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

Elizabeth Luh

October 15, 2019

Abstract

Law enforcement officers may misreport the race of people that they engage with in order to evade detection of racial bias. Using a unique event in Texas where troopers were caught misreporting minority motorists as white, I propose a new test of racial bias in the presence of misreporting that is well-suited to explore the rich heterogeneity in bias behavior. I find bias against all minority motorists, but especially against Hispanic motorists. I disaggregate the bias to the individual trooper level and find that nearly 60% of troopers were engaging in this behavior. Using my trooper-level measure of bias, I identify causal relationships between bias and labor outcomes using a panel data set of trooper employment outcomes. I also find that a rule change to trooper stop recording policy in response to the misreporting was effective in curbing lying behavior in troopers and reducing bias against Hispanic motorists.

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)). 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), Depew et al. (2017)), parole (Anwar and Fang (2015)), and capital punishment (Alesina and Ferrara (2014)). While researchers and policy makers have focused on detecting and measuring racial bias, little attention has been given to the response of law enforcement officers to this heightened scrutiny.

This is important as law enforcement officers have considerable discretion when recording interactions with motorists. Biased officers may be motivated to distort the recorded events or in more extreme cases, completely mis-record events, to appear less biased. In my paper, I focus on a documented instance where Texas troopers were caught purposefully misreporting minority motorists race as white to appear less biased (Collister (2015b)). This was possible because from 2010 to 2015, Texas troopers were allowed to record motorist race based on their own best judgment, allowing for a discrepancy between motorist's actual race and the recorded race. This is a major concern since many current tests of racial bias rely on the assumption that the officer has recorded the incident as factually and impartially as they can.

I am the first to show how this intentional misreporting allowed biased troopers evade detection of racial bias. To create a model of racial bias, I adapt the framework of Anwar and Fang (2006) and Knowles et al. (2001) that allows the misreporting to reduce the cost of over-searching minorities. Using this, I identify bias using the differential misreporting rate by search outcome. The intuition of my model is that misreporting is only profitable when the search ends in failure since misreporting a failed search as white will increase the search success rate for minorities. Thus my measure of bias is the difference in misreporting rate when the search ends in failure failure compared to when the search ends in success. With a statistical model, I show that only biased troopers will misreport in such a manner. Using my model, I find that troopers were engaging in this behavior with all minority motorists in Texas, but most frequently with Hispanic motorists.

Guided by my model, I am able to create a measure bias at the individual, trooper level using the difference in misreporting rate by search outcome. I use an approach similar to Goncalves and Mello (2017), which tests for racial bias with leniency in speeding tickets. With my measure, I test for trooper characteristics associated with bias, such as race and rank, and also show causal effects of racial bias on labor outcomes. Taking advantage of the panel-like structure of my employment characteristics, I also track trooper's career over time in relation to their bias. I further my analysis by merging complaint data to trooper data to test if biased troopers are more likely to have complaints filed against them.

I also exploit the discovery of the misreporting as a unique event study. KXAN, an NBC-affiliated news agency in Austin, published an article revealing this fraudulent practice on November 8th, 2015. Texas' state legislature's response was swift, and by November 15th, a hearing was called where DPS denied any institutional knowledge of this practice. To prevent this behavior from reoccurring again, DPS changed its race recording policy on November 23rd to require troopers to ask driver's to verbally give their race, whereas before, motorist race was recorded based on the trooper's best judgment. I show that this simple rule change had dramatic effects on the search behavior of biased troopers compared to unbiased troopers.

The timing of these events and the structure of the data allows me to answer two additional questions. The first is, how do biased troopers change their behavior after the rule change? The second is, how do biased troopers fare in terms of labor outcomes after misreporting is not allowed? The first question is particularly interesting because misreporting minority motorists reduces the risk of being labeled as biased for the troopers. Thus, I test if the risk of punishment for racial bias is enough to change trooper's search behavior of minority motorists after the policy.

I test my model on a unique data set of 13 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 from 2010 to 2015. I use multiple data driven approaches that leverage the surname and home address of the driver to uncover misreporting in millions of recorded stops between troopers and motorists. Since I cannot perfectly identify when troopers are purposefully misreporting, I rely on the differential misreporting rate by search outcome to measure bias. I further augment my data set using publicly available highway stop data from 2016 - 2017 from Texas DPS and 2019 trooper employment data using publicly available salary data from the Texas Tribune. Additionally, I include complaint data from 2010 to 2015 that I can match to my trooper data using their badge number.

My results show that misreporting distorted search success rates significantly for Asian, Hispanic, and white motorists, with strongest effects on the search success rate for white motorists. While the misreporting increased the search success rates for Black and Hispanic motorists by .1 percentage points and 1 percentage point respectively, the white search success rates decreased by 6 percentage points as a result of the misreporting. I also find that misreporting was most effective and used most frequently with Hispanic motorists, who were 2 percentage points more likely to be misreported when the search ended in failure compared to when the search ended in success. I find that misreporting was used less against black motorists with a differential misreporting rate of only .1 percentage point. Since Asian and black motorists make up a small proportion of stopped motorists in my data set, I focus the rest of my results on Hispanic bias. Surprisingly, I find no significant differences in the misreporting by trooper race, including for Hispanic motorists, which I interpret as evidence of like-race bias. I also find no significant differences in bias on employment outcomes, except for trooper rank with the most biased troopers being unranked, non-probationary troopers.

For my results on bias and trooper labor outcomes, I find that bias is higher for troopers who were hired earlier, which I interpret as a positive relationship between experience and bias. I also find that one standard deviation in bias is associated with $100 more in monthly salary, even after controlling for rank. I also find that one standard deviation is correlated with 1 percentage point higher probability of having a complaint filed against you, which is an increase of 15%. Taking advantage of the panel structure of the data, I identify a causal relationship between bias and trooper's labor outcomes both before and after rule change. I do not find any significant effects of bias on employment outcomes prior to the rule change. But, I do find that one standard deviation of bias was associated with a 1% chance higher probability of having a complaint filed against the trooper.

After the rule change, my results show adverse consequences for biased officers' labor outcomes. I find that one standard deviation of bias reduces salary increases by $30, even when controlling for rank in 2015. I also find that the probability of increasing in rank falls by 5 percentage points for every standard deviation in bias. I do not find any significant effects on trooper's probability of staying in the force after the rule change. To the best of my knowledge, my results are the first to be able to track career progression by bias.

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 and can only measure average bias. My paper is the first to my knowledge to test for misreporting behavior in troopers and to explicitly link this behavior to 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 finally conclude in Section 5.

2 Background

2.1 Literature Review

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, where a successful search is defined as a search where illicit contraband is found. Such *statistical discrimination*, where race is useful as an indicator for whether the motorist is actually guilty, is legal in the U.S. Thus distinguishing if a motorist search is driven by statistical discrimination or racial bias of the law enforcement officer is the fundamental question for policy makers and researchers.

Attributing observed disparities in the treatment of citizens of different race to bias is difficult for a few 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 variables, such as driver demeanor, which can affect the outcome of trooper-driver interactions, are unobservable to researchers. Since researchers have no way for controlling for all these omitted variables, identifying racial bias on the individual level is difficult 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.

One approach that much of the past literature has used to distinguish between racial prejudice and statistical discrimination is 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 for minority motorists carrying contraband is smaller than the likelihood for white motorists. In other words, if racial prejudice is the reason for racial profiling, then 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. Despite this straightforward approach, the application of this test is difficult since the researcher can never observe the marginal motorists. Identifying the marginal motorist requires knowing all the variables that could influence a trooper's decision to search, which is impossible.

Past empirical studies have acknowledged this issue when testing for racial bias (Anwar and Fang (2006), Knowles et al. (2001), Antonovics and Knight (2009)). 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 identify 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 cannot measure the magnitude of bias on the individual trooper level.

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. Another method by West (2018), uses 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 in teachers. This type of motivation and incentives is key to understanding why troopers may choose to misreport 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 searches where the motivation for bias may be different.

Given that mine is possibly the first to document such severity of misreporting behavior in troopers, this may provide motivation to reexamine past work in racial bias in motorist stops. With misreporting, past literature may still be under-detecting the existence of bias.

2.2 Misreporting and Highway Troopers in Texas

Texas Highway Patrol is a division of the Texas Department of Public Safety, which is responsible for enforcing state traffic laws and commercial vehicle regulation on highways of Texas. They currently employ over 2800 troopers in Texas divided across 6 regions in Texas, with a separate region for their headquarters in Austin. The department is responsible for licensing of drivers, vehicle inspections, and handgun licensing.

To become a trooper, a person must complete recruit school or transfer from prior law enforcement service. New hires spend some at least one year as probationary troopers before being permanent assignments. After the one year probationary period, troopers take their final exam and are promoted to trooper.

With every four years, troopers can be promoted to different level of trooper classes and to different ranks, which include salary increases. Salary amounts are determined by years in the force and rank. Ranks or classes of troopers are similar to military ranks and go from trooper, corporal, sergeant, lieutenant, captain, and major. In general, only troopers in good standing (no complaints, no disciplinary actions, no demotions) are promoted. Unlike other state police agencies, Texas legislature sets the salary of troopers, rather than the individual agencies. With each salary promotion, troopers can be moved to different stations across the state to fill availability. Troopers are allowed to have some say in the choice of where they are stationed after significant changes in DPS in 2012. Prior to 2012, station assignment was based on availability and need.

Due to Texas' proximity to the Mexican border, Texas Highway Patrol heavily participates in increased law enforcement along the shared border. Since 2014, DPS has sent troopers from across Texas to the border to serve for approximately one week through various operations. Often, these operations are multi-department efforts such as Operation Strong Safety, which was conducted jointly between DPS, Texas State Guard, and the National Guard (Benen (2014)). The main goal of the operation was to reduce drug trafficking and undocumented immigration across the Texas-Mexico border (HSIN (HSIN)).

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.

On November 8th, 2015, a KXAN published the results of their investigation of DPS, which found that troopers were " inaccurately recording the race of large numbers of minority drivers, mostly Hispanic, as white" (Collister (2015b)). Texas troopers were already under scrutiny due to the death of Sandra Bland in jail after being pulled over for failing to signal a lane change (Sanchez (2015)). One week after the misreporting was uncovered, the House Committee on County Affairs held a hearing where DPS blamed the error on a computer glitch. As a result of the hearing, DPS changed its policies to require troopers to ask drivers to provide their race, rather than recording it based on the trooper's best judgment. This policy went in effect by November 23rd; as a result of the policy, the percent of white motorists being stopped fell from 18% to 4% by 2016 (Collister (2015a)).

An important result of the KXAN investigation was that misreporting was also found in other law enforcement departments in Texas, namely the Houston and Austin police departments. Thus, it is not out of the question to test for possible misreporting behavior in police or trooper forces in other state and law enforcement agencies. This raises the question if whether past reports and research of racial bias are possible under-measuring and under-detecting the existence of bias. Less than a month after the publication of the article, DPS changed its policies to require Texan troopers to now ask drivers for their race rather than using their own best judgment (Oyeniyi (2015)).

Misreporting 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 misreporting 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 $\theta \in (0,1)$.[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 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 misreporting

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:

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

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:

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

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 $c_{m,t} < c_{W,t}$.

[1] 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.

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 $\theta^*(m,i) < \theta^*(W,i)$.

After observing the search outcome, G, the trooper decides whether to misreport the motorist of race M as a motorist of race W. misreporting 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. Intuitively, the trooper would be more likely to be caught misreporting if she misreported successful searches as these may end in arrest or other high-profile outcomes. This is why $\mu > 0$ when the search is successful. The trooper will misreport 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.
A trooper can hide unsuccessful searches of motorists of race M by misreporting the motorist as W. By misreporting the M motorist as W, the trooper incurs the cost, $c_{W,t}$ plus the misreporting cost $\mu_{MW,t}$. The trooper has a misreporting 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 $m \neq m\prime$

This is a more sophisticated form of bias because troopers who exhibit this form of bias, take action to hide or reduce the appearance of bias.

3.2 Theoretical Implications

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

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

where $F^m_n(\theta)$ and $F^m_n(\theta)$ is the cumulative distribution for motorists not carrying contraband and motorist carrying contraband, respectively.

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

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

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

$$\gamma_{M,t}^O = \pi_M[1 - F_g^M(\theta^\mu)] + (1 - \pi_M)[1 - F_n^M(\theta^\mu)] \tag{7}$$

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

$$\begin{aligned} \gamma_{W,t}^O = \pi_W[1 - F_g^W(\theta^*)] + (1 - \pi_W)(1 - F_n^W(\theta^*)] \\ +(1 - \pi_M)[F_n^M(\theta^*) - F_n^M(\theta^\mu)] \end{aligned} \tag{8}$$

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

$$S_{M,t}^O = \frac{\pi_M[1 - F_g^M(\theta^\mu)]}{\pi_M[1 - F_g^M(\theta^\mu)] + (1 - \pi_M)[1 - F_n^M(\theta^\mu)]} \tag{9}$$

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

$$S_{W,t}^O = \frac{\pi_W[1 - F_g^W(\theta^*)]}{\pi_W[1 - F_g^W(\theta^*)] + (1 - \pi_W)(1 - F_n^W(\theta^*)] + (1 - \pi_M)[F_n^M(\theta^*) - F_n^M(\theta^\mu)]} \tag{10}$$

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

$$\phi_{M,t} = \frac{(1 - \pi_M)[F_n^M(\theta^\mu) - F_n^M(\theta_M^*)]}{\pi_m[1 - F_g^m(\theta^*)] + (1 - \pi_m)(1 - F_n^m(\theta^*)]} \tag{11}$$

The second definition of racial bias is based on the misreporting rate of trooper t. Given the trooper is not naively biased, she will not misreport 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.[2]

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 $\theta_{MW,t}^\mu$ s.t eq (4) holds. Then we can test for bias by simply testing if the trooper misreports at all, or if $\eta_{M,i} = 0$ and $\eta_{W,i} = 0$.

[2]Note, all troopers with sophisticated racial bias are also naively biased, but not all naively bias troopers are sophisticated.

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 misreporting 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:

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

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 misreport 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 misreport unsuccessful searches. Suppose a trooper t misreports 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 misreporting 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 Eq (6), Eq (9), Eq (10), and Eq (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:

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

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 naive 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 Data

3.4.1 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 use the Texas portion of the SOPP data because Texas was the only state where troopers were caught misreporting. 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 misreporting behavior prior to 2010. The data contains detailed information on the stop such as latitude and longitude of the stop, make and model of the car, the owner of the car, 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. Pierson et al. (2017) courteously provided the raw version of the data, which had the driver's full name and home address. This becomes important when I do the driver race estimation. I drop Native American and Middle Eastern motorists, which is about 30,000 observations. For reasons I explain in the race correction section, I also only keep male drivers. Overall, the subset of the data I use contains about 9 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. Because of this, I restrict my definition of search success to only include searches due to probable cause or driver consent.

I also augment the SOPP data with 2016 - 2017 highway stop data from the Texas Department of Public Safety. This data has identical information to the SOPP data, but does not have the driver's full name or addresses in order to protect the privacy of the driver's in the data set. The new data set contains additional information such as whether the driver was a fugitive, the sergeant in charge of the area of the stop, the alleged speed, the judge assigned to the case, and the court date and location. I also drop the female driver's from this data set to maintain consistency with the SOPP data. Since the stops occurred after the misreporting was revealed in November 2015, I take the driver's races as given. My primary purpose for including the publicly available data is to measure trooper's change in stop behavior after the publication of the article.

3.4.2 Race Estimation

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), Freedman et al. (2018)). 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 'estimated' race as Hispanic.[3] For example, Figure 5 shows an actual ticket from a stop. The driver, with last name Mendez, is pulled over for speeding by Officer Salinas and is recorded as a white, male driver. Since, the probability this driver 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 since married women tend to change their last names to that of their husband's. Thus, I only keep male drivers in my sample.

The second method I employ is geocoding analysis, which I only use on to uncover black drivers "because at least half of black Americans continue to live in predominantly black neighborhoods (Fiscella and Fremont (2006), Glaeser and Vigdor (2001))." 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 (67%), I correct the race as "black." I use 67% since Fiscella and Fremont (2006) found that with "block

[3] As a robustness check, I raise the threshold to higher levels in later sections.

groups where more than two-thirds of the residents were black... 89 percent were classified correctly."[4] 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 computationally expensive so I restrict this analysis to only drivers who live in Texas, which is approximately 90% of the stop data.

For every stop in the data set, I only use one race correction method. This is to prevent the estimated race from depending on the order of the race correction. For example, if a driver with a Hispanic last name who lived in a predominantly black block FIPS area was mis-recorded as white, then he would be corrected as black or Hispanic depending on if I used the geocoding or surname analysis first. There are only 427 drivers who were recorded as white with a Hispanic surname living in a predominantly black neighborhood so there's no significant difference if I were to use both methods or changed the order of applying the analysis. I also only correct the races of motorists originally recorded as white or unknown. 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.

Lastly, I show the how the misreporting affected the observed stop rates of motorists by race by comparing stop rates before and after the DPS rule change. Figure 4 shows the raw, time trend of the stop rate by driver race using the recorded races for Asian, black, Hispanic, and white motorists from 2010 - 2017. The dashed, vertical line indicates the year-month of Sandra Bland's stop. During the time period after her stop and the publication of the article, Texas troopers appeared to significantly increase their misreporting behavior. The second vertical line denotes year-month of the publication of the article and DPS' rule change. After November 2015, the Hispanic stop rate rises to over 40% while the white stop rate falls to nearly equal levels. I observe no noticeable changes for the Asian or black motorist stop rate before and after November 2015. Since only Hispanic motorists have enough observations that are misreported, I focus the rest of my analysis on bias on Hispanic motorists. Since I also cannot discern if the spike in misreporting prior to November 2015 is due to her death or whether there was a shift in behavior, I only use stops from January 2010 to June 2015 for the rest of my analysis.

3.4.3 Trooper Employment Data

The employment data is from the Texas Department of Public Safety. Unfortunately, DPS only has this information for employees after 2013. If a trooper left DPS prior to 2013, I do not have his or her

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

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. I have approximately 2,789 unique troopers of which I can match 2,578 to the stop data.

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. My final number of observations is 7,685,007 after dropping observations after June 2015 for reasons listed in the prior section.

3.5 Descriptive Statistics

I present summary statistics of motorist characteristics in Table 1 using the estimated races. On average, most motorists stopped are white, but this pattern doesn't carry over to searches. Instead I find that conditional on being stopped, Hispanics motorists are searched the most at nearly 40% followed by white motorists at nearly 39%. Black motorists also show a higher search rate compared to stop rate with a difference of 9.5%. The only non-white race of motorist that is stopped at a higher rate than the search rate are Asian motorists. I also find that certain stop characteristics, such as having a Midnight stop, an older car and a luxury brand card are also more likely to be searched compared to the stop rate.

Table 2 shows summary statistics of troopers. Of the 2,701 troopers I was able to match to the data, approximately 60% are white, 30% are Hispanic, and almost 9% are black. The last one percent is composed of Asian, American Indian, and other race troopers. The force is predominantly male at 94%. By trooper race, I find that white troopers make up most of the searches at 64%, followed by Hispanic troopers at 21%. I find that only white troopers search at a higher rate compared to the stop rate while black and Hispanic troopers search at a lower rate. I also find that troopers less experienced troopers searched more than more experienced troopers since the average hire year for searches was greater than the average hire year for stops.

In the bottom part of the table, I break down the stop and search statistics by trooper position, with rank listed in decreasing order.[5] Ranked officers make up only 20% of the highway patrol. I find as rank increases, troopers are less likely to search. Using the rank of Captain as an example, the interpretation of the probabilities is "If the trooper is a captain, then captains conduct 0% of total searches." I find that troopers make up most of the stops and searches at 70% and 72% respectively.

[5] I excluded the rank of major as only two troopers were majors and they conducted no searches and only four stops during 2010 to 2015.

4 Empirical Results

Past tests for racial prejudice

Before showing the results of my test of bias on the data, it is important to highlight the effect of misreporting on past statistical tests of bias, notably, Knowles et al. (2001). Under their test, if troopers were biased against Hispanic motorists, then the search success rate for Hispanic motorists would be lower than the search success rate for white motorists. In Table 3, Column (1) shows the search success rate by motorist race using the recorded races. From Column (1), I find Hispanic motorists have the lowest success rate of 30.7% compared to the white, Asian, and black search success rate, at approximately 42.0%. Therefore, KPT's test would conclude that Hispanic motorists are the most biased against. When I use the *estimated* races, I find different search success rates, which are shown in Column 2. Using the estimated races and applying KPT's test again, I instead find that Asian, black, and Hispanic motorists are biased against and that the magnitude of bias against Hispanics is actually much larger. This is driven by the white search success rate risking by 6%. The race correction also slightly changes the Hispanic search success rate, which falls significantly by 1%. I find that the race correction does not significantly change the black search success rate.

Table 3 also shows the difference in the effectiveness of each race correction method. For black motorists, which uses the geocoding analysis, only 318 of the searches are corrected. In contrast, after applying the surname analysis, the total number of searches for Hispanic searches increases from 23,868 to 56,530, which is an increase of 76%. The total number of searches of Asian motorists increases from 1,287 to 1,530, which is a 20% increase.

This also shows that misreporting reduces the appearance of racial bias for troopers by reducing the search success rate of white motorists. While I find no large increases in the search success rate after race correction for black and Hispanic motorist search success rate, the white search success rate has risen significantly by 6% to nearly 50%. This implies that trooper's also differentially misreport based on search outcome, which becomes the basis of my own racial bias test.

My test for racial prejudice

Thus far, I have shown strong evidence that misreporting not only distorts the observed stop and search rate of motorists, but it also leads to the wrong conclusions of tests for racial bias. Troopers were systematically misreporting races of motorists, especially Hispanics, to appear less likely to stop and search minority motorists (Collister (2015a)). Using my model, I show this behavior is linked to bias by comparing the rate of misreporting conditional on the search outcome and by comparing the search success rate using the observed races and the estimated races. Table 4 shows the results of my main test of bias, which compares the probability of misreporting conditional on search outcome. The key identification of my test is that biased troopers should differentially misreport based on search outcome; specifically, the trooper should misreport only when the search ends in failure. Since my race estimation method cannot perfectly estimate driver's race or identify when a trooper is misreporting versus making a mistake, I rely on the differential misreporting behavior across search outcome to measure bias. The key assumption is that any other possible reason for trooper error or driver race mis-identification will occur at equal rates across search failure or search success. Then, if the trooper is biased, then I can interpret higher rates of misreporting when the search ends in failure compared to when the search ends in success as bias.

Columns (1) and (2) show the misreporting rate conditional on search outcome. According to my model, if troopers are indeed misreporting in a biased manner, then the $Pr(Mismatch|Failure) > Pr(Mismatch|Success)$. Biased troopers should only misreport searches that end in failure with the intuition that misreporting successful searches raises the probability of be being caught misreporting. Indeed, I find that troopers are 8% more likely to misreport Asian motorists when the search ends in failure than when the search ends in success. For black motorists, I find that black motorists are 0.4% more likely to be misreported as white when the search ends in failure compared to when the search ends in success, which is a small, but significant difference. For Hispanic motorists, I find that failed searches are 2.1% more likely to be misreported compared to successful searches.

In order to control for confounding variables, such as seasonality and county characteristics, I regress the following equation:

$$I(Mismatch_{i,t}) = \alpha + \beta_1 I(Failure)_{i,t} + X_{i,c,t}\gamma + \epsilon_{j,t} \tag{14}$$

where $I(Mismatch_{i,t})$ is an indicator variable equal to one if the recorded race did not equal the estimated race for stop i at time t. $I(Failure)$ is an indicator variable equal to one if the search ends in in failure and equal to zero if the search ends in success. $X_{c,t}$ are county fixed effects, year trends and county specific time trends. I also control for seasonality by including month fixed effects. The coefficient of interest is β_1. If a trooper is biased, then he is more likely to misreport the minority motorist as white when the search ends in failure thus this coefficient will be greater than 0 if troopers are misreporting in a biased manner. I run this regression separately for Asian, black, and Hispanic motorists using the estimated races.

From my results in Table 5, I find that the rate of misreporting was 6% more likely to occur when the search ended in failure for Asian motorists, 0.3% more likely for black motorists, and 2% more likely for Hispanic motorists. These results are robust to county and month fixed effects. The results for Asian and black motorists are robust when including year fixed effects, but the results for Hispanic motorists. are not. This is not surprising given the time trend observed in Figure 4, which shows that misreporting was falling steadily during 2010-2015. These results are slightly smaller, but similar to the rates shown in Table 4. Since most of the misreporting is driven by Hispanic motorists, I limit my officer-level analysis to only Hispanic motorists as I do not have a sufficient number of misreported Asian or black stops to control for possible confounders.

To measure the magnitude of Hispanic bias for each officer, I use Eq. (14), but allow for each trooper to have his own disproportionate misreporting rate depending on failure. The more biased the trooper is, the more he will misreport motorist of race-R. I rely on the difference because my method of race correction also corrects successful searches so the differential rate of misreporting based on search outcome will identify bias.The equation to identify each trooper's bias is:

$$I(Mismatch)_{i,j,t} = \alpha + \beta_1^j I(Failure)_{i,t} + \delta_j + X_{i,c,t}\gamma + \epsilon_{j,t} \tag{15}$$

Each β_1^j will measure each officer's differential misreporting behavior based on search outcome. A positive estimate indicates bias against Hispanics. Since trooper's with more searches will have a more precise estimate of bias than troopers with few searches, I exclude troopers with less than 5 searches. To include officers who only search in one county while also controlling for differences across counties, I include the same controls from 14 in $X_{i,c,t}$. These characteristics include median income, percentage Hispanic, percentage black, employment rate, percentage with high school diploma, and population size. I also include month specific fixed effects. I show the distribution of β_1^j in Figure 6.

One notable characteristic of the distribution of bias is the heterogeneity in the measures of bias for troopers. While most troopers are concentrated at no bias, I find a large mass with positive levels of bias, with a mean and a median of 0.12 and 0.05, respectively. This means the average trooper is 12% more likely to have the estimated race not match the recorded race when the search ends in failure compared to success. Officers at the 90th percentile of the distribution are 66 percent more likely to misreport with failed searches compared to successful searches while at the bottom 10th percentile are 33% more likely to misreport successful searches than failed searches.

Another important characteristic of the distribution is the negative side. Troopers here are more likely to misreport when the search ends in success compared to failure. This can occur for the following reasons. First, some troopers may only misreport when the search ends in success for reasons that may or may not be related to bias (non-compliers). If that is the case, then few troopers engage in this behavior, as evidenced by the small mass on the left hand side of the distribution (with 934 troopers with bias less than 0 and 1,236 with bias greater than or equal to zero). Second, the race estimation method is an estimate of trooper's bias, thus there will be troopers who are negative. Given the smoothness of the negative side compared to the 'hump' on the positive side, troopers with 'negative bias' may just show the natural distribution of bias across Texas troopers.

Next, I use the publication of the news article by KXAN revealing the misreporting as a natural experiment. The article was published in November 8th, 2015, a hearing was conducted by November 18th, and by November 23rd, DPS changed its policies to now require troopers to ask drivers for their race. I can test the effect on stop behavior of troopers after the changes are implemented by augmenting the SOPP data with the publicly available stop data from DPS. Since I observe changes in misreporting behavior from June 2015 to November 2015, I will only use stops preceding July 2015 to January 2010 as my pre-data. For my post data, I am using the publicly available data, which has the recorded driver's races but no driver's names or addresses. If my measure of bias is correct, troopers who were using misreporting to hide their bias should have the greatest changes in stop behavior with Hispanic and white motorists. Biased troopers are more likely to misreport the search if the search was unsuccessful. Thus, I should observe a negative relationship between search success rates and motorists being recorded as white relative to trooper's bias prior to November 2015 and no effect after November 2015.

I test this using:

$$\begin{aligned} I(RecRace = White_{i,j,t}) = \alpha + \sum_{\substack{t=2010 \\ t\neq 2015}}^{2017} \Big[\beta_3^t Hisp\ Bias_j \times I(Year = t) + \\ \beta_4^t I(Failure_{i,j,t}) \times I(Year = t) + \beta_5^t Hisp\ Bias_j \times I(Year = t) \times I(Failure_{i,j,t})\Big] \\ + X_{i,c,t}\gamma + \epsilon_{i,t} \end{aligned} \tag{16}$$

I use the *recorded* race rather than the estimated race because the recorded races will show the greatest change after 2015 if my measure, $Hisp\ Bias_j$, actually quantifies officer j's bias, where $Hisp\ Bias_j$ is derived from Eq.(13). $I(Failure)$ is an indicator variable if stop i ends in failure. The primary coefficient of interest is β_5^t for $t > 2015$, which is the interaction between officer level bias pre-2015, the search outcome, and the years after the changes were implemented. β_5^t will reflect the differential probability in being recorded as white when the search ends in failure for biased officers. If my measure is capturing bias, then officers with higher levels of bias should change their search behavior more than officers with low level of bias.

I use a different estimation strategy than with Eq. (14) since I cannot test for misreporting using the publicly available data because the publicly available data doesn't contain driver's names or driver's home addresses. But, even if I cannot measure misreporting directly, DPS changes its policies to require troopers to verbally ask for driver's race, thus the recorded race should be the estimated race after November 2015. Thus, I can assume that stops after November 2015 are recorded correctly.

Figure 7 shows the results for both Hispanic and white motorists. If my measure of officer-level bias is correct, then the coefficient for β_5^t for $t < 2015$ should be positive and should go to zero for $t > 2015$ for white motorists. The intuition being that biased officers, will disproportionately misreport Hispanic motorists as white depending on the search outcome, and are therefore more likely to record the motorist as white when the search ends in failure compared to unbiased officers prior to the publication of the article. Specifically, a standard deviation away from the mean level of bias leads to a 8% increase in the probability of being recorded white when the search ends in failure from 2010 - 2014. But, from 2016 - 2017, the coefficients are not significantly different from zero. For Hispanic motorists, I find the opposite pattern. Motorists are 9% less likely to be recorded as Hispanic if the search ended in failure. But, by 2015, there is no significant difference in the probability of being recorded as Hispanic depending on the search outcome.

This reveals two important characteristics of trooper behavior. The first is that the biased troopers complied to the rule change and began correctly reporting motorists race after 2015. While DPS was not clear as to how they would enforce their new policy, it appears effective in changing biased trooper's behavior. The second is that motorists who self-identified as Hispanic were the ones predominantly being misreported. This is especially important since Hispanic is technically an ethnicity and many driver's may identify as white rather than Hispanic (Lopez et al. (2017)).

Robustness Checks

To ensure that the relationship between my 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), and 85%, and re-estimate my trooper level measure of bias.[6] Figure 8 shows the distribution for each cutoff. These distributions show a similar shape; it also shows that even with tighter thresholds, the distribution of officer level bias is robust.

Since my main estimation of Hispanic motorists relies on the distribution of Hispanic surnames in the United States from the 2000 census, I test if my results are driven by the unequal distribution of last names across search outcome. I randomly assign the probability a last name is Hispanic based off the normal distribution. Using different Z-score cutoffs, I re-estimate bias using Equation (14). I repeat this procedure a thousand times to get the distribution of average bias in Figure 9. The average level of bias is -0.04, which is far less than my estimate in Table 5. In fact, that natural distribution implies troopers

[6] I also use 95% as a threshold, but at 95%, there are only 1,242 Hispanic surnames compared to the 4,647 surnames at 85%. Most of the bias estimates were concentrated at no bias, making the other densities hard to see on the graph.

on average are *less* likely to misreport unsuccessful searches by 4 percentage points. Thus, I can reject the null hypothesis that my results are merely driven by the distribution of last names.

I also test if biased troopers were able to effectively evade detection of bias using Knowles et al. (2001) and Becker's test of bias. I regress:

$$\begin{aligned} Y_{icjt} = &\alpha + \beta_1 Hisp\ Bias_j + DriverRace_i \boldsymbol{\beta_2} + DriverRace_i \times Hisp\ Bias_j \boldsymbol{\beta_3} \\ &+ \gamma_m + \delta_c + \epsilon_{icjt} \end{aligned} \tag{17}$$

separately using the recorded races and the estimated races. The outcome of interest is whether a search was conducted and if the search was successful conditional on search. I control for officer level Hispanic bias along with the recorded (estimated) race of the driver. The variable of interest is $\boldsymbol{\beta_3}$, which shows the differential probability of search or success for each driver race compared to white motorists.

Table **??** shows the results using the recorded races in Panel A and the estimated races in Panel B. From Panel A Column(1), biased troopers appeared less likely to search black and Hispanic motorists by .2 to .3 percentage points compared to white motorists. Biased troopers also appeared to have significantly higher search success rates, specifically Hispanic motorists, with one standard deviation of bias being associated with a 12 percentage point higher search success rate than white motorists. I find similar positive effects for black motorists with 3 percentage points higher search success rate associated with one standard deviation in bias.

In contrast, Panel B, shows different results using the estimated races. While the estimated races are not the true race of the motorists, these results give some insight in the true search behavior of troopers by driver's race. My results show no significant difference in the probability search by driver race compared to the results using the recorded races. In contrast, I find different results using the estimated races on the probability of search success. I find a negative, but not significant, coefficient for the effect of Hispanic bias on the Hispanic motorist search success rate where one standard deviation is associated with a one percentage point decrease in the probability of search success for Hispanic motorists compared to white motorists. This coefficient is significantly different than the results in Panel A Column (2). The results from the estimated races and recorded races show that biased officers appeared to have a much higher search success rate when searching minority motorists compared to unbiased officers, but in actuality, they were equally good if not worse than unbiased officers.

4.1 Bias and Peers - unfinished

Another important question is whether troopers are biased because of their own preferences, the environment they work in, or because they adopt the preferences of their peers. This is especially important in the context of Texas since troopers may have different incentives (for example: reducing drug trafficking, reducing undocumented immigration, ensuring border protection) to search motorists depending on which county they are stationed in and understanding the underlying mechanism of bias is helpful for informing policymakers in how to counteract bias.

To measure bias on the work city level, I create officer by county measures of bias using Eq. (13) and use the mean level of bias within the county as my new outcome of interest.[7] I first test to see if counties with high averages of officer level of bias are correlated with any county characteristics, which are measured using data from the American Community Survey (ACS). Specifically, I regress:

$$\begin{aligned} Hisp\ Bias_c = \alpha + \beta_1 PercBlack_c + \beta_2 PercHisp_c + \beta_3 Perc\ no\ health\ ins_c + \\ \beta_4 Perc\ HS\ Diploma_c + \beta_5 Median\ HH\ income_c + \beta_6 Perc\ Employed_c + \\ \beta_7 Perc\ Age > 16_c + \beta_8 Total\ Pop_c + \beta_9 I(Border = 1)_c + \epsilon_c \end{aligned} \tag{18}$$

where $Hisp\ Bias_c$ is the mean trooper level of bias within a county using each trooper's search history within the county to construct the trooper-county average bias. All of the explanatory variables are from the 5-year ACS from 2010 - 2015 except $I(Border = 1)$, which is an indicator variable for if the county is defined as county that borders Mexico as defined by the Le Paz Agreement. A positive coefficient for $\beta_1 - \beta_8$ imply higher instances of that trait are higher levels of bias. I limit my regression to only counties with at least five troopers, which reduces the number of counties from 254 to 131.

Table 7 shows the results. I find that counties with higher mean levels bias have lower percentage employed and smaller population sizes. This implies that counties with low population and low level employment are more likely to have higher level of bias. A ten percentage decrease in employment is correlated with a .8 standard deviation decrease in the mean level of bias within the county. I also find that border counties have a 0.1 standard deviation higher mean level of bias within the county compared to non-border counties.

I also test whether troopers change their behavior based on their coworkers or peers within the county. For example, a good trooper behaves unbiased in County 1, but is transferred to County 2, where all troopers misreport in a biased manner. Thus, this good trooper changes his behavior to match his coworkers. If this is true, then a trooper's own bias should increase if he moves to a more biased county. In order to test this, I focus on the 532 troopers who move across to different work city from 2013 to 2015.

To ensure my results from Table 7 are not driven just by county characteristics, I use a similar

[7] I do not use the Eq. (15) to estimate each officer-county measure of bias since most officers do not have enough searches per county to estimate their county-specific measure of bias with a regression.

approach as Eq.(16) but use the county's mean level of Hisp Bias instead:

$$\begin{aligned}
I(RecRace = White_{i,c,t}) = \alpha + \beta_0 Hisp\ Bias_c + \beta_1 I(Failure_{i,c,t}) + \\
\beta_2 Hisp\ Bias_c \times I(Failure_{i,c,t}) + \sum_{\substack{t=2010 \\ t\neq 2015}}^{2017} \Big[\beta_3^t Hisp\ Bias_c \times I(Year = t) + \\
\beta_4^t I(Failure_{i,c,t}) \times I(Year = t) + \beta_5^t Hisp\ Bias_c \times I(Year = t) \times I(Failure_{i,c,t})\Big] \\
+ X_{i,c,t}\gamma + \epsilon_{i,t}
\end{aligned} \tag{19}$$

where the dependent variable is the search success rate for motorist of race-H. I use the recorded race again for reasons listed in the previous section. The coefficient of interest is β_t; if $Bias_c$ is truly capturing the county level of bias and not another county characteristics, then I expect β_t for $t > 2015$ to be significantly different for counties with high mean level of bias from 2010 - 2015. A key assumption in this regression is that there are no differential changes in population given a county's level of bias and that the composition of motorists who could be stopped also does not change with a county's level of bias across time. In other words, the only change across counties and time should be the county's own level of bias or coefficient of variation.

Figures ?? shows the results for white and Hispanic motorists together with respect to the county level estimate of bias from 2010 to 2015. For white motorists, I find a downward trend showing that one standard deviation of bias equates to nearly 10% more likely to record the motorist as white for failed searches from 2010 - 2011. This difference disappears from 2012 to 2014, but in 2015, one standard deviation of bias led to being nearly 10% *less* likely to record the motorist as white for failed searches. I find the complete opposite effect for Hispanic motorists, as seen by the near mirror trends of both races.

Bias and Trooper Characteristics

The next section is to understand if any trooper characteristics are related to bias. One major contribution of this paper is to be able to generate trooper-level estimates of discrimination. In this section, I will test whether bias varies by trooper demographics. Additionally I will address how discrimination varies with other employment characteristics such as promotions, salary, and officer transfers.

Table 8 shows column 3 of Table 4 broken down by trooper race. To control for the heterogeneity in motorist stops by counties across Texas and to control for the county specific time trends, I use a regression with county by year fixed effects. Using a model similar to Eq. (14), I run the regression:

$$\begin{gathered} I(Mismatch)_{i,j,t} = \alpha + \beta_1 I(Failure)_i + I(TroopRace_j)\beta_{2j} + \\ I(Failure) \times I(TroopRace_j)\beta_{3j} + X_{c,t} + \epsilon_{i,j,t} \end{gathered} \tag{20}$$

where I now control for the race of trooper j. The coefficient of interest is β_{3j}, which is the differential misreporting behavior by search outcome for black and Hispanic troopers. To control for the county specific time trends, I also include county by year fixed effects. If white troopers are more biased against Hispanic motorists, then I $\beta_{3,black}$ and $\beta_{3,Hisp}$ should be negative since I omit white troopers.

I report my results in Table 8. I find that black troopers are 2.5% less likely to misreport black motorists compared to white troopers, but this coefficient is not significant. This is a power problem since the only 236, or 8.5% of troopers are black. I find similar results for black troopers and Hispanic motorists. Black troopers are nearly 8% less likely to misreport Hispanic motorists, but this result is insignificant again, but the standard errors are smaller relative to the coefficient estimate. For Hispanic motorists, I find no significant difference in their misreporting behavior relative to white motorists. Surprisingly, Hispanic troopers are as equally biased as white troopers against Hispanic motorists , which is contrary to past tests of bias (Goncalves and Mello (2017), Antonovics and Knight (2009)).

I also test whether biased behavior varies by rank in Table 9. Since I have few troopers with rank greater than corporal, I combine sergeants, lieutenants, and majors into one category and use troopers as the omitted category. Since some troopers change rank during the time period, I use the maximum rank reported in the employment data. I find no significant differences in the amount of bias across trooper rank.

Next, I test if employment outcomes, such as salary and experience, are related to bias where experience is measured using the hire year of the trooper. I omit troopers with negative bias since I am testing the relationship between having Hispanic bias compared to having no bias. This also focuses the interpretation of the results and provides insight in how DPS was responding to bias during that time. Were biased troopers being promoted more than unbiased troopers? Do biased troopers get paid more? Does bias increase with experience?

I show my results in Table 10. I find positive correlations of trooper salary and experience to Hispanic bias, but these results are not significant. For trooper rank, I find that troopers with rank Sergeant or higher have Hispanic bias 0.2 standard deviations higher than troopers. No other rank has significant different levels of bias compared to troopers.

To examine the relationship between trooper bias and trooper's career across time, I divide the trooper's career into two sections: pre-2013, and 2014-2015. This has a few advantages; first, I do not have trooper employment data prior to 2013 so 2013 is the earliest year I can use. Second, the measure of bias has high variance since it's measured on the differential misreporting behavior across search outcomes. Therefore, officers with few searches have high variance. By dividing the trooper's career into two sections rather than by year, my estimate of bias is more efficient and more consistent. Lastly, with the panel-like structure, I can test if changes in employment outcomes are related to bias, specifically outcomes such as increasing in rank, moving cities, and leaving the force. Moving cities is a proxy for salary. Rather than a salary increase, a trooper can be compensated for good behavior by being stationed at a preferred city. I omit troopers with negative bias again for the same reasons as before.

I show my results in Table 11. I regress the likelihood of leaving the force, moving cities, and increasing rank on the standardized measure of bias from the first half of the trooper's career including controls for trooper's rank before 2014 and for their work city. I find that the probability of leaving the force increases by 2% for every 1 standard deviation increase in Hispanic bias, but this result is not significant. I find no relationship between an officer's measure of Hispanic bias prior to 2014 has no effect on increasing in officer's rank or moving cities in 2014 and 2015. Not only are the point estimates insignificant with large standard errors, but the estimates are also close to zero.

I also break column 3 from Table 11 into each rank. I regress the probability of increasing rank for each rank of trooper interacted with the level of bias measured from stops conducted 2010 - 2013. From my results in Table 12, I find no evidence that more biased troopers are more or less likely to be promoted regardless of rank.

I next test if biased troopers are also perform worst in other aspects of their job by using complaint data obtained from DPS. The results in Table 13 show no relationship between trooper level bias and the probability of receiving a complaint. While the coefficient is positive, it is not significant and the point estimate is close to zero. But, from the 1,055 complaints, only 166 included the trooper's badge number. When I restrict my analysis to sustained complaints, I still find similar results. The analysis is severely hampered by the lack of trooper identification for a majority of the complaints.

I also test to see how the employment outcomes of troopers were affected by the publication of the article. I use publicly available 2019 salary data published by the Texas Tribune. My results in 14 show one standard deviation above the mean level of bias leads to a 5% increase in the probability of leaving the force by 2019. I do not find any significant effect in having a higher salary or in the probability of ranking up, but I do find a positive coefficient. These results are robust even when including troopers with negative bias.

Negatively Biased Troopers - not done

Another important question is how to consider the troopers with negative measure of bias. Are trooper characteristics significantly different for troopers with negative bias compared to troopers with no bias ($Hisp\ Bias = 0$)? From the density plot in Figure 4 the 2,319 troopers, only 456 are biased, and 320 have zero bias. I first test if trooper characteristics vary significantly for troopers with negative bias compared to troopers with no bias. My results in Table 15 show that troopers with negative bias were less likely to be black troopers compared to troopers with zero bias. This is unsurprising since black troopers tended to be less biased, albeit not significantly so, compared to white troopers. When including the trooper rank, I find that probationary troopers were also less likely to be biased, which is surprising given the results in Table 10, which finds higher, but insignificant, levels of Hispanic bias for probationary troopers. I also find that corporals were more likely to also have negative bias, which is significant at the 10% level. In column (3), I find that troopers with negative bias are not more likely to be paid more after controlling for trooper rank. I also find no correlation with experience.

One major concern of troopers with negative bias is that the negative bias is produced by errors in the race correction. This would be a major concern since this would imply similar issues even with troopers with positive bias. One way for this to occur is if troopers with negative bias have fewer searches and their misreporting measure is governed by over-correction of successes, which occurs by random. Columns 4 and 5 of Table 15 show that troopers with negative bias in fact search more in general and search more Hispanic motorists compared to troopers with no bias.

Further, if there was a systematic issue with the race correction, then troopers with negative bias would misreport significantly more when searches with Hispanic motorists ended in success. To test this, I regress a version of Eq. 14 but include a triple interaction indicating whether the trooper's level of bias is negative. Specifically:

$$\begin{aligned} I(RecRace = White)_{i,j,t} = \alpha + \beta_0 I(CorrRace = Hispanic)_i + \beta_1 I(Failure)_i + \\ \beta_3 I(Hisp\ Bias_j < 0) + \beta_4 I(CorrRace = Hispanic)_i \times I(Failure)_{i,j,t} + \\ \beta_5 I(CorrRace = Hispanic)_i \times I(Hisp\ Bias_j < 0) + \beta_6 I(Failure)_i \times I(Hisp\ Bias_j < 0) \\ \beta_7 I(CorrRace = Hispanic)_i \times I(Hisp\ Bias_j < 0) \times I(Failure) + \\ X_{i,j,c,t}\gamma + \epsilon_{i,j,t} \end{aligned} \tag{21}$$

Instead, as seen in Table **??**, troopers with negative bias are not significantly less likely to misreport Hispanics when the search ends in failure compared to troopers with zero bias. Conversely, for troopers with positive bias, the point estimate is larger at 0.09. This means that troopers with positive bias are 9% more likely to record the Hispanic trooper as white when the search ends in failure compared to troopers with no bias.

5 Conclusion

In this paper, I use the misreporting behavior of troopers to uncover their taste towards searching driver's by motorist race. I develop a new statistical model to use this misreporting 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 to 2015 merged with Texas Highway Patrol employee administrative data. During this time period, Texas troopers were misreporting 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 estimated, misreported and unmisreported, I am able to develop a comprehensive test for racial bias.

In my results, I find that black 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 misreporting many Hispanic motorists as white, troopers appear less successful in searching white motorists and are able to avoid being labeled biased.

Using my measure of bias, I find strong peer effects on trooper bias with trooper bias increasing by nearly one standard deviation when moving to a county with one standard deviation higher bias. This implies that even good troopers may behave badly as a result of environment rather than from their own tastes and preferences, highlighting the importance of institutions and social norms on officer bias. Further research is needed to examine whether changing your coworkers may be more effective and efficient in curbing biased behavior compared to officer-level interventions.

I also test whether bias affects employment outcomes from 2010 to 2015 and also after the publication of the article. I find that biased troopers tend to be paid more than their unbiased peers. This relationship reverses after 2015 where I find that troopers who were biased prior to 2015 were paid over $400 less than their unbiased peers.

Misreporting may not just be present in law enforcement. In any sort of scenario where racial profiling is illegal, this may induce agents to misreport the race of the biased group to appear less biased. For example in mortgage lending, mortgage lenders may misreport the race of applicants to appear less biased. This misreporting 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*.

Benen, S. (2014, July). Rick Perry's 'Operation Strong Safety. *MSNBC*.

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.

Depew, B., O. Eren, and N. Mocan (2017). Judges, juveniles, and in-group bias. *Journal of Law and Economics*.

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*.

Freedman, M., E. Owens, and S. Bohn (2018). Immigration, employment opportunities, and criminal behavior. *American Economic Journal: Economic Policy*.

Glaeser, E. L. and J. L. Vigdor (2001). Racial segregation in the Census 2000: Promising news. *Brookings Institute Center on Urban & Metropolitan Policy*.

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*.

HSIN. HSIN supports border security operations during Operation Strong Safety.

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*.

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

Lopez, M. H., A. Gonzalez-Barrera, and G. López (2017). Hispanic identity fades across generations as immigrant connections fall away. *Pew Research Center*.

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.

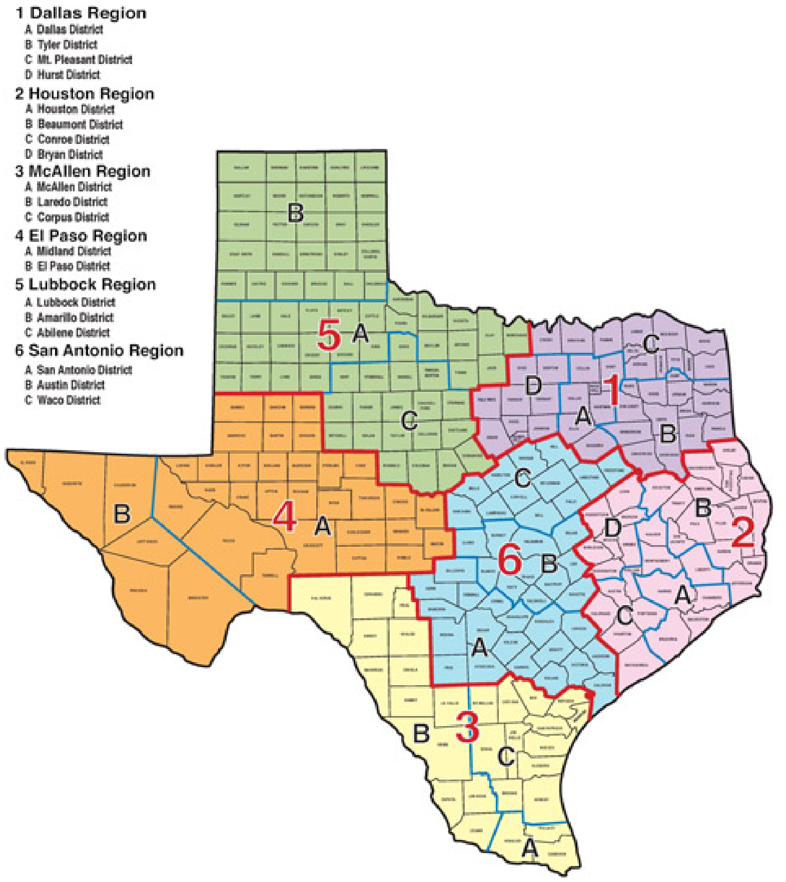
Sanchez, R. (2015). Who was Sandra Bland? *CNN*.

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

TxDPS (2018). Complaint investigation and resolution.

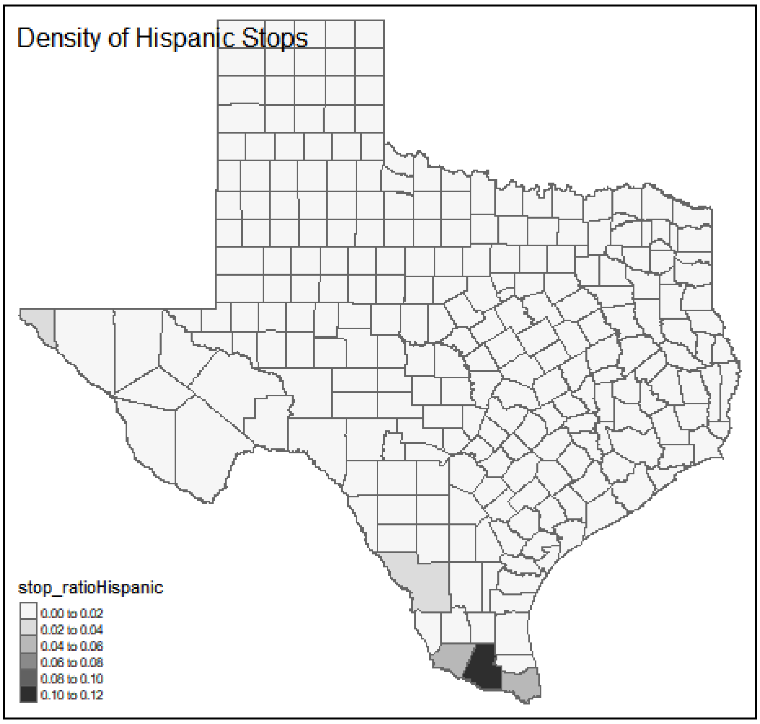
West, J. (2018). Racial bias in police investigations.

Figure 1: Trooper Division Map



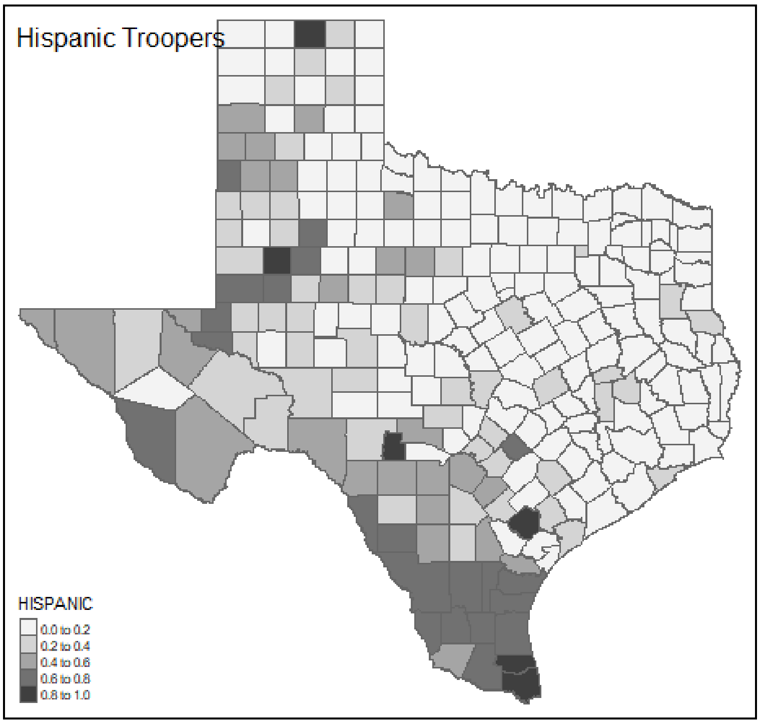
Source: Texas Department of Public Safety. The 7th region is not shown on the map, but its jurisidiction is limited to only Austin, TX.

Figure 2: Hispanic Motorist Stop Density by County



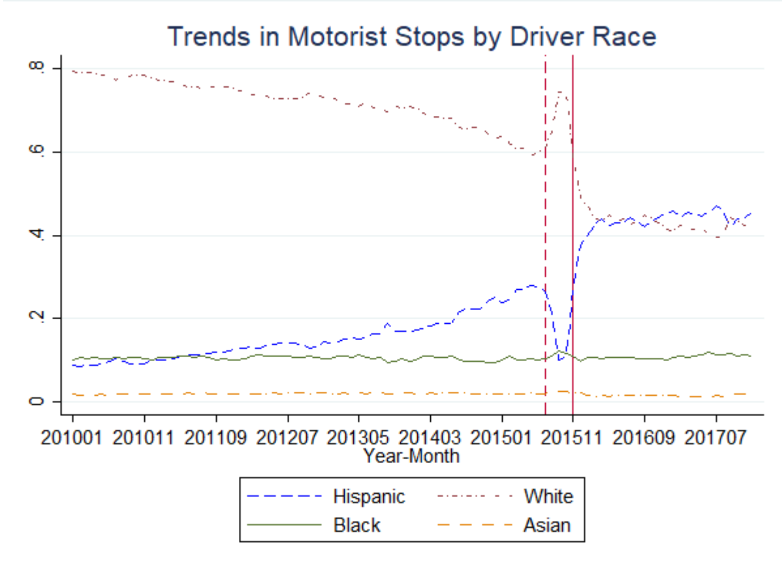
Unweighted means are shown using data from 2010 - 2015. It shows the percentage of Hispanic stops per county using the estimated races. Darker regions imply higher percentages.

Figure 3: Hispanic Trooper Density by County



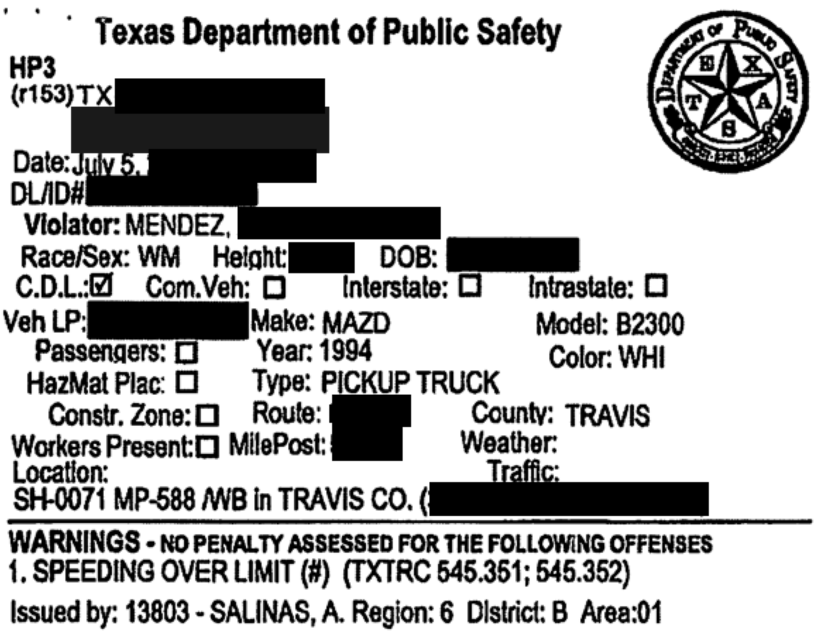
Unweighted mean of stop rate by driver race for each year-month time period from January 2010 to June 2015. Each cell shows the percentage of stops made by Hispanic troopers. Darker regions imply higher percentages.

Figure 4: Monthly Stop Rate by Driver's Race using Recorded Races



Dot-dash line shows the recorded Asian stop rate, solid line shows the recorded black motorist stop rate; dashed line shows the Hispanic motorist stop rate using the recorded races, and the dotted line shows the stop rate for white motorists using the recorded races. All stops are for a given month-year from January 2010 to December 2017 with July 2015 - November 2015 omitted. The vertical red line indicates the year-month the article was released.

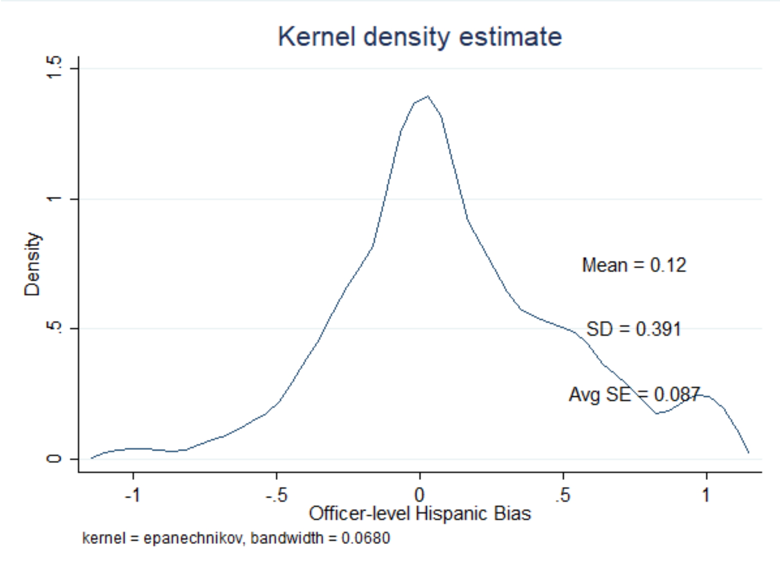
Figure 5: Example of misreported Highway Ticket



Texas Department of Public Safety
HP3
(r153) TX [redacted]
Date: July 5, [redacted]
DL/ID# [redacted]
Violator: MENDEZ, [redacted]
Race/Sex: WM Height: [redacted] 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: [redacted] County: TRAVIS
Workers Present: ☐ MilePost: [redacted] Weather:
Location: Traffic:
SH-0071 MP-588 WB in TRAVIS CO. ([redacted]

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 6: Distribution of officer level measure of Hispanic Bias



Kernel density distribution of officer-level Hispanic bias. The figure plots each officer's β^j from the regression $I(Mismatch_{i,t}) = \alpha + \beta^j I(Failure)_{i,t} + \delta_j + X_{i,c,t}\gamma + \epsilon_{j,t}$. Mean reports the average β_1^j and Avg S.E. reports the average standard error for each β^j.

Figure 7: Natural Experiment - White and Hispanic Motorists

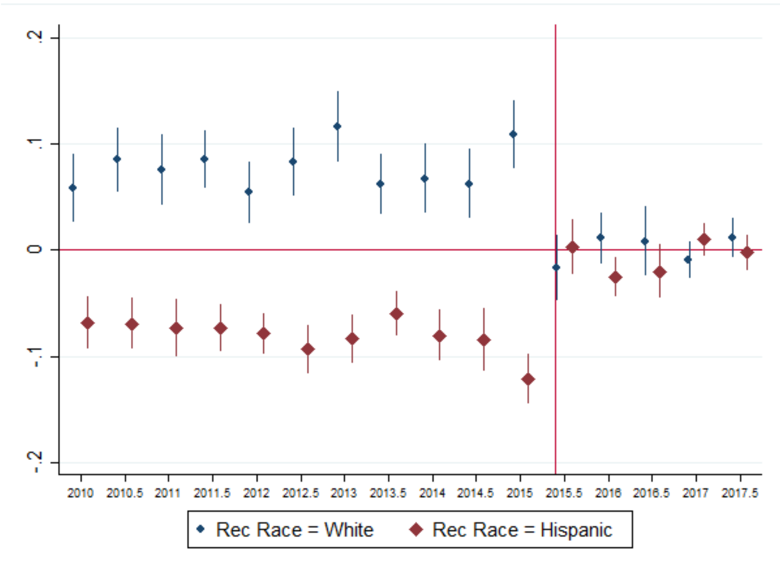
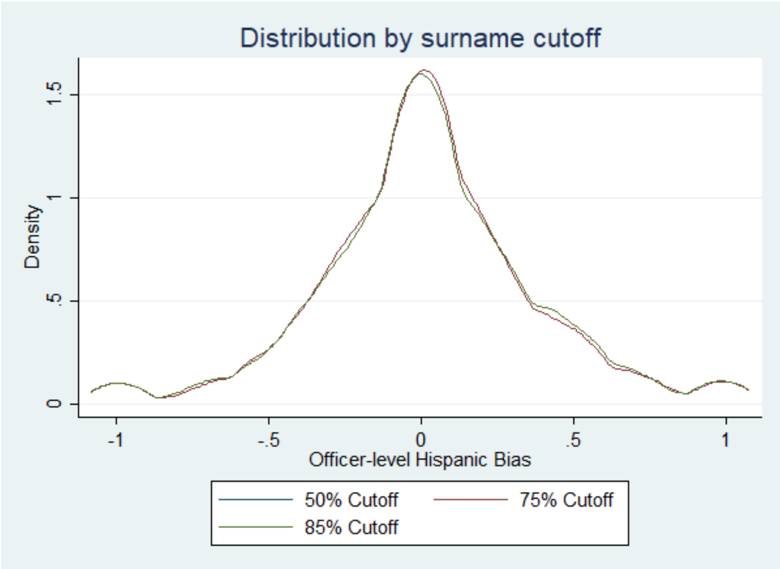


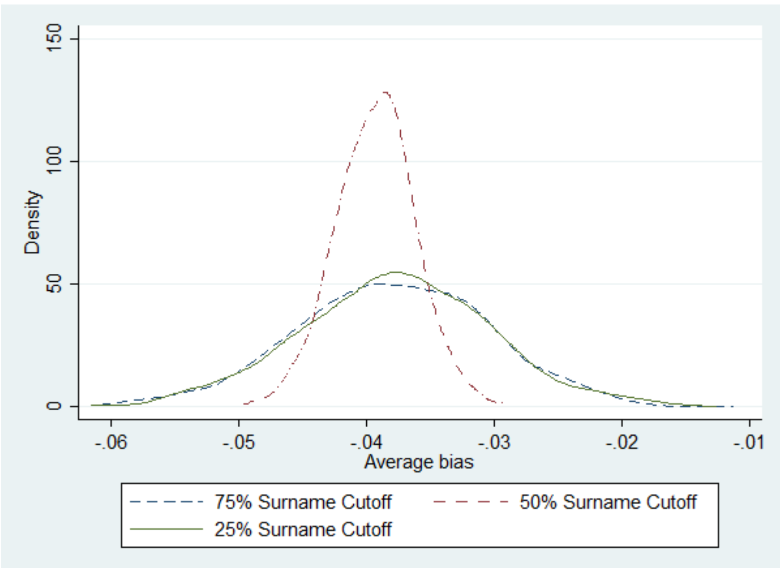
Figure plots the coefficient of interaction $I(Failure_i) \times Hisp\ Bias_j \times I(Year\ Half = t)$, β_5^t, and with 5% confidence intervals. $Hisp\ Bias_j$ is standardized. Points to the right of the vertical line are after the article publication. Diamond points are when the dependent variable is $I(RecRace = White)$ and circle points are when the dependent variable is $I(RecRace = Hispanic)$.

Figure 8: KPT and misreporting Measure of Bias with different thresholds



Each density shows the officer level bias using different levels of surname cutoff and weighted by the total searches.The estimate of bias is from each officer's β^j from the regression $I(Mismatch_{i,t}) = \alpha + \beta^j I(Failure)_{i,t} + \delta_j + X_{i,c,t}\gamma + \epsilon_{j,t}$.

Figure 9: Monte Carlo Simulation of Bias



Each density shows the average officer level bias using different normal distributions from Eq (14).

Table 1: Mean of Variables Related to Drivers

Driver Characteristics	(1) All	(2) Searches Only	(1)-(2) Δ
Estimated Asian	.022	.011	.011
	(.147)	(.103)	(0)
Estimated Black	.104	.199	-.095
	(.306)	(.4)	(.001)
Estimated Hispanic	.341	.399	-.058
	(.474)	(.49)	(.001)
Estimated White	.53	.389	.141
	(.499)	(.488)	(.001)
Midnight	.086	.131	-.045
	(.281)	(.337)	(.001)
Owner Driver	.202	.142	.06
	(.401)	(.349)	(.001)
Texas Driver	.9	.848	.052
	(.299)	(.359)	(.001)
Old Car	.304	.426	-.122
	(.46)	(.494)	(.001)
New Car	.338	.179	.159
	(.473)	(.384)	(.001)
Luxury Car	.085	.102	-.017
	(.279)	(.302)	(.001)
Observations	8045487	145730	

Standard deviations are in parentheses. Unweighted means are shown. Stops made from July 2015 to December 2015 were omitted. Midnight is defined as a stop from 12 am to 6 am. Owner information was missing for 38% of the stops. Vehicle was coded as old if made 10 or more years before the stop year and coded as new if made within 3 years of the stop year. Vehicle was considered luxury car if on the Forbes list of top 20 best selling luxury vehicles in 2010. Only 54.5% of the stops were geocoded to a Texas block FIPS.

Table 2: Mean of Variables Related to Troopers

Troopers' Characteristics	(1) All Stops	(2) Searches Only
Black	.087	.04
	(.282)	(.207)
Hispanic	.287	.205
	(.453)	(.404)
White	.606	.633
	(.489)	(.482)
Male	.946	.979
	(.226)	(.142)
Hire Year	2004	2006
	(7.244)	(4.676)
Trooper Rank		
Captain	.007	0
	(.084)	(.018)
Lieutenant	.023	.004
	(.15)	(.059)
Sergeant	.125	.063
	(.33)	(.242
Corporal	.1	.104
	(.3)	(.305)
Trooper	.697	.723
	(.46)	(.447)
Probationary Trooper	.018	.004
	(.133)	(.063)
No Rank	.031	.102
	(.173)	(.303)
Total Troopers	2,701	

Notes: Only merged observations are shown. Trooper rank uses the highest rank the trooper obtained during 2010 - 2015. Stops from July 2015 to December 2015 were omitted. Stops are considered mismatched if the recorded race does not equal the corrected race. 10.5% of the troopers in the employment data were not matched to the stop data. 22% of the troopers in the stop data were not matched to the trooper employment data.

Table 3: Search Success Rates across Driver's Race

	Search Success Rate		
	(1)	(2)	(3)
	Recorded	Estimated	Δ
Driver Race			
Asian	.426	0.403	0.023
	(.495)	(.491)	(0.018)
	1287	1530	-243
Black	.422	.421	.001
	(.494)	(.494)	(.004)
	27999	28401	-318
Hispanic	.307	.297	.01
	(.461)	(.457)	(.004)
	23868	56530	-32662
White	.421	.482	-.061
	(.494)	(.5)	(.003)
	82451	55631	26736

Notes: Unweighted means are shown. Standard deviations are in the parentheses. Columns 1-3 use data from January 2010 - June 2015. Row 3, 6, 9 show the total number of searches using the recorded and estimated races respectively.

Table 4: Difference in Misrecording Rate by Search Success

Estimated Driver Race	$Pr(Mismatch \mid Failure)$	$Pr(Mismatch \mid Success)$	Δ
Asian	.192	.113	.079
	(.394)	(.316)	(.018)
Black	.018	.014	.004
	(.133)	(.117)	(.001)
Hispanic	.578	.557	.021
	(.494)	(.497)	(.004)

Notes: Unweighted means are shown stops from January 2010 to June 2015. Standard deviations are in parantheses. Mismatch is defined as 1 if the recorded race does not equal the estimated race. Search is defined as success if the trooper found contraband (drugs, weapons, high amounts of currency, and drug paraphernalia)

Table 5: Main Test of bias

	(1)	(2)	(3)	(4)	(5)	(6)
	Asian Motorists		Black Motorists		Hisp Motorists	
I(Failure)	0.057^{**}	0.054^{**}	0.003^{*}	0.003^{*}	0.016^{**}	-0.010
	(0.022)	(0.022)	(0.002)	(0.002)	(0.008)	(0.007)
County, Month FE	X	X	X	X	X	X
County, Year, Month FE		X		X		X
Observations	1488	1488	28454	28454	56463	56463
F	6.623	6.062	3.542	3.458	4.376	1.994

Notes: Dependent variable is an indicator variable equal to one if the recorded race of the motorist in stop i is does not equal the estimated race. Column 1 and 3 have only county fixed effects, Column 2 and 4 have year and county FE. The F-test reports the joint hypothesis test that variables $I(Failure)$ through the fixed effects are equal to zero. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015 * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 6: My measure of Bias and the Becker test of bias

	(A) Recorded Races		(B) Estimated Races	
	(1) I(Vehicle Searched)	(2) I(Success)	(1) I(Vehicle Searched)	(2) I(Success)
Recorded Asian	-0.008^{***}	-0.015		
	(0.001)	(0.017)		
Recorded Black	0.021^{***}	-0.008		
	(0.001)	(0.008)		
Recorded Hispanic	0.009^{***}	-0.022^{***}		
	(0.001)	(0.007)		
Recorded Asian x Hisp Bias	0.001^{***}	0.039		
	(0.000)	(0.025)		
Recorded Black x Hisp Bias	-0.003^{***}	0.033^{***}		
	(0.001)	(0.008)		
Recorded Hispanic x Hisp Bias	-0.002^{***}	0.119^{***}		
	(0.000)	(0.012)		
Hisp Bias	-0.003^{***}	-0.036^{***}	-0.002^{***}	-0.008
	(0.000)	(0.007)	(0.000)	(0.007)
Estimated Asian			-0.005^{***}	-0.078^{***}
			(0.001)	(0.016)
Estimated Black			0.023^{***}	-0.049^{***}
			(0.001)	(0.007)
Estimated Hispanic			0.012^{***}	-0.110^{***}
			(0.001)	(0.008)
Estimated Asian x Hisp Bias			0.001^{**}	0.007
			(0.000)	(0.021)
Estimated Black x Hisp Bias			-0.003^{***}	0.006
			(0.001)	(0.008)
Estimated Hispanic x Hisp Bias			-0.002^{***}	-0.010
			(0.000)	(0.007)
Constant	0.018^{***}	0.401^{***}	0.015^{***}	0.449^{***}
	(0.000)	(0.002)	(0.000)	(0.004)
Observations	6570205	143306	6570205	143306

Notes: Hisp Bias is the normalized measure of Hispanic bias for each trooper. The measure for Hispanic bias comes from Eq. (15). The regression includes county FE, month FE. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 7: Bias and County Characteristics

	(1) Mean Bias	
Violent Crime Rate	16.87	(14.20)
% Hisp	-1.83	(1.31)
% Black	-0.82	(1.17)
% no health ins	1.66	(2.93)
% HS diploma	-0.27	(2.81)
Median HH inc (10000s)	0.04	(0.12)
% Employed	-8.32***	(2.98)
% older than 16	-2.50	(4.86)
Population (100000s)	-0.05***	(0.02)
Border County	1.35***	(0.52)
Constant	9.99**	(4.33)
Observations	131	
R^2	0.251	
F	6.98	

Notes: Dependent variable is mean bias of trooper's level of bias within the county, weighted by the trooper's number of stops. Counties with less than 5 troopers were excluded. Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 8: Hispanic Bias and Trooper Race

	(1)	(2)	(3)	(4)	(5)	(6)
	Asian Motorist		Black Motorist		Hispanic Motorist	
I(Failure)	0.0470*	0.0434*	0.0035*	0.0033*	0.0141*	−0.0070
	(0.0245)	(0.0237)	(0.0020)	(0.0020)	(0.0082)	(0.0077)
Failure X Black Troopers	−0.0029	−0.0006	−0.0016	−0.0013	−0.0257	−0.0218
	(0.0588)	(0.0593)	(0.0048)	(0.0048)	(0.0607)	(0.0519)
Failure X Hispanic Troopers	0.0842	0.0809	−0.0022	−0.0020	0.0199	0.0041
	(0.0585)	(0.0583)	(0.0049)	(0.0049)	(0.0156)	(0.0141)
Observations	1354	1354	26377	26377	51059	51059

Notes: Dependent variable is an indicator variable equal to 1 if the recorded race does not equal to the estimated race. The regression includes county FE, year FE, and month 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. Each regression is run separately for motorists of each race, where race is identified using the estimated race. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 9: Hispanic Bias and Trooper Rank

	(1) cheat	(2) cheat	(3) cheat
I(Failure)	0.050^{**}	0.004^{**}	0.023^{***}
	(0.023)	(0.002)	(0.009)
I(Failure)xCorporal	0.005	-0.012^{**}	-0.016
	(0.053)	(0.006)	(0.018)
I(Failure)xSergeant+	0.102	0.006	-0.015
	(0.096)	(0.007)	(0.023)
I(Failure)xProb. Troop		-0.004	-0.104
		(0.030)	(0.083)
Observations	1354	26377	51055

Notes: Dependent variable is an indicator variable equal to 1 if the recorded race does not equal to the estimated race. 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. The omitted trooper rank is trooper. Each regression is run separately for motorists of each race, where race is identified using the estimated race. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 10: Hisp Bias on Labor Outcomes

	(1) Standardized values of (posbias)	(2) Standardized values of (posbias)	(3) Standardized values of (po
Experience	0.02^{**}	0.02^{**}	
	(0.01)	(0.01)	
Black	-0.01	-0.02	0.00
	(0.18)	(0.17)	(0.18)
Hispanic	0.13	0.15	0.14
	(0.11)	(0.11)	(0.10)
Probationary Troop		0.31	
		(0.30)	
Corporal		-0.29^{**}	
		(0.14)	
Sergeant+		0.15	
		(0.12)	
Salary (1000s)			0.12^{**}
			(0.05)
Observations	1037	1037	1037

Notes: Regression includes controls for the work city and is clustered at the work city level. Troopers with negative levels of bias are omitte
Troopers with rank equal to or higher than sergeant (lieutenant, major, captain) were grouped into "Sergeant +". Salary is monthly salary r

Table 11: Hisp Bias on Labor Outcomes - Panel Results

	(1) Pr(Left Force)	(2) Pr(Moved Cities)	(3) Pr(RankUp)	(4) Δ Salary
Hispanic bias	-0.000	-0.014	0.008	-0.007
	(0.013)	(0.015)	(0.017)	(0.012)
Probationary	-0.067^{*}	0.191	-0.834^{***}	0.159^{***}
Trooper	(0.038)	(0.133)	(0.054)	(0.035)
Corporal	-0.023	-0.075	-0.121^{*}	0.032
	(0.037)	(0.062)	(0.062)	(0.040)
Sergeant+	-0.012	0.055	0.012	0.037
	(0.054)	(0.076)	(0.070)	(0.053)
Black	0.055	-0.107	0.064	-0.214^{**}
	(0.100)	(0.100)	(0.095)	(0.108)
Hisp	-0.094	0.142	-0.053	-0.096
	(0.073)	(0.165)	(0.166)	(0.104)
Observations	818	766	766	766

Notes: Troopers with negative levels of bias are omitted from the regression. Regression has robust standard errors. Dependent variable is the officer level measure of bias from Eq. (15) using only stops from 2010 to 2013. Employment outcomes are from 2013 and 2014. Troopers with rank equal to or higher than sergeant (lieutenant, major, captain) were grouped into "Sergeant +". Omitted categories are white for trooper race and trooper for trooper rank. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 12: Hisp Bias on Labor Outcomes - Transition Matrix

	(1)
Prob Trooper	-0.000
	(0.011)
Corporal	0.021
	(0.050)
Sergeant	-0.014
	(0.051)
Observations	766

Notes: Dependent variable is the probability of increasing in rank conditional on being the rank observed in the row. Each variable in the row is the reported rank of the trooper in 2013 interacted with the trooper's level of bias. Troopers with rank equal to or higher than sergeant (lieutenant, major, captain) were grouped into "Sergeant +". Regression has robust standard errors. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 13: Hisp Bias on Complaints

	(1) Complained	(2) Sustained
Hisp Bias	0.010*	0.012**
	(0.006)	(0.005)
Constant	0.057***	0.054***
	(0.005)	(0.005)
Observations	2041	2041
R^2	0.002	0.003
F	3.284	4.756

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

Table 14: Hisp Bias on Labor Outcomes - after 2015

	(1) Prob(Left Force)	(2) Salary Difference	(3) Prob(Rank Up)
HispBias	0.018	-0.038**	-0.048***
	(0.013)	(0.015)	(0.016)
Black	0.054	0.067	0.141
	(0.057)	(0.113)	(0.132)
Hispanic	0.052*	0.142	0.208*
	(0.030)	(0.101)	(0.120)
Prob. Troop		1.472***	
		(0.228)	
Corporal		-0.141***	
		(0.032)	
Sergeant		0.562***	
		(0.044)	
Lieutenant		1.131***	
		(0.219)	
Observations	1032	816	816

Notes: Regression has robust standard errors show in parentheses and uses 2019 employment data posted publicly by the Texas Tribune. Includes controls for the trooper's gender. Each trooper is weighted by their total number of searches conducted from January 2010 to June 2015. Black and Hispanic are indicator variables equal to one if the trooper is black or Hispanic, respectively, and equal to one otherwise. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 15: Negative Bias and Trooper Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
			I(Hisp Bias < 0)			
Black	0.003	0.002				
	(0.045)	(0.045)				
Hispanic	−0.025	−0.023				
	(0.024)	(0.024)				
Prob. Troop		0.010	−0.037			
		(0.073)	(0.078)			
Corporal		−0.032	−0.021			
		(0.032)	(0.032)			
Sergeant+		−0.006	0.030			
		(0.035)	(0.041)			
Salary			−0.029*			
			(0.017)			
Experience				−0.003*		
				(0.002)		
Total Searches					0.000	
					(0.000)	
Total Searches Hisp						0.000
						(0.000)
N	716	716	716	716	852	852

Notes: Dependent variable is an indicator variable equal to one if the trooper has negative bias and 0 if he has no bias. Regression uses employment data 2013 - 2015. Omitted category for trooper rank is trooper and the omitted category for trooper race is white. Salary is monthly salary in thousands of dollars.* $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$