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

(Latest Version Available Here)

Elizabeth Luh

October 30, 2019

Abstract

Troopers may misreport the race of people that they engage with in order to evade detection of racial bias. 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 using a unique event in Texas where troopers were caught deliberately misreporting minority motorists as white, . 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. My results show misreporting was used effectively to evade detection of bias, with bias having no causal effects on labor market outcomes when the misreporting was possible. I also find that a rule change to trooper stop recording policy in response to the misreporting led to biased troopers being less likely to be promoted and had lower salary growth. This rule was also effective in curbing lying behavior in troopers and reducing bias against Hispanic motorists.

1 Introduction

In the United States, disparate treatment of individuals by race in the criminal justice system is a pervasive and prevalent issue. Black and Hispanic motorists are more likely to be stopped and searched in traffic and highway stops despite having the lowest rates of contraband recovered (Harris, 1999; Makarechi, 2016). Recent research has also linked these disparities to racial bias and discrimination in other steps 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). This has raised concern by policy makers and the general public alike, especially with the recent, highly publicized deadly interactions between motorists and police.[1] According to a series of surveys in 2017, titled *Discrimination in America*, 27% of Latinos and 50% of blacks surveyed felt personally discriminated against when interacting with police compared to only 10% of white respondents.[2]

Despite the focus of policy makers and researchers on testing for racial bias and discrimination in the criminal justice system, little attention has been given to the response of law enforcement officers to this heightened scrutiny. These behavioral responses are important as law enforcement officers have considerable discretion when recording interactions. Discriminating officers may find it more costly to be fair in their treatment of individuals by race and instead misreport their interactions with motorists to appear less discriminatory. Thus, any policies targeting discrimination may miss the most discriminatory officers who need the most discipline and training. Even more concerning, with widespread misreporting, agencies and institutions could be underestimating the presence of discrimination within their department.

In my paper, I use a documented instance where Texas Highway Police Officers (henceforth "troopers") were caught misreporting minority motorists' race as white to appear less biased (Collister, 2015b). This misreporting 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 motorists' actual race and the recorded race. Troopers were able to miscategorize a portion of their minority motorist searches as white, thus decreasing the observed search rate of non-white motorists (Collister, 2015b). My objective for this paper is to link this discriminatory misreporting behavior to racial bias and to construct an individual measure of bias using the frequency of the misreporting behavior.

Studying racial bias in the presence of misreporting is advantageous for several reasons. One concern

[1] For example, Sandra Bland in 2015, Philando Castile in 2016, and Terence Crutcher in 2016.
[2] See https://cdn1.sph.harvard.edu/wp-content/uploads/sites/94/2018/01/NPR-RWJF-HSPH-Discrimination-Final-Summary.pdf for details.

when attributing observed disparities in the treatment of citizens of different race to bias is that many variables, such as driver demeanor, which can affect the outcome of trooper-motorist interactions, are not observable to researchers. In the context of this paper, troopers choose to misreport a portion of their searches to disguise their less justifiable searches, creating an observable distinction between racially biased searches and non-biased searches. Another issue is that assessing trooper's racial bias using only the search behavior is always relative to their peers, which only allows measures of relative bias or aggregate bias. With misreporting, the frequency of misreporting of each trooper provides an absolute measure of bias that is independent of the bias of his peers. Further, these repeated observations of misreporting within trooper allows me to create a measure of racial bias that is on the individual, trooper level.

The first step of this paper is to prove that misreporting can be use as a measure of racial bias. The major challenge is proving that statistical discrimination, or the difference in criminality by motorist race, is unrelated to trooper's misreporting behavior. To prove misreporting is linked to racial bias, I adapt a framework similar to Knowles et al. (2001) and Anwar and Fang (2006) where I show that only troopers biased against minority motorists have incentive to misreport minority searches as white. Further, unbiased troopers have no incentive to misreport at all. Specifically, my model shows that biased troopers will only misreport their failed, minority searches thereby appearing less biased.

The second step is to determine which searches in the data were purposefully misreported. Troopers may accidentally misreport motorists' race for reasons aside from bias (i.e poor visibility) and some motorists who appear non-white in the data, may self-identify as white. While the data can provide estimates of the driver's race, the true race and whether the misreporting was purposeful is unobserved. To deal with this, I use the different rates of mismatch between the estimated race and the observed race by search outcome for each trooper. From my model, unbiased troopers' rate of mismatch should not differ across search outcome. Whereas for biased troopers, the rate of mismatch will be higher for failed searches than for successful searches. Therefore troopers' bias is the difference in mismatch rate by search outcome. To test the identification of my measure of trooper bias, I exploit the discovery of the misreporting and the subsequent rule change at the end of 2015 requiring troopers to verbally ask motorists' for their race. I show that biased troopers were statistically more likely than unbiased troopers to record a minority motorist as white when the search ended in failure prior to the rule change, but this difference disappears after the rule change. These results suggest that troopers who I identify as biased were using indeed using misreporting in a biased manner and that they were misreporting motorists who did not self-identify as white.

Using a simple regression framework, I find that misreporting was used most frequently against Hispanic motorists, who were 3.5 percent more likely to be misreported when the search ended in failure compared to when the search ended in success. I estimate that over 27,000 searches were mismatched for

Hispanic motorists, which is more than half of all estimated-Hispanic searches. Misreporting was used more effectively against black and Asian motorists with a differential misreporting rate of 20 percent and 40 percent respectively. But, Asian and black motorists make up a small proportion of mismatched motorists in my data set, I focus the rest of my results on Hispanic bias.[3]

My main contribution is to provide an estimate of racial bias at the individual, trooper level and to identify causal effects of bias on trooper's labor market outcomes. Specifically, I calculate the trooper's differential likelihood of misreporting a Hispanic motorist as white depending on the search outcome and use this differential likelihood as the trooper's measure of bias. I find that the average trooper is 17% more likely to misreport when the search ends in failure compared to success. Correlating trooper characteristics to bias, I find that lower paid, recently hired troopers are less racially biased. Surprisingly, I find that racial bias is not significantly correlated across trooper race.

I find that prior to the discovery of misreporting, bias had no effect on labor market outcomes for troopers who remained in the force. After the discovery of misreporting, every standard deviation in bias decreased the likelihood of promotion by 27 percent and decreased salary growth by 8 percent. Since these negative outcomes occur after the rule change, I interpret this as suggestive evidence that bias was effective in shielding biased troopers from punishment for bias.

I then show that misreporting was effective in decreasing the appearance of bias both in the aggregate search statistics and on the individual level. Further, I show that past tests of bias would underestimate the presence of bias in Texas when using the observed races in the data. By comparing average search outcomes by motorists' race with recorded and estimated races, I find 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 Asian and Hispanic motorists by 5 percent and 3.3 percent, respectively; the white search success rates decreased by 12 percent. The disproportionate decrease for white motorists is due to significantly fewer failed white searches compared to non-white searches.

In the remainder of the paper, I use the trooper-level measure of Hispanic to understand how the trooper's work environment affects own bias. How does the demographic composition of peers affect own bias? What characteristics of the work environment affect trooper bias? Do policies need to vary by trooper race to effectively reduce racial bias?

This paper contributes 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 Knowles et al. (2001) and Anwar and Fang (2006) along with Antonovics and Knight (2009). My paper is the first to address the possibility of the

[3]Asian and black motorists make up only $< 2\%$ of all mismatched searches in the data set.

data being purposefully misreported to hide bias and to use this misreporting to differentiate between statistical discrimination and racial bias.

As is well known, racial disparities in aggregate statistics is not evidence of racial bias. For example, if Hispanic motorists are more likely than white motorists to carry contraband, then the aggregate number of searches and stops for Hispanic motorists would be higher even if race was not a factor in the decision. Moreover, troopers attempting to maximize successful searches might racially profile motorists. Such *statistical discrimination*, 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.

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 success 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 observing all of 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.

This paper also contributes to a growing literature of identifying bias and discrimination on the individual level. 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. Using the individual officer's difference in leniency across motorist race, they identify discrimination by comparing these lenient officers to non-lenient officers. This is vastly different from my paper in that I study searches where the motivation for differential treatment may be different and they do not link their measure of discrimination to racial bias.

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.

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. In Section 4, I explain my data construction. Section 5 shows my empirical results and other testable implications of my model. I finally conclude in Section 6.

2 Background

2.1 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. Figure 1 shows the division map across Texas.

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, 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

Suppose we have troopers and motorists: motorists of race m and troopers of race-t. I assume troopers and motorists are of race M (minority) and W (white). 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 $\theta \in (0,1)$.[4] θ is drawn from the distribution $f_g^m(.)$ if

[4]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,

the driver does carry contraband and from $f_n^m(.)$ if the driver does not carry contraband. For simplicity, I assume that the two densities are continuous and satisfy the strict monotone likelihood ratio property (MLRP). This implies that the $\frac{f_g^m}{f_i^m}$ is strictly increasing in θ. Intuitively, this property implies that a higher index of θ implies a higher probability of driver guilt. The MLRP also implies that the cumulative distribution, $F_g^m(.)$ stochastically dominates $F_n^m(.)$. In other words, motorists who carry contraband are more likely to appear more suspicious, or signal higher θ's. I also assume that $\frac{f_g^m}{f_n^m} \to 0$ as $\theta \to 1$ as some motorists may be obviously guilty.

3.1 The Search Decision

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 $c_{m,t}$. For simplicity, I assume the trooper receives a benefit normalized to one if the driver is guilty so that $0 < c_{m,t} < 1$.

G denotes 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:

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

Based on the expected payoff of searching, trooper of race-t will search a motorist of race-m with observed signal θ if and only if:

$$\Pr(G = 1|m, \theta) \geq c_{m,t} \quad (2)$$

The trooper has a search threshold θ^* where (2) holds with equality.

Similar to Knowles et al. (2001) and Anwar and Fang (2006), I define racial bias and statistical discrimination using the search cost, $c_{m,t}$.

Definition 1. A trooper of race-t exhibits "naive racial bias" against motorist of race M if $c_{M,t} < c_{W,t}$.

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

Troopers may use a different search threshold for race M motorist than with race W motorist depending on the distribution of θ. For example, if $c_{M,t} = c_{W,t}$, but $\pi_M > \pi_W$, it will be optimal for the trooper to have a lower search threshold for race M motorists and search race M motorists at a higher

whether the vehicle is rented, and the attitude of the driver.

frequency. Another reason may be due to the different underlying distributions of $f_g^m(.)$ and $f_n^m(.)$ across motorist race.

3.2 The Misreporting Decision

With probability $v(c_{M,t}, c_{W,t})$, troopers biased against race M motorists can be punished for racial bias, where $v(c_{M,t}, c_{W,t}) = |c_{W,t} - c_{M,t}|$. The more naively biased a trooper is, or the greater the difference of $c_{W,t} - c_{M,t}$, the higher likelihood the trooper will be punished for racial bias.[5]

Troopers can choose to misreport some searches to reduce the risk of being punished. The cost of misreporting is $\mu(\theta, G)$, where $\mu(\theta, G)$ is increasing in θ and G. Specifically, misreporting is too costly when the search ends in success, or $\mu(\theta, G = 1) > 1$.[6]

The trooper decides whether to misreport the motorist of race M as a motorist of race W after observing the search outcome. The trooper will misreport a race M motorists as W if and only if:

$$c_{M,t} + \mu_{M,t}(\theta, G) \leq c_{W,t} \tag{3}$$

which implies the trooper will only misreport if the misreporting reduces $v(c_{M,t}, c_{W,t})$. Misreporting the search decreases the probability of punishment by $\mu_{M,t}(\theta, G)$. From Eq.(3):

Proposition 1. *The trooper has a misreporting threshold, $\theta^{\mu}_{M,t}$ s.t Eq. (3) holds with equality. Moreover, only troopers exhibiting naive bias against race M motorists will misreport failed searches of race M motorists as race W.*

In my model, misreporting has a small cost. Therefore if a trooper is unbiased and $c_{M,t} = c_{W,t}$, no θ exists such that Eq. (3) holds. Proposition 1 also implies that the cost of misreporting the marginal motorist in failed search is equal to the difference in the search costs across motorist race.

Proposition 2. *Troopers will misreport motorists of characteristics (M, θ) if and only if $\theta \in (\theta^*, \theta^{\mu})$ and if and only if the search ends in failure ($G = 1$).*

Proposition 2 comes from the fact that troopers must first search a motorist to determine if she will misreport the motorist. Thus, only motorists with a $\theta \in (\theta^*, \theta^{\mu})$ will be misreported.

[5] For simplicity, I derive the empirical implications of the model assuming troopers are biased against race M motorists, but the results hold if the bias is reversed.

[6] In the appendix, I show my propositions in the subsequent sections will hold the misreporting cost is only a function of θ. In successful searches, another person is more likely to see the record of the search (i.e in court) so the trooper only misreports when the search ends in failure.

3.3 Theoretical Implications

Given the search threshold from Eq. (2), troopers' average search rate of race m motorists is:

$$\gamma_{m,t} = \pi_m[1 - F_g^m(\theta^*)] + (1 - \pi_m)(1 - F_n^m(\theta^*)] \tag{4}$$

For a race t trooper, the *average search success rate* of race m motorist is:

$$S_{m,t} = \frac{\pi_m[1 - F_g^m(\theta^*)]}{\pi_m[1 - F_g^m(\theta^*)] + (1 - \pi_m)[1 - F_n^m(\theta^*)]} \tag{5}$$

With misreporting, the $\gamma_{m,t}$ and $S_{m,t}$ are unobservable to the researcher. Let $\gamma_{m,t}^O$ and $S_{m,t}^O$ denote the observed search rate and search success rate, respectively, of race m motorist.

$$\gamma_{M,t}^O = \pi_M[1 - F_g^M(\theta^*)] + (1 - \pi_M)[1 - F_n^M(\theta^\mu)] \tag{6}$$

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

From Eq. (3), only race M motorists are misreported as W. Specifically, Proposition 2 implies that only race M motorists of $\theta > \theta^\mu$ will be correctly recorded as race M. This reduces the observed search rates of race M motorists so that $\gamma_{M,t}^O < \gamma_{M,t}$. The misreporting also raises the race W search rate so that $\gamma_{W,t}^O > \gamma_{W,t}$.

This affects the observed search success rates for race M and W motorists:

$$S_{M,t}^O = \frac{\pi_M[1 - F_g^M(\theta^\mu)]}{\pi_M[1 - F_g^M(\theta^\mu)] + (1 - \pi_M)[1 - F_n^M(\theta^\mu)]} \tag{8}$$

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

From Proposition 1, unsuccessful race M searches are more likely to be misreported. This increases the observed search success rates for race M motorists and decreases the observed success rate for race M so that $S_{M,t}^O > S_{M,t}$ and $S_{W,t}^O < S_{W,t}$.

The misreporting rate is:

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

which is the portion of race M motorists with $\theta \in (\theta^*, \theta^\mu)$ who are not guilty, derived from Proposition 2. The magnitude of $\phi_{M,t}$ is also informative on the magnitude of bias. 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 $\theta^*_{M,i} < \theta^*_{M,j}$. From Proposition 2, this implies that:

$$\Rightarrow \theta^\mu_{M,i} - \theta^*_{M,i} > \theta^\mu_{M,j} - \theta^*_{M,j}$$
$$\Rightarrow \phi_{M,i} > \phi_{M,j}$$

Thus, since trooper i is more biased than trooper j, trooper i also misreports a higher portion of race M searches than trooper j.

Proposition 3. *If a trooper exhibits sophisticated racial bias against race M motorists, then the trooper will misreport race M motorists, $\phi_{M,t} > 0$. Further the misreporting rate, $\phi_{M,t}$, is the magnitude of bias against race M motorist.*

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.

4 Data

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. Including searches where the motorist is arrested prior to searching the vehicle 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.

4.2 Trooper Employment Data

The employment data is from the Texas Department of Public Safety, which I obtained using a Freedom of Information Act (FOIA). 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 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 further the time period of my trooper employment data by adding 2019 trooper employment data, publicly available on the Texas Tribune Salary website. I link the 2019 employment data to my trooper data using the full name of the trooper. I include this data as a measure of trooper's long-term employment outcomes.

I also include trooper complaint data from 2010 - 2015 via a FOIA. The complaint data contains information on the date the incident occurred, the date the complaint was received, the allegation of the complaint, the employee's name, the trooper's badge number (if applicable), and the investigator of the complaint. The badge number is not always included due to Texas' privacy laws. Out of the original 1,873 complaints, only 334 had the trooper's badge number in the complaint.

I merge the stop data to the trooper data 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.

4.3 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.[7] 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

[7] 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."[8] 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 misreported 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.

4.4 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

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

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.[9] 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 Empirical Results

My test for racial prejudice

I first test if troopers in my data are indeed misreporting motorists as white differentially by search outcome. Table 4 shows the results of this test, 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 mismatch between the estimated race and the observed race is driven by bias, then the probability of mismatch should be greater when the search ends in failure compared to success.

[9] 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.

Columns (1) and (2) of Table 4 show the misreporting rate conditional on search outcome. 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. Since I observe bias with both black and Asian motorists, this also shows that my method of uncovering bias is robust to not only the race estimation technology, but is not unique to just Hispanic motorists, since I detect bias against Asian motorists using the same race estimation technique.

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

$$I(Mismatch_{i,t}) = \alpha + \beta_1 I(Failure)_{i,t} + X_i\gamma + c + t + c * t + m + \epsilon_{i,t} \tag{11}$$

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_i is vector of controls of the stop, such as vehicle characteristics, whether the driver is in-state, and if the owner is the driver. I also include year trends, county fixed effects, 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. (11), 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 j's bias is:

$$I(Mismatch)_{i,j,t} = \alpha + \beta_1^j I(Failure)_{i,t} + \delta_j + X_{i,c}\gamma + m + \epsilon_{i,j,t} \tag{12}$$

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 11 in $X_{i,c}$. 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:

$$I(RecRace = White_{i,c,j,t}) = \alpha + \sum_{t=2010}^{2017.5} \Big[\beta_3^t Hisp\ Bias_j \times I(Year = t) + \beta_4^t I(Failure_{i,c,j,t}) \times I(Year = t) \\ + \beta_5^t Hisp\ Bias_j \times I(Year = t) \times I(Failure_{i,c,j,t})\Big] + \boldsymbol{X}_{i,c,t}\gamma + \epsilon_{i,t} \tag{13}$$

I use the *recorded* race rather than the estimated race because the recorded races will show the greatest change after June 2015 if my measure, $Hisp\ Bias_j$, actually quantifies officer j's bias, where $Hisp\ Bias_j$ is derived from Eq. (**??**). $I(Failure)$ 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. (11) 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 motorists recorded as white. 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)).

Past tests for racial prejudice

Misreporting can distort the results when using 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.

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.[10] 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 (11). 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 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 (1957)'s test of bias. I regress:

$$\begin{aligned} Y_{icjt} =& \alpha + \beta_1 Hisp\ Bias_j + DriverRace_i \boldsymbol{\beta_2} + DriverRace_i \times Hisp\ Bias_j \boldsymbol{\beta_3} \\ & + \gamma_m + \delta_c + \epsilon_{icjt} \end{aligned} \tag{14}$$

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 6 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.

[10] 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.

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.

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 7 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. (11), I run the regression:

$$\begin{aligned} I(Mismatch)_{i,j,t} = \alpha + \beta_1 I(Failure)_i + I(TroopRace_j)\beta_{2j} + \\ I(Failure) \times I(TroopRace_j)\beta_{3j} + X_{c,t} + \epsilon_{i,j,t} \end{aligned} \tag{15}$$

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 7. 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 8. Since I have few troopers with rank greater than corporal, I combine sergeants, lieutenants, majors, and captain 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 9. 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 10. 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 decompose column 3 from Table 10 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 11, 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 12 show a positive relationship between trooper level bias and the probability of receiving a complaint. One standard deviation of bias is associated with a one percentage point higher likelihood in having a complaint filed against the trooper. This is possibly an underestimate of the actual association of bias and complaints since much of the complaint data repressed the trooper badge number from the stop. 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, significant results.

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 13 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.

Bias and the Work Environment

Another important question is whether troopers are biased because of their own preferences or because of the environment they work in. Many characteristics of the environment can affect own bias, such as social norms, demographic and labor characteristics of the environment, or peers. Additionally, in the context of Texas, troopers may have different incentives (for example: reducing drug trafficking, reducing undocumented immigration, ensuring border protection) to search motorists depending on which county they patrol in. Understanding the underlying mechanism of bias is helpful for informing policymakers in how to target policy to counteract bias.

To construct measures of peer composition within county, I restrict my sample of troopers to those who have conducted 75% of their searches within the same county. This excludes troopers who patrol multiple counties who may not be present enough within the county to be considered a peer. Using these non-mover troopers, I construct estimates of the demographic and labor composition of troopers working within the county.

I measure the correlation of county's trooper composition to county characteristics in Table 15. The county characteristic data is from the 2011-2015 American Community Survey. I also include the average, annual violent crime rate at the county level from the University of Wisconsin Population Health Institute. My results in Table 15 show that assignment of troopers by race to counties is non-random. Specifically, Hispanic troopers are more likely to be patrolling in counties with high Hispanic populations, black troopers are more likely to patrol in counties with high white population, and white troopers are more likely to patrol in areas with high black population, high percentage of Latin American foreign born, lower percentage of uninsured, and lower median income. The high estimate of the constant for percentage of white troopers in Column (3) also shows that white troopers tend to be assigned to the same counties.

To determine if the composition of peers within the county has any causal effects on trooper's own bias, I use troopers who have conducted at least 5 searches in two different counties and were excluded when constructing the county level average bias. These troopers were excluded from the measurement of peer composition in Table 15. I create a measure of Hispanic bias for each trooper-county pair. I have 283 troopers with enough observations across multiple counties to measure bias for each trooper on the county level. Of these 283 troopers, 16 are black, 80 are Hispanic, and 182 are white. I regress trooper bias on peer composition variables with officer fixed effects to identify the causal effects of moving.

My results in Figure 11 show differential causal effects of peer composition on own bias by trooper race. I find that a higher proportion of white peers increases own bias significantly for Hispanic troopers. I also find positive effects for black troopers, but I do not have enough statistical power for these estimates to be significant. Interestingly, black troopers bias increases when patrolling in counties with more white troopers or Hispanic troopers. This implies that black troopers bias decreases when patrolling in counties with a higher fraction of black troopers. Changes in average salary and experience of peers have no causal effects on trooper's bias, but the race composition of ranked officers has causal effects on Hispanic trooper's bias. Specifically, a 10 percentage point increase in the percent of white troopers increases Hispanic trooper bias by 0.2 standard deviations.

I also test if county characteristics have a causal effect on own bias. Figure 12 show the causal effects of county characteristics on own bias. I find that Hispanic and white troopers' bias are not affected by county characteristics. Black troopers bias rises significantly when moving counties with higher percentage of Latin American Foreign born. This result aligns with Figure 11, since from Table 15, counties with higher percent of Latin American foreign born are more likely to have white peers. Black troopers bias increases when moving to counties with lower rates of uninsured population and higher median income, which are also counties with higher percentages of white troopers.

Lastly, I test whether peer bias affects own bias. To create a measure of county-level bias, I exclude troopers who move to different counties during the time period and keep only troopers who have conducted 75% of their searches within one county.[11] Troopers who patrol in the same county patrol within the same environment. Any underlying county characteristics that affect trooper's search motivation will be the same for troopers patrolling within the same county. I measure county level bias by taking the average level of bias of troopers within the county. I show the distribution of average bias by county in Figure 10. My results show that county level bias has a positive mean level of bias of 0.108, which is slightly lower than the average, trooper-level bias.

My results in Figure 13 show the correlation of own bias to county level bias without trooper fixed in effects in the first graph and with trooper fixed effects in the second graph. The results with no trooper fixed effects shows a positive correlation between county bias and Hispanic trooper's own bias. But, these results are not significant after including trooper fixed effects. The inclusion of trooper fixed effects allows for a causal interpretation of county bias on own bias. My results show county bias has no effect on own bias since the estimates are small and insignificant even when separating the results by race.

The combination of all of these results suggests the effect of the environment on trooper's bias varies by trooper race. I find that white trooper's bias is impervious to changes in characteristics in the environment, including peer composition, peer bias, and county characteristics. On the other hand, Hispanic and black troopers show varying levels of sensitivity to the composition of peers. Notably, Hispanic troopers increase in own bias when patrolling in counties with more white troopers and this effect is amplified when the percentage of ranked, white officers increases. For black troopers, I find suggestive, but not significant, evidence, that an in increase in white trooper peers increases own bias.

In terms of policy implications,

6 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 apply my model using a rich data set of Texas highway stop data from 2010 to 2017 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, and misreported stops by search outcome, I am able to develop a comprehensive test for racial bias.

In my results, I find misreporting was used effectively to evade detection of bias. When troopers were able to misreport motorists' race, I find no significant effects of bias on trooper's probability of ranking

[11] In the appendix, I test if troopers who conduct 75% of their searches within the same county are significantly different in Hispanic bias than troopers who switch between multiple counties and I find no significant difference.

up or growth in salary. After misreporting became impossible due to a rule change at the end of 2015, I find that biased troopers were less likely to be promoted and were paid $30 less for every one standard deviation of bias than their unbiased peers.

I find that after the rule change, troopers no longer differentially misreported minority motorists as white if the search ended in failure. This rule change was relatively costless in terms of application and enforceability but was able to stop troopers from misreporting motorists. Without the ability to hide bias, I find that the least biased troopers changed their search behavior towards Hispanic motorists.

Using my measure of bias, I find the county-level average Hispanic bias is more correlated to trooper-level bias than work-city bias. I use this as evidence that the environment the trooper patrols in affects their own bias more than the environment the trooper reports to. My results show that policy targeted towards changing the peer composition of troopers within the county may be able to change the average level of bias within the county.

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.

7 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*.

Becker, G. (1957). *The economics of discrimination.* University of Chicago Press.

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*.

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

Makarechi, K. (2016). What the data really says about police and racial bias. *Vanity Fair*.

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.

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 16 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 9, 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 16 show that troopers with negative bias in fact search more in general and search more Hispanic motorists compared to troopers with no bias.

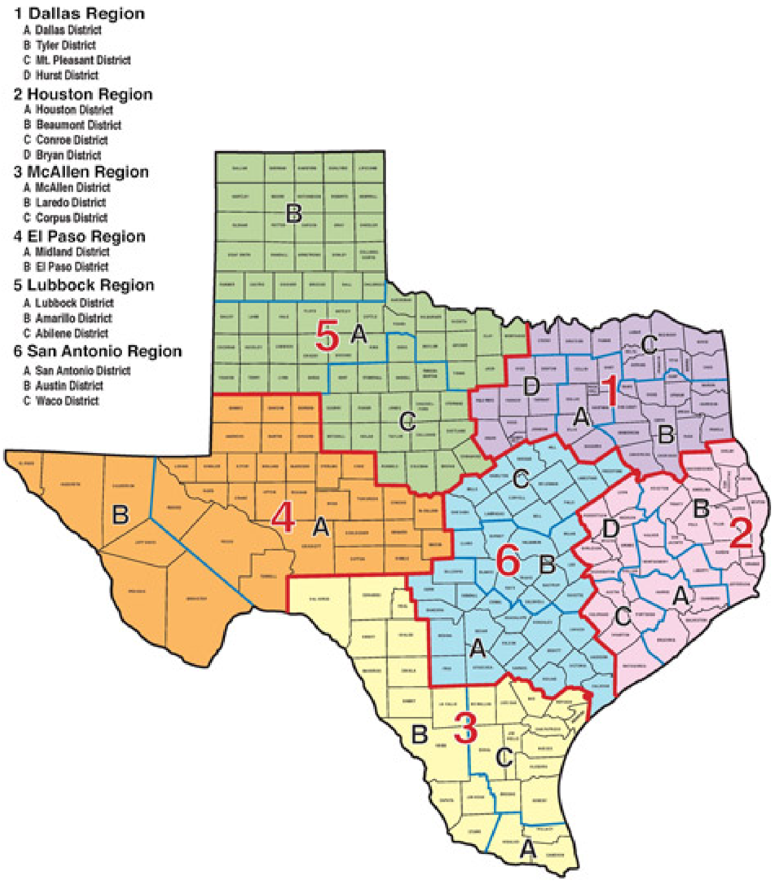
Further, if there was a systematic issue with the race correction, then troopers with negative bias would misreport significantly more when searches with Hispanic motorists ended in success. To test this, I regress a version of Eq. 11 but include a triple interaction indicating whether the trooper's level of bias is negative. Specifically:

$$\begin{aligned}
I(RecRace = White)_{i,j,t} = \alpha + \beta_0 I(CorrRace = Hispanic)_i + \beta_1 I(Failure)_i + \\
\beta_3 I(Hisp\ Bias_j < 0) + \beta_4 I(CorrRace = Hispanic)_i \times I(Failure)_{i,j,t} + \\
\beta_5 I(CorrRace = Hispanic)_i \times I(Hisp\ Bias_j < 0) + \beta_6 I(Failure)_i \times I(Hisp\ Bias_j < 0) \\
\beta_7 I(CorrRace = Hispanic)_i \times I(Hisp\ Bias_j < 0) \times I(Failure) + \\
X_{i,j,c,t}\gamma + \epsilon_{i,j,t}
\end{aligned} \tag{16}$$

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.

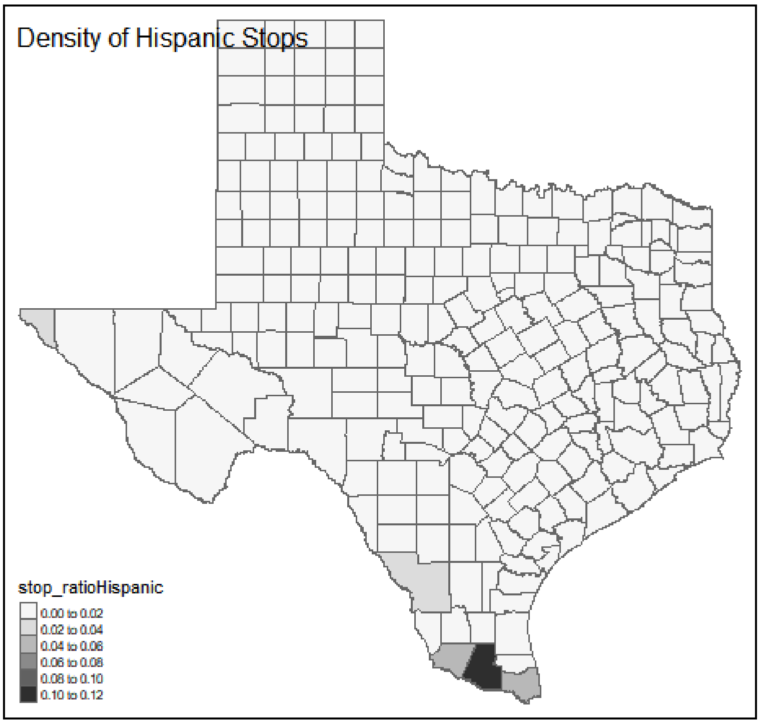
7.1 Figures

Figure 1: Trooper Division Map



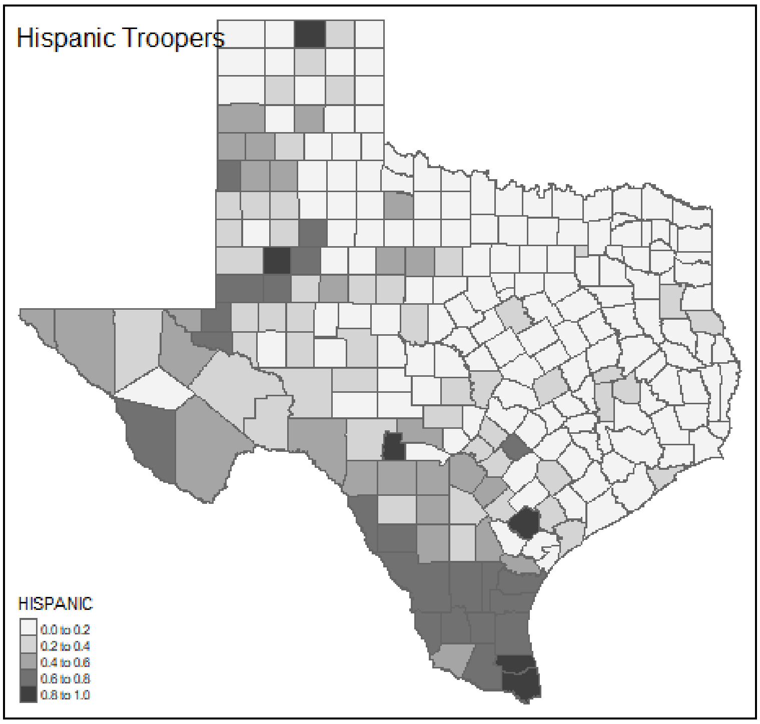
Notes: Source: Texas Department of Public Safety. The 7th region is not shown on the map, but its jurisidiction is limited to only Austin, TX.

Figure 2: Hispanic Motorist Stop Density by County



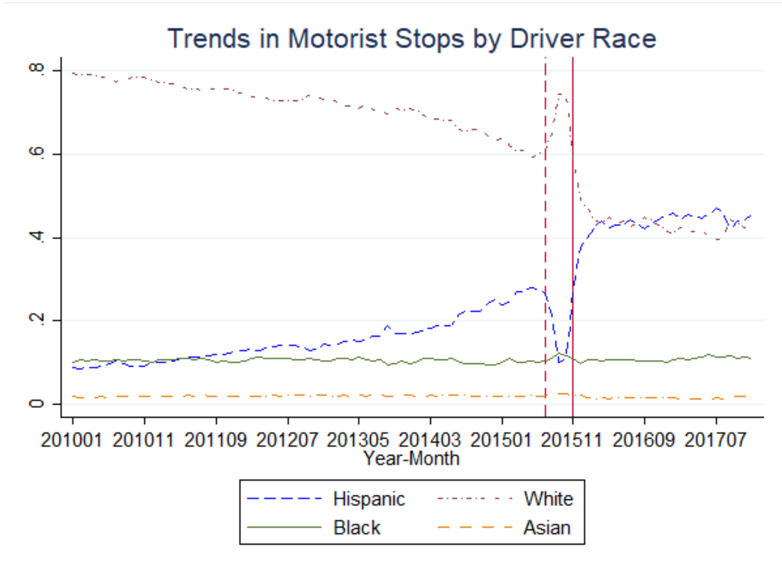
Notes: Unweighted means are shown using data from 2010 - 2015. It shows the percentage of Hispanic stops per county using the estimated races. Darker regions imply higher percentages.

Figure 3: Hispanic Trooper Density by County



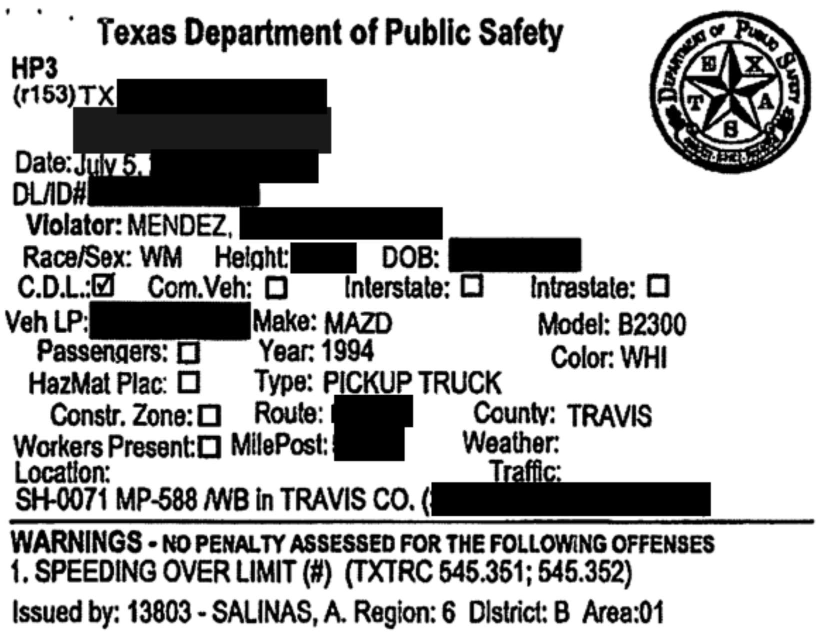
Notes: Unweighted mean of stop rate by driver race for each year-month time period from January 2010 to June 2015. Each cell shows the percentage of stops made by Hispanic troopers. Darker regions imply higher percentages.

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



Notes: 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. Average stop rates for a given month-year from January 2010 to December 2017 are shown. The vertical red line indicates the year-month the article was released. The dashed red line indicates the year-month Sandra Bland died after a trooper stop.

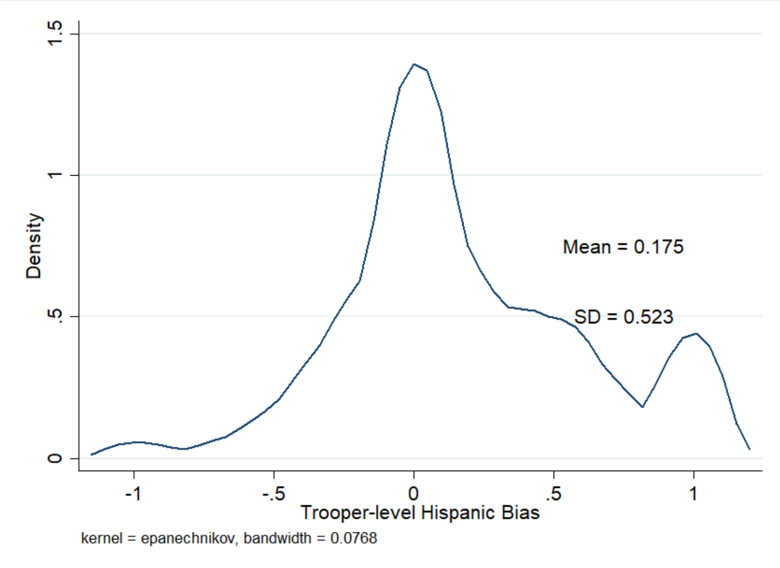
Figure 5: Example of misreported Highway Ticket



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]

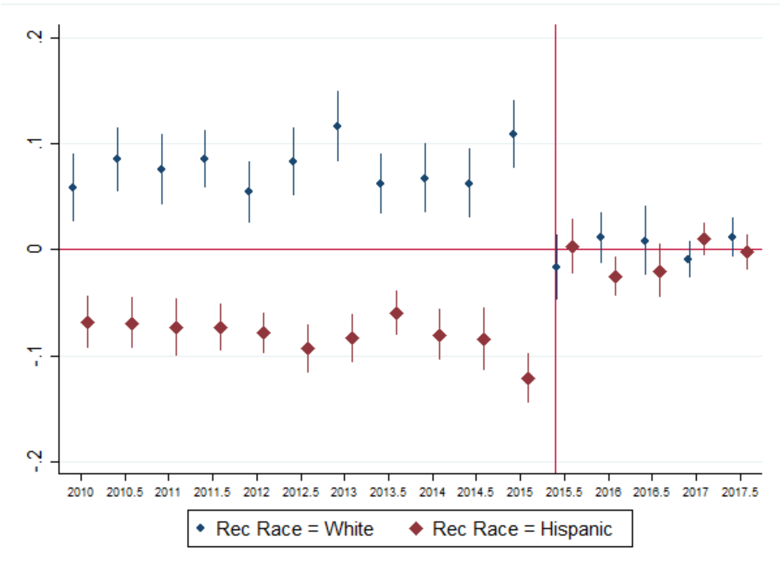
WARNINGS - NO PENALTY ASSESSED FOR THE FOLLOWING OFFENSES
1. SPEEDING OVER LIMIT (#) (TXTRC 545.351; 545.352)
Issued by: 13803 - SALINAS, A. Region: 6 District: B Area: 01

Figure 6: Distribution of officer level measure of Hispanic Bias



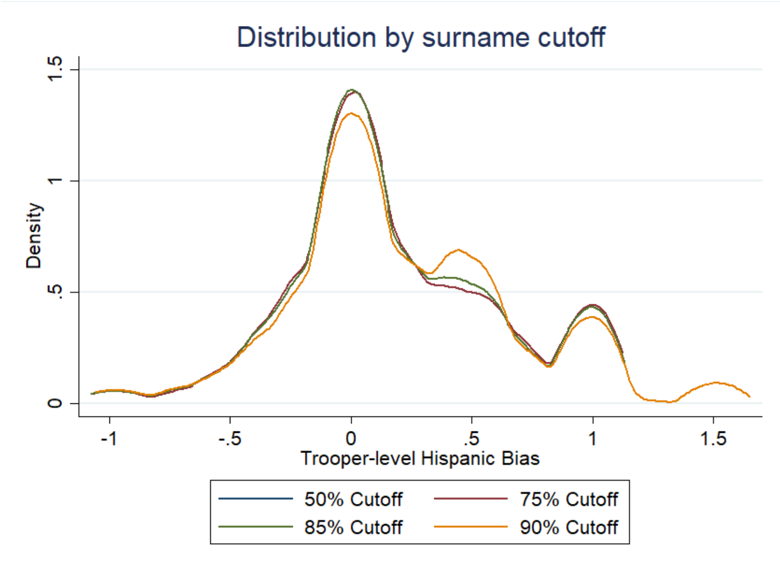
Notes: Kernel density distribution of officer-level Hispanic bias. The figure plots each officer's β^j from the regression $I(Mismatch_{i,t}) = \alpha + \beta^j I(Failure)_{i,t} + \delta_j + X_{i,c,t}\gamma + \epsilon_{j,t}$. Mean reports the average β_1^j and Avg S.E. reports the average standard error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

Figure 7: Natural Experiment - White and Hispanic Motorists



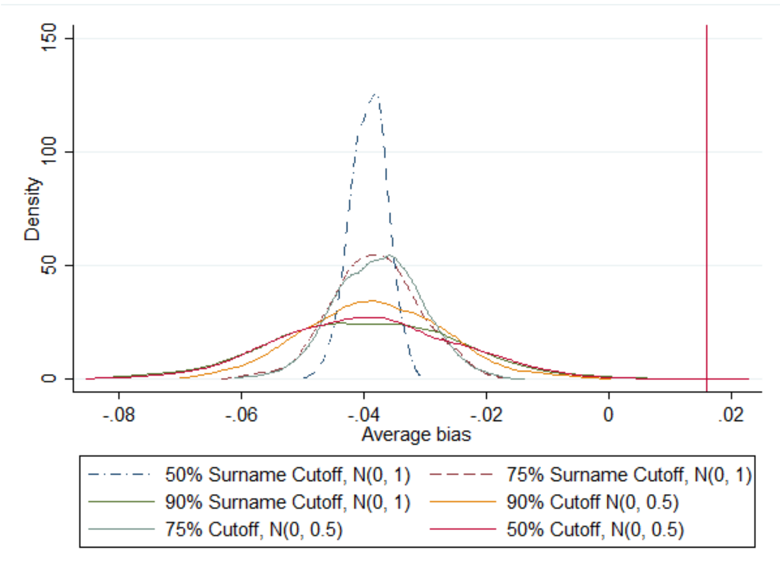
Notes: Figure plots the coefficient of interaction $I(Failure_i) \times Hisp\ Bias_j \times I(Year\ Half = t)$, β_5^t, and with 5% confidence intervals. $Hisp\ Bias_j$ is standardized. Points to the right of the vertical line are after the article publication. Diamond points are when the dependent variable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



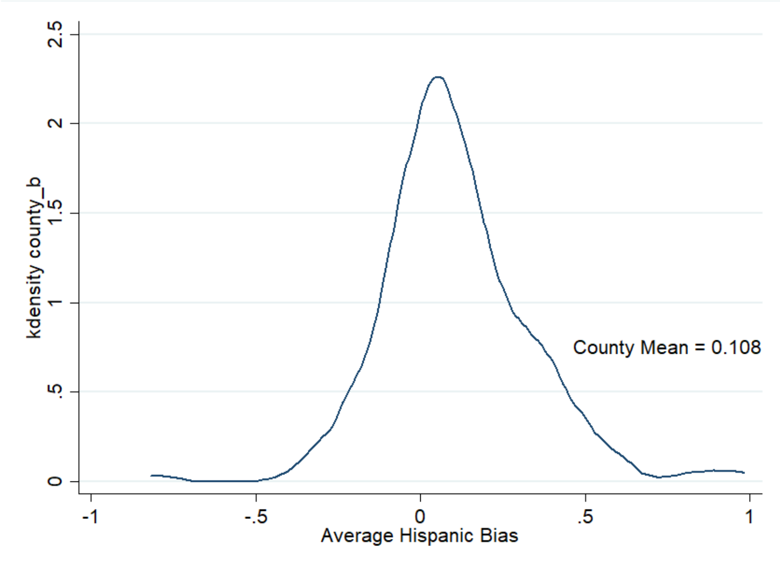
Notes: Each density shows the officer level bias using different levels of surname cutoff and weighted by the total searches.The estimate of bias is from each officer's β^j from the regression $I(Mismatch_{i,t}) = \alpha + \beta^j I(Failure)_{i,t} + \delta_j + X_{i,c,t}\gamma + \epsilon_{j,t}$.

Figure 9: Monte Carlo Simulation of Bias



Notes: Each density shows the average officer level bias using different normal distributions from Eq (11).

Figure 10: Distribution of County Hispanic Bias

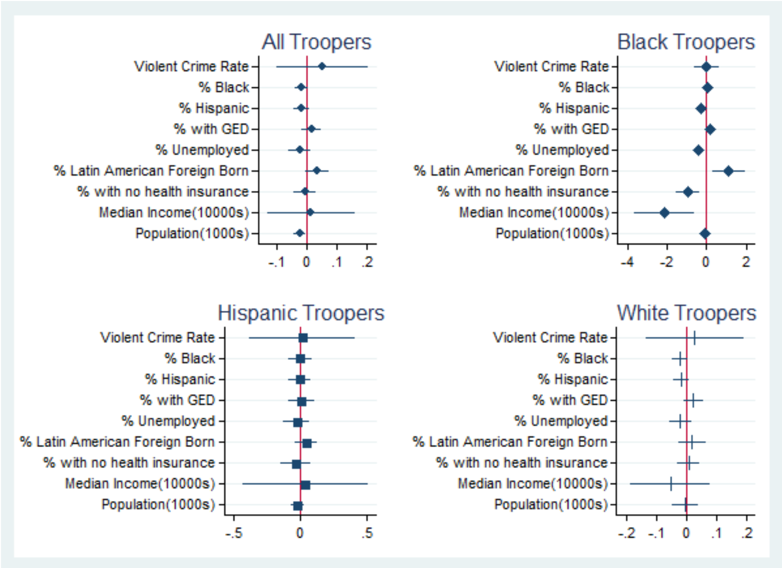


Notes: Kernel density distribution of county-level and work city level Hispanic bias. The figure plots each geographic area's average level of bias aggregated from the trooper's level of bias. Each geographic area only contains troopers who conducted over 75% of searches within the county and did not move work cities from 2013 - 2015. The density has 238 counties and 203 different work cities with 1,250 troopers.

Figure 11: Causal Effects of Peer Composition on Own Bias

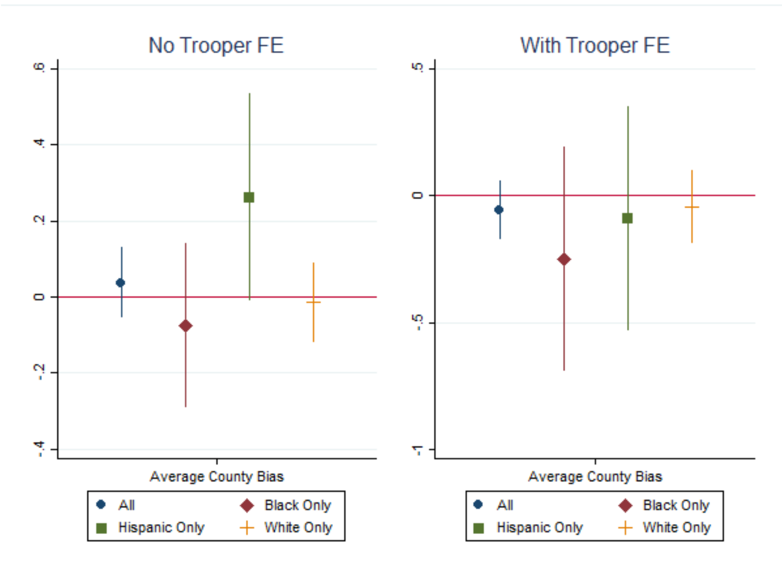
Notes: Each graph shows the coefficient estimate with 5% confidence intervals for the variable in the x-axis. Regression is run separately for all troopers and by race with robust standard errors.

Figure 12: Causal Effects of County Characteristics on Own Bias



Notes: Each graph shows the coefficient estimate with 5% confidence intervals for the variable in the x-axis. Regression is run separately for all troopers and by race with robust standard errors.

Figure 13: Effects of Peer Bias on Own Bias



Notes: Each graph shows the coefficient estimate with 5% confidence intervals for the variable in the x-axis. Regression is run separately for all troopers and by race with robust standard errors.

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.

7.2 Tables

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	$Pr(Mismatch\|Failure)$	$Pr(Mismatch\|Success)$	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		Hispanic Motorists	
I(Failure)	0.056**	0.052**	0.003*	0.003*	0.018**	−0.009
	(0.021)	(0.021)	(0.002)	(0.002)	(0.008)	(0.007)
Constant	0.126***	0.128***	0.014***	0.015***	0.559***	0.578***
	(0.013)	(0.013)	(0.001)	(0.001)	(0.005)	(0.005)
County, Month FE	X	X	X	X	X	X
Year FE		X		X		X
Observations	1520	1520	29031	29031	58098	58098
F	6.679	5.965	3.699	3.420	5.850	1.692

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 estimat 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 th variables $I(Failure)$ through the fixed effects are equal to zero. Standard errors are clustered at the county level. Regression uses data fro 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 Hispanic Bias	0.001^{***}	0.039		
	(0.000)	(0.025)		
Recorded Black x Hispanic Bias	-0.003^{***}	0.033^{***}		
	(0.001)	(0.008)		
Recorded Hispanic x Hispanic Bias	-0.002^{***}	0.119^{***}		
	(0.000)	(0.012)		
Hispanic 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 Hispanic Bias			0.001^{**}	0.007
			(0.000)	(0.021)
Estimated Black x Hispanic Bias			-0.003^{***}	0.006
			(0.001)	(0.008)
Estimated Hispanic x Hispanic 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: Hispanic Bias is the normalized measure of Hispanic bias for each trooper. The measure for Hispanic bias comes from Eq. (12). 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: 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)
Month FE, County FE	X	X	X	X	X	X
Year FE		X		X		X
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. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. HispanicTroopers 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 8: Hispanic Bias and Trooper Rank

	(1)	(2)	(3)
	Asian Motorists	Black Motorists	Hispanic Motorists
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)xProbationary 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 and month 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 9: Correlates of Hispanic Bias

	(1)	(2)	(3)	(4)	(5)
			Hispanic Bias		
Experience	0.02**	0.02**			
	(0.01)	(0.01)			
Black	−0.01	−0.02	0.00	−0.01	0.02
	(0.18)	(0.17)	(0.18)	(0.18)	(0.18)
Hispanic	0.13	0.15	0.14	0.15	0.17
	(0.11)	(0.11)	(0.10)	(0.11)	(0.11)
Probationary Troop		0.31		0.40	0.20
		(0.30)		(0.32)	(0.30)
Corporal		−0.29**		−0.28**	−0.23*
		(0.14)		(0.14)	(0.14)
Sergeant+		0.15		0.07	0.23*
		(0.12)		(0.14)	(0.13)
Salary (1000s)			0.12**	0.14**	
			(0.05)	(0.06)	
Observations	1037	1037	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 omitted from the regression. Dependent variable is the officer level measure of bias from Eq. (12). Troopers with rank equal to or higher than sergeant (lieutenant, major, captain) were grouped into "Sergeant +". Salary is monthly salary measured in thousands of dollars. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 10: Hispanic Bias on Labor Outcomes - Panel Results

	(1)	(2)	(3)	(4)
	(Pr(Left Force)	(Pr(Moved Cities)	(Pr(RankUp)	Salary Difference
Hisp bias	0.065***	−0.002	0.008	−0.007
	(0.022)	(0.019)	(0.017)	(0.012)
Prob. Troop	−0.137**	0.214	−0.834***	0.159***
	(0.068)	(0.141)	(0.054)	(0.035)
Corporal	−0.047	−0.054	−0.121*	0.032
	(0.080)	(0.068)	(0.062)	(0.040)
Sergeant	0.287***	0.256	0.012	0.037
	(0.083)	(0.157)	(0.070)	(0.053)
Black Trooper	0.146	−0.188	0.064	−0.214**
	(0.205)	(0.201)	(0.095)	(0.108)
Hisp Trooper	−0.236**	0.171	−0.053	−0.096
	(0.110)	(0.220)	(0.166)	(0.104)
Constant	0.093	0.037	0.913***	0.806***
	(0.093)	(0.108)	(0.069)	(0.041)
Observations	818	559	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. (12) 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 11: Hispanic Bias on Labor Outcomes - Transition Matrix

	(1)
Probationary 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 12: Hispanic Bias on Complaints

	(1) Complained	(2) Sustained
Hispanic 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 13: Hispanic 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)	
Constant	0.237***	0.502***	0.176
	(0.070)	(0.096)	(0.143)
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 14: County Characteristics and Hispanic Bias

	(1) $Hispanic\ Bias_{county}$	
Violent Crime Rate	-4.25	(8.70)
% Hisp	-1.46	(1.06)
% Black	-0.77	(1.18)
% no health ins	1.23	(2.52)
% HS diploma	0.36	(2.31)
Median HH inc (10000s)	-0.08	(0.11)
% Employed	-5.25*	(3.01)
% older than 16	0.46	(3.91)
Population (100000s)	0.01	(0.01)
Border County	0.31	(0.27)
Constant	5.39	(3.79)
Observations	187	
R^2	0.077	
F	2.85	

Notes: Dependent variable is officer j's level of Hispanic Bias. Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

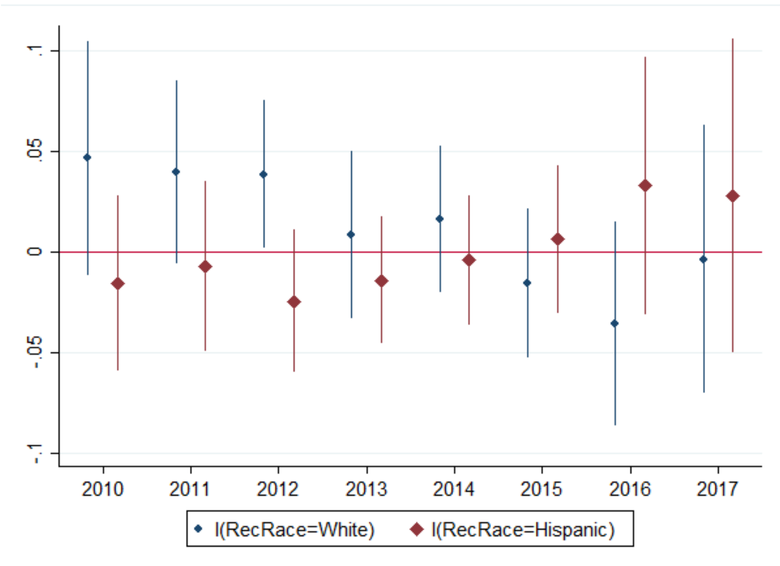
Table 15: Correlations of Trooper and County Characteristics

	(1) % Black Troopers	(2) % Hispanic Troopers	(3) % White Troopers
Violent Crime Rate	-1.047***	1.169	-0.148
	(0.372)	(1.134)	(1.232)
% Black	-0.548***	0.139	0.494***
	(0.075)	(0.114)	(0.136)
% Hisp	-0.516***	1.092***	-0.484***
	(0.064)	(0.092)	(0.117)
% HS diploma	0.104	0.528	-0.611**
	(0.099)	(0.328)	(0.309)
% Unemployed	0.147	0.020	-0.152
	(0.109)	(0.215)	(0.247)
laratio	0.008	-0.546***	0.541***
	(0.065)	(0.181)	(0.188)
% no health ins	0.361***	-0.008	-0.268
	(0.108)	(0.291)	(0.299)
Median HH inc (10000s)	0.016***	0.011	-0.027***
	(0.005)	(0.009)	(0.009)
Population (100000s)	0.004***	-0.002**	-0.002*
	(0.001)	(0.001)	(0.001)
Constant	0.334***	-0.521***	1.061***
	(0.078)	(0.191)	(0.208)
Observations	180	180	180
R^2	0.546	0.747	0.674
F	27.325	62.312	49.335

Notes: Dependent variable is in the column and is constructed from troopers who conduct at least 75% of their searches within that county.
Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

7.3 Appendix Tables and Figures

Figure 14: Natural Experiment (County) - White and Hispanic Motorists



Notes: Figure plots the coefficient of interaction $I(Failure_i) \times Hisp\ Bias_{county} \times I(Year\ Half = t)$, β_5^t, and with 5% confidence intervals. $Hisp\ Bias_{county}$ 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 $I(RecRace = Hispanic)$.

Table 16: 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.* $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$