initial margin of whether to even write a ticket.

In Figure 5, we evaluate how the characteristics of an officer's stops vary with whether the officer is lenient or not, where both variables have been residualized with location-time fixed effects. The top left panel of the figure shows that officer lenience is not predictive of his share of tickets written to minorities. The top right panel shows that officer lenience is uncorrelated with whether a driver's race is missing, and the bottom left panel shows that officer lenience has only a small, though significant, correlation with the likelihood that a ticket is for speeding. The bottom right panel shows the relationship between officer lenience and the average daily number of tickets. For this figure we calculate both measures at the annual level, during which officers write most of their tickets in one county, allowing us to control for county-by-year fixed effects. We find that whether or not an officer is lenient is not predictive of the number of tickets written per day.

To further test for selection on observables, Table 3 estimates how officer lenience varies with driver characteristics. The outcome for all regressions is the indicator for whether the stopping officer is identified as lenient. The F-tests report a joint test of the hypothesis that all driver characteristics have zero correlation with officer lenience. Column (1) reports results with no controls for location or time. Here officer lenience varies significantly with driver characteristics. Hispanic drivers and in-state-license drivers are ticketed in areas where officers are less lenient to everyone. Columns (2) and (3) restrict attention to variation within location and location plus time, respectively. With these controls, officer type varies much less significantly with driver characteristics. A joint F-test fails to reject at 10% significance that all driver characteristics are equal to zero. Columns (4) and (5) report results for our GPS'ed sample. Both with and without fixed effects for the road-segment of the stop, we find that our indicator for officer lenience is uncorrelated with driver characteristics.[15]

[15]In Appendix Table A.4, we consider a similar set of randomization checks, where the outcome is the officer's share of tickets at 9 MPH over rather than the indicator for lenience. The results are similar to Table 3. However, in Column (3) we reject the null of no relationship between the outcome and observables at the 5% level. We believe this correlation is due to our inability to perfectly control for officer assignment. When we restrict attention to our GPS sample in Columns (4) and (5), we continue to have no relationship between observables and officer lenience.

5 Results

Our first use of non-lenient officers is to test whether minorities truly drive faster than white drivers. The bottom two panels of Figure 4 report the distribution of ticketed speeds for non-lenient officers. Unconditional on any covariates, minorities drive 1.5 MPH faster than whites. However, when controlling for county and individual covariates, this disparity shrinks to 0.39 MPH, and the disparity is barely perceptible visually. The majority of the reduction comes from accounting for county fixed effects, since minorities tend to drive in counties in which all drivers are stopped at faster speeds. The fact that the county-specific disparity is so small suggests that the racial disparity in discounting cannot be explained by differences in driving speed. In Appendix Table (A.3), we show that this small gap is consistent across various specifications for time and location controls.

Figure 6 and Table 4 report the results of the difference-in-differences test of discrimination. The figure reports regression coefficients from both a specification with no controls and our preferred specification with individual covariates and fixed effects for county by year by month by shift by highway. As indicated by the interaction variable for white drivers and lenient officers encountered at 9 MPH over, white drivers are significantly more likely to receive a discount than minority drivers. Off a mean probability of 45%, white drivers stopped by lenient officers are encountered at the bunch point 6-8.4pp more often than minorities, and this disparity persists regardless of the specification. In Columns (4) and (5) of Table 4, we perform the same regression for the restricted sample with GPS ticket location. The results continue when allowing for stretch-of-road fixed effects, though the coefficient is a slightly smaller 5.5pp.

The interpretation of these coefficients tell us how much more likely a *lenient* officer is to discount a white driver. To calculate a differential probability of discount by an average officer, we use the fact that two-thirds of tickets are written by lenient officers and scale accordingly, finding that an average encounter leads to a 4pp higher discount probability for white drivers, off a base of 30%.

The interaction coefficients for speeds above 9 MPH shown in Figure 6 indicate where minority drivers are disproportionately being ticketed, and thus the speeds at which white drivers are being differentially discounted. The interaction coefficient is negative and significant for all speeds between 12 and 20 MPH, suggesting that at these speeds minorities are less likely to receive a break.

A natural question to ask is how this estimate aggregates to a total cost of discrimination. Every year, about 590,000 speeding tickets are given to drivers in Florida for 9 MPH over or greater, 240,000 of which are given to black and Hispanic drivers. The jump from 9 MPH over to 10 MPH over leads to a $75 fine increase. Using our estimate that minority drivers are 4 percentage points less likely to be discounted, we calculate the cost of discrimination toward minority drivers to be $720,000 per year. Scaled up to the entire US population, that figure increases to $11.3 million.[16]

Officer-level results are reported in Figure 7. The figure displays the across-officer distribution of the interaction coefficient $\hat{\beta}_3^j$, where non-lenient officers are assigned $\hat{\beta}_3^j = 0$. The line represents a kernel density plot of our measure of discrimination against minority drivers, so that the farther right an officer is in the distribution of discrimination, the greater his level of discrimination. The unit of our measure is probability difference in percentage points. An officer whose discrimination against minorities is 0.1, for example, is 10 percentage points more likely to offer a fine reduction to a white than a minority driver. The percentiles of officer discrimination are also reported in Appendix Table A.5.

The first fact to note is the substantial heterogeneity in discrimination across officers. While the modal officer practices no discrimination, we find a large mass of officers with positive discrimination. Officers at the 10th and 90th percentiles of discrimination have a 14 percentage point difference in their racial disparity. When calculating their lenience toward minorities as a share of their lenience toward whites, officers at the 90th percentile are more than 40% less likely to discount minorities.

The second notable fact is that the median level of discrimination is quite small, three percentage points off a base of 30%. While this disparity is comparable to the black-white wage gap (Neal and Johnson, 1996), it is possible that the officer in question is not aware of such a disparity. A large literature has explored the role of implicit bias as a source of discrimination (Greenwald and Krieger, 2006; Banks *et al.*, 2006), and in many cases the individual in question is not aware of his bias. We believe that for the median officer our results are consistent with such a theory. However, for higher percentiles of the distribution, it is hard to explain large gaps in treatment as a practice that is imperceptible to the officer.

[16] Florida's 2016 population is 20.6 million, and the US population is 323.1 million, so we multiply our figure by 323.1/20.6

An officer at the 75th percentile has a 6.8pp difference in treatment, and this gap nearly doubles to 12.8pp at the 90th percentile.

Even under a data-generating process in which officers all have the same true discrimination, our estimates would have a distribution due to sampling error. This scenario, however, cannot explain the heterogeneity we find. The average standard error for an officer's $\hat{\beta}_3^j$ is 0.014 – less than one-fourth the standard deviation of $\hat{\beta}_3^j$ across officers, 0.068. In the scenario in which true discrimination is uniform, these numbers would be similar in magnitude. We thus conclude that the majority of the variation is due to true officer differences in discrimination rather than estimation error. [17]

5.1 Share of Officers Who Are Discriminatory

Another approach to understanding the variance in discrimination across officers is to estimate what share of officers are discriminatory. We know that each officer's discrimination measure is an additive function of his true discrimination plus estimation error, $\hat{\theta}_j = \theta_j + \epsilon_j$, where ϵ_j is asymptotically normally distributed and σ_j^2 is estimated in the officer-level regression. We can assume an officer's discrimination can take on a finite set of values on a fine grid, $\theta_j \in \{\theta^k\}$,[18] and calculate the likelihood of observing each officer's discrimination measure $\hat{\theta}_j$ given the noise in the measure and the true distribution $f(\theta_k)$:

$$\text{Prob}(\hat{\Theta}_j = \hat{\theta}_j) = \sum_{\{\theta_k\}} f(\theta_k) \cdot \text{Prob}(\epsilon_j = \theta_k - \hat{\theta}_j)$$

We then estimate $\{f(\theta_k)\}$ by maximum likelihood. Using this approach, and calculating $1 - \hat{F}(0)$ as the share, we find that 41% (CI 38.5-43.7%) of officers are discriminatory.[19] In contrast, we find that only 7% (CI 5.6-8.7%) of officers have $\theta_j < 0$, i.e., practice reverse discrimination.[20]

[17] One way to calculate officer heterogeneity's accounting for noise is to do a Bayes shrinkage procedure. When we replicate the approach of Aaronson *et al.* (2007), our distribution of discrimination looks nearly identical to the unshrunk version.

[18] The grid is 99 points spanning the 1st to 99th percentiles of the empirical distribution of $\hat{\theta}_j$.

[19] Confidence intervals are calculated through bootstrapping by performing 100 draws of the set $\{\hat{\theta}_j\}$ and performing MLE on each draw.

[20] This approach is a discretized version of a deconvolution procedure (Delaigle *et al.*, 2008). Doing the continuous deconvolution leads to an identical estimate for the share of officers who are discriminatory.

6 Robustness Checks and Alternative Explanations

In this section we report various specification and robustness checks to evaluate the strength of our findings. In particular, we consider various explanations of our findings that are not officer racial bias.

In Section 4, we reported various specification checks for the randomization of officer lenience. An additional test for the random assignment of officer to driver is that officer *discrimination* is not correlated with driver characteristics. We report such regressions in Appendix Table A.6. As before, Column (1) reports the regression with no controls, and the F-test indicates that some driver characteristics are correlated with officer discrimination, statewide. All other regressions, which include controls for county, report no relationship between officer discrimination and driver demographics.

In Table 5 we report the primary difference-in-differences results with various changes in the regression specification, with Column (1) re-reporting the baseline specification. In Column (2), we conduct a split-sample analysis where we calculate whether an officer is lenient using a randomly-selected 20% of officers' tickets, which we exclude from the regression. In Column (3), lenience is calculated separately for each officer's year of ticketing, allowing for changes in officer behavior over a career. In Column (4), we calculate an officer's measure of lenience using the Frandsen (2017) test for manipulation of a discrete running variable (designed for testing the validity of the regression discontinuity research design). In Column (5), we re-weight the set of observations so that the "share" minority in each county is the same. This approach is borrowed from Anwar and Fang (2006) and accounts for the possibility that officers differ across counties in their lenience, which could be correlated with minority status. In Column (6), we interact officer lenience with all driver characteristics, testing that lenience towards whites is not confounded by lenience towards observable non-race characteristics.

One feature of the data discussed earlier is that the histogram of ticketed speeds exhibits jumps at multiples of five, and we argue that this heaping is due to officers not using a radar gun and writing an approximate speed for the driver. In Column (7) of Appendix Table 5, we find that our baseline regression is essentially unchanged when restricting to the sample of tickets from a radar gun.

In all these specifications, the interaction coefficient between officer lenient and driver race is significant and quantitatively similar to the baseline specification. The largest disparity is evident in the re-weighted specification, where the coefficient reduces from 6.8pp to 5.5pp. This difference suggests that some of the gap in treatment between whites and minorities is due to minorities disproportionately driving in counties where officers are less lenient overall. These differences across counties could be due to differences in how much drivers exceed the speed limit. In our model in Section 8, we explicitly account for the possibility that counties and races differ in speeds and continue to find a disparity in discounting between races.

6.1 Selection into the Data

As we state in Section 2, our data are constrained by the fact that we do not observe interactions that do not result in a ticket. One concern is that differences on the margin of whether to give a ticket vary across officers and that this difference may make our estimates of officer-level discrimination inconsistent.

We do not believe that this issue is a serious concern in our setting. In Section 4 we show that officer lenience is only very weakly correlated with the frequency of tickets written and, in Section 6, that discrimination does not correlate with the share of tickets written for minorities.

We further believe that any discrimination on the stopping margin would likely bias our results toward finding less discrimination in discounting. To see this argument, imagine a minority driver who is on the margin of being ticketed, such that if he were white he would have been let off with a warning. This driver appears in our data only because he is a minority. Because he is at this margin, it is very likely the officer will give him a discount. Therefore, discrimination on the ticketing margin places too many minority drivers in our sample who are disproportionately at the discount point. Thus, the disparity in discounting would be even greater without a hypothetical disparity in ticketing.[21]

In Appendix Section B, we formalize this logic with a simple selection model that allows for officer differences in propensity to let drivers off with a warning. Using this model, we implement a sample selection correction, as in Heckman (1979), that accounts for officer-by-

[21]As pointed out in Brock *et al.* (2012), it is not necessarily the case that an individual at the margin of appearing in the data is guaranteed a certain treatment once in the data. In light of their argument, our selection correction procedure allows for an arbitrary relationship between an individual's propensity to be ticketed and propensity to be discounted.

race differences in propensity to appear in the data. We reports the results of this regression in Table 6. Column (1) reports our baseline regression, and Column (2) implements the sample-selection correction. The results look identical after making this correction.

6.2 Racial Difference in Requesting a Break

One key insight of our analysis is that while whites and minorities do not seem to be differentially exposed to police through traffic enforcement, the quality of the interaction can vary significantly. This insight has also been made by research that documents racial differences in the quality of police-civilian interactions (Najdowski, 2011; Najdowski *et al.*, 2015; Trinkner and Goff, 2016; Voigt *et al.*, 2017).

However, differences in the quality of the interaction leave open the possibility that white drivers are actually more likely to request a break than minorities. If officers are open to requests for a discount, this difference in solicitations could generate a disparity in lenience. As in most discrimination studies, we do not have direct information on the quality and content of the interaction between officer and driver, so we cannot directly test for whether drivers differ in their propensity to request a break.

We do not believe, however, that differences in requests for a break can explain the disparity in discounting we observe. For a given level of lenience toward whites, we still see differences in discrimination across officers. If officers are simply receiving solicitations for a discount from the drivers (and whites ask more often), we should expect that for a given level of lenience toward whites, lenience toward minorities is a fixed fraction of that lenience. This pattern is not borne out in the data. We find that 50% of the variance in discrimination across officers remains after conditioning for lenience against whites.

Relative to existing studies in the discrimination literature, one strength of our data is that individuals can be linked across tickets, allowing us to evaluate whether there are individual-level differences in propensity to receive a discount. We probe this question further in Columns (3)-(5) of Table 6. To do so, we restrict attention to individuals with at least two tickets. Column (3) presents a regression of discounting on individual characteristics, and Column (4) adds officer fixed effects. This addition increases the R^2 from 0.318 to 0.527. In contrast, the further addition of individual-fixed effects in Column (5) only increases the R^2 to 0.542. This small increase shows that, beyond individual covariates and the stopping officer, the specific individual has little explanatory power for whether a discount is given, indicating that individual differences in propensity to request a break is likely not a substantial factor in the disparity in discounting.

6.3 Statistical Discrimination v. Taste-Based Discrimination

Throughout the paper, we have defined racial bias as the differential treatment of drivers by race who are stopped for the same speed. This definition is not innocuous, as there may be some reasons for differential treatment unrelated to observed driving speed that, while contentious in their use, are not specifically racial animus. For example, officers may choose who to discount on the basis of how individuals respond after the stop: some drivers may be more deterrable and speed less after a harsh ticket; others may respond by contesting the ticket in court. Our baseline regressions show that officers differentiate between white and minority drivers after controlling for previous tickets, suggesting that the observed disparity does not reflect statistical discrimination on the level of criminality. However, these estimates do not rule out racial differences in the *responsiveness* to the ticket.

In Appendix Section C, we present a simple test for whether officers are attempting to minimize court contesting or maximize deterrence, which we report in Table 7. To evaluate the impact of a discounted ticket, we instrument for receiving a discount using the stopping officer's persistent (leave-out) level of lenience.[22]. Our test then follows the logic of Heckman *et al.* (2010) and claims that non-linearities in the relationship between the outcome and the propensity score reflect sorting of individuals on the basis of their responsiveness.

We find no evidence that officers choose who to discount on the basis of deterrability: the impact of a discount on future speeding is positive but constant across levels of officer lenience. However, we do find that officers choose who to discount based on whether they will contest their ticket in court: among officers who are not very lenient, the marginal impact of giving a driver a discount is a large reduction in likelihood of contesting the ticket. In contrast, more lenient officers have a marginal impact of a discount on court contestation that is significantly smaller, suggesting that more responsive drivers are discounted first. We then perform in Column (5) a hit-rate test similar to Arnold *et al.* (2018) and find that

[22]This procedure is very commonly used in the criminal justice literature when judges differ in their punitiveness (Kling, 2006; Dobbie and Song, 2015) We use this approach to evaluate how individuals respond to their ticket in a follow-up paper, Goncalves and Mello (2017).

officers' statistical discrimination on court contestation cannot explain the racial disparity in discounting.

7 Applications of Officer Heterogeneity

Relative to the literature, our central contribution is the ability to generate officer-level estimates of discrimination, as presented in Section 5. The first insight we gain from this distribution is that discrimination varies greatly from officer to officer. However, estimating the degree of discrimination of individual officers allows us to address various previously unanswerable questions. How does discrimination vary by officer demographics? Are early measures of discrimination predictive of long-term discrimination? And which personnel policies can mitigate the effect of discrimination? We answer the first two questions in this section and the third in Section 8.

7.1 Do Officer Characteristics Predict Discrimination?

Given an officer-level measure of racial discrimination, a natural question is how it correlates with other officer characteristics and behaviors. We can tackle this question using the personnel records collected from the FDLE and the FHP.

The left panel of Figure 8 shows how our measure of discrimination varies by officer race. Perhaps consistent with intuition, white officers are much more likely to be discriminatory against minority drivers, with a greater rightward skewness in their distribution. However, minority officers are still, on average, discriminatory against minority drivers. Among black officers, a very small percentage are discriminatory in favor of minority drivers. Some of the disparity in discrimination across officer race is driven by minority officers being less likely to be lenient overall. This fact is due in part to minority officers working in troops in which all officers are less lenient. In the right panel of Figure 8, we show the distribution of discrimination only for lenient officers. The white officers' distribution continues to be shifted farther to the right.

The ability to identify discrimination separately by officer race is another advance beyond the previous literature. Several benchmarking papers detect bias using comparisons across officer race (Anwar and Fang, 2006; Antonovics and Knight, 2009; Price and Wolfers, 2010; Anbarci and Lee, 2014). With such an approach, we can know that some race of officers is acting in a discriminatory manner, but not which group. With our method, we can see the magnitude of discrimination separately for each officer race.

In Table 8, we present regressions of officer-level discrimination on officer characteristics. Here we have disaggregated officer discrimination to be calculated separately against black drivers and Hispanic drivers[23]. All observations are weighted by the variance of the noise in our estimate of the officer's bias.

As with the density plots, the clear takeaway from the regressions is that minority officers are more lenient toward minority drivers, as we might expect. Female officers appear less biased against black drivers and marginally less biased against Hispanic drivers. Officers with more years experience are more discriminatory against Hispanic drivers, though the standard errors are large. There appears to be no relationship between officer discrimination and level of education, number of civilian complaints, or number of use-of-force incidents.

While some officer demographics are predictive of discrimination, we are also interested in the usability of our measures of discrimination to predict other officer behavior. A growing literature is interested in identifying the factors that can predict officer misconduct (Chalfin *et al.*, 2016). Here we ask whether our measures of lenience and discrimination can be used to predict an officer's propensity to receive a civilian complaint or use force on the job. To make the analysis at the officer-level – but still account for differences in years and locations worked – we run regressions of the following form:

$$Y_i = \alpha_0 + \alpha_1 \cdot \text{Lenience}_i + \alpha_2 \cdot \text{Bias}_i + X_i \cdot \beta + \sum_k \text{District}_i^k + \sum_k \text{Year}_i^k + \epsilon_i$$

where Y_i is an outcome of either receiving a civilian complaint or using force. District_i^k is an indicator for an officer ever working in District k in the years 2011-2016, and Year_i^k indicates whether an officer appears in our traffic data in year k. X_i are other officer-level characteristics.

The results, reported in Table A.7, indicate that lenience is statistically predictive of both civilian complaints and use of force. An increase of one standard deviation in lenience (25% change in discounting) correlates to 0.19 fewer civilian complaints and a 5.5% decreased

[23]Specifically, we run $S_{ij}^9 = \beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Hispanic}_i + \beta_3^j \cdot \text{Lenient}_j + \beta_B^j \cdot \text{Black}_i \cdot \text{Lenient}_j + \beta_H^j \cdot \text{Hispanic}_i \cdot \text{Lenient}_j + X_{ij}\gamma + \epsilon_{ij}$. We take -$\beta_B^j$ and -β_H^j to be our measures of discrimination against black and hispanic drivers, respectively.

likelihood of receiving any complaints. Similarly, a one SD increase in lenience is associated with 0.06 fewer incidents of force and 3% lower likelihood of any force. Black officers are less likely to engage in force, as are older officers. Female officers are less likely to receive complaints but just as likely as male officers to use force. Discrimination against minorities seems to be positively related to force and complaints, though the standard errors are too large to say conclusively.[24]

7.2 Are Our Estimates Usable by a Police Department?

We argued above that the central value of estimating the distribution of discrimination is its use for conducting policy. Knowing who is discriminatory is crucial for identifying who to train or discipline. Given this motivation, a natural question is whether the measure we have constructed for each officer is actually usable by a department to identify discriminatory officers. Specifically, we ask whether an individual's discrimination– as calculated from his first 100 tickets, which the median officer writes in 400 calendar days– is close to his measure from the full sample.

To calculate the early measure of discrimination, we first predict whether a ticket is going to be at the discount point using only our sample of non-lenient officers, fitting $E(S^9_{ij}|X_{ij}) = X_{ij}\beta$. We then calculate $\epsilon_{ij} = S^9_{ij} - X_{ij}\hat{\beta}$ for each ticket, including those by lenient officers. Then, we take each officer's first 100 tickets and calculate discrimination as the difference in residuals across his white and minority drivers.

$$D^{\text{early}}_j = \overline{\epsilon^{\text{white}}_{ij}} - \overline{\epsilon^{\text{min}}_{ij}}$$

We report in Table 9 the relationship between this early measure and our full-sample estimates of discrimination. We find that D^{early}_j has significant value for policy. Its correlation with our full measure β^j_3 is 0.45. The top panel reports how the percentiles of the two distributions correspond. Among the 2% of officers with the most discrimination in our early

[24]Our estimates of lenience and discrimination are both measured with error, leading to attenuation in the relationship between these measures and misconduct. To attempt to account for this error, we also do a split-sample instrumental variables procedure. We divide each officer's data randomly in half and estimate their bias and lenience for each sample. We then use one estimate as an instrument for the other. Doing so, we find the coefficients on discrimination increase overall in magnitude, though the standard errors remain too large to definitively say whether there is a true relationship.

measure, the median percentile in the full sample is 3.2. The 5% and 10% most discriminatory also mostly consist of officers who are discriminatory in the full sample. However, the 95th percentile of "early discrimination" officers are quite nondiscriminatory when calculated in the full sample. This fact implies that some officers who are discriminatory in their early ticketing grow out of this practice in later years.

This "mistake" in the early measure is confirmed in the bottom panel of Table 9, which reports Type-I and Type-II error in identifying career-wide discriminators. Among the 398 (25%) officers whose early measure indicates discrimination with 95% confidence, 32.2% are found not to be discriminatory at 5% significance in the full sample. Restricting attention to officers whose z-statistic in the early measure exceeds 3 (99.8% confidence) barely reduces Type-I error, to 28%. The stubbornness of this error suggests that the early measures are somewhat incorrect – not because of imprecision, but because officers change in their ticketing practice past their first year in policing. The Type-II error column indicates the share of officers who are found in the full sample to be discriminatory at the 5% level but were not detected in the early measure. This number is greater than 50% in all columns, suggesting that early detection can catch no more than half of discriminatory officers.

Taken together, these calculations suggest that our early measure can be useful for identifying officers for training as part of an early-warning system (Walker *et al.*, 2000). However, we caution against disciplining or removing officers on the basis of our early measures, as they often identify officers who are non-discriminatory in the totality of their careers. An early warning system is also not a panacea, as it fails to identify more than 50% of officers who will practice discrimination in their later careers.

8 Model and Counterfactuals

One of the central motivations of our paper is the need to understand how various personnel policies affect the aggregate disparity in treatment between whites and minorities. We have argued that the key input into the outcome of these policies is the distribution of discrimination across officers. To perform counterfactual analyses, however, we need to know both how driver speeds are generated and how officers then choose to discount these speeds. To do so, we present a simple model that allows us to simultaneously estimate officers' taste parameters for each racial group and speed parameters for each race-by-county. Doing so allows us to perform counterfactuals that change the distribution of discrimination across officers.

Individual i drives at a speed s that is drawn from a Poisson distribution $P_{\lambda_i}(s)$, where λ_i is a function of the county-by-race of the driver and other demographics $Z^{(1)}$:

$$\lambda_i = \lambda_{rc} + \gamma Z^{(1)}$$

where we include in $Z^{(1)}$ the driver's gender, age, and number of tickets in the previous three years. Within a county, officers and drivers match randomly with each other. If the driver is stopped for a speed s at or below the discount point x_d, the officer charges s. If $s > x_d$, the officer has the choice to discount the driver to x_d. He makes this decision by weighing a cost to discounting, which we impose to have the form $c(s) = b \cdot s$, against the "value" of discounting, $t_{ij} = t_{rj} + \alpha Z_i^{(2)} + \epsilon_{ij}$, where t_{rj} depends on the officer identity and driver race, $Z^{(2)}$ are driver demographics, and ϵ_{ij} is a standard normal random variable reflecting differences in preference not captured by driver demographics. Thus, the driver has her speed reduced to x_d if

$$t_{rj} + \alpha Z_i^{(2)} + \epsilon_{ij} > a + b \cdot s_i$$

In addition to the $Z^{(1)}$ demographics, $Z^{(2)}$ includes the share of drivers in a county who are minorities. We include this share to account for the possibility that officers change their behavior depending on the racial mix of the county's drivers.

Two simplifications of the model should be discussed here. First, we do not allow the driver's distribution of speeds to respond to the lenience of the officers in their county. We are comfortable in making this restriction because we find that there is no cross-sectional relationship between the county lenience rate and the speeds charged.[25]

Second, we provide no micro-foundation for an officers' decision to discount a driver. In Appendix Section C, we provide a series of tests for identifying what the officer is maximizing.

[25]In Goncalves and Mello (2017), we find that drivers do respond ex-post to receiving a harsh ticket by speeding less. This should lead to a steady-state relationship between lenience and the frequency of traffic tickets. However, the magnitude of the deterrence effect is small enough that the racial gaps in the counterfactuals would not be meaningfully impacted. For example, in the 11 years of our sample, if all minority drivers were treated as white drivers, there would only be about 70 more car accidents and fewer than one more death.

However, for the purposes of conducting the counterfactuals, it suffices to identify differences across officers in their propensity to discount.

8.1 Identification

In principle, our model can be identified using only aggregate information, as if all data came from one officer and one county. Intuitively, the tickets provide 40 moments (for each potential speed) to estimate three parameters (discount slope, preference for discounting, and true speed). Such an estimation approach relies heavily on the functional form assumptions of a Poisson speed distribution.

In practice, our estimation is similar to our difference-in-differences regressions, in that it relies heavily on the heterogeneity across officers in discount lenience. While all officers' data enter the maximum likelihood equations, the speed parameters are primarily identified using officers who exhibit no lenience, from which we get an estimate of the true distribution of speeds. To do so, we strongly rely on the assumption that officers and drivers are randomly sorting within a county, allowing us to suppose that the underlying distribution of speeds are the same for non-lenient and lenient officers.

Our estimation also depends heavily on the smoothness and parameterization of the underlying speed distribution. Any excess mass at the bunch point is taken to be lenience on the part of the officer. As argued earlier, we believe this assumption is valid, and drivers are not systematically choosing to bunch below the fine increase.

We estimate the model via maximum likelihood. The model parameters to be identified are the 67×2 county-race speeds λ_{rc}; 3 demographic speed parameters γ; 1592×2 officer average racial preferences, t_{rj}; 4 demographic preference parameters α; and the slope of the cost function b, totaling 3,326 parameters. Details of how the estimation is carried out in practice are provided in Appendix Section D.

8.2 Model Estimates

The results of the model estimation are reported in Appendix Table A.8. Because the estimates are closely aligned to the findings from our difference-in-differences approach, we leave our full discussion of these estimates to Appendix Section D.1. In short, we find that the average officer practices substantial lenience, with a significant variance across officers.

Off a baseline of 35.7% likelihood of discounting a driver from 10 MPH to 9 MPH, the average officer is 2pp less likely to discount minority drivers. We find that minorities drive significantly faster than white drivers, as do males, younger drivers, and drivers with previous tickets. The average officer is also more generous to female drivers, old drivers, and drivers with fewer previous tickets. They are also less lenient to all drivers when ticketing in a county with more minorities.

8.2.1 Decomposing the Gap in Discounting

A first-order question in the study of discrimination is the extent to which an aggregate racial disparity can be explained by the measured amount of bias. Table 10 seeks to answer this question by decomposing the measured racial discounting disparity into discrimination by officers, sorting of officers across counties, and differential speeding by racial groups. We do so by simulating the model with different restrictions on the behavior and location of the officers. In each simulation, drivers are randomly re-assigned a new officer from their county and drawn a new speed s from their individual specific distribution $P_{\lambda_i}(s)$. If the driver's speed is above the discount point, the officer draws a preference shock ϵ and gives the driver a discount to 9 MPH over if $t_{ij} + \epsilon > b \cdot x$. Standard errors are calculated by iterating the simulation 100 times, as explained in Appendix Section D.2.

The "Baseline" row of Table 10 shows how the charged speeds of drivers appear in a simulation of the model that does not change any of the parameters of the model. All of the decompositions are benchmarked to this baseline. In the "No Discrimination" row, we remove discrimination by making each officer treat minority drivers like they treat white drivers. This restriction reduces the gap in discounting by 25%. In the "No Sorting" row, drivers and officers match randomly from throughout the entire state rather than the initial county. Here we find that 28% of the gap in speeding is removed, consistent with the earlier finding that officers tend to be more lenient overall in neighborhoods with fewer minorities. Removing both sorting and discrimination, the gap in speeding is reduced by 45%. The remaining gap is due exclusively to the fact that minorities are driving faster speeds. In the second panel of Table 10 we report the same decompositions, where the gap is conditional on the county of the stop. Removing the sorting of officers no longer has any effect, since that only leads to differences *across* counties. Further, notice that over 80% of the within-county disparity can be explained by discrimination, leaving only about 17% of the disparity to be explained by differences in speeding across races. In Appendix Table A.9, we perform these same calculations, where the outcome of interest is the average speed rather than share discounted.

8.3 Policy Counterfactuals

Reported in Table 11, we now use the estimates to conduct a series of policy counterfactuals to explore how best to curb discrimination in speeding tickets. The results of these counterfactuals are compared relative to a baseline simulation, reported in the first row, that retains the empirical pool of officers and their distribution across counties. As with Table 10, the calculation of standard errors is discussed in Appendix Section D.

8.3.1 Firing and Hiring

We first consider the most direct policy for mitigating the disparity in treatment: removing the most discriminatory officers. We take officers in the 95th percentile and above of discrimination and remove them from the pool of officers. This cutoff removes officers with a difference in discounting of 16 percentage points or greater between whites and minorities. For symmetry, we also remove officers who reverse discriminate by that amount (comprising only 0.4% of officers).

The statewide disparity in treatment barely changes in response to removing these officers, falling by less than 4%. The lack of effectiveness from this policy partly stems from the fact that discriminatory officers are on average very lenient. When they are removed, drivers are left to be stopped by officers who, while less discriminatory, are also less lenient overall. This fact can be seen by noting that the average discount rate goes down for both white and minority drivers.

The next counterfactuals we consider are increased hiring of minority and female officers. Given our earlier finding that minority and female officers exhibit lower levels of discrimination, we should expect that increasing their presence might lead to lower levels of aggregate bias. We calculate this counterfactual by re-simulating which officer each driver draws, taken from within his county, where the probability of drawing a minority or female officer is exogenously changed. Consistent with our intuition, the gap in probability of discount declines, though very modestly. Increasing the share of female officers from 8% to 18% of the force leads to a 7.5% reduction in the discount gap. An increase in minority officers from the empirical share of 35% to 45% reduces the gap by 13.5%.

Demographic policies have been suggested in the past as a possibility for systemically changing police behavior, particularly toward poor and minority communities. Donohue III and Levitt (2001) find that an increase in minority officers leads to an increase in arrests of white offenders, no effect on non-white offenders, and vice versa for an increase in white officers. Our results, though only counterfactuals, are consistent with their findings.

8.3.2 Resorting Officers

The final counterfactuals we consider are to reassign officers to specific areas based on their behavior and the share of minorities in each county. Officers are assigned to troops, which patrol 6-10 counties. Within the troops, officers regularly vary in which locations they patrol. It may be potentially feasible for a senior officer to, for example, change the assignment of officers such that minorities face less biased officers. The bottom two rows of Table 11 present the results of such a policy. Column (1) sorts officers within a troop such that the least biased officers are in counties with the most minorities. Column (2) sorts officers within a troop such that the most lenient officers are in counties with the most minorities.

Surprisingly, sorting officers to expose minorities to the least discriminatory has a very small effect on the treatment gap. The least biased officers are also not very lenient on average, dampening the impact of their equal discounting across races and reducing the gap in discounting by only 11%. Much more effective in reducing the gap in treatment is assigning the most *lenient* officers to minority counties. This policy reduces the treatment gap by 86%.

In short, the counterfactual analyses highlight the importance of absolute lenience as a consideration separate from discrimination. The policy aimed at exposing minorities to lenience is much more effective than removing overall bias through firing biased officers or hiring minority and female officers.

8.4 Caveats

Our simplified modeling framework and counterfactuals are meant to be suggestive of how the racial treatment gap might change when various personnel policies are considered. That being said, many caveats must be recognized. We are not taking a strong normative stance on the social welfare function, and the only outcome we consider is the statewide disparity in discounting. Other outcomes could be relevant to the policy makers's problem that we do not consider here.

For example, increasing lenience uniformly may lead to increased speeding, which we show to be the case in a separate study (Goncalves and Mello, 2017). Changing leniency standards may also lead officers to give drivers verbal warnings rather than a reduced charge. A full consideration of the welfare impact of the ensuing policies would likely consider additional outcomes, such as the speeding response to changes in enforcement (Gehrsitz, 2017; Goncalves and Mello, 2017; Chalfin and McCrary, 2017) and the tradeoff between the level and inequality in lenience.

One additional concern is that officers will change their lenience behavior in response to being reassigned counties. We address this concern in part by allowing officer behavior to vary by the share of drivers who are minorities, though it is important to note that officers may respond in other ways.

9 Conclusion

The large racial disparities in the criminal justice system have led many to claim discrimination as the root cause. We argue in this paper that identifying discrimination at the level of the individual criminal justice agent is crucial for understanding the best policy for mitigating the disparities in outcomes. We study speeding tickets and the choice of officers to discount drivers to a speed just below an onerous punishment.

By using a bunching estimator approach that allows for officer-by-race measures of lenience in tickets, we can explore the entire distribution of both lenience and discrimination on the part of officers. We find that 83% of the gap in discounting can be attributed to discrimination. The rest of the gap is due to underlying differences in driving speeds across races. Officers are very heterogeneous in their degree of discrimination, with 40% of members explaining the entirety of the aggregate discrimination.

We explore whether discrimination is predictable by regressing individual officers' bias on demographic and personnel characteristics. We find that officers tend to favor their own race, and female officers are less biased on average. Personnel information, such as failing an entry exam, receiving civilian complaints, and seeking a promotion, are not strongly informative about bias.

Using a model of driver speeding and officer decision-making, we confirm that while minorities drive faster on average, our officer-level estimates of bias are not confounded by differences in speeding across groups. We find that setting discrimination to zero across officers fails to remove the majority of the treatment gap, due to the fact that minorities tend to live in regions where officers are less lenient toward all drivers. Because of this fact, policies directed at reducing discrimination directly have only a modest effect on the treatment gap. Policies that instead target officers' lenience, by reassigning lenient officers to minority neighborhoods, are much more effective at reducing the aggregate treatment disparity. These counterfactuals highlight our central argument, that the impacts of various policy reforms will depend crucially on the *distribution across officers* in their degrees of discrimination.

References

Aaronson, D., Barrow, L. and Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, **25** (1), 95–135.

Abrams, D. S., Bertrand, M. and Mullainathan, S. (2012). Do judges vary in their treatment of race? *The Journal of Legal Studies*, **41** (2), 347–383.

Agan, A. Y. and Starr, S. B. (2016). Ban the box, criminal records, and statistical discrimination: A field experiment.

Aizer, A. and Doyle Jr, J. J. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, **130** (2), 759–803.

Anbarci, N. and Lee, J. (2014). Detecting Racial Bias in Speed Discounting: Evidence from Speeding Tickets in Boston. *International Review of Law and Economics*, **38**, 11–24.

Antonovics, K. and Knight, B. (2009). A New Look at Racial Profiling: Evidence from the Boston Police Department. *Review of Economics and Statistics*, **91** (1), 163–177.

Anwar, S., Bayer, P. and Hjalmarsson, R. (2012). The impact of jury race in criminal trials. *The Quarterly Journal of Economics*, **127** (2), 1017–1055.

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

Arnold, D., Dobbie, W. and Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics*, **133** (4), 1885–1932.

Arrow, K. (1973). The theory of discrimination. *Discrimination in labor markets*, **3** (10), 3–33.

Banks, R. R., Eberhardt, J. L. and Ross, L. (2006). Discrimination and implicit bias in a racially unequal society. *California Law Review*, **94** (4), 1169–1190.

Becker, G. S. (1957). *The Economics of Discrimination: an Economic View of Racial Discrimination*. University of Chicago.

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review*, **94** (4), 991–1013.

Best, M. C., Cloyne, J., Ilzetzki, E. and Kleven, H. J. (2015). Interest rates, debt and intertemporal allocation: Evidence from notched mortgage contracts in the united kingdom.

Björklund, A. and Moffitt, R. (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics*, pp. 42–49.

Brock, W. A., Cooley, J., Durlauf, S. N. and Navarro, S. (2012). On the observational implications of taste-based discrimination in racial profiling. *Journal of Econometrics*, **166** (1), 66–78.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J. and Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *The American Economic Review*, **106** (5), 124–127.

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

Chetty, R., Friedman, J. N., Olsen, T. and Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *The Quarterly Journal of Economics*, **126** (2), 749–804.

Coviello, D. and Persico, N. (2013). An Economic Analysis of Black-White Disparities in NYPD's Stop and Frisk Program. *NBER Working Paper*, pp. 1–28.

Dee, T., Dobbie, W., Jacob, B. and Rockoff, J. (2016). The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examination. *NBER Working Paper*, pp. 1–67.

Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *The Annals of Statistics*, pp. 665–685.

Diamond, R. and Persson, P. (2016). The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests. *SIEPR Discussion Paper*, **16-003**, 1–70.

Dobbie, W. and Song, J. (2015). Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *The American Economic Review*, **105** (3), 1272–1311.

Donohue, J. (2014). An Empirical Evaluation of the Connecticut Death Penalty System Since 1973: Are There Unlawful Racial, Gender, and Geographic Disparities? *Journal of Empirical Legal Studies*, **11** (4), 637–696.

Donohue III, J. J. and Levitt, S. D. (2001). The impact of race on policing and arrests. *The Journal of Law and Economics*, **44** (2), 367–394.

Doyle, J. (2007). Child protection and child outcomes: Measuring the effects of foster care. *The American Economic Review*, **97** (5), 1583–1610.

Doyle Jr, J. J. (2008). Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care. *Journal of Political Economy*, **116** (4), 746–770.

Edelman, B., Luca, M. and Svirsky, D. (2017). Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics*, **9** (2), 1–22.

Frandsen, B. R. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. In *Regression Discontinuity Designs: Theory and Applications*, Emerald Publishing Limited, pp. 281–315.

Fryer, R. G. (2018). An Empirical Analysis of Racial Differences in Police Use of Force. *NBER Working Paper*, **22399**, 1–63.

Gehrsitz, M. (2017). Speeding, punishment, and recidivism: Evidence from a regression discontinuity design. *The Journal of Law and Economics*, **60** (3), 497–528.

Goldin, C. and Rouse, C. (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review*, **90** (4), 715–741.

Goncalves, F. and Mello, S. (2017). Does the punishment fit the crime? speeding fines and recidivism. *Unpublished Manuscript*, pp. 1–48.

Greenwald, A. G. and Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, **94** (4), 945–967.

Grogger, J. and Ridgeway, G. (2006). Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness. *Journal of the American Statistical Association*, **101** (475), 878–887.

Heckman, J. J. (1979). Sample selection as a specification error. *Econometrica*, **47** (1), 153–161.

—, Schmierer, D. and Urzua, S. (2010). Testing the correlated random coefficient model. *Journal of Econometrics*, **158** (2), 177–203.

Horrace, W. C. and Rohlin, S. M. (2016). How Dark Is Dark? Bright Lights, Big City, Racial Profiling. *Review of Economics and Statistics*, **98** (2), 226–232.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, **8**, 435–464.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *The American economic review*, **96** (3), 863–876.

Knowles, J., Persico, N. and Todd, P. (2001). Racial Bias in Motor Vehicle Searches: Theory and Evidence. *Journal of Political Economy*, **109** (1), 203–229.

Lange, J. E., Johnson, M. B. and Voas, R. B. (2005). Testing the racial profiling hypothesis for seemingly disparate traffic stops on the new jersey turnpike. *Justice Quarterly*, **22** (2), 193–223.

Mueller-Smith, M. (2014). The criminal and labor market impacts of incarceration. *Unpublished Working Paper.*

Najdowski, C. J. (2011). Stereotype threat in criminal interrogations: Why innocent black suspects are at risk for confessing falsely. *Psychology, Public Policy, and Law*, **17** (4), 562.

—, Bottoms, B. L. and Goff, P. A. (2015). Stereotype threat and racial differences in citizens' experiences of police encounters. *Law and Human Behavior*, **39** (5), 463.

Neal, D. A. and Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *Journal of political Economy*, **104** (5), 869–895.

Persico, N. (2009). Racial Profiling? Detecting Bias Using Statistical Evidence. *Annual Review of Economics*, **1** (1), 229–254.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review*, **62** (4), 659–661.

Price, J. and Wolfers, J. (2010). Racial Discrimination Among NBA Referees. *Quarterly Journal of Economics*, **125** (4), 1859–1887.

Rehavi, M. M. and Starr, S. B. (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy*, **122** (6), 1320–1354.

Ridgeway, G. and MacDonald, J. (2010). Methods for Assessing Racially Biased Policing. In S. Rice and M. White (eds.), *Race, Ethnicity, and Policing New and Essential Readings*, pp. 180–204.

— and MacDonald, J. M. (2009). Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. *Journal of the American Statistical Association*, **104** (486), 661–668.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, **2** (3), 180–212.

Smith, W., Tomaskovic-Devey, D., Zingraff, M., Mason, H. M., Warren, P. and Wright, C. (2004). The North Carolina Highway Traffic Study. *NCJRS Grant Report*, pp. 1–407.

Trinkner, R. and Goff, P. A. (2016). The color of safety: The psychology of race & policing. *The SAGE Handbook of Global Policing*, pp. 61–81.

Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., Jurgens, D., Jurafsky, D. and Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences*, p. 201702413.

Walker, S., Alpert, G. P. and Kenney, D. J. (2000). Early warning systems for police: Concept, history, and issues. *Police Quarterly*, **3** (2), 132–152.

West, J. (2018). Racial Bias in Police Investigations. *Working Paper*, pp. 1–36.

Table 1: Summary Statistics

	(1) White	(2) Black	(3) Hispanic	(4) Total
Driver Female	0.362	0.402	0.305	0.356
	(0.481)	(0.490)	(0.461)	(0.479)
Age	37.256	34.228	34.267	36.006
	(14.850)	(12.139)	(11.992)	(13.838)
Florida License	0.825	0.851	0.896	0.846
	(0.380)	(0.356)	(0.306)	(0.361)
Zip Code Income	52.819	37.772	44.375	48.096
	(51.675)	(29.912)	(41.444)	(46.464)
Citations in Past Year	0.288	0.427	0.408	0.341
	(0.721)	(0.909)	(0.877)	(0.799)
MPH Over	15.560	16.658	18.334	16.404
	(6.524)	(7.033)	(6.988)	(6.825)
Discount	0.343	0.314	0.204	0.306
	(0.475)	(0.464)	(0.403)	(0.461)
Fine Amount	182.060	187.999	197.436	186.636
	(76.130)	(80.366)	(80.401)	(78.154)
Share	0.584	0.184	0.231	1
N	667086	210272	264270	1141628

Notes: Standard deviations in parentheses. Zip code income is missing for 42% of White stops, 40% of Black stops, 37% of Hispanic stops. To account for the fact that a large share of fine amounts are missing or zero in our data, we impute the fine amount with the modal non-zero fine for each county $\times$ speed over the limit cell.

Table 2: Characteristics of Cited Drivers Relative to Other Data Sources

	(1) Citations	(2) ACS - Any	(3) ACS - Drivers	(4) Crash - Any	(5) Crash - Injury
Female	0.359	0.514	0.474	0.424	0.441
Age	34.979	47.667	41.755	39.692	39.812
White	0.588	0.649	0.632	0.569	0.588
Black	0.177	0.143	0.142	0.193	0.197
Hispanic	0.234	0.208	0.226	0.238	0.216

Notes: ACS data are from 2006-2010 include individuals aged 16 or older and sampling weights are used. To keep the data from the same years, we restrict attention to citations and crashes for the years 2006-2010.

Table 3: Officer Lenience Randomization Check

	Full Sample			GPS Sample	
	(1)	(2)	(3)	(4)	(5)
	Lenient	Lenient	Lenient	Lenient	Lenient
Driver Black	-0.000928	0.00211	-0.00215	-0.00464	0.00112
	(0.0257)	(0.00435)	(0.00423)	(0.00637)	(0.00465)
Driver Hispanic	-0.120	0.0000806	-0.00116	-0.0181	-0.00291
	(0.0469)	(0.00503)	(0.00487)	(0.0141)	(0.00470)
Driver Female	0.0302	0.00456	0.00317	0.00343	0.00137
	(0.00736)	(0.00268)	(0.00228)	(0.00249)	(0.00309)
Florida License	-0.131	0.00143	0.000531	0.00826	-0.00587
	(0.0402)	(0.00349)	(0.00386)	(0.00824)	(0.00438)
Driver Age	-0.446	0.0483	-0.0682	0.0584	-0.133
	(0.273)	(0.153)	(0.146)	(0.105)	(0.0988)
1 Prior Ticket	-0.0129	0.000478	-0.000193	0.00122	0.00396
	(0.0101)	(0.00125)	(0.000975)	(0.00311)	(0.00341)
2+ Prior Tickets	-0.0343	0.000607	0.00181	0.000902	-0.00183
	(0.0214)	(0.00193)	(0.00171)	(0.00237)	(0.00402)
Log Zip Code Income	-0.0113	0.00630	-0.00279	-0.000713	0.00184
	(0.0140)	(0.00344)	(0.00238)	(0.00273)	(0.00365)
F-test	0	.359	.144	.564	.324
Mean	.305	.305	.304	.323	.326
Location FE		X			
Location + Time FE			X	X	
GPS FE					X
Observations	1141628	1141628	1079250	125040	135553

Notes: All regressions includes vehicle type fixed effects and county fixed effects. The F-test reports the joint hypothesis test that variables Driver Black through Log Zip Code Income are zero. Standard errors are clustered at the county level. "Location FE" includes county by highway fixed effects. "Location + Time FE" includes county by highway by year by month by day of the week by shift fixed effects. "GPS FE" includes road segment by year by month by day of the week by shift fixed effects.

Table 4: Difference-in-Difference Results

	Full Sample			GPS Sample	
	(1)	(2)	(3)	(4)	(5)
	Discount	Discount	Discount	Discount	Discount
Driver White	0.00126	-0.0205	-0.0119	-0.00766	-0.00650
	(0.000326)	(0.00626)	(0.00610)	(0.00436)	(0.00370)
Officer Lenient	0.396	0.304	0.297	0.243	0.199
	(0.0355)	(0.0377)	(0.0453)	(0.0192)	(0.0321)
Driver White	0.0840	0.0671	0.0620	0.0683	0.0549
× Officer Lenient	(0.0167)	(0.0111)	(0.0105)	(0.00841)	(0.00672)
Mean	.305	.305	.305	.32	.32
Covariates		X	X	X	X
Location FE		X			
Location + Time FE			X	X	
GPS FE					X
Observations	1141628	1141628	1079250	125040	125040

Notes: Table reports linear probability estimates where the outcome variable is whether an individual is ticketed for 9 MPH over the limit, as in Equation (3). Standard errors are clustered at the county level. "GPS FE" includes road segment by year by month by day of the week by shift fixed effects.

Table 5: Alternative Difference-in-Differences Specifications

	(1) Discount	(2) Discount	(3) Discount	(4) Discount	(5) Discount	(6) Discount	(7) Discount
Driver White	-0.0114	-0.0123	-0.00872	-0.0212	-0.00286	-0.00411	0.0116
	(0.00551)	(0.00615)	(0.00496)	(0.00863)	(0.00125)	(0.00620)	(0.00656)
Officer Lenient	0.300	0.279	0.357	0.216	0.318	0.185	0.147
	(0.0371)	(0.0387)	(0.0340)	(0.0427)	(0.0359)	(0.0341)	(0.0289)
Driver White	0.0694	0.0699	0.0696	0.0763	0.0552	0.0589	0.0563
× Officer Lenient	(0.0102)	(0.0105)	(0.0104)	(0.0109)	(0.00622)	(0.0108)	(0.0108)
Specification	Baseline	Split-Sample	Lenience by Year	Frandsen (2017) Test	Re-weighted	Lenience-Cov. Interaction	Radar Gun Sample
Difference		0	0	.006	-.015	-.011	-.014
		(.014)	(.014)	(.014)	(.011)	(.014)	(.014)
Mean	.31	.31	.31	.31	.311	.31	.381
R2	.344	.337	.379	.317	.339	.346	.328
N	1124513	898956	1124513	1124513	1116461	1124513	100705

Notes: All regressions include vehicle type fixed effects and fixed effects for county-year-month. Standard errors are clustered at the county level. The baseline specification is the same regression as Column (3) from Table 4. Column (2) reports a regression where a random sample of 20% of the data is used to estimate whether an officer is lenient, and the remaining 80% is used in the regression. Column (3) allows officer lenient/non-lenient to vary by year. Column (4) identifies officer lenient/non-lenient using the test from Frandsen (2017) for manipulation of a discrete running variable. Column (5) reweights the observations so that the relative weight given to minority drivers is equalized across county-year-month. Column (6) interacts officer lenient/non-lenient with all observable demographics of drivers. Column (7) restricts attention to a sample of tickets where the officer reports that he/she used a radar gun to identify the driver's speed.

Table 6: Alternative Interpretations

	Section 6.1		Section 6.2		
	(1)	(2)	(3)	(4)	(5)
	Discount	Discount	Discount	Discount	Discount
White	-0.0137	-0.0135	0.0203	0.0189	
	(0.00501)	(0.00522)	(0.00469)	(0.00343)	
Lenient	0.291	0.289			
	(0.0388)	(0.0398)			
Driver White	0.0651	0.0657			
× Officer Lenient	(0.00884)	(0.00874)			
Heckman Correction		0.0172			
		(0.0196)			
Individual Demographics	X	X	X	X	X
Location + Time FE	X	X	X	X	X
Officer FE				X	X
Individual FE					X
Mean	.31	.31	.288	.284	.278
R2	.377	.377	.318	.527	.542
N	1124513	1124513	189629	181769	172810

Notes: Column (1) reports the same regression as the baseline regression in Table 4. Column (2) reports the same regression with the addition of the Heckman Correction term, as explained in Section 6.1 and Appendix Section B. Columns (3)-(5) correspond to Section 6.2. Column (3) restricts attention to drivers with two or more tickets and regresses discounting on individual demographics and county-year-month fixed effects. Column (4) reports the same regression with the addition of officer fixed effects. Column (5) additionally includes driver-level fixed effects.

Table 7: Alternative Interpretations, Section 6.3

	(1) Recidivism	(2) Recidivism	(3) Court	(4) Court	(5) Court
P(Discount)	0.00683	0.00887	-0.120	-0.153	-0.110
	(0.00364)	(0.00582)	(0.0221)	(0.0232)	(0.0219)
Driver Minority	0.00465	0.00465	0.0451	0.0451	0.0519
	(0.00218)	(0.00218)	(0.00393)	(0.00392)	(0.00690)
P(Discount)^2		-0.00245		0.0403	
		(0.00764)		(0.0171)	
P(discount) $\times$ Driver Minority					-0.0235
					(0.0113)
Location + Time FE	X	X	X	X	X
Mean	.104	.104	.379	.379	.379
R2	.024	.024	.376	.376	.376
N	844422	844422	844422	844422	844422

Notes: Columns (1)-(2) use as outcome whether an individual receives another speeding ticket in Florida in the following year. Column (1) regresses recidivism on driver demographics and the propensity score for receiving a discount. The propensity score uses driver demographics and an instrument for officer lenience interacted with driver race, as explained in Appendix Section C. Column (2) additionally includes a quadratic term for the propensity score. Columns (3) and (4) are analogous to Columns (1) and (2), where the outcome is whether the driver contests the ticket in court. Column (5) regresses court contestation on propensity score, where propensity score is also interacted with driver race. For all regressions, we restrict attention to in-state drivers with a ticket at 9 MPH or over for whom we have a court record of whether the driver contested.

Table 8: Predicting Officer Bias

	(1) White Lenience	(2) Black Bias	(3) Hispanic Bias
Black Officer	-0.034	-0.017	-0.023
	(0.020)	(0.003)	(0.004)
Hispanic Officer	-0.042	-0.007	-0.016
	(0.018)	(0.004)	(0.004)
Other Race	0.004	0.014	0.001
	(0.047)	(0.013)	(0.011)
Female Officer	-0.050	-0.009	-0.006
	(0.025)	(0.003)	(0.004)
Age (/10)	0.010	0.001	0.002
	(0.011)	(0.002)	(0.002)
Experience (/10)	0.154	0.010	0.015
	(0.047)	(0.008)	(0.009)
Failed Entrance Exam	0.029	-0.003	0.001
	(0.024)	(0.003)	(0.004)
Any College	-0.014	-0.001	-0.001
	(0.016)	(0.003)	(0.003)
Number of Complaints	-0.010	-0.001	-0.000
	(0.004)	(0.000)	(0.000)
Use of Force Incidents	-0.006	-0.000	0.000
	(0.006)	(0.001)	(0.001)
Mean	.289	.03	.043
Observations	1,402	1,402	1,402
R2	.316	.127	.129

Notes: Robust standard errors in parentheses. Outcomes are derived from the regression $S^9_{ij} = \beta_0 + \beta_1 \cdot \text{Black}_i + \beta_2 \cdot \text{Hispanic}_i + \beta^j_3 \cdot \text{Lenient}_j + \beta^j_B \cdot \text{Black}_i \cdot \text{Lenient}_j + \beta^j_H \cdot \text{Hispanic}_i \cdot \text{Lenient}_j + X_{ij}\gamma + \epsilon_{ij}$. White Lenience is calculated as $\beta_0 + \beta^j_3\text{Lenient}_j$. Black Bias and Hispanic Bias are calculated as $\beta^j_B \cdot \text{Lenient}_j$ and $\beta^j_H \cdot \text{Lenient}_j$, respectively. The sample of officers is reduced from 1591 to 1402 because of the restriction that each officer stop both black and Hispanic drivers.

Table 9: Early Discrimination

Early Measure Cutoff	Full Sample Percentiles		
	(1) N	(2) Median	(3) 95th percentile
2% most discriminatory	28	3.2	23.6
5% most discriminatory	76	6.5	60.2
10% most discriminatory	153	9.2	82.6

Early Sample	Full Sample		
	(1) N	(2) Type I Error	(3) Type II Error
$\hat{\theta}_j > 1.96 \cdot \text{SE}(\hat{\theta}_j)$	398	32.2%	54.6%
$\hat{\theta}_j > 2.33 \cdot \text{SE}(\hat{\theta}_j)$	329	31.0%	61.8%
$\hat{\theta}_j > 3 \cdot \text{SE}(\hat{\theta}_j)$	236	28.0%	71.4%

Notes: This table presents the relationship between early measures of discrimination (using first 100 tickets) and discrimination using all an officer's data. The first panel reports how different cutoffs in the percentile of early discrimination translate to percentiles in the full sample. For example, the median percentile of full-sample discrimination for an officer who is in the top 2% of early discrimination is 3.2. The 95th percentile among those from the early 2% cutoff is 23.6. The bottom panel reports how often the early measures mislabels an officer as discriminatory and how often it misses a discriminatory officer. Type-I Error reports the percentage of officers identified as discriminatory in the early sample who are *not* discriminatory at the 5% level in the full sample. Type-II Error reports the percentage of officers who are discriminatory in the full sample at the 5% level who are not identified as discriminatory in the early sample.

Table 10: Discounting Gap Decomposition

	State-Wide Disparity			
	(1) White Mean (MPH)	(2) Minority Mean	(3) Difference	(4) Percent
Baseline	0.347	0.266	-0.081	100
	(0.001)	(0.001)	(0.001)	
No Discrimination	0.347	0.286	-0.061	75.553
	(0.001)	(0.001)	(0.001)	(0.010)
No Sorting	0.327	0.269	-0.059	72.045
	(0.001)	(0.001)	(0.001)	(0.014)
Neither	0.327	0.291	-0.037	45.016
	(0.001)	(0.001)	(0.001)	(0.012)

	County-Level Disparity			
	(1) White Mean (MPH)	(2) Minority Mean	(3) Difference	(4) Percent
Baseline	0.347	0.321	-0.027	100
	(0.001)	(0.001)	(0.001)	
No Discrimination	0.347	0.343	-0.005	17.903
	(0.001)	(0.001)	(0.001)	(0.033)
No Sorting	0.327	0.300	-0.027	100.888
	(0.001)	(0.001)	(0.001)	(0.043)
Neither	0.327	0.322	-0.005	17.656
	(0.001)	(0.001)	(0.001)	(0.035)

Notes: Table presents how the racial gap in discounting and changes without bias and sorting of officers across counties. The probability gap is the probability of being discounted if you are at the speed right above the jump in fine. Both gaps are the minority drivers' outcome minus white drivers' outcome. No bias is calculated by assigning each officer's preferences toward minorities to be the same as his preference to whites. No sorting is calculated by simulating a new draw of officers for each driver, where the draw is done at the state level. The county-level disparities reweight the minority observations so that the "share" minority is identical across counties.

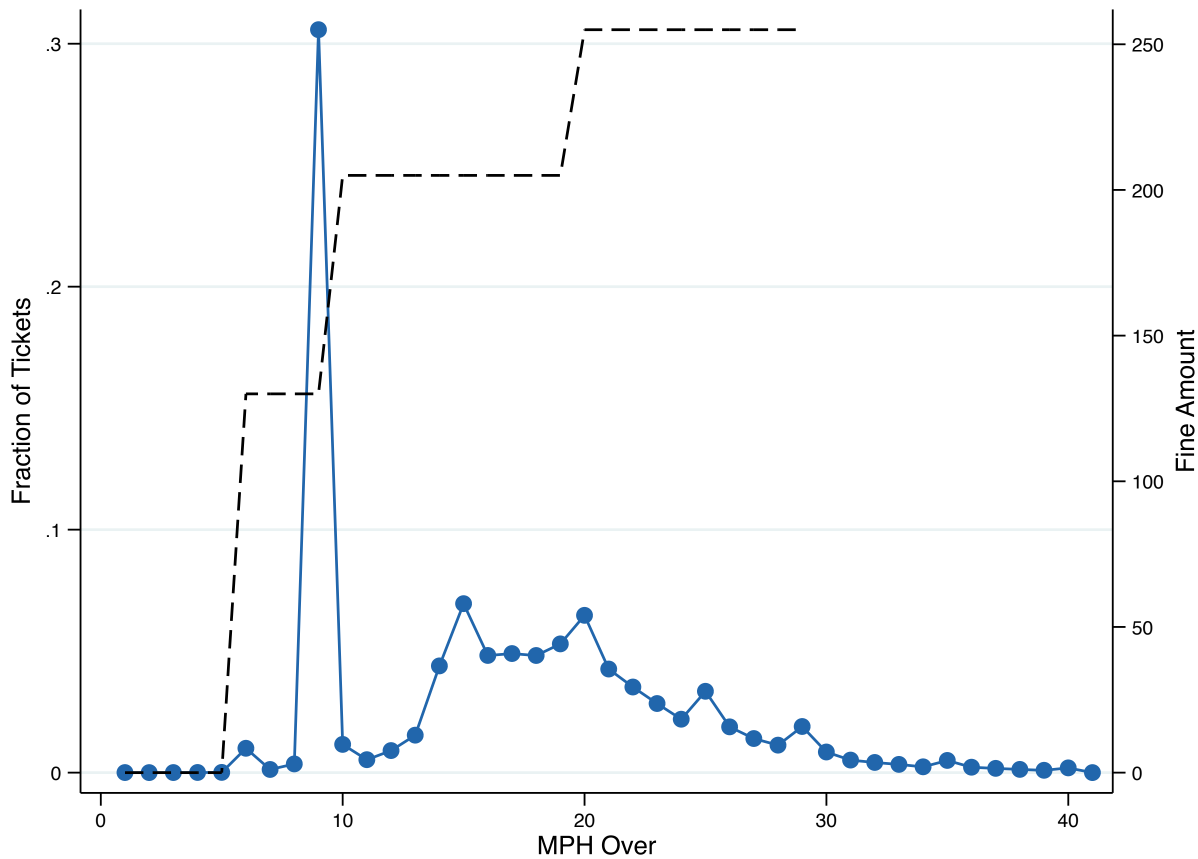
Table 11: Model Counterfactuals

	Hiring & Firing			
	(1) White Mean	(2) Minority Mean	(3) Difference	(4) Percent
Baseline	0.3473	0.2661	-0.0813	100
	(0.0007)	(0.0007)	(0.0010)	
Fire 5% Most Biased Officers	0.3440	0.2655	-0.0785	96.5733
	(0.0013)	(0.0012)	(0.0012)	(0.0154)
Increase Female Share 10pp	0.3423	0.2671	-0.0752	92.4890
(Base of 8%)	(0.0007)	(0.0008)	(0.0011)	(0.0131)
Increase Minority Share 10pp	0.3057	0.2354	-0.0703	86.5128
(Base of 35%)	(0.0016)	(0.0018)	(0.0024)	(0.0293)

	Resorting Officers			
	(1) White Mean	(2) Minority Mean	(3) Difference	(4) Percent
Exposing Minorities	0.3327	0.2602	-0.0725	89.1587
To *Least Biased*	(0.0022)	(0.0018)	(0.0017)	(0.0212)
Exposing Minorities	0.2989	0.2879	-0.0110	13.5403
To *Most Lenient*	(0.0008)	(0.0010)	(0.0011)	(0.0129)

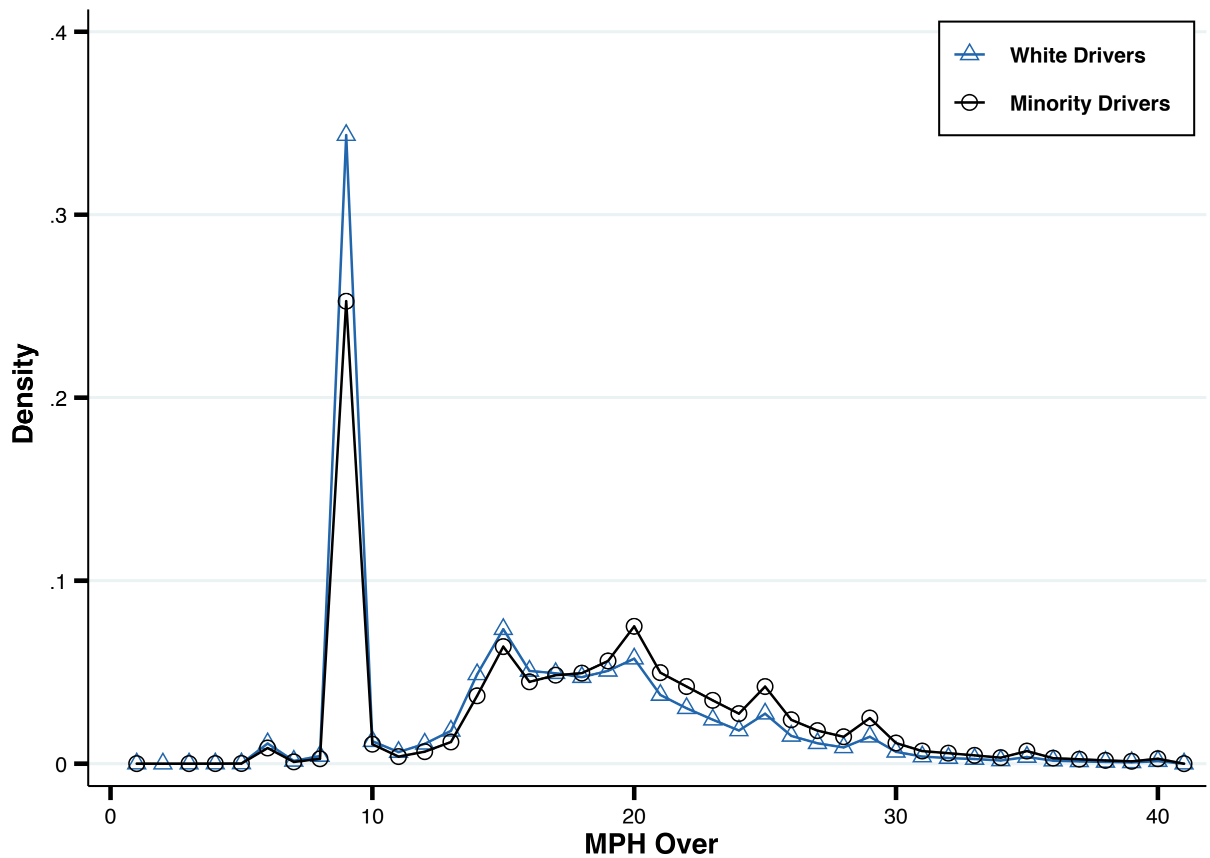
Notes: Results are reporting the probability of being ticketed 9MPH over, where the averages are statewide. In the bottom panel of counterfactuals, officers are resorted *within* troops.

Figure 1: Distribution of Charged Speeds and Fine Schedule



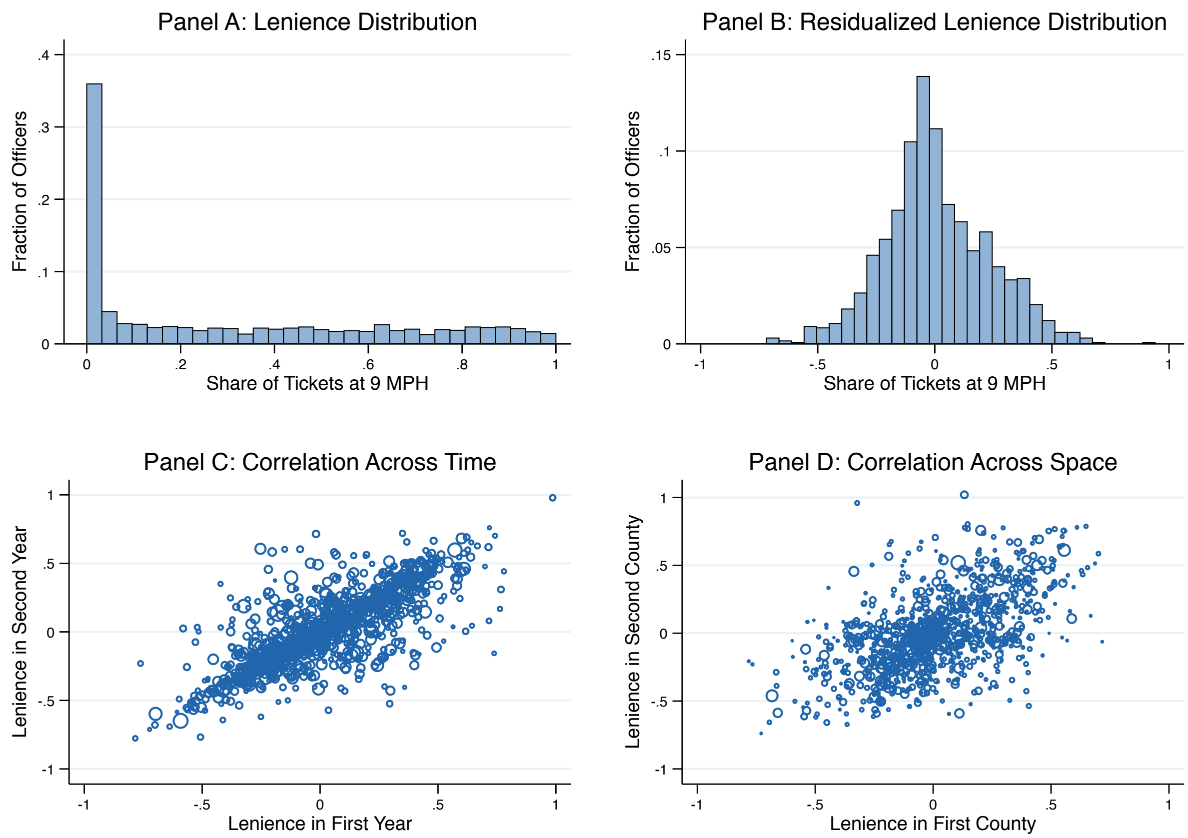
Notes: Connected line shows histogram of tickets. Dashed line plots fine schedule for Broward County. 30 MPH over is felony speeding and carries a fine to be determined following a court appearance.

Figure 2: Charged Speed Distributions by Driver Race



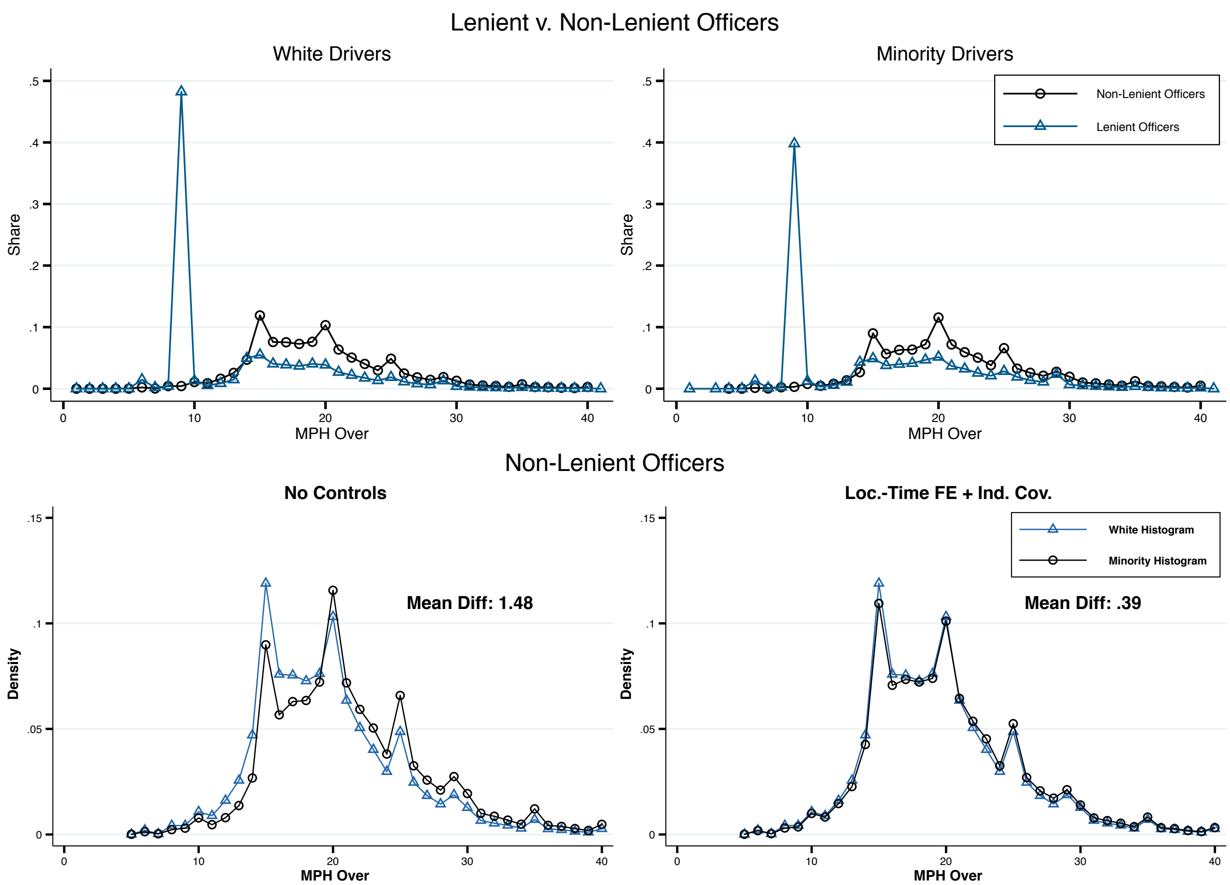
Notes: Connected line shows histogram of ticketed speeds, separately by driver race. 34.3% of tickets to white drivers are given at 9 MPH over compared to 25.2% of tickets for minority drivers.

Figure 3: Evidence of Officer Lenience



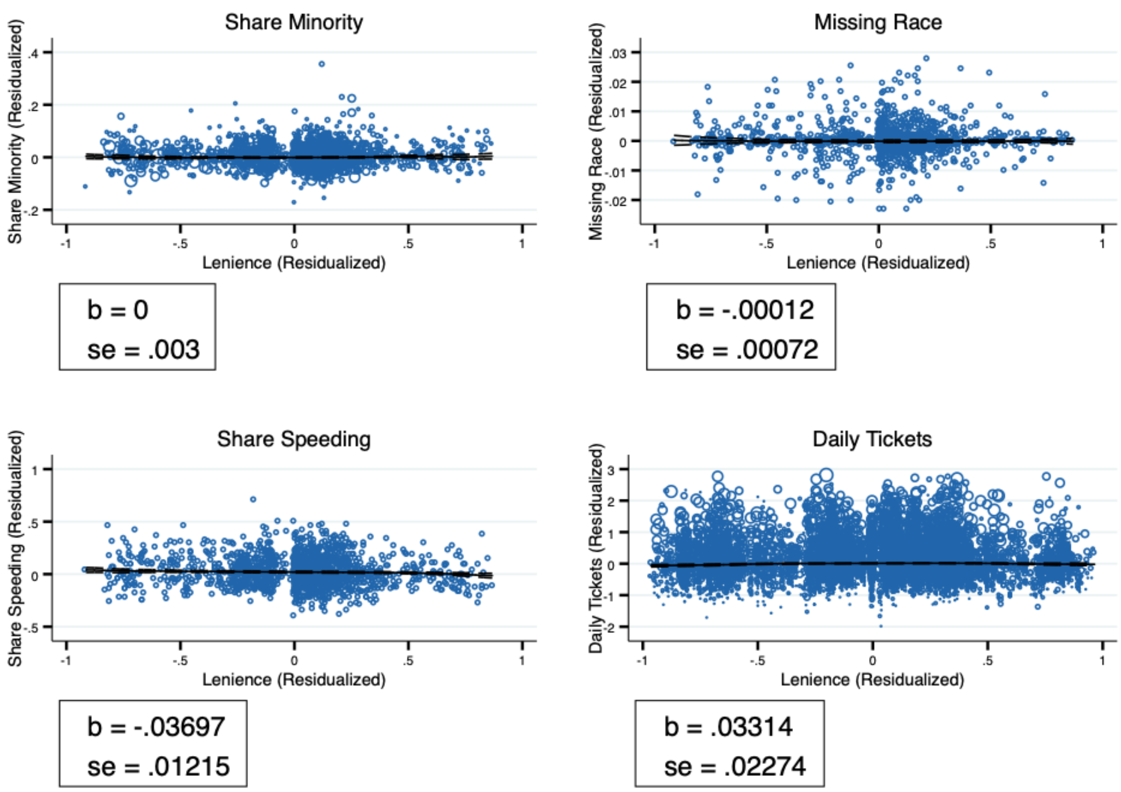
Notes: Panel A plots the across-officer distribution of lenience, calculated as the share of tickets given for 9 MPH over the limit. Panel B plots the across-officer distribution of residualized lenience. Panel C plots officers' residualized lenience in the years with the most and second most citations. Panel D plots the residualized lenience in the county with the most and second most citations. Estimates residualized by conditioning on county fixed effects, speed zone fixed effects, year and month fixed effects, and day of week fixed effects.

Figure 4: Difference-in-Difference Raw Data Plot



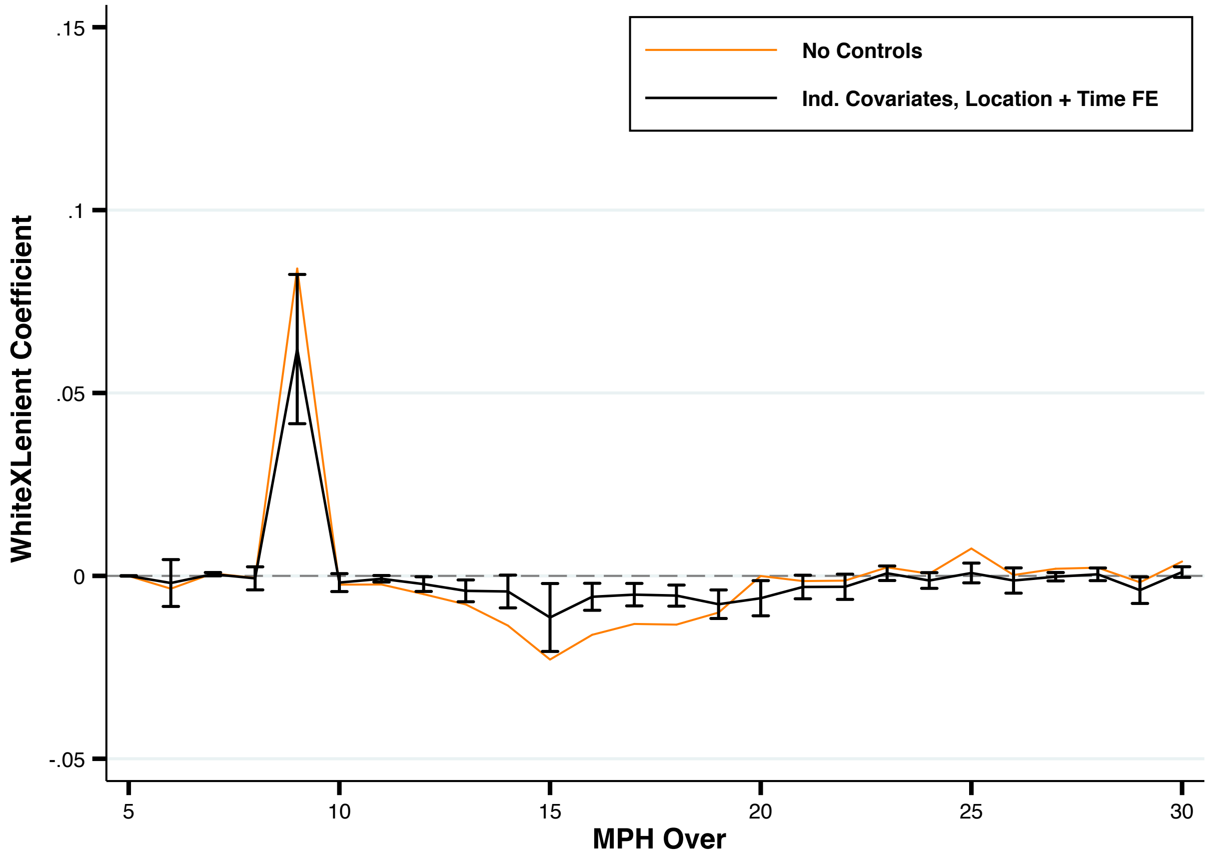
Notes: The top left figure plots the histograms of speeds for white drivers, separately for stops made by lenient and non-lenient officers. The top right figure plots the same histograms of speeds for minority drivers, separately by officer lenience. The bottom left figure plots the histograms for speeds ticketed by non-lenient officers, separately for white and minority drivers. The bottom right figure plots the histograms of speeds ticketed by non-lenient officers separately by race, where we have controlled for other demographics and county-year-month fixed effects. Specifically, for each speed, we regress whether an individual is ticketed at that speed, controlling for minority driver and all other demographics and county-year-month fixed effects. The white histogram is the same as the bottom left figure, and the minority histogram is the white histogram with the addition of minority regression coefficient for each speed.

Figure 5: Officer Lenience and Stop Characteristics



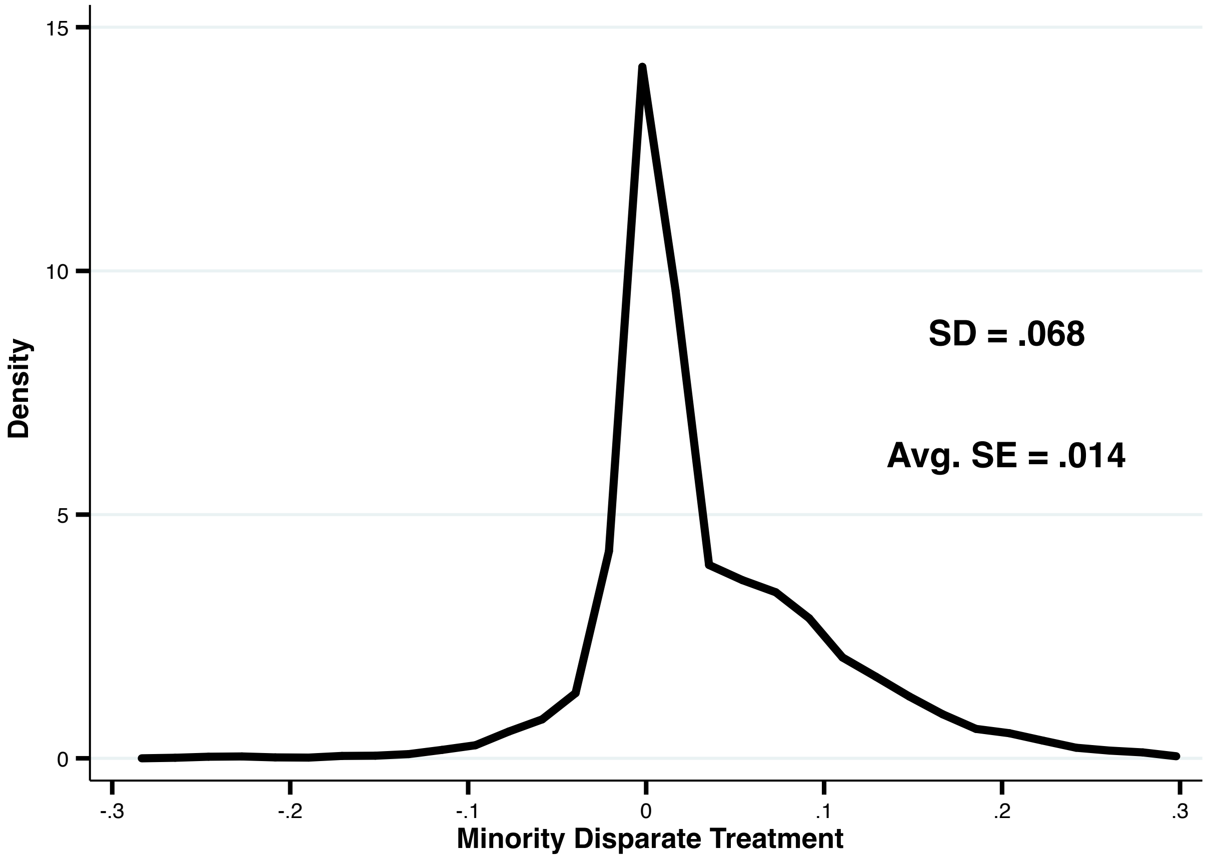
Notes: Figure plots the relationship between officer lenience and various characteristics of the officers' stops, where both officer's lenience and the stop characteristic have been residualized to remove location-time fixed effects. By officer lenience here we mean the indicator for whether an officer has more than 2% of tickets charge at 9mph over. The top left panel plots officer lenience against his share of tickets given to minority drivers, the top right the share of tickets with race missing, and the bottom left the share of tickets that are for speeding. For the bottom right figure, we calculate the number of daily tickets for each officer-by-year, and similarly calculate whether an officer is lenient in each year. We residualize both with county-by-year fixed effects.

Figure 6: Difference-in-Difference Results



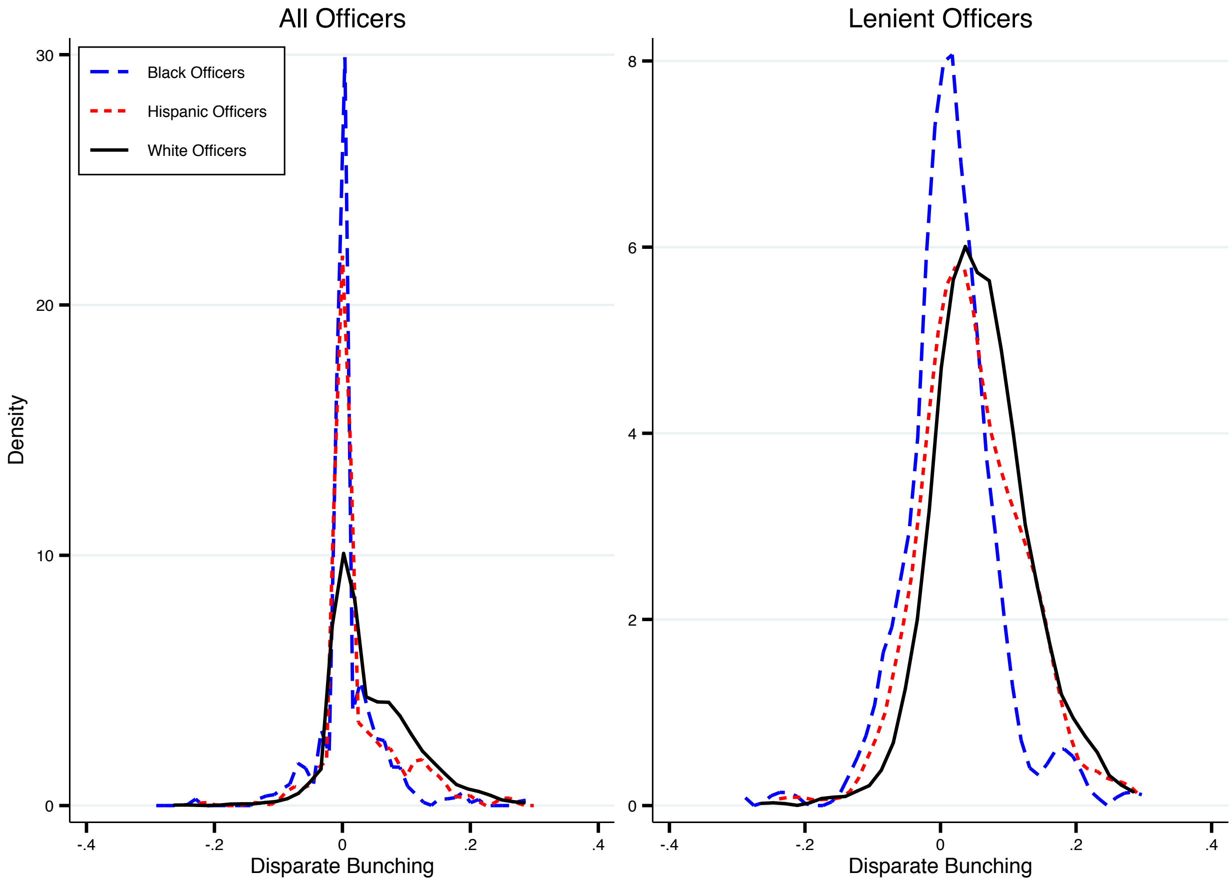
Notes: Figure plots the difference-in-difference regression results for each speed. The y-axis plots the interaction between driver being white and the officer being lenient. Standard errors are at the 5% level.

Figure 7: Difference-in-Differences Officer-Level Results



Notes: Figure plots each officer's β_3^j from the regression
$S_{ij}^9 = \beta_0 + \beta_1 \cdot \text{White}_i + \beta_2^j \cdot \text{Lenient}_j + \beta_3^j \cdot \text{White}_i \cdot \text{Lenient}_j + X_{ij}\gamma + \epsilon_{ij}$.
Officers who are non-lenient are assigned $\beta_3^j = 0$. SD reports the standard deviation across β_3^j, and Avg SE. reports the average standard error for each individual β_3^j.

Figure 8: Officer-Level Results



Notes: Left figure plots the discrimination coefficient β_3^j for all officers. Right figure plots the discrimination coefficient for all lenient officers.

Appendix: For Online Publication Only

A Data Appendix

A.1 Citations Data

Our data cover the universe of citations written by the Florida Highway Patrol for the years 2005-2015, comprising 2,614,119 observations. We make several restrictions that reduce the number of observations:

1. speeding is the primary citation (2,124,692 observations)
2. no crash is involved (2,123,311 observations, 99.9% of previous sample)
3. speed is between 0 and 40 over the limit (2,109,258, 99.3%)
4. posted speed limit is between 25MPH and 75MPH (2,107,933, 99.9%)
5. citations not from an airplane (2,103,923, 99.8%)
6. race/ethnicity is not missing (1,759,257, 83.6%)
7. race/ethnicity is white, black or Hispanic (1,671,089, 95.0%)
8. not missing driver's license state, gender, or age (1,667,558, 99.8%)
9. officer is identifiable (1,215,588, 72.9%)
10. officer has at least 100 tickets, and at least 20 for minorities and 20 for whites (1,174,284, 96.6%)
11. driver has no more than 20 citations in Florida for period 2005-2015 (1,141,628, 97.2%)

A.2 Linking Offenses to Personnel Information

Officers enter their information by hand onto each speeding ticket, leading to inconsistencies in how their names are recorded. Some names are misspelled, and sometimes officers place only their last name and first initial. The Florida Department of Law Enforcement (FDLE) maintains a record of each certified officer in the state, along with demographic information. We link these using each officer's last name and first three letters of first name (if available on ticket) using a fuzzy match algorithm in Stata (reclink). We restrict attention to officers who are unique up to last name and first three letters of first name in the FDLE data. Among tickets where only the first initial is listed, we keep matches where the last name and first initial of an officer are unique in the FDLE data. Of the 2,124,692 speeding tickets in our data, 504,644 match successfully to the FDLE data.

A.3 Hours and Shifts of Tickets

Officers manually enter time of day, and there are several inconsistencies in how these are recorded. Most officers use either a 12-hour time and clarify AM versus PM, and others use 24-hour military time. Some officers regularly use 12 hour time and do not record AM versus PM. We set these times to be missing.

The FHP has three shifts, 6am to 2pm, 2pm to 10pm, and 10pm to 6am. We record these directly from the hour of the ticket if it is properly recorded above. If there is no correct hour of day, we take a two-week moving average of the officer's modal shift for his citations and impute the shift. For the remaining tickets we leave shift as missing. Of the 1.6 million initial speeding citations, 692,416 have shift missing, and 413,560 remain missing after the imputation procedure.

B Accounting for Stopping Margin Selection

As discussed in Section 6, one concern we face is that we do not observe interactions that do not result in a ticket. Therefore, officer differences in lenience and discrimination on whether to give a ticket may bias our estimates of discrimination on whether to give a discount. Here we write down a simple selection model to discuss the potential bias from selection into the data and present a procedure to correct our estimates for officer-by-race differences in ticketing.

Consider a model of ticketing where there is a first margin of whether or not a driver is ticketed at all:

$$
\begin{aligned}
D_{ij}^{*} &= \theta_j^W + \theta_j^M \cdot M_i + \epsilon_{ij} \\
Z_{ij} &= \alpha_j^W + \alpha_j^M \cdot M_i + \eta_{ij}
\end{aligned}
$$

D_{ij}^{*} is a latent variable for whether the driver receives a discount, and Z_{ij} is a latent variable for whether the officer tickets the driver at all, where we assume $\eta_{it} \sim N(0,1)$. We observe D_{ij} if Z_{ij} crosses zero and the officer chooses to ticket the driver:

$$
D_{ij} = \begin{cases} \mathbb{I}(D_{ij}^{*} \geq 0) & \text{if } Z_{ij} \geq 0 \\ \text{missing} & \text{otherwise} \end{cases}
$$

Therefore, the comparison we make to determine the degree of discrimination is based on the difference in discounting among observed drivers[26]:

$$
\begin{aligned}
\hat{\theta}_j^M &= E[D_{ij}^{*}|M_i = 1, Z_{ij} > 0] - E[D_{ij}^{*}|M_i = 0, Z_{ij} > 0] \\
&= \theta_j^M + E[\epsilon_{ij}|\eta_{ij} > -\alpha_j^W - \alpha_j^M] - E[\epsilon_{ij}|\eta_{ij} > -\alpha_j^M]
\end{aligned}
$$

If there's a difference in treatment in the first margin ($\alpha_j^M \neq 0$) and $corr(\epsilon_{ij}, \eta_{ij}) \neq 0$, then our estimate of θ_j^M will be inconsistent. In particular, if $\alpha_j^M > 0$ (discrimination in ticketing) and $\text{corr}(\epsilon_{ij}, \eta_{ij}) < 0$ (drivers more likely to be ticketed are less likely to be discounted), then the error term above will be positive, suggesting that our measure of discrimination will be biased toward zero.

To deal with the issue of potential correlation between ticketing on the first margin and discounting, we will use an approach similar to the Heckman (1979) correction. Imagine that all officers working in the same county and year face the same number of drivers of a certain race on a given day of work, N_r. Officers choose whether or not to write a ticket for the driver, Z_{ij}, and thus the daily rate of tickets for that officer for that race-county-year is $N_{rj} = N_r \cdot P(Z_{ij} = 1)$.

Under the presumption that all officers in the same county-year face the same quantity of drivers who could potentially be ticketed for speeding, we can compare officers to calculate their propensity to give a ticket. Within each county-year-race, we calculate the average

[26] We abstract here from the lenient v. non-lenient approach from the main text as well as including observable characteristics. However, when implementing the correction procedure we return to both.

daily number of tickets given by each officer. To account for large right-tail values, we allow the 95th percentile across officers of N_{rj} for each county-year-race to be our value for N_r. Then for each officer-race-county-year, $P(Z_{ij} = 1) = \frac{N_{rj}}{N_r}$, which we call P_{ij}. Using this value, we can identify the expectation for the error term η_{ij} in the ticketing equation for each driver:

$$\begin{aligned}
P_{ij} &= Pr(\alpha_j^W + \alpha_j^M \cdot M_i + \eta_{ij} \geq 0) \\
&= \Phi(\alpha_j^W + \alpha_j^M \cdot M_i) \\
\implies E(\eta_{ij}|Z_{ij} = 1) &= \frac{\phi(\alpha_j^W + \alpha_j^M \cdot M_i)}{\Phi(\alpha_j^W + \alpha_j^M \cdot M_i)} \\
&= \frac{\phi(\Phi^{-1}(P_{ij}))}{P_{ij}}
\end{aligned}$$

Note that in the uncorrected approach, the conditional expectation of the error term is potentially nonzero because of a correlation with the ticketing error term:

$$\begin{aligned}
E(\epsilon_{ij}|\eta_{ij} > -\alpha_j^W - \alpha_j^M \cdot M_i) &= \rho \cdot E(\eta|\eta_{ij} > -\alpha_j^W - \alpha_j^M \cdot M_i) \\
&= \rho E(\eta_{ij}|Z_{ij} = 1)
\end{aligned}$$

Therefore, we can address the potential selection into the data using our officer-county-year-race-specific expected value for the ticketing error term, which we call the Heckman Correction term, and re-run the main regression with this addition. The results of this procedure are presented in Table 6. Column (1) presents again the baseline regression, and Column (2) presents the same regression with the additional Heckman Correction term. The addition of the correction does not change the value of the interaction term on Driver White and Officer Lenient or any other coefficients, suggesting that our result is not due to any issues with sample selection. This finding should not be surprising, as we found in the bottom right panel of Figure 5 that officer lenience is uncorrelated with ticketing frequency.

C Testing for Statistical Discrimination

Our paper argues that racial disparities in officer lenience reflect bias. However, a compelling alternative explanation is that officers are using race as a signal for an unobserved driver type. Our baseline regressions show that officers differentiate between white and minority drivers after controlling for previous tickets, suggesting that the observed disparity does not reflect statistical discrimination on the level of criminality. However, officers may be sorting individuals on how they *respond* to a discount. For example, officers may be trying to identify drivers who will react to a harsh ticket by speeding less in the future. Alternatively, they may choose to discount a particular driver because they are likely to respond by not contesting the ticket. To formalize these stories, imagine that drivers who are stopped for speeding have some outcome after the ticket, Y_i, that depends on whether a discount D_i is given:

$$Y_i = X_i\beta + \alpha_i D_i + \epsilon_i$$

Whether or not they speed, or contest the ticket, is potentially a function of the treatment given to them by the current stopping officer. As throughout the paper, the officer chooses whether to give a discount, and he does so on the basis of demographics, but also potentially other unobservables:

$$D_i \;\; = \;\; \mathbb{1}\big(Z_{ij}\theta - v_i \geq 0\big)$$

where Z_{ij} is written to encapsulate both the individual covariates X_i and an instrument for discounting based on the officer identity, which we discuss below. The story we are interested in testing is whether officers choose who to discount on the basis of how Y_i responds. In other words, do we have $\alpha_i \perp\!\!\!\perp D_i | X_i$ or not? Heckman *et al.* (2010) provide a number of tests for whether there is such a correlation, from which we borrow directly below. In particular, they show that a lack of correlation between discounting and treatment effect implies a linear relationship between the outcome and propensity score for treatment. To see this, we first reformulate the discount equation:

$$\mathbb{1}(D = 1) = \mathbb{1}(v \leq Z_{ij}\theta) = \mathbb{1}(F_v(v) \leq F_v(Z_{ij}\theta)) = \mathbb{1}(U_d \leq P(Z_{ij}))$$

where U_d is a uniform random variable and $P(z_{ij}) = Pr(D = 1|Z_{ij} = z_{ij})$ is the propensity score. The marginal treatment effect is defined as the treatment effect for an individual at a given propensity to be treated (Björklund and Moffitt, 1987):

$$MTE(x, u_d) = E(\alpha_i | X = x, U_d = u_d)$$

The conditional expectation of Y_i as a function of X_i and Z_i can then be written as a function of the marginal treatment effects:

$$\begin{aligned} E(Y|Z=z) &= X_i\beta + E(\alpha_i D_i|z) \\ &= X_i\beta + E(\alpha_i D_i|P(z)) \\ &= X_i\beta + E(\alpha_i|D=1, P(z)) \cdot p \\ &= X_i\beta + \int_0^p E(\alpha_i|U_d = u_d)du_d \end{aligned}$$

Under no correlation between α_i and D_i, then $E(\alpha_i|U_d = u_d) = E(\alpha_i)$. Therefore, the conditional expectation of Y_i should be linear in $P(z)$:

$$\begin{aligned} \frac{\partial E(Y|Z=z)}{\partial P(z)} &= E(\alpha_i|U_d = P(z)) \\ &= E(\alpha_i) \quad \text{under} \quad \alpha_i \perp\!\!\!\perp D_i|X_i \end{aligned}$$

Therefore, a test for the linearity of Y_i in $P(Z)$ tells us whether officers are sorting individuals on the basis of their treatment effect of D_i on Y_i. Under linearity, the marginal treatment effects of individuals with different propensities to be treated (in our case, stopped by different officers) will be the same.

The instrument Z_{ij} we use for whether an individual receives a discount is based on the identify of the officer and is a leave-out measure of the officer's propensity to give a discount:

$$Z_{ij} = \frac{1}{N_j - 1} \sum_{k \in \mathcal{J} \setminus i} D_k$$

where N_j is the number of individuals stopped by officer j. This average-lenience-of-treater instrumenting design has been used in various settings to study the effect of criminal sentence length (Kling, 2006; Mueller-Smith, 2014), bankruptcy protection (Dobbie and Song, 2015), foster care (Doyle, 2007; Doyle Jr, 2008), and juvenile incarceration (Aizer and Doyle Jr, 2015).

We then calculate an individual's propensity to receive a discount based on their stopping officer and demographic characteristics. Because an officer's lenience can vary with the race of the driver, we interact the instrument with driver race:

$$P(Z,X) = X_i\gamma + \theta^0 Z_{ij} + \theta^M \text{DriverMinority}_i Z_{ij}$$

We then run regressions of Y_i on specifications that are linear and quadratic in $P(z,x)$, where the outcomes we consider are whether a driver receives another ticket in the year following the FHP stop[27] and whether the driver contests the ticket.

The results of this analysis are presented in Table 7. We restrict attention to in-state drivers with a ticket at 9 MPH or over for whom we have a court record of whether the driver contested. These restrictions leave us with 844,422 tickets. The first two columns treat an individual's recidivism as the outcome. In Column (1) we see that an increase in the probability of receiving a discount increases an individual's likelihood of recidivating.[28] However, the quadratic in the second column is insignificant. Though not shown, a specification that includes a cubic in the propensity score also has insignificant higher terms.

Columns (3) through (5) use as an outcome whether the driver contests the ticket in court. As with recidivism, we find an effect of receiving a discount: drivers stopped by officers who are more likely to give discounts are less likely to contest their ticket. However, when we add a quadratic term in Column (4), we find a non-linear relationship, with the quadratic having a significant positive coefficient. Drivers stopped by very harsh officers have a larger marginal response to discounting than drivers stopped by less harsh officers.

The intuition for this result is the following: imagine an officer who is very lenient toward his drivers. If he is going to be harsh to one driver, he will pick someone who is not very responsive to a harsh ticket and will not contest. We will thus see that officer have a small effect of discounting on contesting. In contrast, imagine an officer who is harsh toward nearly all drivers. If he is going to give a break to someone, that discount should give him a large return in reduced court time. We should thus expect a large contest response among that officer's drivers. Our findings are thus consistent with the story that officers do try to

[27] The recidivism of the driver is calculated as an indicator for whether they receive any traffic ticket in the state of Florida in the following year. We link drivers by driver's license number. More information is available in Goncalves and Mello (2017).

[28] Though we do not report it here, the first-stage coefficient on the instrument is close to 1 and slightly smaller for minority drivers. The first-stage relationship is essentially linear, indicating that any non-linearity in the reduced form regressions presented here are not due to differences in the strength of the instrument at different levels.

identify driver's propensity to not contest their ticket.

While we therefore do find evidence of statistical discrimination on court contest response, our primary objective is to determine whether any form of statistical discrimination can explain the disparity we observe between whites and minorities. To do so, we implement a test based on Arnold *et al.* (2018). They implement the logic of the Becker (1957) hit-rate test in the random-judge design and show that, under no discrimination, the impact of a treatment should be the same at the margin across racial groups. To conduct this test, we interact the propensity score with the race of driver in Column (5). Doing so, we find that the marginal effect of a discount on contesting is statistically larger for minorities than for white drivers, indicating that the discrimination we observe cannot be explained by sorting on contest response.

D Notes on Model Estimation

While the setup of the model is simple, the non-parametric identification of the distribution of officer bias and the distribution county-by-race speeds leads to a significant number of parameters to be identified. We estimate the model through maximum likelihood, programmed in Matlab. We provide the program with the gradient vector and utilize "fminunc" with a quasi-newton search algorithm option. The variance matrix of the parameters is calculated as the inverse of the information matrix, which we calculate as the variance of the score functions.

One issue to note is that the log likelihood function is essentially flat for certain regions of the parameter space for some officer preference parameter values. This flatness occurs because some officers have no (all) drivers at the bunch point, consistent with an infinitely negative (positive) "t." The optimization algorithm reaches values that are large in magnitude. However, because the score function is essentially flat at these large values, the parameters's standard errors are extremely large.

To deal with this issue, we treat these parameters (specifically, the t estimates for officers with $P(Discount \mid X = 10) < .02$ or $P(Discount \mid X = 10) > .98$) as known and set their variances to be zero.

D.1 Model Estimates Discussion

Table A.8 presents estimates of the model parameters. Columns in the top panel present the mean and variance of each class of parameters, broken down by race, and the final column compares differences across racial groups in the mean parameter estimates. The slope parameter is positive and significant at 0.0395. Consistent with our intuition, officers face an upward-sloping cost with respect to speed, meaning that tickets are less likely to be discounted the higher the observed speed. The parameter t represents an officer's mean valuation of a racial group. We find both significant heterogeneity and a significant disparity across whites and minorities in how officers value discounting drivers, with officers' mean valuation for whites being 0.0275 higher than for minorities.

While the values of t are by themselves hard to interpret, the racial differences in treatment are more easily understood in terms of the probability of discount (i.e., fine reduction). $Pr(Discount|E(Z), j, X = 10)$ represents the likelihood of receiving a reduced ticket if the driver is at the speed right above the bunching speed, where, besides race, the driver has the average demographics Z. Consistent with the reduced-form evidence, the average officer is substantially lenient, with a large variance across officers. Officers are 3.3 percentage points less likely to discount minorities than whites, off a baseline of 35.7% likelihood of discount. Figure A.2 further shows this disparity, highlighting how racial bias results in a decreased mass of officers with very high lenience and an increase in mass of officers with very low lenience. Figure A.3 shows how the disparity only arises among officers with some degree of lenience.

The λ estimates tell us how races-by-counties differ in their underlying speeds prior to officers' choice of lenience. As we found in Section 5 when restricting our attention to non-lenient officers, model estimates suggest that minorities on average drive significantly faster than whites, on the order of 0.5 to 0.7 MPH. Figure A.4 presents this gap by county, showing that minority speeds stochastically dominate white speeds. These results are in line with previous studies of highway patrol ticketing, which argue that much of the gap in ticketing between whites and minorities can be explained by higher speeds by minorities (Smith *et al.*, 2004; Lange *et al.*, 2005). However, these previous studies and the news coverage that followed implicitly argued that the racial difference in speeds rules out the presence of bias by officers. Our study highlights how this thinking is incorrect by showing that disparities in driving and racial bias coexist in our setting. As shown in Figure A.5, the distribution of bias across officers looks very similar to the distribution found in our reduced form estimates from Section 5.

The bottom panel of Table A.8 presents the demographic-specific speed and preference parameters. Female drivers, older drivers, and those with fewer tickets all drive slower speeds on average and are more likely to be discounted. The effect of county minority share indicates that officers are less likely to discount everybody in a more minority neighborhood, regardless of the race of the stopped individual.

We report in Figure A.6 various estimates of model fit to the data. For each panel, we construct the model statistics by simulating 100 times and averaging across iterations. The top left panel compares the aggregate histograms of speeds. The top right panel compares the average ticketed speeds by race-county. The bottom left panel compares the share of tickets at 9 MPH over by officer-race. The bottom right panel compares the racial disparity in bunching at 9 MPH over by officer. In all cases, the model estimates match very closely with the true data.

D.2 Counterfactuals

Here we provide information on how the counterfactuals and their standard errors are calculated. There are several sources of uncertainty in the estimation that lead to standard errors on our calculations: 1) uncertainty of our parameter estimates, 2) randomness of the matching between officers and drivers, 3) randomness in the speed draws for the drivers, and 4) randomness in the officers' decisions to discount. We therefore calculate standard errors through a sampling procedure as follows:

- Draw a sample of parameters $\theta^{(1)} \sim N(\hat{\theta}, \hat{\Sigma})$, where $\hat{\theta}$ and $\hat{\Sigma}$ are our parameter point estimates and variance matrix, respectively.

- Within each county, randomly match officers and drivers. In the baseline estimation, the probability of encountering an officer is the share of tickets in the data which that officer gave. All the counterfactuals consist of changing the distribution of officers being matched.

- Drivers draw a speed from their Poisson distribution, $s \sim P_{\lambda_i}$.

- We draw a set of $\epsilon_{ij} \sim N(0,1)$ for all stops, and an officer discounts her driver if $t_{rj} + \alpha Z^{(2)} + \epsilon_{ij} > b \cdot s$.

- Iterate 500 times.

Then, our estimates and standard errors for the racial gaps in each counterfactual are the average and standard deviation across all iterations.

Here we describe explicitly how each counterfactual is performed:

- Decomposition with no sorting: Rather than matching drivers and officers randomly within a county, they are matched randomly across the entire state.

- Decomposition with no bias: Identical to the baseline, officers and drivers are matched randomly within a county. Officers preferences for minority drivers is set to be their white preference, t_{wj}.

- Firing 5% most discriminatory officers: Calculate $P_j^{\text{bias}} \equiv Pr_j(Discount \mid X = 10, E(Z), r = w) - Pr_j(Discount \mid X = 10, E(Z), r = m)$, and find the 5th percentile for the entire state and "remove" all officers below this threshold. We also remove officers that cross the same threshold of discrimination against *white* drivers. The probability of an individual encountering a specific officer is that officer's share of tickets among the remaining officers.

- Hiring more minority officers: We increase the share from 35% to 45%. We do so by proportionately increasing the number of minority officers in each county. e.g. a county that previously was 10% minority officers is now 16% minority. The distribution of officer tastes t_{rj} is the same as the existing distribution *within* officer race. The procedure is identical for increasing the share of female officers.

- Re-assigning officers based on discrimination: Within a troop, officers are ranked based on their discrimination. In the county of that troop with the most minorities, the lowest-ranked officers are assigned. The second-most minority county receives the next-least discriminatory officers, and so on. Officers write as many tickets as in the true data, so some officers may write tickets in two counties that are adjacent in their share minority. The procedure is identical when assigning officers based on their lenience, where the most lenient officers are assigned to the most minority neighborhoods.

Table A.1: Racial Disparity in Speeding

	Full Sample					GPS Sample	
	(1) MPH Over	(2) MPH Over	(3) MPH Over	(4) MPH Over	(5) MPH Over	(6) MPH Over	(7) MPH Over
Driver Black	1.073	0.809	0.728	0.637	0.622	0.890	0.782
	(0.268)	(0.0877)	(0.0844)	(0.0832)	(0.0759)	(0.0703)	(0.0751)
Driver Hispanic	2.765	0.875	0.793	0.648	0.652	1.027	0.764
	(0.526)	(0.128)	(0.134)	(0.137)	(0.135)	(0.214)	(0.134)
Driver Female				-0.619	-0.563	-0.436	-0.379
				(0.0453)	(0.0403)	(0.0600)	(0.0572)
FL License				-0.183	-0.353	-0.685	-0.534
				(0.0810)	(0.0808)	(0.152)	(0.127)
Driver Age				-0.0443	-0.0421	-0.0378	-0.0338
				(0.00135)	(0.00130)	(0.00164)	(0.00199)
1 Prior Ticket					0.281	0.268	0.285
					(0.0243)	(0.0507)	(0.0682)
2+ Prior Tickets					0.799	0.682	0.740
					(0.0379)	(0.0760)	(0.0637)
Log Zip Code Income					0.123	0.0843	0.0313
					(0.0501)	(0.0443)	(0.0478)
Mean	16.554	16.554	16.587	16.587	16.587	16.027	16.027
Vehicle FE					X	X	X
Location FE		X				X	
Location + Time FE			X	X	X	X	
GPS FE							X
Observations	1124513	1124513	1063227	1063227	1063227	123516	123516

Notes: Table reports regressions where the outcome is the speed for which the individual is ticketed. ”Location FE” are fixed effects at the county by posted speed limit. ”Location + Time FE” are fixed effects at the county by posted speed limit by year by month by day of week by hour fixed effects. ”GPS FE” are fixed effects at the road segment by posted speed limit by year by month by day of week by hour fixed effects. GPS sample are tickets with the GPS location available. Standard errors are clustered at the county level.

Table A.2: Racial Disparity in Discounting

	Full Sample					GPS Sample	
	(1) Discount	(2) Discount	(3) Discount	(4) Discount	(5) Discount	(6) Discount	(7) Discount
Driver Black	-0.0316	-0.0270	-0.0241	-0.0218	-0.0229	-0.0378	-0.0310
	(0.0179)	(0.00454)	(0.00480)	(0.00488)	(0.00456)	(0.00665)	(0.00616)
Driver Hispanic	-0.143	-0.0401	-0.0392	-0.0345	-0.0357	-0.0559	-0.0378
	(0.0331)	(0.00883)	(0.00939)	(0.00892)	(0.00883)	(0.0120)	(0.00859)
Driver Female				0.0288	0.0269	0.0198	0.0179
				(0.00434)	(0.00407)	(0.00402)	(0.00395)
FL License				0.00806	0.0143	0.0308	0.0191
				(0.00404)	(0.00438)	(0.00805)	(0.00883)
Driver Age				0.00136	0.00128	0.00122	0.000999
				(0.000244)	(0.000234)	(0.000203)	(0.000206)
1 Prior Ticket					-0.0121	-0.0102	-0.0129
					(0.00257)	(0.00376)	(0.00450)
2+ Prior Tickets					-0.0294	-0.0219	-0.0274
					(0.00581)	(0.00638)	(0.00719)
Log Zip Code Income					-0.00950	-0.00381	-0.00112
					(0.00217)	(0.00336)	(0.00445)
Mean	.31	.31	.309	.309	.309	.324	.324
Vehicle FE					X	X	X
Location FE		X				X	
Location + Time FE			X	X	X	X	
GPS FE							X
Observations	1124513	1124513	1063227	1063227	1063227	123516	123516

Notes: Table reports regressions where the outcome is an indicator for the individual being ticketed at 9MPH over the limit. "Location FE" are fixed effects at the county by posted speed limit. "Location + Time FE" are fixed effects at the county by posted speed limit by year by month by day of week by hour fixed effects. "GPS FE" are fixed effects at the road segment by year by month by day of week by hour fixed effects. GPS sample are tickets with the GPS location available. Standard errors are clustered at the county level.

Table A.3: Racial Disparity in Speeding, Non-lenient Officers

	Full Sample					GPS Sample	
	(1) MPH Over	(2) MPH Over	(3) MPH Over	(4) MPH Over	(5) MPH Over	(6) MPH Over	(7) MPH Over
Driver Black	1.230	0.702	0.637	0.516	0.487	0.730	0.584
	(0.184)	(0.160)	(0.148)	(0.135)	(0.129)	(0.148)	(0.203)
Driver Hispanic	1.578	0.485	0.418	0.264	0.259	0.423	0.378
	(0.277)	(0.0677)	(0.0614)	(0.0668)	(0.0741)	(0.150)	(0.171)
Driver Female				-0.523	-0.466	-0.423	-0.323
				(0.0770)	(0.0705)	(0.112)	(0.0999)
FL License				-0.211	-0.378	-0.575	-0.530
				(0.0839)	(0.0855)	(0.165)	(0.199)
Driver Age				-0.0454	-0.0429	-0.0330	-0.0329
				(0.00264)	(0.00244)	(0.00254)	(0.00257)
1 Prior Ticket					0.267	0.256	0.195
					(0.0309)	(0.0790)	(0.106)
2+ Prior Tickets					0.708	0.669	0.626
					(0.0444)	(0.116)	(0.138)
Log Zip Code Income					0.0240	0.0576	0.0440
					(0.0497)	(0.0759)	(0.0725)
Mean	20.378	20.378	20.403	20.403	20.403	20.024	20.024
Vehicle FE					X	X	X
Location FE		X				X	
Location + Time FE			X	X	X	X	
GPS FE							X
Observations	366146	366146	348275	348275	348275	30285	30285

Notes: Table reports regressions where the outcome is the speed for which the individual is ticketed, restricting attention only to non-lenient officers. "Location FE" are fixed effects at the county by posted speed limit. "Location + Time FE" are fixed effects at the county by posted speed limit by year by month by day of week by hour fixed effects. "GPS FE" are fixed effects at the road segment by year by month by day of week by hour fixed effects. Standard errors are clustered at the county level.

Table A.4: Officer Lenience Randomization Check

	Full Sample			GPS Sample	
	(1) Lenience	(2) Lenience	(3) Lenience	(4) Lenience	(5) Lenience
Driver Black	0.000635	0.00166	-0.000493	-0.00704	0.000622
	(0.0166)	(0.00292)	(0.00337)	(0.00549)	(0.00550)
Driver Hispanic	-0.0900	-0.00594	-0.00666	-0.0224	-0.00255
	(0.0287)	(0.00497)	(0.00462)	(0.0130)	(0.00338)
Driver Female	0.0188	0.00423	0.00251	0.00148	0.00163
	(0.00428)	(0.00217)	(0.00181)	(0.00132)	(0.00201)
Florida License	-0.0774	0.000511	0.000401	0.00769	-0.00196
	(0.0252)	(0.00346)	(0.00317)	(0.00806)	(0.00422)
Driver Age	-0.214	0.216	0.0744	0.134	0.0371
	(0.162)	(0.129)	(0.119)	(0.103)	(0.0741)
1 Prior Ticket	-0.0113	-0.000750	-0.000483	0.00244	0.0000311
	(0.00645)	(0.000896)	(0.000948)	(0.00144)	(0.00223)
2+ Prior Tickets	-0.0235	-0.00132	-0.0000563	0.00443	0.00212
	(0.0125)	(0.00145)	(0.00149)	(0.00225)	(0.00371)
Log Zip Code Income	-0.00361	0.00276	-0.00415	-0.000165	-0.000452
	(0.00855)	(0.00295)	(0.00195)	(0.00312)	(0.00211)
F-test	0	.616	.039	.419	.946
Mean	.31	.31	.309	.326	.33
Location FE		X			
Location + Time FE			X	X	
GPS FE					X
Observations	1139734	1139734	1077412	124916	135427

Notes: All regressions includes vehicle type fixed effects and county fixed effects. The F-test reports the joint hypothesis test that variables Driver Black through Log Zip Code Income are zero. Standard errors are clustered at the county level. ”Location FE” includes county by highway fixed effects. ”Location + Time FE” includes county by highway by year by month by day of the week by shift fixed effects. ”GPS FE” includes road segment by year by month by day of the week by shift fixed effects.

Table A.5: Difference-in-Differences Officer-Level Results

	Discrimination Percentile					
	(1) 10 %	(2) 25%	(3) 50%	(4) 75%	(5) 90%	(6) N
All Officers	-0.0113	0.0000	0.0053	0.0681	0.1275	1591
White Officers	-0.0076	0.0000	0.0231	0.0835	0.1386	1591
Black Officers	-0.0339	0.0000	0.0000	0.0228	0.0637	1591
Hispanic Officers	-0.0112	0.0000	0.0000	0.0404	0.1199	1591

Notes: Table reports percentiles of the distribution of officer-level discrimination, as calculated from Equation (4).

Table A.6: Officer Discrimination Randomization Check

	Full Sample			GPS Sample	
	(1) Discrimination	(2) Disc	(3) Disc	(4) Disc	(5) Disc
Driver Black	0.000888	0.00152	0.000661	0.00233	0.00000407
	(0.00190)	(0.000604)	(0.000553)	(0.00182)	(0.000880)
Driver Hispanic	-0.00617	0.000164	-0.000751	-0.000986	-0.000444
	(0.00319)	(0.000849)	(0.000668)	(0.00258)	(0.000699)
Driver Female	0.00124	-0.000176	-0.000154	0.000703	0.000367
	(0.000615)	(0.000181)	(0.000202)	(0.000523)	(0.000431)
Florida License	-0.0111	-0.000238	-0.000563	-0.0000257	0.0000737
	(0.00304)	(0.000715)	(0.000657)	(0.00227)	(0.000606)
Driver Age	-0.0139	-0.00259	-0.000300	-0.0122	-0.00549
	(0.0274)	(0.0169)	(0.0156)	(0.0253)	(0.0153)
1 Prior Ticket	-0.000769	0.0000154	0.0000520	0.000813	0.000519
	(0.000693)	(0.000174)	(0.000178)	(0.000773)	(0.000515)
2+ Prior Tickets	-0.00198	0.0000737	0.000178	0.00114	0.000326
	(0.00144)	(0.000223)	(0.000215)	(0.000886)	(0.000511)
Log Zip Code Income	0.00115	0.00150	0.0000372	0.000848	0.0000294
	(0.00128)	(0.000721)	(0.000377)	(0.00110)	(0.000641)
F-test	0	.175	.148	.221	.687
Mean	.305	.305	.304	.323	.323
Location FE		X			
Location + Time FE			X	X	
GPS FE					X
Observations	1141628	1141628	1079250	125040	125040

Notes: All regressions includes vehicle type fixed effects and county fixed effects. The F-test reports the joint hypothesis test that variables Driver Black through Log Zip Code Income are zero. Standard errors are clustered at the county level. "Location FE" includes county by highway fixed effects. "Location + Time FE" includes county by highway by year by month by day of the week by shift fixed effects. "GPS FE" includes road segment by county by highway by year by month by day of the week by shift fixed effects.

Table A.7: Predicting Officer Complaints/Force

	(1) # Complaints	(2) Any Complaints	(3) # Use of Force	(4) Any Use of Force
Lenience	-0.622	-0.184	-0.184	-0.108
	(0.206)	(0.0546)	(0.152)	(0.0494)
Discrimination	-0.247	0.134	0.0642	-0.0698
	(0.574)	(0.192)	(0.433)	(0.166)
Black	0.111	0.00131	-0.196	-0.0863
	(0.176)	(0.0401)	(0.0908)	(0.0335)
Hispanic	-0.00981	0.0165	0.0309	0.00569
	(0.144)	(0.0372)	(0.0993)	(0.0373)
Other	0.178	0.0242	-0.234	-0.0727
	(0.380)	(0.0996)	(0.181)	(0.0938)
Female	-0.295	-0.109	-0.0149	0.0125
	(0.158)	(0.0496)	(0.104)	(0.0443)
Age	-0.120	0.163	-0.736	-0.194
	(0.332)	(0.0900)	(0.212)	(0.0793)
Age Squared	0.0167	-0.0249	0.0611	0.0138
	(0.0478)	(0.0131)	(0.0266)	(0.0108)
Experience	-0.0633	-0.0800	-0.554	-0.00694
	(0.414)	(0.130)	(0.331)	(0.117)
Exp Squared	-0.0209	0.0350	-0.00580	-0.00173
	(0.0766)	(0.0249)	(0.0457)	(0.0195)
Failed Entrance Exam	0.259	0.0432	-0.106	-0.00330
	(0.205)	(0.0483)	(0.110)	(0.0458)
Any College	-0.183	-0.0250	0.103	0.0134
	(0.104)	(0.0293)	(0.0946)	(0.0264)
Sought Promotion	-0.194	-0.0664	-0.0405	0.0239
	(0.113)	(0.0294)	(0.0884)	(0.0277)
Mean	1.26	.551	.559	.294
Observations	1402	1402	1402	1402
Regression	OLS	OLS	OLS	OLS

Notes: Heteroskedasticity-robust standard errors in parentheses. Column title indicates the dependent variable. Data is at the officer level. Regressions have indicator variables for years when and districts where the officer worked.

Table A.8: Model Parameter Estimates

	White			Minority			
	(1) μ	(2) σ^2	(3) # Param	(4) μ	(5) σ^2	(6) # Param	(7) Mean Diff
b, slope	0.0395	—	1	—	—	—	—
	(0.0006)						
t, officer valuations	-0.2824	4.5876	1591	-0.3099	4.2300	1591	0.0275
	(0.0031)	(0.1627)		(0.0035)	(0.1500)		(0.0046)
λ, speeds	20.5058	2.7202	67	20.9833	2.1300	67	-0.4775
	(0.0517)	(0.4735)		(0.0407)	(0.3708)		(0.0658)
Pr(Discount | E(Z), j)	0.3745	0.1283	1591	0.3547	0.1204	1591	0.0198
	(0.0007)	(0.0000)		(0.0008)	0.0000		(0.0011)

	Speed Parameters γ		Preference Parameters α	
	(1)	(2)	(3)	(4)
Female	-0.4813	(0.0087)	0.1353	(0.0036)
Age	-0.0453	(0.0003)	0.0057	(0.0001)
Previous Tickets	0.1868	(0.0027)	-0.0388	(0.0013)
County Minority Share			-1.8714	(0.0270)

Notes: This table presents estimates of the model introduced in section 8. b is the slope parameter for how officers weight the speed of drivers in choosing to discount, t is each officer's mean valuation of a racial group in choosing to discount, and λ is the poisson speed parameter for each race by county. $Pr(Discount \mid E(Z), j) = \Phi(t_{rj} + E(Z)\alpha - 10b)$, i.e. the probability of being discounted when driving right above the bunch point for an average driver. The variances are empirical variances of the estimates, not adjusted for sampling error.

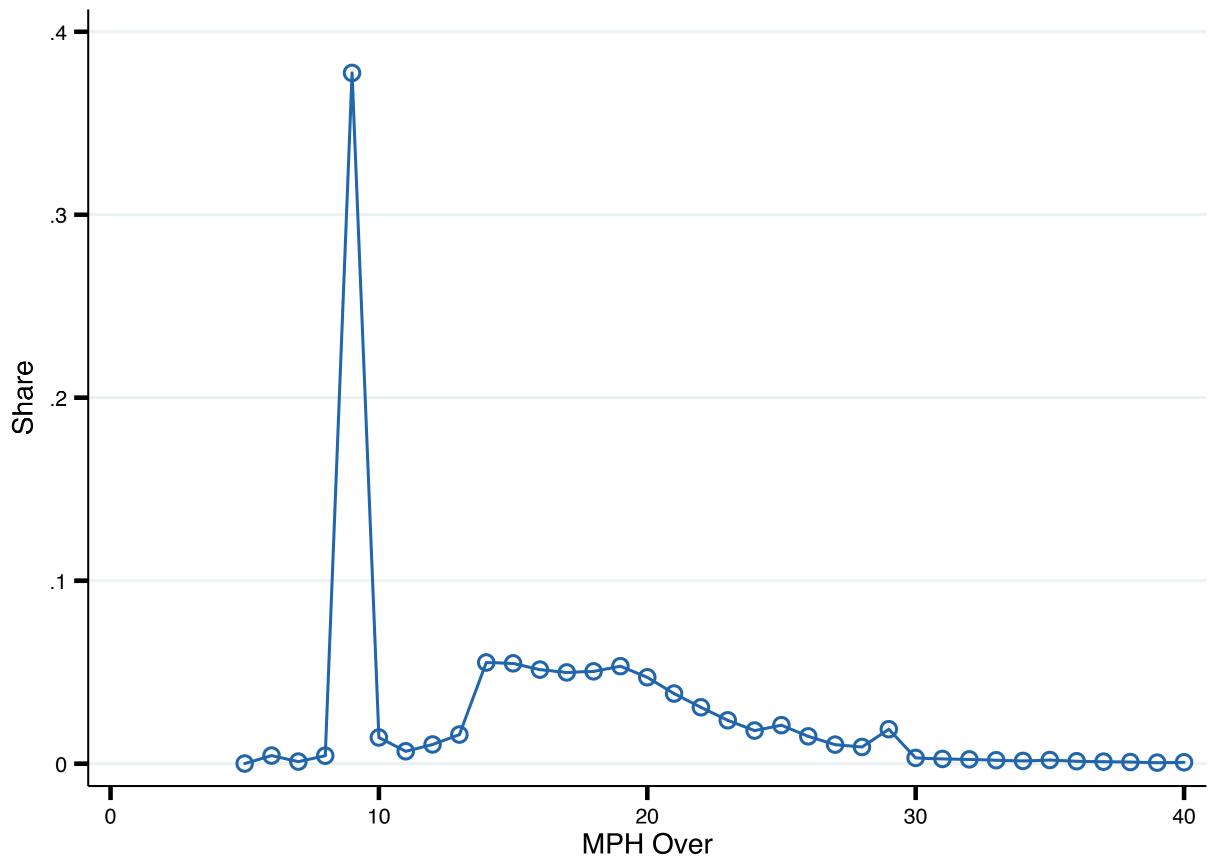
Table A.9: Speed Gap Decomposition

	State-Wide Disparity			
	(1) White Mean (MPH)	(2) Minority Mean	(3) Difference	(4) Percent
Baseline	15.531	17.296	1.764	100
	(0.009)	(0.011)	(0.014)	
No Discrimination	15.530	17.087	1.557	88.244
	(0.008)	(0.012)	(0.013)	(0.014)
No Sorting	15.645	17.166	1.521	86.193
	(0.009)	(0.012)	(0.015)	(0.015)
Neither	15.644	16.927	1.283	15.644
	(0.009)	(0.012)	(0.014)	0.013

	County-Level Disparity			
	(1) White Mean (MPH)	(2) Minority Mean	(3) Difference	(4) Percent
Baseline	15.531	16.194	0.662	100
	(0.009)	(0.011)	(0.014)	(NaN)
No Discrimination	15.530	15.967	0.436	65.868
	(0.008)	(0.012)	(0.013)	(0.022)
No Sorting	15.645	16.341	0.695	104.980
	(0.009)	(0.012)	(0.015)	(0.027)
Neither	15.644	16.106	0.462	69.714
	(0.009)	(0.012)	(0.014)	(0.024)

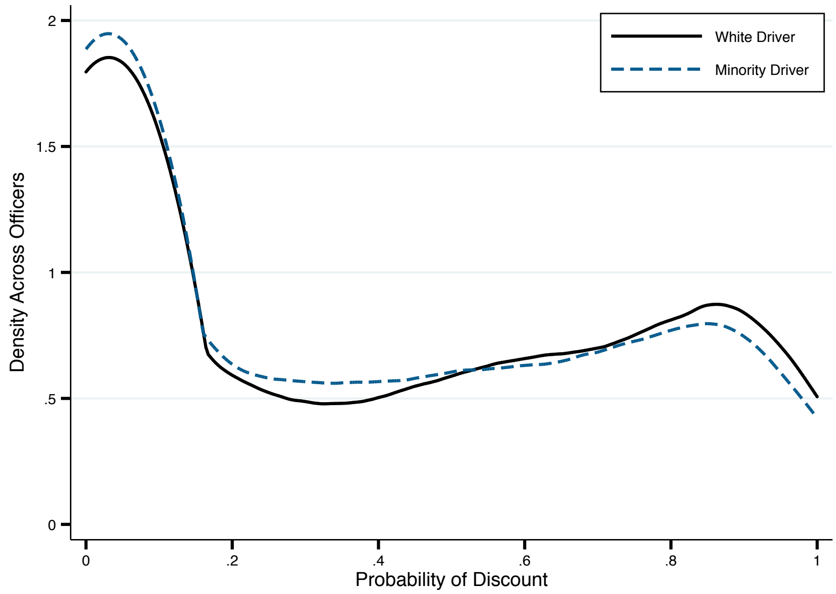
Notes: Table presents how the racial gap in speeds changes without bias and sorting of officers across counties. The gap is the minority drivers' outcome minus white drivers' outcome. No bias is calculated by assigning each officer's preferences toward minorities to be the same as his preference to whites. No sorting is calculated by simulating a new draw of officers for each driver, where the draw is done at the state level.

Figure A.1: Distribution of Charged Speeds for Radar Gun Sample



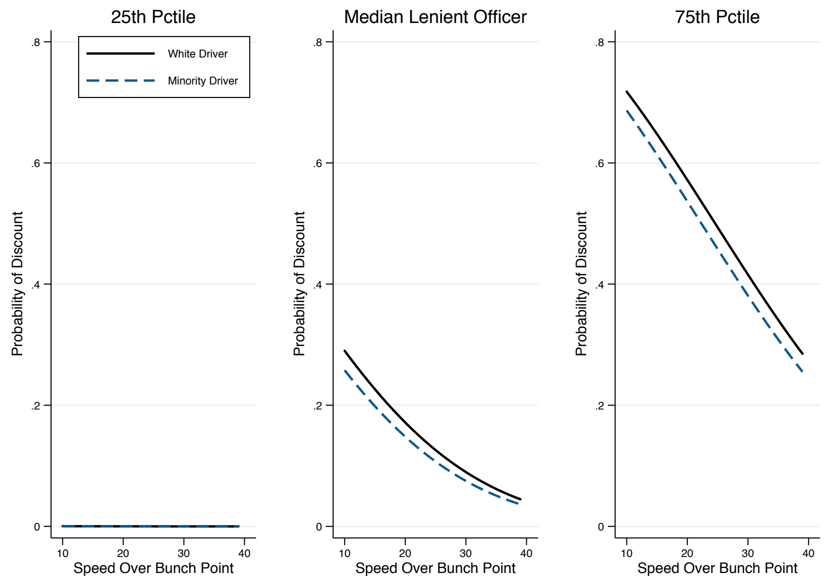
Notes: Line shows histogram of ticketed speeds for observations where the officer records that the speed is detected from a radar gun (N = 101,716).

Figure A.2: Model Estimates: Officer Lenience by Race



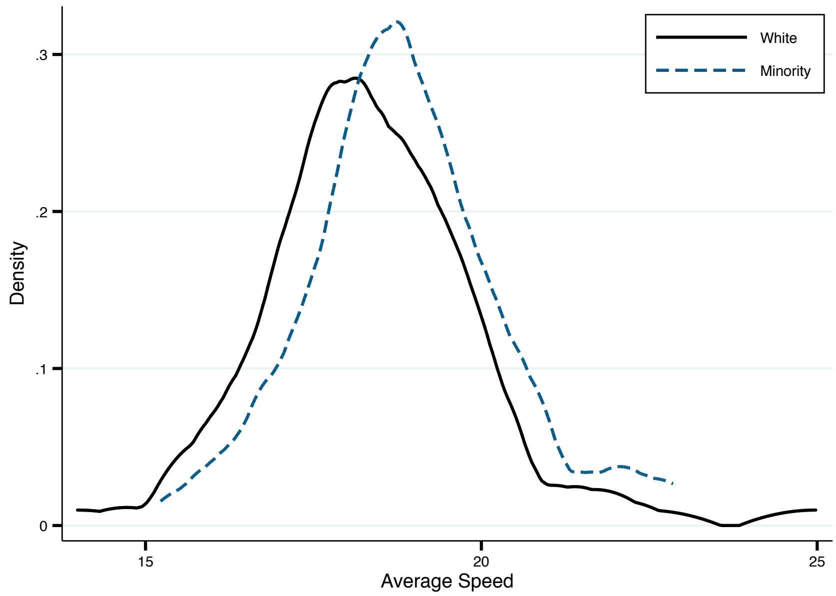
Notes: $P_{rj} \equiv P_j(\text{Discount}|X = 10, \text{ Driver Race} = r, Z = E(Z))$

Figure A.3: Model Estimates: Percentiles of Officer Lenience



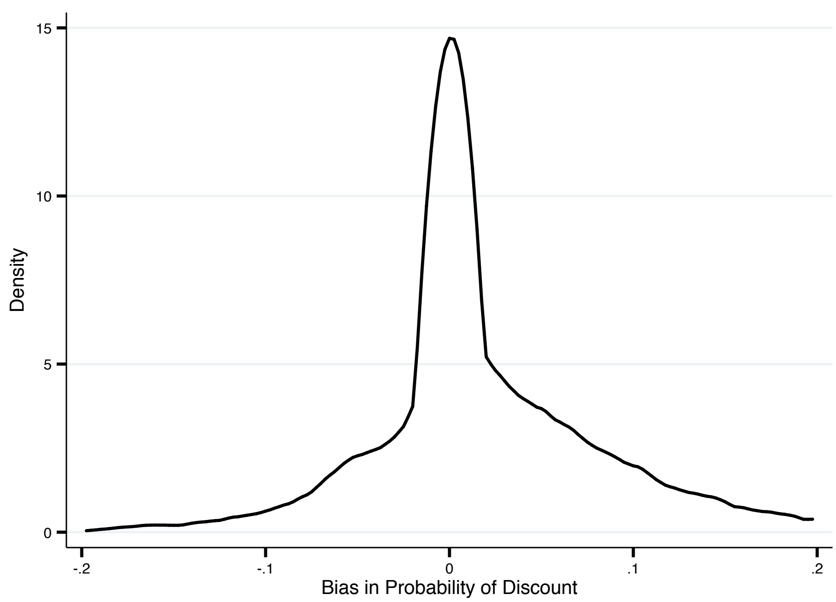
Notes: $P_{rj} \equiv P_j(\text{Discount}|X = 10, \text{ Driver Race} = r, Z = E(Z))$

Figure A.4: Model Estimates: Speed Distribution



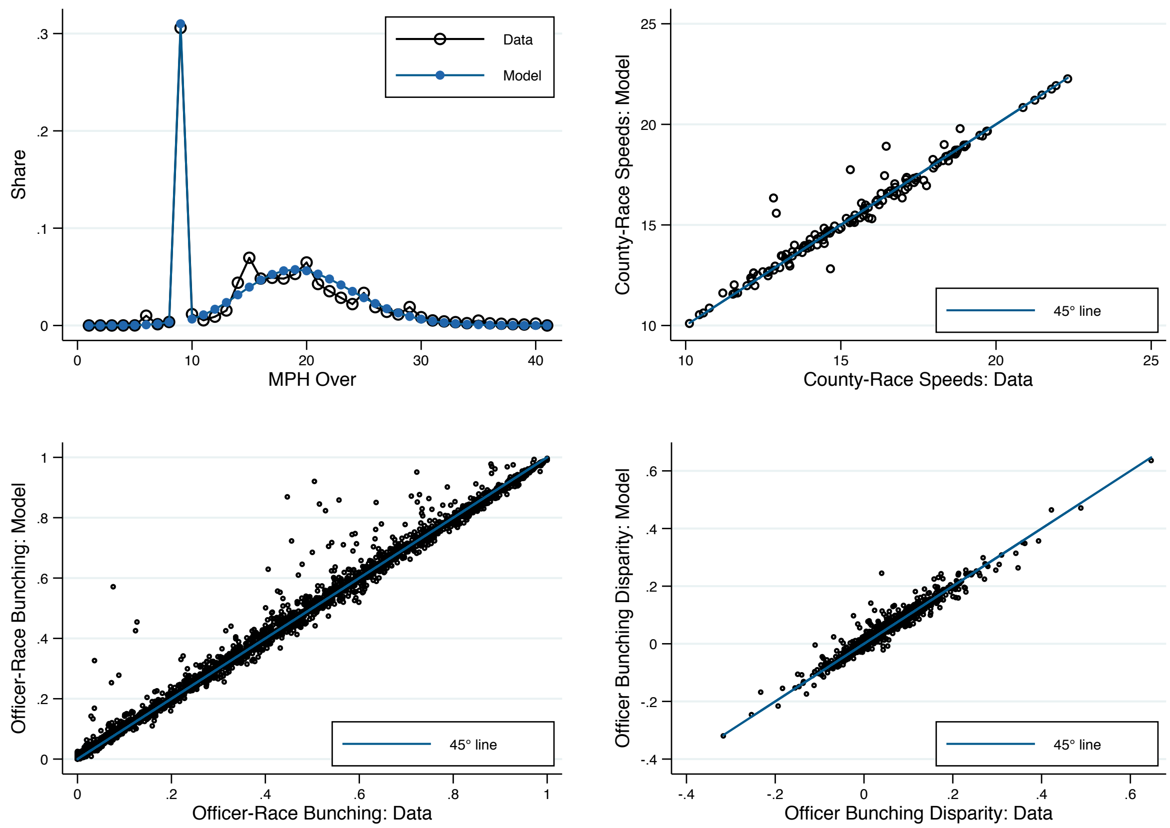
Notes: Figure plots the distribution of speed parameters λ across counties, separately by race of the driver, where individual covariates are set to the average value. In other words, we plot $\lambda = \lambda_{cr} + \gamma E(Z)$

Figure A.5: Model Estimates: Racial Discrimination by Officer



Notes: $P_j(\text{Discount}|X = 10, \text{Driver Race} = \text{White}, Z = E(Z)) - P_j(\text{Discount}|X = 10, \text{Driver Race} = \text{Minority}, Z = E(Z))$

Figure A.6: Model Diagnostic Figures



Notes: Figures compare various model estimates with their counterparts in the true data. Model estimates are found by simulating 100 iterations of the model and calculating averages across iterations. The top left panel compares the aggregate histograms of speeds. The top right panel compares the average ticketed speeds by race-county. The bottom left panel compares the share of tickets at 9 MPH over by officer-race. The bottom right panel compares the racial disparity in bunching at 9 MPH over by officer.