Not So Black and White: Uncovering Racial Bias from Systematically Masked Police Reports

Elizabeth Luh

July 19, 2019

Preliminary and Incomplete – do not cite

Abstract

Biased police officers may purposely mis-record, or mask, the race of citizens that they interact with in order to evade detection. Indeed, journalists uncovered widespread evidence of such masking among Texas Highway troopers from 2010 to 2015. I propose a new test of racial bias in the presence of masking that is more powerful than standard tests and is well-suited to explore the rich heterogeneity in bias. Using various data-driven techniques to detect masking, I estimate that 24% of 130,240 searches were masked, with over half being Hispanic drivers being mis-recorded as white when searches failed to turn up contraband. I find that Hispanic and white troopers are biased against non-white motorists, with Hispanic motorists being treated the most unfairly. Using my model, I also find evidence of institutional racial bias and 'bad apple' troopers across Texas.

1 Introduction

In the United States, racial bias in the criminal justice system is a pervasive and prevalent issue. Disparities in the treatment of citizens of different races have been found in nearly all types of interactions between citizens and law enforcement, from motorist stops for black and Hispanic motorists (Harris (1999)) or in the more extreme cases of alleged excessive use of police force against minorities (Lind (2015))(for example: Michael Brown). These disparities carry over into each step of the justice system from airport screening (Persico and Todd (2005)), ticketing (Anbarci and Lee (2014), Goncalves and Mello (2017)), stop and frisk participation (Coviello and Persico (2013)), bail decisions (Arnold et al. (2018)), sentencing (Shayo and Zussman (2011)), parole (Anwar and Fang (2015)), and in capital punishment (Alesina and Ferrara (2014)). While these disparities may seem driven by racial bias, empirically proving this is difficult for a few major reasons. First, the true proportion of guilty citizens in the sample is unobservable. Researchers do not know the true number of drivers who actually carry contraband. Second, many unobservable variables, such as driver demeanor, can affect the outcome of trooper-driver interactions, are unavailable to researchers. Without this data, researchers have no way for controlling for all these omitted variables, making it very difficult to understand the true motivation behind racial profiling using the available data. Researchers and economists have attempted to answer this question in recent years, by focusing on the stylized interaction between motorists and law enforcement. This raises another problem in that researchers can only observe interactions between motorists and law enforcement conditional on the interaction being recorded (Knox et al. (2019)). As researchers, not only are we unable to know how many motorists a law enforcement officer chooses not to pull over/interact with, but we also rely on the assumption that the officer records the data as factually as he or she can. Given that a law enforcement officer is biased, he or she is motivated to mis-record the interaction to appear less biased.

In this paper, I develop a new statistical test of racial bias that exploits the fraudulent behavior by Texas highway troopers in motorist stops from 2010 - 2015. Troopers were caught purposefully misrecording minority motorists' race as white in a stop, a practice which I call masking or cheating. I exploit this masking behavior to develop a new statistical test which links masking behavior to the magnitude of racial bias of individual troopers. The intuition of my test is that troopers mask to cover up biased behavior. Specifically, I find that troopers were masking minority motorists as white in order to make their minority search success rates appear higher than in reality. Masking gives the trooper direct payoff and also a way to cover up racially biased behavior. I build upon past models of Knowles et al. (2001) (which I'll refer to as KPT) and Anwar and Fang (2006) (which I'll refer to as AF) and incorporate this new dimension of masking behavior of troopers. Using this test, I develop a new test and method of measuring racial bias, which yields clues of trooper preferences and motivations.

This paper is related to the literature on motorist stops and whether the decision to search is influenced by racial bias. There are many earlier contributions to the literature that examine the role of motorist race and trooper race in stop interactions, notably KPT and AF along with Antonovics and Knight (2009). None of the past papers have addressed the possibility of the data being purposefully mis-recorded to hide bias. My paper is the first to my knowledge to test for masking behavior in troopers and to explicitly link this behavior to racial bias.

As is well known, racial disparities in aggregate statistics is not evidence of racial bias. For example, if Hispanic motorists are more likely than white motorists to carry contraband, then the aggregate number of searches and stops for Hispanic motorists would be higher even if race was not a factor in the decision. Moreover, troopers attempting to maximize successful searches might racially profile motorists. Such *statistical discrimination*, which is legal in the U.S., race is an indicator for whether the motorist is actually guilty. Thus distinguishing if a motorist search is driven by statistical discrimination or racial bias of the law enforcement officer becomes the fundamental question for policy makers and researchers.

In my model, I use the distortion in the search success rate from cheating to create a measure of bias. I also apply different statistical tests to determine if this masking is intentional. In the context of this research, troopers used cheating to make their white motorist search success rate lower than the actual rate and to make their minority search success rate higher. I use the frequency of cheating and the effect cheating had on the trooper's success rates to determine the trooper's preferences towards motorist rates. My test can not only detect the existence of bias, but can also measure the magnitude of the bias for the individual trooper.

I test my model on a unique data set of 8 million stops created by combining Texas stop data from the Stanford Open Policing Project (SOPP) with administrative data on Texas highway troopers from the Texas Department of Public Safety for 2010 - 2015. I use multiple data driven approaches that leverage the surname and home address of the driver to uncover masking in millions of recorded interactions between troopers and motorists. I use the level of misrecording of race to develop my measure of masking, or cheating, which I then use to measure bias. My research is the first to use this analysis in the context of highway stops.

Using this data, I am able to make a few contributions to the literature on racial prejudice in law enforcement. First, I find that all troopers, including Hispanic troopers are biased against Hispanic motorists. Further, I find that Hispanic troopers are the most biased overall and black troopers are least biased. Troopers are more biased in their interactions when the motorist is guilty. I strictly define guilty as when a trooper discovers contraband (i.e drugs, illegal arms) conditional on searching a motorist. I also use this biased behavior to uncover trooper preferences towards motorist race. Specifically, I find that Hispanic troopers exhibit same race bias by being biased against Hispanic motorists, white troopers are biased against Hispanic and black motorists, and black troopers are rather unbiased. Comparing my results to past work by Anwar and Fang (2006), Knowles et al. (2001), and Antonovics and Knight (2009), my results are more informative about the magnitude and direction of the bias, using this innovative test. Unlike past papers, I find the existence of like-race bias for Hispanic troopers in motorist stops.

Second, I am able to use my measure of bias to uncover covariates of bias such as quality of trooper and geographic clustering of bias. I find that biased troopers have a higher search success rate compared to unbiased troopers, but this gap falls when searching Hispanic and black motorists. I also find evidence of both institutional bias and 'bad apple' troopers when testing for clustering of bias and biased troopers.

Finally, these results motivate testing for accuracy in the data. The issue of misrecording in law enforcement has been heightened in recent years with instances across the U.S., not only in Texas. The increasing number of cities requiring the use of body camera footage during police stops is in response to the concern of the accuracy in police reporting. This paper shows that without correcting for the misrecording, the amount of bias researchers can detect is severely under-measured. Further, by uncovering this bias, researchers can also use this to create a better test of racial bias.

The rest of the paper is organized as follows. In Section 2, I outline the background of my research. Section 3 outlines my statistical model. Section 4 shows my empirical results and other testable implications of my model. I run robustness checks in Section 5 or conclude.

2 Background

Literature Review

One approach that much of the past literature has used to distinguish between racial prejudice and statistical discrimination is to use Becker's (1957) outcome test. In the context of motorist searches, the intuition is simple: if troopers are profiling minority motorists due to racial bias, then they will continue searching minority motorists even if the likelihood of the motorist carrying contraband is smaller than the likelihood for whites. In other words, if racial prejudice is the reason for racial profiling, than the success rate of the marginal minority motorists will be lower than the success rate for the marginal white motorist. On the other hand, if the racial profiling is the result of statistical discrimination, than the search success rates should be same. The intuition being that the trooper will expend his resources searching the race of motorist that is more likely to be guilty. While the approach is simple, the application of this test is difficult. The researcher can never observe the marginal motorists since this requires knowing all the variables that could influence a trooper's decision to search.

Past empirical studies have acknowledged this issue when testing for racial bias (Anwar and Fang (2006)). Even with rich data, researchers cannot prove conclusively the direction of racial bias, even if they can test for the existence (Knowles et al. (2001), Roland G. Fryer (ming), Anwar and Fang (2006)). Without being able to observe the marginal motorist, researchers cannot definitively say if troopers are biased against a certain race of motorist. Another approach is to assume the trooper's preference and then test to see if this holds empirically (Antonovics and Knight (2009)). The main issue with this is that this can only test for relative bias and not absolute bias.

Other researchers have used alternative identification strategies to overcome the selection bias of troopers in the choice of searching motorists. One example is the 'veil of darkness,' which uses the diminished ability of trooper's to observe the motorist race after sunset. While this reduces the prevalence of selection bias, Kalinowski et al. (2017) and Horrace and Rohlin (2016) still find evidence of endogeneity. West (2018) use the plausibly exogeneous assignment of police officers to traffic accidents to identify a causal relationship between the actions of police officers by driver race. The major drawback to this identification strategy is that the results are context specific and may not apply to a wider range of motorist and officer interactions.

My research is also related to empirical research in cheating behavior. Most of these past papers are in the context of cheating behavior in education, whether in teachers and administrators or in students (Jacob and Levitt (2003)). Jacob and Levitt (2003) found that teacher's cheating behavior was highly responsive to incentives. Schools with high-powered incentives induced cheating behavior. This type of motivation and incentives is key to understanding why troopers may choose to mask motorists in stops.

One past paper that studies law enforcement officer using trooper discretionary power in stops to expose racial bias is in the context of providing citations to speeding drivers. Goncalves and Mello (2017) finds that officers are more likely to be lenient when ticketing speed violations with white drivers compared to non-white drivers, which they argue is proof of biased behavior. This is vastly different from my paper in that I study stops conditional on searches and this cheating behavior is less consequential compared to misrecording races. Given that mine is possibly the first to document such severity of cheating behavior in troopers, this may provide motivation to reexamine past work in racial bias in motorist stops. With masking, past literature may still be under-detecting the existence of bias.

Masking and Highway Troopers in Texas

Texas Highway Patrol is a division of the Texas Department of Public Safety. The patrol's primary duties are to enforce state traffic laws and commercial vehicle regulation on highways of Texas. They currently employ over 2800 troopers in Texas divided across 7 regions in Texas. The department is responsible for licensing of drivers, vehicle inspections, and handgun licensing. In a motorist stop, troopers are allowed to investigate the passenger and the driver. While drivers are not required to answer questions, they are required to provide their driver's license and if arrested, they must also provide their name, residence address, and date of birth. Law enforcement officers may ask for consent to search the vehicle or person, which the driver can grant or deny. "... however, if an officer has probable cause to believe that your vehicle contains evidence of crime, it can be searched without your consent (DPS (DPS))." To search a vehicle without the driver's consent, the trooper must either have: probable cause, arrested the driver *prior* to searching the vehicle, reasonably believes the motorist has weapons, or has a warrant. If the officer believes that the driver or passenger has a weapon, he or she may pat down the person and search the vehicle and the surrounding immediate area. Motorists cannot physically resist a search but can notify the officer that he or she does not consent.

Drivers can report troopers if they feel that troopers behaved inappropriately during a stop and troopers can face repercussions if the claim is substantiated. Troopers badge numbers and names are normally provided and drivers can submit complaints to the department. Upon receipt of a complaint, the department assigns the complaint either he Personnel Complaint Investigations or Division Referrals to investigate the complaint. The investigation can have one of four outcomes: unfounded, exonerated, not sustained, or sustained. A sustained complaint can result one or more of the following: formal written reprimand, disciplinary probation, time off without pay, reduction of salary rate, demotion, and or discharge. A formal complaint "alleges one or more of either an infraction of Department rules, regulations, or polices, or an illegal act (TxDPS (2018))." Racial profiling is considered an illegal act under Article 2.132 in the Code of Criminal Procedures and can be a legitimate reason to file a complaint against the trooper.

In 2015, a KXAN investigation found that DPS troopers were " inaccurately recording the race of large numbers of minority drivers, mostly Hispanic, as white" (Collister (2015b)). After this was uncovered, the the House Committee on County Affairs held a hearing where DPS blamed the error on a computer glitch. As a result of the hearing, the percent of White motorists being stopped fell from 18% to 4% (Collister (2015a)).

An important result of the KXAN investigation was that masking was also found in other law enforcement departments in Texas, namely Houston and Austin police departments. This is important as many previous studies do not account for possible cheating behavior in police or trooper forces. This raises the question if whether past reports and research of racial bias are possible under-measuring and under-detecting the existence of bias. Texan troopers now ask drivers for their race rather than using information off of the driver license or their own best judgment (Oyeniyi (2015)).

Masking is easy in motorist stops compared to other points of the criminal justice system. First, the trooper is not required to ask the driver for his or her race. Instead, the trooper is supposed to infer the race based on observable characteristics of the driver. Second, due to the high frequency of stops, troopers or police officers who participate in masking are not checked for accuracy and are less likely to be caught. Usually, only the driver focuses on the content of the ticket. Third, unless the trooper searches the driver and arrests the driver, no other law enforcement officer will see the recorded race.

3 The Model

In this section, I present a straightforward model on trooper search behavior that yield novel tests of racial bias. Suppose we have troopers and motorists; motorists are of race $m \in \{M, W\}$ and each individual trooper of race-$t \in \{M, W\}$. Suppose that among motorists of race m, a fraction π^m choose to carry contraband. This information is available to the trooper along with other pertinent characteristics that are collapsed to a single index $\theta \in (0, 1)$.[*] If a driver of race m is actually carrying contraband, then $\theta \sim f_g^m(.)$; if the driver isn't carrying contraband, then $\theta \sim f_n^m(.)$. I assume that the two densities are

[*]Some examples of these characteristics are age, height, address, gender, the interior of the vehicle, the smell of the driver, whether the driver is under the influence, whether the license plate is in-state, the time and place of the stop, whether the vehicle is rented, and the attitude of the driver.

continuous and satisfy the strict monotone likelihood ratio property. Intuitively, this property implies that a higher index of θ implies a higher probability of driver guilt.

3.1 Search and Masking

Each trooper of race-t can choose to search a motorist after observing the motorist's vector of characteristics, (m, θ). I assume that a trooper wants to maximize the number of successful searches (searches where illicit contraband is found). When a race-t trooper searches motorist of race-m, she incurs a cost of $c_{m,t}$. If the driver is guilty, the trooper receives a benefit, normalized to one such that the cost of the search, $c_{m,t} \in (0, 1)$.

Let G denote the event that a motorist is guilty of carrying contraband. When a trooper pulls over motorist, she observes m and θ. The ex-ante probability the motorist is guilty conditional on the observed m and θ is:

$$\Pr\left(G = 1|m, \theta\right) = \frac{\pi_m f_g^m\left(\theta\right)}{\pi_m f_g^m\left(\theta\right) + \left(1 - \pi_m\right) f_n^m\left(\theta\right)} \tag{1}$$

From the Monotonic Likelihood Ratio Property, $P(G|m, \theta)$ is monotonically increasing in θ.

The trooper then decides to search the motorist of race-m and signal θ based on the expected payoff of searching such that:

$$max\{P(G|m, \theta) - c_{m,t}; 0\} \tag{2}$$

The first term is the expected benefit of searching that motorist and the second term is the payoff for not searching. Therefore, the optimal decision for a trooper of race-t to search a motorist of race-m with observed signal θ if and only if:

$$\Pr\left(G = 1|m, \theta\right) \geq c_{m,t} \tag{3}$$

The trooper has a search threshold θ^* where (3) holds with equality.

Similar to Knowles et al. (2001) and Anwar and Fang (2006)I define two types of racial prejudice. A trooper exhibits racial prejudice if she has a taste or preference for searching motorists of a certain race, which is modeled using the search cost, $c_{m,t}$.

Definition 1. A trooper of race-t exhibits naive racial bias against motorist of race m if $c_{m,t} < c_{m\prime,t}$.

Next, I define statistical discrimination for troopers if they have no taste for racial bias, but still use a different search criteria for motorists of different races.

Definition 2. An unbiased trooper with $c_{m,t} = c_{m\prime,t}$ exhibits statistical discrimination against race m motorist if $\theta^*(m, i) \neq \theta^*(m\prime, i)$ for $m \neq m\prime$

After observing the search outcome, G, the trooper decides whether to mask the motorist of race M as a motorist of race W. Masking incurs a cost of $\mu(\theta)$, which is a function of all observable characteristics of the motorist, aside from race. If the motorist is guilty, $\mu(\theta) > 0$ and is increasing in θ.; if the motorist is not guilty, $\mu(\theta) < 0$ and increasing in θ. The trooper will mask if and only if:

$$c_{M,t} \geq c_{W,t} + \mu_{MW,t}(\theta) \tag{4}$$

which only holds if the motorist is not guilty.

By masking the M motorist as W, the trooper incurs the cost, $c_{W,t}$ plus the masking cost $\mu_{MW,t}$. The trooper has a masking threshold, $\theta^{\mu}_{M,t}$ s.t (4) holds with equality. And,

$$\theta^{*}_{M,t} < \theta^{\mu}_{M,t}$$

Definition 3. A race-t trooper exhibits sophisticated racial bias against motorist of race m if $\mu_{m,m\prime,i} < 0$ for $m \neq m\prime$

3.2 Equilibrium

For the rest of the section, I assume that only race M motorists are masked as W. The *equilibrium search rate* of trooper t against race M motorist, $\gamma_{M,t}$, is given by:

$$\gamma_{m,t} = \pi_m[1 - F^m_g(\theta^*)] + (1 - \pi_m)(1 - F^m_n(\theta^*)] \tag{5}$$

For a race t trooper, the *equilibrium search success rate* is :

$$S_{m,t} = \frac{\pi_m[1 - F^m_g(\theta^*)]}{\pi_m[1 - F^m_g(\theta^*)] + (1 - \pi_m)[1 - F^m_n(\theta^*)]} \tag{6}$$

With masking, the true search rates and the true search success rates are unobservable to the trooper. I will assume for the rest of the section that troopers only mask race M motorists as race W. Therefore the *observed search rate* for race M motorist is:

$$\gamma^{O}_{M,t} = \pi_M[1 - F^M_g(\theta^{\mu})] + (1 - \pi_M)[1 - F^M_n(\theta^{\mu})] \tag{7}$$

For race t trooper t, the *observed* search rate for race W motorist is:

$$\begin{aligned} \gamma^{O}_{W,t} = \pi_W[1 - F^W_g(\theta^*)] + (1 - \pi_W)(1 - F^W_n(\theta^*)] \\ +(1 - \pi_M)[F^M_n(\theta^*) - F^M_n(\theta^{\mu})] \end{aligned} \tag{8}$$

For race t trooper t, the *observed* search success rate for race M motorist is:

$$S^O_{M,t} = \frac{\pi_M[1 - F^M_g(\theta^\mu)]}{\pi_M[1 - F^M_g(\theta^\mu)] + (1 - \pi_M)[1 - F^M_n(\theta^\mu)]} \tag{9}$$

For race t trooper t, the *observed* search success rate for race W motorist is:

$$S^O_{W,t} = \frac{\pi_W[1 - F^W_g(\theta^*)]}{\pi_W[1 - F^W_g(\theta^*)] + (1 - \pi_W)(1 - F^W_n(\theta^*)] + (1 - \pi_M)[F^M_n(\theta^*) - F^M_n(\theta^\mu)]} \tag{10}$$

For race t trooper t, the masking rate for race M motorist is:

$$\phi_{M,t} = \frac{(1 - \pi_M)[F^M_n(\theta^\mu) - F^M_n(\theta^*_M)]}{\pi_m[1 - F^m_g(\theta^*)] + (1 - \pi_m)(1 - F^m_n(\theta^*)]} \tag{11}$$

The second definition of racial bias is based on the masking rate of trooper t. Given the trooper is not Type 1 bias, she will not mask at all. From Definition 1, if the trooper is unbiased, then $c_{m,t} = c_{m\prime,t}$. Therefore, there is no $\mu_{mm\prime,t}$ s.t eq(4) holds.[†]

Suppose a trooper, t is biased against race-M motorists compared to race-W motorist. Then $c_{M,i} < c_{W,i}$ and there exists a $\theta^\mu_{MW,t}$ s.t eq (4) holds. Then we can test for bias by simply testing if the trooper masks at all, or if $\eta_{M,i} = 0$ and $\eta_{W,i} = 0$.

3.3 Testable Implications

Now I derive some simple tests of the model that I will also test empirically. First, if the troopers are not biased such that $c_M = c_W$, then there is no masking threshold, $\theta^\mu_{MW,t}$ such that $c_{M,i} > c_{W,i} + \mu_{W,i}(\theta)$.

If the trooper is racially biased against race-m motorist, then the assumptions of the MLRP provide an intuitive test of racial bias. From the MLRP, since the $\frac{f^m_g}{f^m_i}$ is strictly increasing in θ and $\frac{f^m_g}{f^m_n} \to 0$ as $\theta \to 1$. The MLRP also implies that the cumulative distribution function F^m_g first order stochastically dominates F^m_n

$$\Rightarrow F^m_g(\cdot) > F^m_n(\cdot)$$

$$\Rightarrow 1 - F^m_g(\cdot) < 1 - F^m_n(\cdot)$$

Proposition 1: Detecting bias: If a trooper is racially biased against race-M motorist, then $\phi_{M,i} > 0$ s.t:

$$\phi_{M,t} > 0 \tag{12}$$

[†]Note, all troopers with Type 2 racial bias are also Type 1 biased, but not all Type 1 bias troopers are Type 2.

Next, it is important to be able to test for the direction of the bias.

Proposition 2: Direction of bias: If a trooper t is racially prejudiced against race M motorist, then he/she will mask race-M motorist as W.

This combined with Proposition 1 also implies that the observed search success rate for race-M motorists will be higher than the actual search success rate since the trooper will only mask unsuccessful searches. Suppose a trooper t masks minority (M) motorists as white (W),

$$S^O_{M,i} > S_{M,i}$$

and

$$S_{W,i} > S^O_{W,i}$$

$$\Rightarrow S_{M,i} - S^O_{M,i} < 0; S_{W,i} - S^O_{W,i} > 0$$

To measure the magnitude of bias, I use the masking rate. Suppose trooper i and trooper j are biased against race M motorist, but trooper i is more biased such that $c_{M,i} < c_{M,j}, c_{W,i} = c_{W,j}$, and $c_{M,t} < c_{W,t}$ for $t \in \{i, j\}$. Since both troopers face the same population of race-M motorist and race-W motorist, then this implies that $\theta^{\mu}_{M,i} > \theta^{\mu}_{M,j}$ and $\theta^{*}_{M,i} < \theta^{\mu}_{M,i}$. From formula (6), formula (9), formula(10), and formula (12), this implies that:

$$\phi_{M,i} > \phi_{M,j}$$

and

$$\Rightarrow S_{M,i} < S_{M,j} \text{ and } S^O_{M,i} > S^O_{M,j}$$
$$\Rightarrow S_{M,i} - S^O_{M,i} < S_{M,j} - S^O_{M,j}$$

Proposition 3: Magnitude of bias: If a trooper t is racially prejudiced against race M motorist, then the magnitude of bias is simply:

$$\phi_{M,t} = Masking\ Rate \tag{13}$$

Suppose the trooper of race t is naively biased against race M motorist such that $c_M < c_W$ while $\eta_{M,t} = 0$. Then from Anwar and Fang (2006):

Proposition 4: If neither race M nor race W troopers exhibit Type 2 racial prejudice, then the ranking of $\gamma_{m,M}$ and $\gamma_{m,w}$ nor the ranking of $S_{m,M}$ and $S_{m,W}$ depends on $m \in \{M, W\}$.

3.4 Discussion of the Model

Resampling in the Model

My model assumes that the proportion of motorists of race m carrying contraband is independent of the race of the trooper. From Figures 1 and 2, I observe that the Hispanic troopers are more likely to be assigned near the Texas-Mexico border, which is where most of the Hispanic motorist stops by county occur. Thus, Hispanic troopers are more likely to stop a Hispanic motorist, compared to a black or white motorist. Another issue is that almost 10% of the Hispanic motorist stops are concentrated in Hidalgo County near the Mexico border and another 15% are in Webb, Cameron, Starr, and El Paso County. The other counties only make up <2% of the stops. Thus troopers stationed in these five counties will disproportionately affect the results. Further, since Hispanic troopers are also more likely to be assigned near the border, Hispanic troopers will be disproportionately affected by this.

To overcome this issue, I resample the motorist data so that every trooper now faces the same pool of motorists within each county in a procedure similar to Anwar and Fang (2006). For example, suppose County 1 has 10 Black motorist stops, 30 Hispanic motorist stops, and 60 white motorist stops. County 2 has 20 black motorist stops, 50 Hispanic stops, and 30 white motorist stops. On average in the state, there are 15% blacks, 40% Hispanics, and 45% whites. I reweight the sample so that in County 1, I keep all the black motorists, randomly select 27 out of the 30 Hispanic motorists, and 30 of the White motorists. I follow the same procedure in County 2. Now both counties are equally likely to have black, Hispanic, and white motorists. Now all troopers, regardless of race, face the same pool of motorists within each county.

Race Correction

I only correct the races of motorists originally recorded as white or unknown. I assume that the trooper is using his or her best judgment if he or she records the race as non-white for the following reasons. This allows me to only correct the races of the motorist once. Otherwise, a motorist with the surname Gomez living in a predominantly census block could be corrected as black or Hispanic depending on if I ran the surname analysis first or the geocoding analysis. I will go into further detail my methodology for correcting the race of the motorist in the subsequent section.

Since the KXAN article was published in November 2015, I also graph the average Hispanic stop rate using the observed and corrected races to see if troopers had any idea of journalists uncovering their masking behavior. I find that prior to the news article's release at the second x-intercept, there was a sizable drop in the Hispanic stop rate using the observed and corrected data. This seems to imply that troopers were masking much more in the months preceding the article's release. Specifically, the observed data shows a 20 percentage point drop from 30% to almost 10% whereas the corrected data only shows a 5% drop. For this reason, I only use data preceding the drop, which is from January 2010 - May 2015 for my analysis.

4 Empirical Results

4.1 Data

4.1.1 Stop Data

The Stanford Open Policing Project (SOPP) has collected over 130 million records from 31 state police agencies (Pierson et al. (2017)). The goal of the project is to analyze detailing interactions between police and the public. This information is freely available on the website.

I only use the Texas portion of the SOPP data. While SOPP provides the data from 2006, Texas troopers were not required to record the driver's last name until 2010, so I cannot test for masking behavior prior to 2010. This data contains detailed information on the stop such as latitude and longitude of the stop, time and date of the stop, the reason for the stop, whether a search was conducted and why, if contraband was found, whether an arrest was made, first initial and last name of the trooper recording the stop, and the badge number of that trooper. The data set also has limited information on the type of contraband found: currency, weapon, and other. The unique feature of this data set is it also contains detailed information on the motorist such as: driver's first and last name, address of the driver, recorded race of the driver, make and model of the car, the owner of the car. This becomes important when I do the race correction. I drop Native American, Asian, and Middle Eastern motorists, which is about 1 million of the observations. I also drop stops where the trooper did not record the race of the driver. For reasons I explain in the race correction analysis section, I also only keep Texas, male drivers. Overall, the subset of the data I use contains about 8 million total stops with 3,509 unique troopers.

In Texas, troopers can legally search a vehicle for many reasons aside from probable cause or driver consent. Some of these situations, such as search incidence to arrest, after the car is impounded, or with a warrant, do not fit the framework of the model. One of the assumptions in my model is that motorists are only guilty through finding contraband. If the motorist is arrested prior to searching the vehicle, then that will bias my results. I restrict my definition of search success to only include searches due to probable cause or driver consent.

4.1.2 Race Correction

I use two main methods supported by past literature on using observable characteristics to determine race. These methods are predominantly used in social science and health research to infer patient race (Fiscella and Fremont (2006)). The first method is to use surname analysis, which works well for Hispanic and Asian surnames. I match the driver surnames in my data to the U.S. Census Surnames data set. If the probability of the last name is Hispanic is greater than a certain threshold (75%), I impute the 'corrected' race as Hispanic.[‡] For example, Figure 6 shows an actual ticket from a stop. The driver, Antonio Tovar Mendez, is pulled over for speeding by Officer Salinas and is recorded as a white Male driver. Using the surnames analysis, the probability Antonio is Hispanic, conditional on his last name, Mendez, is 92%. I then correct his race to Hispanic. The advantage of this method is that the correction is fairly quick and simple. But, the main drawback is that this method is only suitable for Asian and Hispanic names and is less effective with females. Thus, I only keep male drivers in my sample.

The second method I employ is geocoding analysis. This method works best for black drivers "because at least half of black Americans continue to live in predominantly black neighborhoods (Fiscella and Fremont (2006))." I use the recorded address of the driver to geocode to a specific latitude and longitude using geocoder.us. I then use that latitude and longitude to map the address to a Block FIPS code using the FCC block finder. I merge this data with the 2010 American Community Survey. If the percentage of Black population in the area is greater than a certain threshold (75%), I correct the race as "black." This method also has a few disadvantages. First, if the trooper did not record the address of the driver ($< 7\%$ of the data), I can't geocode it. Second, the address is inputed by the trooper, which is prone to spelling and typing errors. For example, I found 116 different spellings of the city "Houston," which is the largest city in Texas. Third, this method is also very computationally expensive so I first restrict this analysis to only drivers who live in Texas, which is almost 90% of the stop data. [§]

The effect of the surname and geocoding analysis on the aggregate statistics for stops is shown in Table **??**. Using these methods, I correct over 758,735 of stops as Hispanic, nearly 6,500 number of stops as black, and over 630,000 as white. This has rather large effects in the aggregate stop statistics for Hispanic and white motorists as nearly 1 million white stops are actually from Hispanic motorists. This also has an effect on the search rates as in the observed statistics, Hispanics are searched only 19.5% of the time while after correction, the search rate more than doubles. The difference between the observed and corrected rates for non-white motorist and white motorists do not add up as some troopers simply reported the motorist race as Unknown rather than putting down a race. Of the 132,283 Unknown race, 132,069 I corrected as Hispanic and 214 I corrected as Black.

[‡] As a robustness check, I raise the threshold to higher levels

[§] I also raise this threshold later as a robustness check

4.1.3 Trooper Employment Data

The employment data is from the Texas Department of Public Safety. Unfortunately, DPS only has this information for employees during 2013 - 2015. If a trooper left DPS prior to 2013, I do not have his or her employment information. For troopers in the data, I have the year the trooper was hired, if he or she left the position and why, the salary for each year, which work city he or she was stationed at, the work position for each year, ethnicity of the trooper, the full name of the trooper, and the badge number. The key limitation for this data is that I do not have employment data prior to 2013. I have approximately 2,788 unique troopers of which I can match 2,429.

I merge these two data sets together using the badge number of the trooper. I can match all but 10% of the stop data to the trooper so I only have 11,819,236 observations. If I match the trooper, then I use the information in the earliest matched year to match to prior years, which is appropriate since certain traits of the trooper, such as sex, name, and race, do not change across time. I infer the work city by using the city with the most number of stops for a trooper in a given year. I also drop troopers who have less than 100 stops from 2010-2015. My final data set consists of 7,365,175 stops. After I resample the data set, my final number of observations is 3,413,650 of which 212,292 observations are after May 2015, which I exclude from certain analyses.

4.2 Descriptive Statistics

I present summary statistics of motorist characteristics in Table 1. On average, most motorists stopped are white and male. For searches, the average driver is still male, but Hispanic and whites are stopped and searched at nearly equal rates. Using the corrected data, of all the searches, 45.5% are for Hispanic motorists followed by white motorists with 38.9%. Columns 2 and 3 compare stops that the trooper misrecords the driver race (column 2) to stops where the trooper correctly records the stop (column 3). Hispanic motorists make up 99.2% of the masked stops with the rest of the .8% being Black motorists. In Columns 5 - 7, I show the same probabilities conditional on being stopped *and searched.* The rate of masking is much lower than the rate of masking for stops except for Hispanic motorists. I do not observe any significant differences in masking for Midnight stops.

Columns 4 and 8 show the difference between the Masked and Non-Masked stops. If the difference is greater than zero, that means that motorists with that trait are more likely to be masked. For stops, you are more likely to be masked if you are male, stopped between 12 - 6 am, own the car, Hispanic, and if you drive an old car. Most of the differences are small and less than 10%, but for Hispanic motorists, the difference is 80%. This means of the Hispanic motorists stopped, only 16.6% are recorded with the correct race. For searches, I observe a similar pattern in the traits most likely to be masked except for having being the owner and having an old car. The differences are also less significant. But, I do observe a similar difference for Hispanic motorists. Only 20% of Hispanic motorist vehicle searches are recorded correctly. For the other non-white races, the motorists are more likely to be recorded correctly than incorrectly.

Table 2 shows summary statistics of troopers. Of the 2,429 troopers I was able to match to the data, approximately 61.4% are white, 30% are Hispanic, and almost 9% are black. I drop Asian troopers, which is less than 1% of the data. The force is predominantly male. By trooper race, I find that white troopers are most likely to search at 2% of the time, followed by Hispanic motorists at 1.5%. For masked stops, Hispanic and white troopers mask at near equal rates of 22.5%. Black troopers mask the least at only 21% of the stops.

In the bottom part of the table, I break down the stop and search statistics by trooper position. Ranked officers are only 20% of the highway patrol. The rank of the officer does not have any significant effect on the search rate, but sergeants are significantly more likely to mask in stop when compared to those ranked below. Using the rank of Captain as an example, the interpretation of the probabilities is "If the trooper is a captain, then the search rate is 5%." I observe similar probabilities for searching regardless of trooper rank between 1.4-2% except for Captains at less than 1%. For masked stops, I find that higher rank is correlated with more masking with the exception of Captains. Captains mask 10.6% of the stops whereas troopers mask 22.3% of the time.

4.3 Past tests for racial prejudice

One main advantage of this data set is that I can use past tests of racial bias using both the observed data, the resampled data, and the corrected data. In Table 3, Column 1 shows the results of Knowles et al. (2001) using the observed, raw data. Under their test, I would find bias against Hispanic and white motorists compared to black motorists. Since Hispanic motorists have a lower success rate of 32.7% compared to white search success rate at 42.4%, I would conclude that Hispanic motorists were the most biased against. In Column 2, I show the search success rates broken down across driver's race using the resampled data. Now I find only bias against Hispanic motorists which have a higher success rate compared to Column 1 and no bias against white motorists which have an equal search success rate with black motorists at 45%. Column 3 shows the search success rate after applying the race correction to both Hispanic and black motorists. Here I find different results from the KPT racial bias test. First, the search success rates for white motorists has risen from 45% to nearly 50%. The Hispanic search success rate has fallen by 2% to just under 40% and the black motorist search success rate has also fallen by a small margin to 44.8%. From Column 4, the difference between the search success rates using the observed and corrected data is significant for Hispanic and white motorists. Importantly, if I used only the observed data, I would have incorrectly found no bias against Black motorists and underestimated the magnitude of bias against Hispanic motorists when compared to white motorists.

4.4 My test for racial prejudice

Thus far, I have shown strong evidence that masking distorts the conclusions from past tests of bias. Specifically, I find that trooper's were systematically biased against non-white, especially Hispanic motorists. Using my model, I test if this behavior is linked to bias by comparing the rate of masking conditional on the search outcome and by comparing the search success rate using the observed data and the corrected data. Table 4, I show the results of my main test of bias, which compares the probability of masking conditional on search outcome. The key identification of my test is that biased troopers should differentially mask based on search outcome. My race correction method can not observe whether the race was incorrectly recorded due to bias or due to human error. Thus, if the masking occurs for any reason aside from bias, then I should observe no difference in the masking rate across search outcome. But, if I observe higher rates of masking for unsuccessful searches compared to successful searches, this would provide evidence of biased behavior as dictated by my model. I observe in Table 4 that the probability of masking is higher when the search ends in failure for non-white motorists and this difference is the highest for Hispanic motorists. Specifically, Hispanic motorists are nearly 4% more likely to be masked when the search ends in failure compared to when the search ends in success. For black motorists, this difference is much smaller but significant at 0.4%.

In Table 3, I break down column 3 of Table 4 by trooper race. If the number is positive and significant, this implies bias against the driver's race in the column. I find evidence of bias against Hispanic motorists from Hispanic and white troopers. I find that white troopers are 4% more likely to mask when the search is unsuccessful compared to successful ones. I also find the distortion to be greatest for white troopers, implying white troopers are more biased against Hispanic motorists than Hispanic troopers. Surprisingly, I still find evidence of significance bias for Hispanic troopers. For black motorists, I find that only white troopers are significantly biased. Specifically, they mask 0.6% more for unsuccessful searches compared to successful searches. For black troopers, I find slight evidence of bias against Hispanic motorists, but the difference is insignificant.

To understand the motivation behind trooper's decision to mask, I examine how masking distorts the observed search success rates. Troopers can mask to increase the observed search success rates for non-white motorists or to decrease the search success for white motorists. The outcome is dependent on the size of the search population of each race of motorist. Table **??** shows the difference between corrected and observed search success rates [¶] I find that the masking increased the Hispanic search success rate by 2% for black and white troopers, but the difference is only significant for white troopers. For Hispanic troopers, the masking only increased the Hispanic search success rate by 1.3%. The greatest effects are seen in the difference in the white search success rate. For Hispanic troopers, masking decreased the observed white search success rate by 5%; for white troopers, the white search success rate fell by 4% due to masking. I find no significant differences for black troopers. These results are in agreement with the results from Table 4 and 3. These results also show that troopers mask in order to appear worse at searching white motorists in order to appear less biased.

To measure the magnitude of bias, I use Eq. (13):

$$\phi(R, t|G = 0) - \phi(R, t|G = 1)$$

where R stand for the driver race, and t stands for trooper race. The more biased the trooper is, the more he will mask motorist of race-R when the search ends in failure, $G = 0$, compared to when the search ends in success, $G = 1$. By relying on the difference in masking rate across search outcome, I difference out the possibility the masking occurred due to human error or to non-bias related reasons.

Using Eq. (13), I create a measure of bias for each individual trooper. In order to ensure that troopers who only search a few times during the time period do not bias the estimates, I only use troopers who have conducted at least four searches from 2010 to 2015. Using this measure, I can test possible covariates of bias such as location, trooper quality, trooper rank, and county characteristics. Figure 8 shows the distribution of my estimated measure of trooper bias against Hispanics. The average is 0.154 with a standard deviation of 0.430. The histogram has negative values because my method of estimating masking behavior cannot capture whether the trooper masked the motorist intentionally or not. This means that some instances of misrecording occur because the trooper makes an honest mistake, or perhaps the motorist is actually white while having a Hispanic sounding last name. The histogram also shows that while there are some biased troopers (right side of the histogram), there are also some troopers who are unbiased and only mask due to human error.

To further test if my definition of bias is truly capturing bias, I compare my measure of bias with KPT's measure of bias. If my measure and their measure are indeed the same, then troopers biased with my measure will oversearch Hispanic motorists compared to white motorists. Table 6 shows this exact

[¶] Tables **??** and **??** show the search success rates across motorist and trooper race of the observed and corrected data.

relationship by regressing

$$Y_{ijt} = \alpha + \beta_1 HispBias_j + \beta_2 I(MotoristRace_i) + \\ \beta_3 HispBias_j \times I(MotoristRace_i) + \gamma_t + \phi_c + \gamma_t \times \phi_c + \epsilon_{ijt} \tag{14}$$

where Y_{ijt} is the outcome of a stop of motorist i with trooper j at time t. $HispBias_j$ is the standardized measure of Hispanic bias measured for trooper j. $I(MotoristRace_i)$ is an indicator for motorist race of the stopped driver, i, with white motorists being the excluded category. I also included year fixed effects with γ_t and county fixed effects with ϕ_c. The two outcomes I am interested are search success and search. My results in Table 6 indeed show agreement with KPT's measure of bias. Namely, one standard deviation increase in my measure of bias increases the likelihood of searching a Hispanic motorist by 0.002%; it also reduces the probability of search success for Hispanic motorists by nearly 0.05%, with both estimates being significant. The effect on search success is not trivial as it decreases the overall search success rate by more than 10%.

I also ensure that the relationship between my measure of bias and KPT's measure of bias is not dependent on my census surname cutoff, I vary the threshold I use in the surname analysis at 50%, 75% (the measure I use throughout my analysis), 85%, and 95%. My results in Figure 10 show that even with tighter thresholds, the troopers who engage in biased masking behavior also have a greater search success rate for white motorists than with Hispanic motorists.

Figure **??** shows the geographic distribution of officer level bias by county across Texas. In order to aggregate the officer-level measure of bias to the county level, I aggregated the measure of bias to the county level using a weighted average of every trooper's estimate of bias at the county level, weighted by the number of searches. Troopers who conduct a small proportion of total searches in a county are weighted the least. Using this method, there are six counties that had no masked searches in the entire time period. I find that most counties near the border have bias estimates of greater than 0.25. Interestingly, counties grouped into the highest estimates of bias (> 0.5) are scattered throughout all of Texas and are not concentrated near the border. I labeled the six most populated cities in Texas and find no correlation to bias.

Comparing the geographic distribution of bias to the distribution of all troopers across Texas also uncovers interesting patterns. First, areas where no bias is observed also tend to be areas where very few troopers are patrolling. I also observe areas with low amounts of trooper with high amounts of bias. I find that areas near the border along with close to large cities have the highest patrol rate.

To test if this pattern is driven by institutional bias versus bad apple troopers, I compare the standard deviation of bias of troopers within counties to the proportion of biased stops to total searches. A biased stop is defined as a stop where the trooper incorrectly recorded the motorist race and had an unsuccessful search. If the county has institutional bias, then the county will have a high proportion of biased stops to total searches and a low standard deviation of bias. If the county has a bad apple trooper, then the estimated standard deviation of bias of the county will be high and the proportion of biased stops to total searches will be low. 9 shows that Texas has a mix of counties with bad apple troopers and institutional bias. Certain counties may have institutional bias versus bad apple troopers due to different social norms. Texas is a large state with 254 counties with variety of diverse characteristics. In Table **??**, I test if any county characteristics are correlated to the probability of the county having institutional bias versus bad apple troopers. I find that having low median income of reduces the probability of having institutional bias or bad apple troopers. I also find that counties with high Hispanic populations are positively correlated with having institutional bias. I find that proportion of having at least a high school diploma or a GED have no correlation with institutional bias or bad apples. I find that the employment rate is positively related to bad apples, but not institutional bias.

4.5 Bias and Trooper Characteristics

Another important aspect of bias is to understand how trooper characteristics correlate to bias. In Table **??**, I test if salary and experience is related to bias where experience is measured using the hire year of the trooper. I use the maximum salary in thousands of the trooper from 2013 to 2015. I find slight evidence that more experience and higher salary increase bias. Each additional year in experience increases trooper's bias against Hispanic motorists by .4 percentage points. Salary has stronger effects on bias. Every thousand dollar increase in salary increases bias by 3 percentage points. These results are not robust if I exclude troopers with less than five searches. The coefficients decrease in size and are no longer significant.

5 Conclusion

In this paper, I use the cheating behavior of troopers to uncover their preferences towards searching driver's by motorist race. I develop a new statistical model to use this cheating behavior to measure trooper's racial bias. Unlike past tests of racial bias, my test can measure the magnitude of bias at the individual trooper level. I test my model using a rich data set of Texas highway stop data from 2010 - 2015 merged with Texas Highway Patrol employee administrative data. During this time period, Texas troopers were masking motorist race of non-White motorists as White for certain stops. By comparing these search and search success rates across motorist race, trooper race, observed and corrected, masked and unmasked, I am able to develop a comprehensive test for racial bias.

In my results, I find that lack troopers are unbiased against non-White motorists and white and Hispanic troopers are biased against Hispanic motorists. White troopers are also biased against black motorists. By masking many Hispanic motorists as white, troopers appear less successful in searching white motorists and are able to avoid being labeled biased.

My paper focuses on not only the search decision of troopers but also on the masking decisions. After a trooper stops a motorist, he or she must decide if to search the motorist. Then after observing the outcome, the trooper also must decide if to accurately report the motorist race. My model provides an intuitive test that the more biased trooper is more likely to mask so that he or she appears less bias. In this analysis, I can determine the bias by being able to closely examine these two decisions. I find that with masking, racial bias is underdetected and undermeasured. Investigating the pervasiveness of masking in not only trooper decisions but also across other aspects of the criminal justice system may provide a more intuitive test of racial bias while uncovering previously unmeasurable racial bias.

Masking may not just be present in law enforcement. In any sort of scenario where racial profiling is illegal, this may induce agents to mask the race of the biased group to appear less biased. For example in mortgage lending, mortgage lenders may mask the race of applicants to appear less biased. This masking may not be limited to just race but is also easily extended to other observable characteristics such as income or educational level.

6 Appendix

References

Alesina, A. and E. L. Ferrara (2014). A test of racial bias in capital sentencing. *American Economic Review*.

Anbarci, N. and J. Lee (2014). Detecting racial bias in speed discounting: evidence from speeding tickets in Boston. *International Review of Law and Economics*.

Antonovics, K. and B. G. Knight (2009). A new look at racial profiling: evidence from the Boston police department. *The Review of Economics and Statistics*.

Anwar, S. and H. Fang (2006). An alternative test of racial prejudice in motor vehicle searches: theory and evidence. *American Economic Review*.

Anwar, S. and H. Fang (2015). Testing for racial prejudice in the parole board release process: theory and evidence. *Journal of Legal Studies*.

Arnold, D., W. Dobbie, and C. S. Yang (2018). Racial bias in bail decisions. *Quarterly Journal of Economics*.

Collister, B. (2015a). DPS troopers getting race right after KXAN investigation. *KXAN*.

Collister, B. (2015b). Texas troopers ticketing Hispanic drivers as white. *KXAN*.

Coviello, D. and N. Persico (2013). An economic analysis of black-white disparities in NYPD's stop and frisk program. Working Paper 18803, NBER, https://www.nber.org/papers/w18803.

DPS. *When stopped by law enforcement*. Texas DPS.

Fiscella, K. and A. M. Fremont (2006). Use of geocoding and surname analysis to estimate race and ethnicity. *HSR: Health Services Research*.

Goncalves, F. and S. Mello (2017). A few bad apples? Racial bias in policing. Working Paper 608, IRS Working Papers, http://arks.princeton.edu/ark:/88435/dsp01z890rw746.

Harris, D. A. (1999). Driving while black: racial profiling on our nation's highways. *American Civil Liberties Union Special Report*.

Horrace, W. C. and S. M. Rohlin (2016). How dark is dark? Bright lights, big city, racial profiling. *The Review of Economics and Statistics*.

Jacob, B. A. and S. D. Levitt (2003). Rotten apples: an investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*.

Kalinowski, J., S. L. Ross, and M. B. Ross (2017). Endogenous driving behavior in veil of darkness tests for racial profiling. Working Papers 2017-017, Human Capital and Economic Opportunity Working Group, https://ideas.repec.org/p/hka/wpaper/2017-017.html.

Knowles, J., N. Persico, and P. Todd (2001). Racial bias in motor vehicle searches: theory and evidence. *Journal of Political Economy*.

Knox, D., W. Lowe, and J. Mummolo (2019). The bias is built n: how administrative records mask racially biased policing. Technical report, SSRN.

Lind, D. (2015). The FBI is trying to get better data on police killings. Here's what we know now. *Vox*.

Oyeniyi, D. (2015). State troopers will now just ask drivers their race. *TexasMonthly*.

Persico, N. and P. E. Todd (2005). Passenger profiling, imperfect screening, and airport security. *American Economic Association Papers and Proceedings*.

Pierson, E., C. Simoiu, J. Overgoor, S. Corbett-Davies, V. Ramachandran, C. Phillips, and S. Goel (2017). A large scale analysis of racial disparities in police stops across the United States.

Roland G. Fryer, J. (forthcoming). An empirical analysis of racial differences in police use of force. Working Paper 22399, NBER, https://www.nber.org/papers/w22399.

Shayo, M. and A. Zussman (2011). Judicial ingroup bias in the shadow of terrorism. *Quarterly Journal of Economics*.

TxDPS (2018). Complaint investigation and resolution.

West, J. (2018). Racial bias in police investigations.

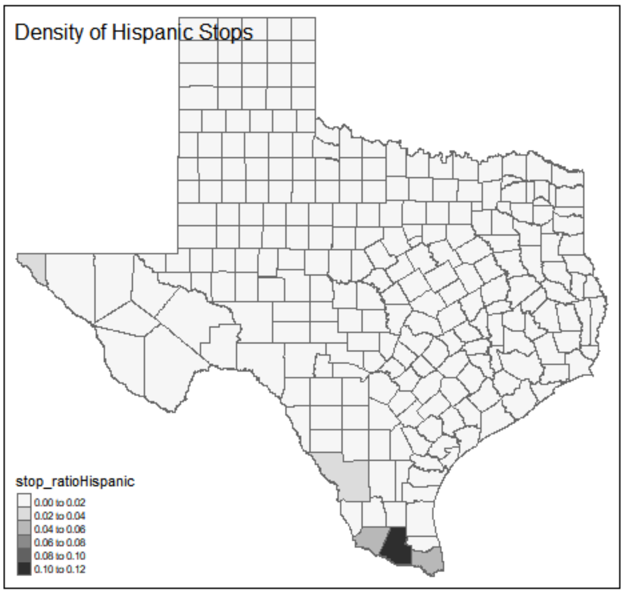


Figure 1: Hispanic Motorist Stop Density by County

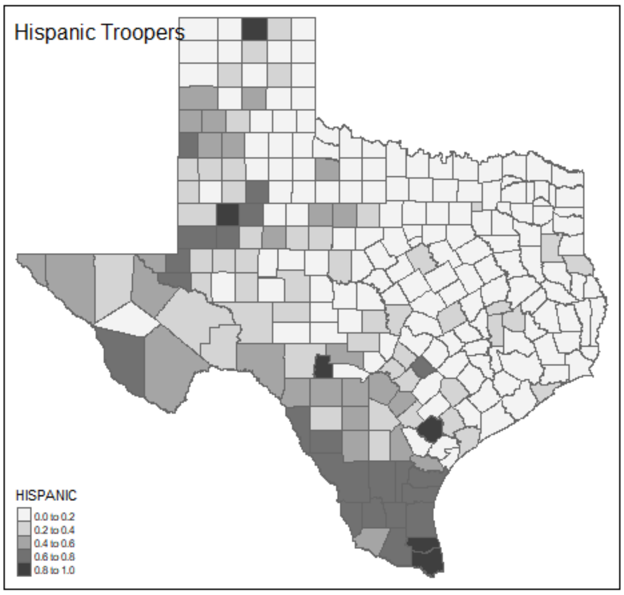


Figure 2: Hispanic Trooper Density by County

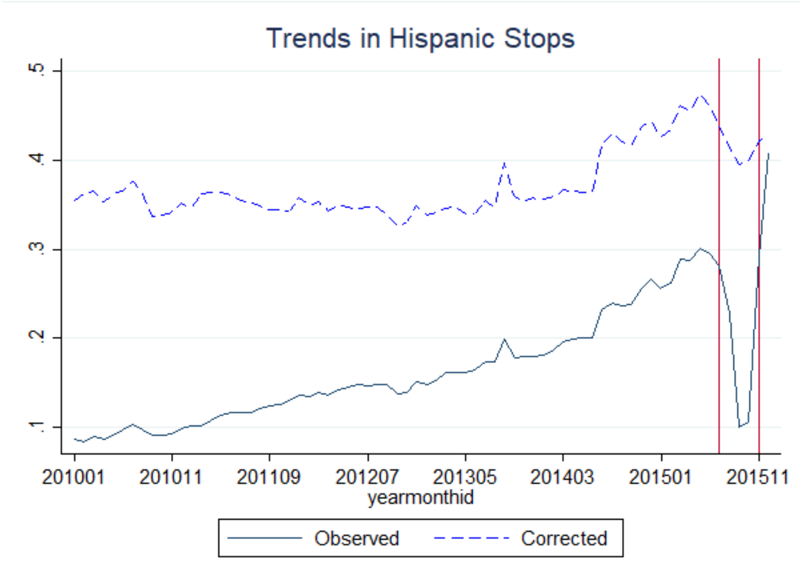


Figure 3: Hispanic Stops Over Time - Observed and Corrected Races

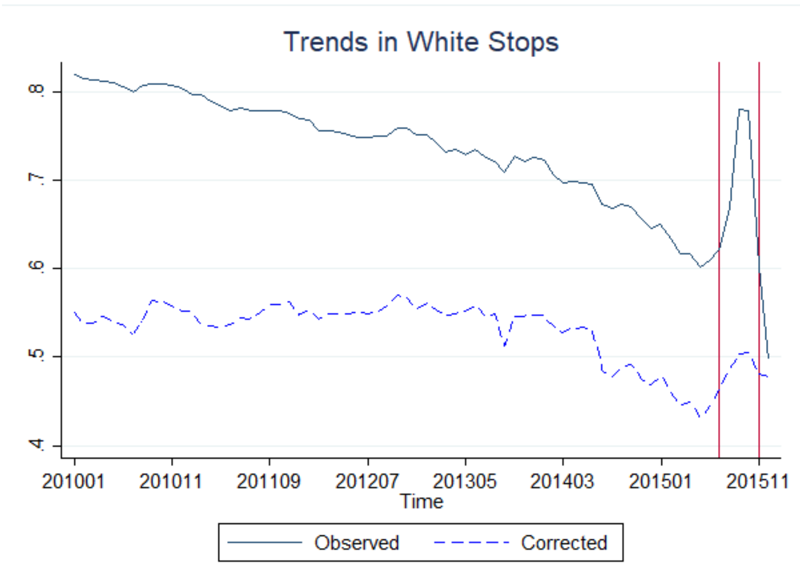


Figure 4: White Stops Over Time - Observed and Corrected Races

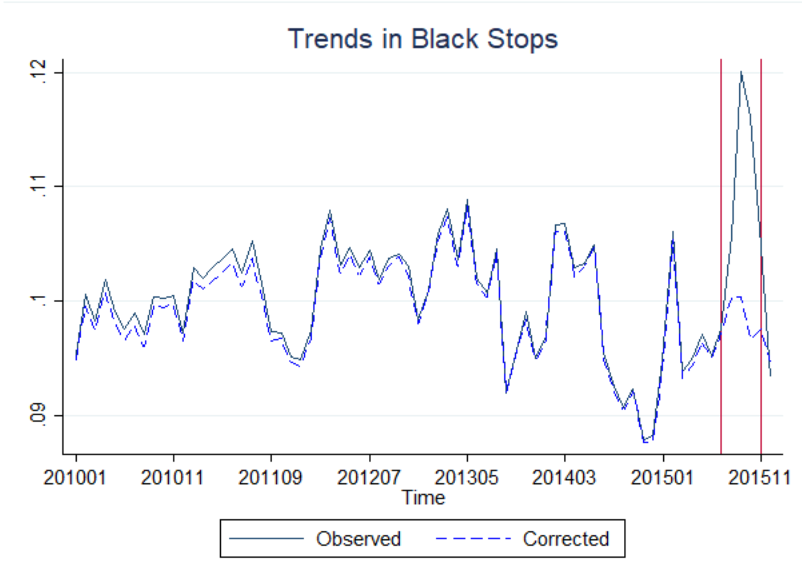
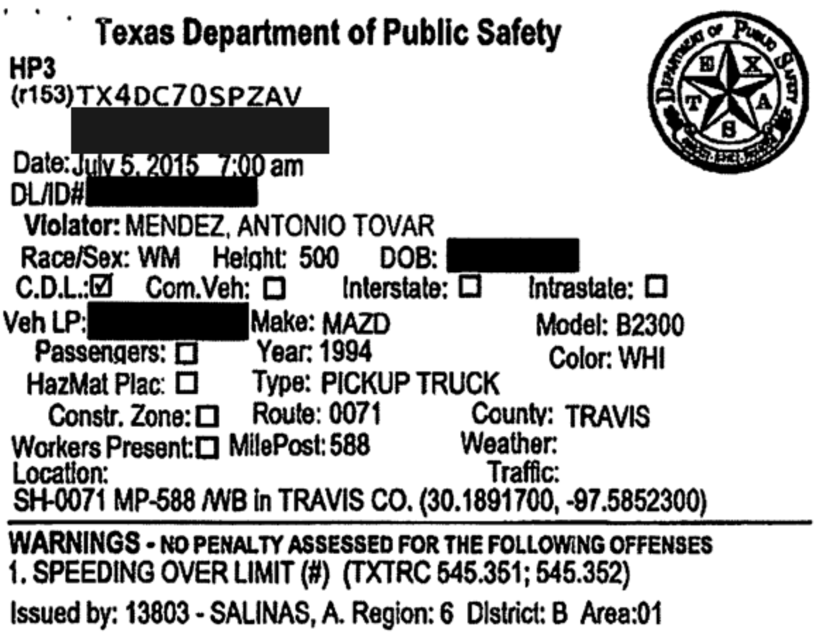


Figure 5: Black Stops Over Time - Observed and Corrected Races

Figure 6: Example of Ticket



Texas Department of Public Safety

HP3
(r153)TX4DC70SPZAV
[redacted]
Date: July 5, 2015 7:00 am
DL/ID# [redacted]
Violator: MENDEZ, ANTONIO TOVAR
Race/Sex: WM Height: 500 DOB: [redacted]
C.D.L.: ☑ Com.Veh: ☐ Interstate: ☐ Intrastate: ☐
Veh LP: [redacted] Make: MAZD Model: B2300
Passengers: ☐ Year: 1994 Color: WHI
HazMat Plac: ☐ Type: PICKUP TRUCK
Constr. Zone: ☐ Route: 0071 County: TRAVIS
Workers Present: ☐ MilePost: 588 Weather:
Location: Traffic:
SH-0071 MP-588 /WB in TRAVIS CO. (30.1891700, -97.5852300)

WARNINGS - NO PENALTY ASSESSED FOR THE FOLLOWING OFFENSES
1. SPEEDING OVER LIMIT (#) (TXTRC 545.351; 545.352)
Issued by: 13803 - SALINAS, A. Region: 6 District: B Area:01

Figure 7: Trends in Masking across Time

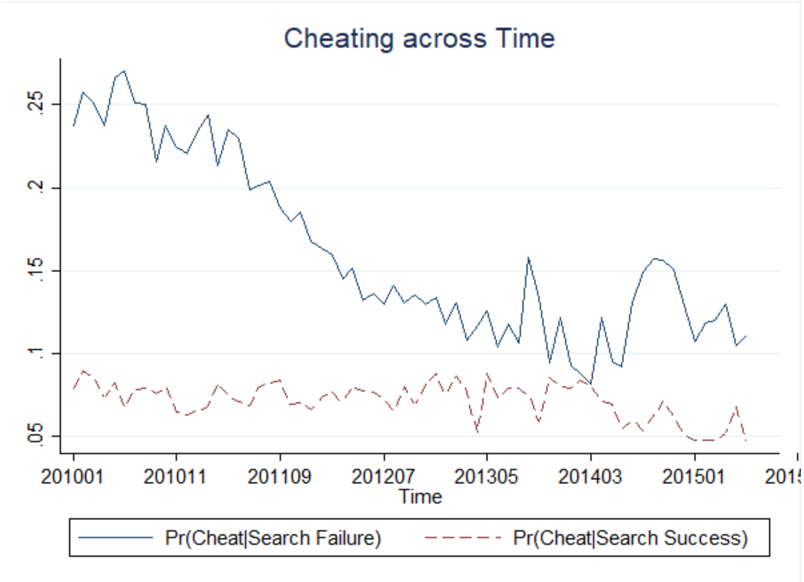


Figure 8: Histogram of Hispanic Bias Measure

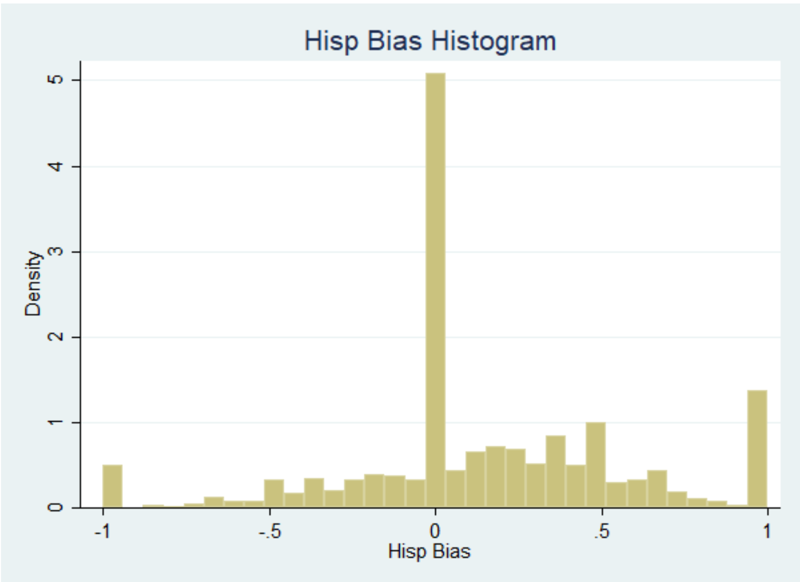
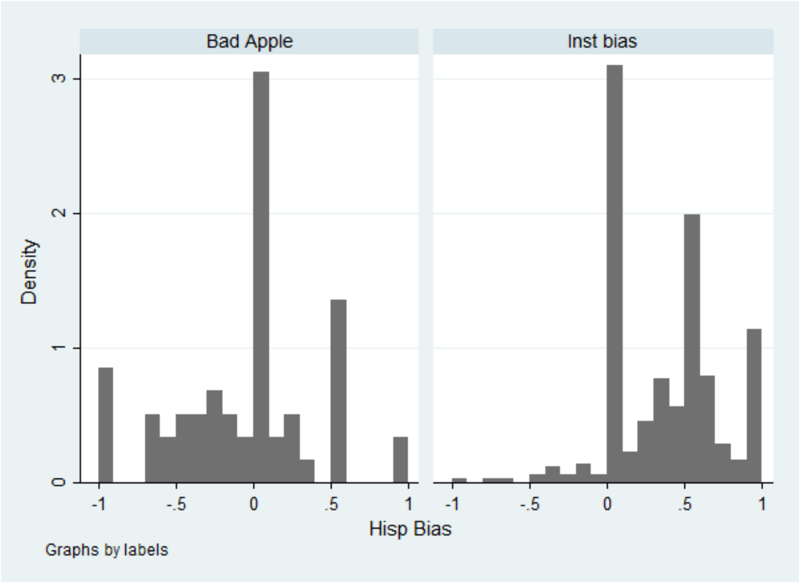
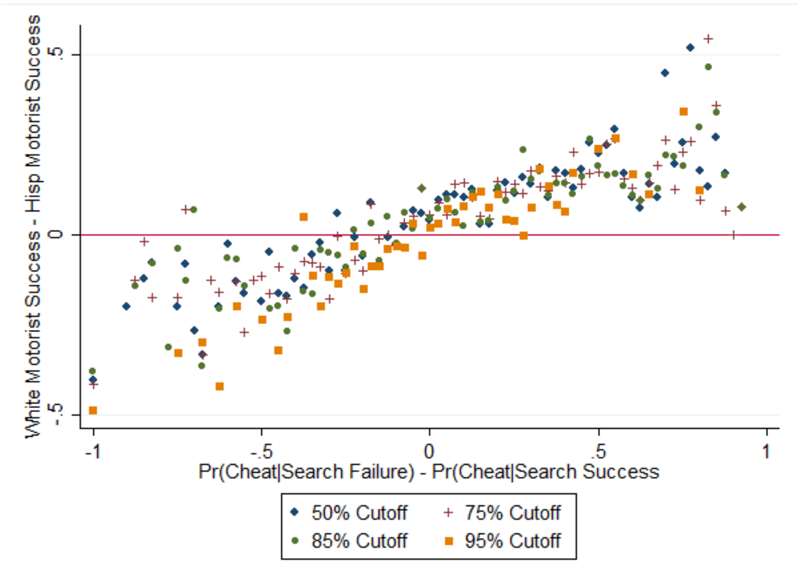


Figure 9: Example of Institutional Bias v. Bad Apple Troopers



The x axis is the measure of Hispanic Bias measured at the trooper x county level with troopers with at least 2 searches within a county. The graph on the left uses observations from county FIPS 48039 and the graph on the right uses observations from county FIPS 48039.

Figure 10: KPT and Masking Measure of Bias with different thresholds



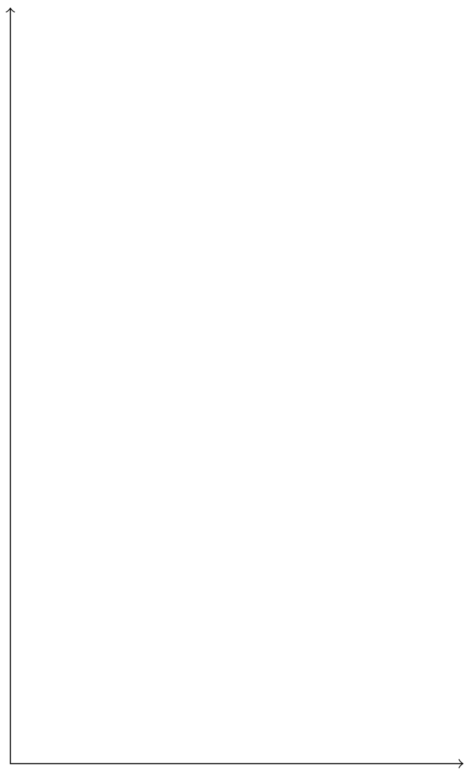


Table 1: Mean of Variables Related to Drivers

Driver Characteristics	All	Searches Only	Δ
Driver Race			
Black	.101	.182	-.081
	(.302)	(.386)	(.001)
Hispanic	.356	.421	-.065
	(.479)	(.494)	(.001)
White	.542	.397	.145
	(.498)	(.489)	(.001)
Stop Characteristics			
Midnight	.088	.136	-.048
	(.283)	(.343)	(.001)
OwnerDriver	.21	.152	.058
	(.408)	(.359)	(.001)
OldCar	.326	.451	-.125
	(.469)	(.498)	(.001)
NewCar	.329	.155	.174
	(.47)	(.362)	(.001)
LuxuryCar	.08	.097	-.017
	(.271)	(.296)	(.001)

Standard deviations are in parentheses. Unweighted means are shown.

Table 2: Mean of Variables Related to Troopers

Troopers' Characteristics	All	Searches	Masked Stops
Trooper Race and Gender			
Black	.087	.009	.143
	(.282)	(.092)	(.35)
Hispanic	.304	.013	.308
	(.46)	(.114)	(.461)
White	.609	.016	.147
	(.488)	(.126)	(.355)
Male	.943	.015	.189
	(.231)	(.121)	(.391)
Trooper Rank			
Captain	.004	.03	.259
	(.065)	(.17)	(.438)
Lieutenant	.014	.024	.326
	(.115)	(.154)	(.469)
Sergeant	.112	.019	.252
	(.315)	(.138)	(.434)
Corporal	.095	.013	.194
	(.293)	(.112)	(.395)
Trooper	.694	.015	.184
	(.461)	(.12)	(.388)
ProbTrooper	.08	.014	.194
	(.271)	(.115)	(.395)
NoRank	.002	.016	.23
	(.039)	(.125)	(.421)
Total Observations	2591		

Only merged observations are shown. Trooper rank uses the highest rank the trooper obtained during 2010 - 2015

Table 3: Search Success Rates across Driver's Race

		Search Success Rate		
	Raw	**Non Border Counties Only**		
	Observed	Observed	Corrected	Δ
Driver Race				
Black	.440	.444	.442	.002
	(.496)	(.497)	(.497)	(.005))
Hispanic	.319	.399	.380	.019
	(.466)	(.490)	(.485)	(.005)
White	.420	.460	.495	-.035
	(.493)	(.498)	(.5)	(0.003)

Unweighted means are shown using the resampled data. Standard deviations are in the parentheses. Columns 2 and 3 exclude border counties. Columns 1-3 use data from January 2010 - June 2015.

Table 4: Difference in Mask Rate by Search Success

Driver Race	$Pr(Mask\|Failure)$	$Pr(Mask\|Success)$	Δ
Black	.014	.009	.005
	(.116)	(.096)	(.001)
Hispanic	.608	.575	.033
	(.488)	(.494)	(.006)

Unweighted means are shown using non-border counties only from January 2010 to June 2015. Standard deviations are in parantheses

Table 5: Hispanic Bias and Trooper Race

	(1) $BlackMotorists$	(2) $HispMotorists$
Hisp Troopers	-0.0018	0.0323^{***}
	(0.0020)	(0.0120)
Black Troopers	-0.0013	-0.0354
	(0.0018)	(0.0306)
Constant	0.0075^{***}	0.3878^{***}
	(0.0004)	(0.0045)
Observations	19126	42696

Dependent variable is the proportion of masked stops that end in failure divided by total searches. The regression includes county FE, year FE, and county x year FE. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. Hisp Troopers and Black Troopers are indicator variables for the trooper's race with white troopers being the omitted category. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 6: Bias and Trooper Quality

	(1) $Pr(Search)$	(2) $Pr(SearchSuccess)$
Hisp Bias	-0.001	-0.136^{***}
	(0.001)	(0.007)
I(Black=1)*Hisp Bias	-0.001	0.013^{*}
	(0.001)	(0.007)
I(Hispanic=1)*Hisp Bias	0.002^{***}	-0.084^{***}
	(0.001)	(0.012)
Constant	0.012^{*}	0.386^{***}
	(0.006)	(0.035)
Observations	5207938	105174

Hisp Bias is the normalized measure of Hispanic bias for each trooper. The regression includes county FE, year FE, and county x year FE along with driver race FE. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. $I(Black = 1)$ and $I(Hispanic = 1)$ are indicator variables for the driver's race. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 7: Hisp Bias on Labor Outcomes

	(1) Hisp Bias	(2) Hisp Bias	(3) Hisp Bias
Experience	0.0011		
	(0.0019)		
Salary		0.0166	
		(0.0140)	
Prob. Troop			-0.2086^{***}
			(0.0796)
Corporal			-0.0432
			(0.0316)
Sergeant			0.0219
			(0.0345)
Lieutenant			-0.0200
			(0.1813)
Observations	1766	1766	1766

Regression has robust standard errors. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 8: Hisp Bias on Labor Outcomes - Panel

	(1) $Pr(LeftForce)$	(2) $Pr(MovedCities)$	(3) $Pr(RankUp)$
Hisp bias	0.0203^{*}	0.0090	-0.0059
	(0.0113)	(0.0154)	(0.0128)
Prob. Troop	-0.1432^{***}	0.0445	0.8252^{***}
	(0.0470)	(0.0695)	(0.0406)
Corporal	0.0660	0.0581	0.0656
	(0.0426)	(0.0437)	(0.0493)
Sergeant	-0.0597	0.0728	-0.0470
	(0.0447)	(0.0866)	(0.0834)
Lieutenant	0.0211	1.0309^{***}	-0.1858^{***}
	(0.0360)	(0.0391)	(0.0335)
Black Trooper	-0.0125	0.0456	-0.0040
	(0.0487)	(0.0659)	(0.0501)
Hisp Trooper	0.0020	0.0218	-0.0141
	(0.0370)	(0.0406)	(0.0336)
Observations	999	999	999

Regression has robust standard errors. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 9: Hisp Bias on Labor Outcomes - Transition Matrix

	(1)	(2)	(3)	(4)	(5)
ProbTrooper	0.021				
	(0.030)				
Trooper		0.065			
		(0.053)			
Corporal			−0.034		
			(0.039)		
Sergent				0.093	
				(0.095)	
Lieutenant					0.000
					(.)
Observations	961	961	961	961	961

Dependent variable is the probability of increasing in rank. Includes work city fixed effects and is clustered at the work city level. Regression has robust standard errors. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 10: Search Decision for Biased Stops

	(1) $HispMotorists$	(2) $BlackMotorists$
Incident to Arrest	−0.0141***	−0.4893***
	(0.0035)	(0.0154)
Inventory	−0.0160***	−0.4933***
	(0.0026)	(0.0122)
Probable Cause	−0.0098***	−0.2811***
	(0.0016)	(0.0124)
Luxury Vehicle	−0.0043***	−0.0030
	(0.0015)	(0.0099)
Owner is Driver	−0.0013	0.0663***
	(0.0020)	(0.0074)
Midnight Stop	0.0013	−0.0312***
	(0.0017)	(0.0082)
Constant	0.0140***	0.5588***
	(0.0010)	(0.0045)
Observations	15008	31938

Dependent variable is the proportion of masked stops that end in failure divided by total searches. The regression includes county FE, year FE, and county x year FE. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. Incident to Arrest, Inventory, and Probable Cause are dummy variables for search reason with Search Consent being the omitted category. Midnight Stop is an indicator variable equal to 1 if the stop was conducted between 10 PM and 6 AM. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

	(1)	(2)
	Inst Bias	Bad Apples
% Hisp	0.18	-10.49***
	(0.51)	(-2.68)
% Black	0.19	-12.20***
	(0.53)	(-2.93)
% no health ins	1.19	2.10
	(1.60)	(0.23)
% HS diploma	0.40	-1.23
	(0.44)	(-0.13)
Median HH inc (10000s)	-0.01	0.01
	(-0.24)	(0.03)
% Employed	-0.96	-1.18
	(-0.98)	(-0.11)
% older than 16	-0.19	-13.72
	(-0.17)	(-0.98)
Population (10000s)	-0.00	-0.02***
	(-1.45)	(-3.48)
Border County	0.28**	1.58
	(2.29)	(1.41)
Observations	173	209
R^2	0.168	0.091
F	3.24	3.23

Inst Bias is an indicator variable equal to one if the proportion of biased stops to total searches is greater than 40% and the standard deviation of bias within the county is less than 0.25. Inst Bias is equal to zero otherwise and considered missing if Bad apple is equal to one. Bad apple is equal to one if the standard deviation of bias is greater than 0.38 and the proportion of biased stops is less than 27% and equal to zero otherwise and if Inst Bias is not equal to one. County level variables are from the 2010 - 2015 ACS. Counties with less than 4 stops annually were excluded.

Regression uses robust standard errors. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$