An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence

Shamena Anwar† Hanming Fang‡

June 16, 2004

Abstract

We exploit a simple but realistic model of trooper behavior to design empirical tests that address the following two questions. Are police monolithic in their search behavior? Is racial profiling in motor vehicle searches motivated by the troopers’ desire for effective policing (statistical discrimination) or by their racial prejudice (racism)? Our tests require data sets with race information about both the motorists and troopers. When applied to vehicle stop and search data from Florida, our tests can soundly reject the null hypothesis that troopers of different races are monolithic in their search behavior; but fail to reject the null hypothesis that no racial groups of troopers are racially prejudiced.

*Preliminary and incomplete. All comments are welcome. We are grateful to Joseph Altonji for his constant support and many useful suggestions. We also thank Richard Zeller at Florida Highway Patrol for providing us with the data, and Jessica Duan for able research assistance in the beginning stage of this project. Fang is grateful to financial support from a Yale University Junior Faculty Fellowship. All remaining errors are our own.

†Department of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520-8268. Email: shamena.anwar@yale.edu

‡Department of Economics, Yale University, P.O. Box 208264, New Haven, CT 06520-8264. Email: hanming.fang@yale.edu
1 Introduction

Black motorists in the United States are much more likely than white motorists to be searched by highway troopers. Several recent lawsuits against state governments have used this racial disparity in treatment as evidence of “racial profiling,” a term that refers to the practice that police officers use race as one of the criteria in deciding whether to search cars. Racial profiling originated with the attempt to interdict the flow of drugs from Miami up Interstate 95 to the cities of Northeast. For example, in 1985 the Florida Department of Highway Safety and Motor Vehicles issued guidelines for police on “The Common Characteristics of Drug Couriers,” in which race/ethnicity was explicitly mentioned as one characteristic (Engel, Calnon and Bernard, 2002). While the initial motivation for such guidelines may be to increase the troopers’ effectiveness in interdicting drugs, it also unfortunately opened up the possibility for troopers to engage in racist practices against minority motorists.

Following the public backlash generated by several cases in the 1990s such as Wilkins v. Maryland State Police [1996] and Chavez v. Illinois State Police [1999], almost all highway patrol departments have denounced race as a criterion in stop and search decisions. But many citizens, especially minorities, are skeptical of this claim: motor vehicle search decisions, by its very nature, are made in the midst of face-to-face interactions, thus it is simply hard to imagine that troopers can block the race and ethnicity information. Moreover, data on trooper searches continue to show that they tend to search a higher proportion of minority motorists than white motorists. As is now well known, however, racial disparities in the aggregate rates of stops and searches do not necessarily imply racial prejudice (see, for example, Knowles, Persico and Todd 2001, Engel, Calnon and Bernard 2002). If, for example, black drivers are more likely than white drivers to carry contrabands, then the aggregate rate of stops and searches would be higher for black drivers even when race were hypothetically invisible to troopers. Moreover, racial profiling may also arise if police attempt to maximize successful searches and race helps predict whether a driver carries contraband. This situation is called statistical discrimination in the terminology of Arrow (1973).

How can we empirically distinguish racism from statistical discrimination? This question has garnered enormous public and academic interest (see, for example, National Research Council 2004), but it is also challenging, partly as a result of data limitations. For example, unless truly random searches are conducted, researchers typically would not observe the true proportion of drivers who
carry contrabands. Some ethnographic studies, for example, Sherman (1980) and Riksheim and Chermak (1993), have shown that, many situational factors, including suspects' demeanor in the police-citizen encounter, influence police behavior. Such data are also typically unavailable.

A seminal paper by Knowles, Persico and Todd (2001, KPT hereafter) developed a simple but elegant theoretical model about motorist and police behavior that suggests an empirical test using data on search outcomes (i.e., the percent of searches in which contraband was found) for each racial group – a statistic typically available to researchers. The primary idea of KPT’s empirical test is the outcome test originated in Becker (1957). It is based on the following intuitive notion. If troopers are profiling minority motorists due to racial prejudice, minorities are likely to be searched even when the returns from searching minorities, i.e., the probabilities of successful searches against minorities, are smaller than whites. More precisely, if racial prejudice is the reason for racial profiling, then the success rate of the marginal minority motorist (i.e., the last minority motorist deemed suspicious enough to be searched) will be lower than the success rate of the marginal white motorist. In contrast, if racial profiling results from statistical discrimination, that is, if the troopers are profiling to maximize the number of successful searches, then the optimality condition would require that the search success rate for the marginal minority motorist be equal to that of the marginal white motorist. While this idea has been well-understood, it is problematic in empirical applications because researchers would never to able to directly observe the search success rate of the marginal motorists; instead we can only observe the average success rate of white and minority searches. Precisely for this reason, KPT proposed a simple model of motorist and police behavior to cleverly circumvents this problem. In their model, motorists differ in their characteristics, including race and possibly other characteristics (that are observable to troopers but may or may not be available to researchers). Troopers decide whether to search motorists and motorists decide whether to carry contrabands. In this “matching pennies”-like model, they show that, if troopers are not racially prejudiced, then all motorists, regardless of their race and other characteristics, would in equilibrium carry contraband with equal probability, thus there is no difference between the marginal and the average search success rates. Their model thus suggests a

---

1The ideas in this paper are inspired from reading KPT, from which we learned a great amount.

2See, also, Ayres (2001) for other applications of the outcome test idea.

3They of course allow the motorists with different characteristics to have different costs and benefits from carrying contrands. These differences, however, only imply that in equilibrium troopers will search motorists with different
simple test based on the comparison of the average search success rate by the race of the motorists. A lower average search success rate implies racial prejudice against that group. Applying their test to a data set of 1,590 searches on a stretch of I-95 in Maryland from January 1995 through January 1999, they found no evidence of racial prejudice against African-American motorists, but Hispanic drivers are indeed prejudiced against.

KPT’s model of motorist and police behavior provides a theoretical rationale for the average search success rate as the basis of empirical test for racial prejudice. Therefore, the validity of the test also hinges on the realism of the model that justifies it. We now present two weaknesses of the model.\(^4\)

First, KPT’s model predict that all motorists for a given race, regardless of their other characteristics that may be observed by the police, will carry contraband with equal probability – the vital prediction that allows them to equate the average search success rate in a given racial group of motorists to the marginal search success rate. This, however, also implies that a motorist’s characteristics other than race should provide no information when a trooper decides whether to search. This implication of the police behavior simply goes against trooper guidelines which require them to base their search decisions on the information the motorist presents to the trooper at the time of the stop, including the motorist’s personal characteristics, their demeanor, and the contents of their vehicle that are in plain view etc. (see, e.g., Sherman 1980 and Riksheim and Chermak 1993). KPT’s basic model assumes that motorists’ characteristics are exogenous, thus rules out the plausible scenario that a motorist’s demeanor when stopped are intimately related to whether or not he or she is carrying contraband.

Second, KPT (and this field of research in general) assume that all troopers’ behavior are monolithic. Due to lack of data on the characteristics, including race information, of troopers, it is characteristics at different rates. In fact, these different search rates provide the necessary deterrence to ensure that all motorists will carry contraband with equal probabilities.

\(^4\)Dharmapala and Ross (2003) also point out that KPT’s test does not generalize if potential drug carriers may not be observed by the police or if there are different levels of drug offense severity. In the first case, the equilibrium of the model may involve a group of motorists carrying drugs with probability one even when they are searched with probability one whenever the troopers observe them (KPT recognized this issue in their footnote 16). If the probability of being a “dealer” is higher for minorities, then the average success rate for minorities should be greater than that for whites under statistical discrimination; and equal average success rates would actually indicate taste discrimination, contrary to KPT’s conclusion. In the second case, KPT’s test has to be modified.
assumed that all troopers have the same racial prejudice against motorists, regardless of their races. While there is no direct evidence on this assumption in the context of highway searches, Donohue and Levitt (2001), in their study on arrest patterns and crime, find that the racial composition of a city’s police force has an important impact on the racial patterns of arrests, suggesting that police behavior (or information they possess) are not monolithic. The consequence of an invalid monolithic trooper behavior assumption is serious. Imagine a world in which minority troopers are racially prejudiced against white motorists, while white troopers are prejudiced against minority motorists. It is possible that when examining the aggregate search outcomes of white and minority troopers, we would reach a conclusion that the police as a whole is not racially prejudiced. But this seriously underestimates the harassment experienced by both white and minority motorists.

In this paper, we develop an alternative model of motorist and police behavior in which troopers are allowed to behave differently depending on their race and the race of the motorists they interact with. Our model does not yield the convenient, but in our view unrealistic implication that all drivers of the same race carry contraband with the same probability. As a result, the distinction between average and marginal search success rate becomes, yet again, the central issue in the empirical determination of racial prejudice versus statistical discrimination. Our model follows the spirit of labor market statistical discrimination model (see, e.g., Coate and Loury 1993). Police officers observe noisy but informative signals about whether or not a driver carries contraband when they decide if a search is warranted. Guilty drivers, i.e., drivers who actually carry contrabands, are more likely than innocent drivers to generate suspicious signals. A police officer incurs a cost of search that depends on both his/her own race and the race of the motorist. Troopers of a particular race, say \( r_p \), are said to be racially prejudiced if their cost of searching motorists depend on the race of the motorist. The police force exhibit non-monolithic behavior if the cost of searching motorists of a given race, say \( r_m \), depend on the race of the troopers. Troopers are assumed make their search decisions to maximize the number of successful searches (or arrests). The optimal decision of an race-\( r_p \) police officer in deciding whether a race-\( r_m \) motorist should be searched satisfies a threshold property: the motorist should be searched if and only if his/her posterior probability of being guilty exceeds the cost of of race-\( r_p \) officer against race-\( r_m \) motorists, \( t (r_m; r_p) \). We show that the police officers exhibit monolithic behavior if and only if both the search rate and average search success rate against any given race of motorists are independent of the race of troopers. Moreover,
if none of the racial groups of the troopers are racially prejudiced, then the ranking of search rates, and average search success rate as well, for a given race of motorists across the race of troopers should not depend on the race of the motorists. That is, if troopers of race $r_p$ have a higher search rate against race-$r_m$ motorists than troopers of race $r'_p$, race-$r_p$ troopers should also have a higher search rate against race-$r'_m$ motorists than race-$r'_p$ troopers. We use these theoretical predictions of the model to design empirical tests for both monolithic behavior and racial prejudice. Another nice feature of our model is that it could potentially be refuted by the data we have available.

The implementation of our empirical tests relies on data sets that have race information on both the troopers and motorists. While such data has not been used in earlier empirical studies on racial profiling, we are able to obtain it from the Florida Highway Patrol which contains information on all vehicle stops and searches conducted on Florida highways between January 2000 and November 2001, together with the demographics of the trooper that conducted each stop and search. In implementing our empirical tests, we found strong evidence that the Florida Highway Patrol troopers do not exhibit monolithic behavior, but we fail to reject the null hypothesis that none of the racial groups of troopers are racially prejudiced.

The remainder of the paper is structured as follows. Section 2 presents and analyzes our model of trooper search behavior, and propose empirical tests based on the theoretical predictions of the model; Section 3 describes the data set from the Florida Highway Patrol, presents our test results, and contrasts our results with those using KPT’s test; Section 4 concludes. In Appendix A we present a simple equilibrium model of drug carrying behavior to show that our focus on trooper behavior in Section 2 is not problematic.

2 The Model

We now present a simple model of trooper search behavior that underlines the empirical work in Section 3.2. There is a continuum of troopers (interchangeably, police officers) and motorists (interchangeably, drivers). Let $r_m$ and $r_p \in \{M, W\}$ denote the race of the motorists and the troopers respectively, where $M$ stands for minorities and $W$ for whites.\(^5\) Suppose that among

\(^5\)In the empirical part of the paper, we will examine three racial or ethnic groups, whites, blacks, and Hispanics. For now, we group blacks and Hispanics as minorities for the ease of exposition.
motorists of race \( r_m \in \{M,W\} \), a fraction \( \pi^{r_m} \in (0,1) \) of them carry contraband.\textsuperscript{6,7}

When deciding whether to search a vehicle, the police officer observes the race of the motorist. Besides race of the driver, many other characteristics of the motorist may be potentially used by the officer in the decision to search cars. Such characteristics may include, for example, the gender, age and residence address of the driver, the interior of the vehicle that is in the trooper’s view, the smell from the driver or the vehicle, whether the driver is intoxicated, the demeanor of the driver in answering the trooper’s questions, the make of the car, whether the car has an out-of-state plate, whether the car is rented or owned, location and time of the stop, as well as the serious-ness of the reason for the stop, etc.\textsuperscript{8} Note that while the police officer observes all the characteristics in the decision to search, a researcher will typically have access to a small subset of them. We assume, however, that the police officer will use a single-dimensional index \( \theta \in [0,1] \) that summarizes the likelihood that a driver may be carrying contraband. We assume that, if a driver of race \( r_m \in \{M,W\} \) actually carries contraband, then the index \( \theta \) is randomly drawn from a continuous probability density distribution \( f_g^{r_m}(\cdot) \); and if a race \( r_m \) driver does not carry contraband, \( \theta \) would be randomly drawn from \( f_n^{r_m}(\cdot) \).\textsuperscript{9} Without loss of generality, we can assume that the two densities \( f_g^{r} \) and \( f_n^{r} \) satisfy strict monotone likelihood ratio property (MLRP), i.e., for \( r_m \in \{M,W\} \),

**MLRP:** \( f_{g}^{r_m}(\theta) / f_{n}^{r_m}(\theta) \) is strictly increasing in \( \theta \).

The MLRP property on the signal distributions essentially means that a higher index \( \theta \) is a signal that a driver is more likely guilty.\textsuperscript{10} To the extent that there are obviously guilty drivers,\textsuperscript{6} For the purpose of deriving our empirical test, we will assume that \( \pi^{r} \) is exogenous. For an equilibrium model in which \( \pi^{r} \) is endogenously determined, see Appendix A.\textsuperscript{7} A trooper must first stop the motorist prior to a search. Examining the possibility of racial prejudice in highway stops is beyond the scope of this paper. In our analysis, we will take the sample of cars that are stopped as our population and focus solely on determining racial prejudice in troopers’ search decisions. The presence, or lack thereof, of racial prejudice at the stop level should not affect our conclusions.\textsuperscript{8} The questions the trooper will ask the motorist are typically focused on where the motorist is headed and the purpose of their visit. In listening to the response the trooper will try to discern how nervous or defensive the motorist is, and how logical the motorist’s response is.\textsuperscript{9} The subscripts \( g \) and \( n \) stand for “guilty” and “not guilty” respectively.\textsuperscript{10} For any one dimensional index \( \theta \), we can always reorder them according to their likelihood ratio \( f_{g}(\theta) / f_{n}(\theta) \) in
for example, if illicit drug is in plain view, we assume that:

\textbf{Unbounded Likelihood Ratio:} \( f_{g}^{\text{rm}} (\theta) / f_{n}^{\text{rm}} (\theta) \to +\infty \text{ as } \theta \to 1. \)

MLRP also implies that the cumulative distribution functions \( F_{g}^{\text{rm}} (\cdot) \) first order stochastically dominates \( F_{n}^{\text{rm}} (\cdot) \), which implies that drivers who actually carry contraband are more likely to generate higher and thus more suspicious signals. We think single dimensional index formulation summarizes information available to troopers when making search decisions on the highway in a simple but realistic manner.

Each police officer can choose to search a vehicle after observing the driver’s vector \( (r_{m}, \theta) \), where \( r_{m} \) is the driver’s race and \( \theta \) is the single-dimensional index that summarizes all other characteristics observed during the stop. We assume that a trooper wants to maximize the total number of convictions (or the number of drivers found carrying illicit contrabands) minus a cost of searching cars.\(^{11}\)

Let \( t(r_{m}; r_{p}) \) be the cost of a police officer with race \( r_{p} \) searching a motorist \( r_{m} \), where \( r_{p}, r_{m} \in \{ M, W \} \). We normalize the benefit of each arrest to equal one, and scale the cost of search to be a fraction of the benefit, so that \( t(r_{m}; r_{p}) \in (0, 1) \) for all \( r_{m}, r_{p} \). It is worth emphasizing that, different from KPT, we allow the troopers’ cost of searching a vehicle to depend on the races of both the motorist and the officer, thus we can directly confront the possibility that the police officers may not be monolithic in their search behavior.

Let \( G \) denote the event that the motorist searched is found with illicit drugs in the vehicle. When a police officer observes a motorist of race \( r_{m} \) and signal \( \theta \), the posterior probability that such a motorist may be guilty of carrying contraband, \( \Pr(G|r_{m}, \theta) \), is obtained via Bayes’ rule:

\[
\Pr(G|r_{m}, \theta) = \frac{\pi^{r_{m}} f_{g}^{r_{m}} (\theta)}{\pi^{r_{m}} f_{g}^{r_{m}} (\theta) + (1 - \pi^{r_{m}}) f_{n}^{r_{m}} (\theta)}. \]

It immediately follows from the MLRP that \( \Pr(G|r_{m}, \theta) \) is monotonically increasing in \( \theta \). From the unbounded likelihood ration assumption, \( \Pr(G|r_{m}, \theta) \to 1 \text{ as } \theta \to 1. \)

\(^{11}\)This is also the police objective postulated in KPT. It is a plausible assumption because awards (such as Trooper of the Month honors) and/or promotion decisions are partly based on troopers’ success in catching motorists with contrabands.
The decision problem faced by a police officer of race $r_p$ when facing a motorist with race $r_m$ and signal $\theta$ is thus as follows:

$$\max \{ \Pr (G|r_m, \theta) - t(r_m; r_p) ; 0 \}$$

where the first term is the expected benefit from searching such a motorist and the second term is the benefit from not searching normalized at zero. Thus the optimal decision for a trooper of race $r_p$ is to search a race-$r_m$ motorist with signal $\theta$ if and only if

$$\Pr (G|r_m, \theta) \geq t(r_m; r_p).$$

From the monotonicity of $\Pr (G|r_m, \theta)$ in $\theta$, we thus conclude:

**Proposition 1** A race-$r_p$ police officer will search a race-$r_m$ motorist if and only if

$$\theta \geq \theta^* (r_m; r_p)$$

where $\theta^* (r_m; r_p)$ is uniquely determined by

$$\Pr (G|r_m, \theta^* (r_m; r_p)) = t(r_m; r_p).$$

Moreover, the search threshold $\theta^* (r_m; r_p)$ is monotonically increasing in $t(r_m; r_p)$.

Proposition 1 says that the probability of search success for the marginal motorist is equal to the cost of search. Any infra-marginal motorist would have a higher search success probability. In what follows, we will refer to $\theta^* (r_m; r_p)$ as the *equilibrium search criterion* of race-$r_p$ police officers against race-$r_m$ motorists. Define the *equilibrium search probability* of race-$r_p$ police officers against race-$r_m$ motorists as $\gamma (r_m; r_p)$, given by

$$\gamma (r_m; r_p) = \pi^{r_m} \left[ 1 - F_{g}^{r_m} (\theta^* (r_m; r_p)) \right] + (1 - \pi^{r_m}) \left[ 1 - F_{n}^{r_m} (\theta^* (r_m; r_p)) \right]. \quad (1)$$

The *equilibrium average search success rate* of race-$r_p$ police officers against race-$r_m$ motorists, denoted by $S (r_m; r_p)$, is given by

$$S (r_m; r_p) = \frac{\pi^{r_m} \left[ 1 - F_{g}^{r_m} (\theta^* (r_m; r_p)) \right]}{\pi^{r_m} \left[ 1 - F_{g}^{r_m} (\theta^* (r_m; r_p)) \right] + (1 - \pi^{r_m}) \left[ 1 - F_{n}^{r_m} (\theta^* (r_m; r_p)) \right].} \quad (2)$$

We now introduce three definitions. First, a police officer of race $r_p$ is defined to be *racially prejudiced* if he or she exhibits a preference for searching motorists of one race. Following KPT, we model this preference in the cost of searching motorists.
**Definition 1** A police officer of race \( r_p \) is racially prejudiced, or has a taste for discrimination, if \( t(M; r_p) \neq t(W; r_p) \).

Next, we say that the police does not exhibit **monolithic behavior** if officers of different races do not use the same search criterion when dealing with motorists of some race.

**Definition 2** The police officers do not exhibit monolithic behavior if \( t(r_m; M) \neq t(r_m; W) \) for some \( r_m \in \{M, W\} \).

Note that a monolithic police force does not mean that they are not racially prejudiced: it could be that police officers of both races are equally prejudiced against some race of motorists. Likewise, a non-monolithic police force does not necessarily imply that some racial group of troopers are racially prejudiced: it could be that each group of troopers have the same search cost against all groups of motorists, but that cost depends on the race of the trooper.

Finally, we say that race-\( r_p \) police officers exhibit **statistical discrimination** if they have no taste for discrimination and yet they use different search criterion against motorists with different races.

**Definition 3** Assume \( t(M; r_p) = t(W; r_p) \). Then race-\( r_p \) police officers exhibit statistical discrimination if \( \theta^* (M; r_p) \neq \theta^* (W; r_p) \).

Now we derive some simple implications of the model that will serve as the basis of our empirical test. First, note that if the police officers are monolithic, then their cost of searching any given race of motorists are the same regardless of the race of the officer. Thus Proposition 1 implies that both race of officers will use the same search criterion against a given race of motorists. Thus following from the formula for the search rate (1) and average search success rate (2), we have:

**Proposition 2** If the police officers exhibit monolithic behavior, then \( \gamma (r_m; M) = \gamma (r_m; W) \) and \( S (r_m; M) = S (r_m; W) \) for all \( r_m \in \{M, W\} \).

Next, if none of the police officers are racially prejudiced, then the ranking of \( t(r_m; M) \) and \( t(r_m; W) \) does not depend on the motorist’s race \( r_m \). If \( t(M; M) = t(W; M) < t(M; W) = t(W; W) \) for both \( r_m \), then the search criteria of race-M police officers applying to both races of motorists will be lower than those applied by race-W officers, which implies that race-M officers’
search rates will be higher for both races of motorists. Similarly, if \( t(M; M) = t(W; M) > t(M; W) = t(W; W) \), then race-\( M \) officers’ search rates will be lower for both races of motorists than race-\( W \) officers. In fact we have the following stronger result:

**Proposition 3** If neither race-\( M \) nor race-\( W \) of police officers exhibit racial prejudice, then neither

the ranking of \( \gamma(r_m; M) \) and \( \gamma(r_m; W) \) nor the ranking of average search success rates \( S(r_m; M) \) and \( S(r_m; W) \) depends on \( r_m \in \{M, W\} \). Moreover, for any \( r_m \), the ranking of \( \gamma(r_m; M) \) and \( \gamma(r_m; W) \) should be the exact opposite of the ranking of \( S(r_m; M) \) and \( S(r_m; W) \).

Now we explain the ranking independence result regarding the average search success rate under

the null hypothesis of no racial prejudice. Suppose without loss of generality that \( t(M; M) = t(W; M) < t(M; W) = t(W; W) \). The there will be the following ranking on the search criteria \( \theta^*(W; M) < \theta^*(W; W) \) and \( \theta^*(M; M) < \theta^*(M; W) \). The average search success rate with a search criterion \( \theta^* \) against race-\( r_m \) motorist is simply

\[
\frac{\pi^{\text{rm}} [1 - F^{\text{rm}}_g (\theta^*)]}{\pi^{\text{rm}} [1 - F^{\text{rm}}_g (\theta^*)] + (1 - \pi^{\text{rm}}) [1 - F^{\text{rm}}_n (\theta^*)]}
\]

and one can show that it is strictly increasing in \( \theta^* \). Thus we have \( S(W; M) < S(W; W) \) and
\( S(M; M) < S(M; W) \). That is, the ranking of \( S(r_m; M) \) and \( S(r_m; W) \) does not depend on \( r_m \).

The contrapositive of Proposition 3 is simply that: if the ranking of \( \gamma(r_m; M) \) and \( \gamma(r_m; W) \),
or the ranking of \( S(r_m; M) \) and \( S(r_m; W) \), depend on \( r_m \), then at least one racial group of the

---

\(^{12}\)To see this, note that it will be strictly increasing in \( \theta^* \) if and only if

\[
H(\theta) = \frac{1 - F^{\text{rm}}_g (\theta^*)}{1 - F^{\text{rm}}_n (\theta^*)}
\]

is strictly increasing in \( \theta^* \). Note that

\[
H'(\theta^*) = \frac{-f^{\text{rm}}_g (\theta^*) [1 - F^{\text{rm}}_n (\theta^*)] + [1 - F^{\text{rm}}_g (\theta^*)] f^{\text{rm}}_n (\theta^*)}{[1 - F^{\text{rm}}_n (\theta^*)]^2}
\]

\[
= \frac{-f^{\text{rm}}_g (\theta^*) \int_{\theta^*}^1 f^{\text{rm}}_n (\theta) d\theta + f^{\text{rm}}_n (\theta^*) \int_0^1 f^{\text{rm}}_g (\theta) d\theta}{[1 - F^{\text{rm}}_n (\theta^*)]^2}
\]

\[
= \frac{\int_{\theta^*}^1 [f^{\text{rm}}_g (\theta^*) f^{\text{rm}}_n (\theta) - f^{\text{rm}}_n (\theta^*) f^{\text{rm}}_n (\theta)] d\theta}{[1 - F^{\text{rm}}_n (\theta^*)]^2}.
\]

From MLRP, we know that, for all \( \theta > \theta^* \),

\[
\frac{f^W_g (\theta)}{f^W_n (\theta)} > \frac{f^W_g (\theta^*)}{f^W_n (\theta^*)},
\]

thus the integrand in the numerator is always positive. Thus \( H'(\theta^*) > 0 \).
troopers exhibit racial prejudice. Without further assumptions, it is not possible to determine which group of troopers are racially prejudiced.

2.1 Empirical Tests

2.1.1 Test for Monolithic Trooper Behavior

Proposition 2 suggests a test for whether troopers of different races exhibit monolithic search behavior that is implementable even when researchers have no access to the signals observed by troopers \( \theta \) in making their search decisions. Under the null hypothesis that police officers exhibit monolithic behavior, then, for any race of the drivers, the search rates and average search success rates against drivers of that race should be independent of the race of the troopers. That is, under the null hypothesis of monolithic trooper behavior, we must have, for all \( r_m \in \{M,W\} \),

\[
\gamma (r_m; M) = \gamma (r_m; W),
\]

(3)

\[
S (r_m; M) = S (r_m; W).
\]

(4)

Any evidence in violation of any of these equalities would reject the null hypothesis.

It is worth pointing out that both equalities (3) and (4) hold if and only if the null hypothesis is true. This claim is straightforward for equalities (3), but needs some explanation for equalities (4). Without loss of generality, let \( r_m = W \), and let \( t (W; W) > t (W; M) \). That is, suppose that white troopers experience a higher search cost than minority troopers in searching white motorists. It then immediately follows from Proposition 1 that \( \theta^* (W; W) > \theta^* (W; M) \), i.e., white troopers will use a more strict search criterion than minority troopers when searching white motorists. The arguments we used to show Proposition 3 immediately implies that \( S (W; W) > S (W; M) \). Thus the test using either (3) or (4) has an asymptotic power of one. It also suggests that in principle, our model can be refuted. For example, a situation in which equalities (3) hold, but equalities (4) fail would indicate that the data is inconsistent with the predictions of our model.

2.1.2 Test for Racial Prejudice

Proposition 3 suggests a test for whether some racial groups of the troopers exhibit racial prejudice in their search behavior. Under the null hypothesis that none of the racial groups of the troopers have racial prejudice, it must be true that both the ranking of search rates for a
given race of motorists \( r_m \) across the races of troopers \( \gamma(r_m; M) \) and \( \gamma(r_m; W) \), and the ranking of average search success rates \( S(r_m; M) \) and \( S(r_m; W) \), do not depend on \( r_m \in \{M, W\} \). The null hypothesis will be rejected if the rankings of \( \gamma(r_m; M) \) and \( \gamma(r_m; W) \) or \( S(r_m; M) \) and \( S(r_m; W) \), depend on the race of the motorists \( r_m \). Moreover, any time there is some motorist race \( r_m \) such that the ranking of \( \gamma(r_m; M) \) and \( \gamma(r_m; W) \) is not exactly opposite to the ranking of \( S(r_m; M) \) and \( S(r_m; W) \), our model is refuted by the data.

This test, however, has an asymptotic power less than one. That is, one may fail to reject the null hypothesis even when it is false. To see this, suppose that the truth is \( t(M; M) = t(W; M) < t(M; W) < t(W; W) \). That is, race-\( M \) officers are not racially prejudiced, but race-\( W \) officers are prejudiced against minorities (race-\( W \) officers’ cost of searching minority motorists are smaller). In this case, race-\( W \) officers will apply higher search criteria toward both races of motorists, thus the race-\( W \) officers’ search rates will be lower regardless of the race of the motorists. Thus the null would not be rejected even it is false, and we commit the type-II error. This is a clear weakness of this test. On the other hand, if we do find evidence against the null hypothesis, we are confident that at least one racial group of the troopers are racially prejudiced.\(^{13}\)

### 2.2 Discussion of Two Key Assumptions

We made two key assumptions in the description of the model that play important roles in our empirical methodology.

**Assumption on the Pool of Motorists Faced by Troopers of Different Races.** In the model, we assume that the fraction of race-\( r_m \) motorists carrying contrabands is \( \pi^{r_m} \in (0, 1) \), which does not depend on the race of the troopers. That is, we assumed that the pools of motorists faced by troopers of different races are the same. This assumption may not be empirically valid if white and minority troopers are systematically assigned to patrol in different locations or time of day (indeed, our raw data indicated this is the case).

We now propose an empirical method that can resolve this problem even when the raw data does not satisfy this condition. For illustration purpose, suppose that there are two trooper stations

\(^{13}\)If we were to willing to assume that the signal distributions \( f^r_{g_{m}} \) and \( f^r_{n_{m}} \) do not depend on \( r_m \), then one can derive more powerful tests for racial prejudice. But we think such restrictions are too strong to be realistic in empirical applications.
1 and 2, each with 100 officers. Suppose that in trooper station 1, 80 officers are whites and 20 are minorities; while in station 2, 60 officers are whites and 40 are minorities. Thus on average 70 percent of the troopers are whites and 30 percent of the troopers are minorities. If the motorists that drive through the patrol areas of stations 1 and 2 differ in their characteristics, then the assumption that on average white and minority troopers face the same pool of motorists may be invalid. To deal with this issue, we create artificial samples in the following way. We keep all the minority officers (20 of them) in station 1, but randomly (with replacement) select 47 out of the 80 white officers. Similarly, we keep all the white officers (60 of them) in station 2, but randomly (with replacement) select 26 out of the 40 minority officers. Thus we create an artificial sample of 107 white officers and 46 minority officers. Among the 153 officers in the artificial sample, (roughly) 70 percent of them are whites and 30 percent are minorities, and they are equally likely to be assigned to stations 1 and 2. We can calculate the various search rates and average search success rates in this artificial sample. To alleviate the sampling error, we use independent resampling to create a list of such artificial data sets.

This subsampling method can effectively make sure that, when we calculate the search rates and average search success rates, the white and minority officers in the sample are assigned to different trooper stations with equal probability. Thus on average, white and minority officers are facing the same pool of motorists. However, we still need to assume that troopers do not have station specific priors about $\pi^{r_m}$. That is, station 1 troopers’ belief about the proportion of guilty race-$r_m$ drivers in its patrol area is the same as station 2 troopers’ belief.

**Assumption on the Signal Distributions.** In the model, we allow the signal distributions $f_{g}^{r_m}$ and $f_{n}^{r_m}$ to be specific to the racial group of the drivers. This flexibility is important if we intend to use our model as a basis for empirical test. To the extent that black and white drivers may exhibit different characteristics in their encounters with highway troopers, imposing $f_{g}^{M}, f_{n}^{M}$ to be equal to $f_{g}^{W}$ and $f_{n}^{W}$ respectively would be very strong and may be empirically implausible. Also note that, since $\theta$ is most likely not observable by researchers, we do not want to impose parametric distributional assumptions.

Despite this flexibility, our formulation does assume that the signals of race $r_m$ motorists are drawn from the same distributions independent of police officers’ race. For example, we do not allow that the signal about a black driver observed by a black officer will be drawn from the same
distribution as that by a white officer. This would be a suspicious assumption, for example, if the stops and searches occur in local streets. As argued in Donohue and Levitt (2001), a black community may be more willing to cooperate with a black officer, thus black officers may obtain more information about a black motorist on the streets. However, we maintain that this is a realistic assumption in highway searches. When stopping a black driver on highways, a trooper typically does not have any other citizens to rely on for additional information. Thus any informational advantage that black troopers have about black motorists may not be applicable on the highways. Thus as long as white and black troopers observe the same list of characteristics and summarize them in the same way, this is a valid assumption.

3 Empirical Results

3.1 Data Description

We now apply the tests described above to data from the Florida State Highway Patrol. The Florida data is composed of two parts. The first is the traffic data set that consists of all the stops and searches conducted on all Florida highways from January 2000 to November 2001. For each of the stops in the data set, it includes (among other things) the date, exact time, county, driver’s race, gender, ethnicity, age, reason for stop, whether a search was conducted, rationale for search, type of contraband seized, and the ID number of the trooper who conducted the stop and/or search. This part of the data is similar to those used in earlier studies of racial profiling (e.g. KPT 2001 and Gross and Barnes 2002). The unique feature of our data set is the second part, which is the personnel data that contains information on each of the troopers that conducted the stops and searches in the traffic data set, including their ID number, date of birth, date of hiring, race, gender, rank, and base troop station. We merge the traffic data and the personnel data by the unique trooper ID number that appears in both data sets. The merged data set thus provides information about the demographics of the trooper that made each stop and search. After eliminating cases in which there was missing information on the demographics of the trooper that conducted the stop, we end up with 906,339 stops and 8,976 searches conducted by a total of 1,312 troopers.

Even though KPT have data on the stops, they did not use them in their analysis. Gross and Barnes (2002) provided some basic statistics about the stop data.
Florida State Highway Patrol troopers are assigned to one of the ten trooper stations. Except for trooper station K, which is in charge of the Florida Turnpike, all other stations cover fixed counties. Figure 1 shows the coverage area of different troop stations.

3.2 Empirical Findings

3.2.1 Descriptive Statistics

Table 1 summarizes the means of the variables related to the motorists in our sample. Of the 906,339 stops we observe, 66.5 percent were carried out against white motorists, 17.3 percent against Hispanic motorists, and 16.2 percent against blacks. In all race categories, male motorists account for at least 67 percent of the stopped motorists for all race categories. Among all the motorists that are stopped, 48 percent were in 16-30 age group, 33.6 percent in 31-45 age group and 18.3 percent were older than 46. Close to 90 percent of stopped motorists have in-state license plates; and close to 70 percent of the stops were conducted in the day time (defined to be 6am to 6pm).

Of the 8,976 searches we observe, 54.6 percent were performed on white motorists, 23.4 percent on Hispanic motorists, and 22.1 percent against blacks. In all race categories, more than 80 percent of searchers were performed on male motorists; and overall, 84.8 percent of searches were against male drivers. Among the motorists that were searched, 58.4 percent were in 16-30 age group, 31.7 percent in 31-45 age group and only 9.9 percent. Vehicles with in-state plates accounts for 85.7 percent of the searches; and 52.5 percent of the searches were conducted at night (recall 30.3 percent of the stops were at night). 79.2 percent of searches were not successful (they yielded nothing). Drugs were the most common contraband seized in successful searches (15.1 percent of total searches), followed by alcohol/tobacco (2.1 percent) and drug paraphernalia (1.5 percent).

Table 2 summarizes the means of variables related to the troopers in our sample. The first column shows that in our data, Blacks, Hispanic and whites account for 14, 11, and 75 percent of the troopers respectively. The second column shows that white troopers conducted 73 percent of all stops and 86 percent of all searches. The corresponding numbers for black troopers are 16 and 4.6 percent; for Hispanic troopers are 11.4 and 9.5 percent.

We also eliminated cases where the race of the motorist and trooper was not either white, black, or Hispanic, since there are not enough observations for the other racial groups to consider them.
Figure 1: Troop Station Coverage Map
<table>
<thead>
<tr>
<th>Motorists’ Characteristics</th>
<th>Stops</th>
<th></th>
<th>Searches</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All by Sex</td>
<td></td>
<td>All by Sex</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stops</td>
<td>Female</td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>.162</td>
<td>.327</td>
<td>.673</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>.173</td>
<td>.225</td>
<td>.775</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>.665</td>
<td>.319</td>
<td>.681</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>.304</td>
<td>1.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>.697</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Age:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-30</td>
<td>.481</td>
<td></td>
<td>.584</td>
<td></td>
</tr>
<tr>
<td>31-45</td>
<td>.336</td>
<td></td>
<td>.317</td>
<td></td>
</tr>
<tr>
<td>46+</td>
<td>.183</td>
<td></td>
<td>.099</td>
<td></td>
</tr>
<tr>
<td>License Plate:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-state</td>
<td>.899</td>
<td></td>
<td>.857</td>
<td></td>
</tr>
<tr>
<td>Out-of-state</td>
<td>.101</td>
<td></td>
<td>.143</td>
<td></td>
</tr>
<tr>
<td>Time:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day (6am-6pm)</td>
<td>.697</td>
<td></td>
<td>.475</td>
<td></td>
</tr>
<tr>
<td>Night</td>
<td>.303</td>
<td></td>
<td>.525</td>
<td></td>
</tr>
<tr>
<td>Contraband Seized:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td>.792</td>
<td></td>
</tr>
<tr>
<td>Drugs</td>
<td></td>
<td></td>
<td>.151</td>
<td></td>
</tr>
<tr>
<td>Paraphernalia</td>
<td></td>
<td></td>
<td>.015</td>
<td></td>
</tr>
<tr>
<td>Currency</td>
<td></td>
<td></td>
<td>.003</td>
<td></td>
</tr>
<tr>
<td>Vehicles</td>
<td></td>
<td></td>
<td>.010</td>
<td></td>
</tr>
<tr>
<td>Alcohol/Tobacco</td>
<td></td>
<td></td>
<td>.021</td>
<td></td>
</tr>
<tr>
<td>Weapons</td>
<td></td>
<td></td>
<td>.006</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td>.003</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Means of Variables Related to Motorists
<table>
<thead>
<tr>
<th>Troopers’ Characteristics</th>
<th>Troopers All Troopers</th>
<th>Stops All By Sex</th>
<th>Searches All By Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>.141</td>
<td>.160 .119 .881</td>
<td>.046 .065 .935</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.105</td>
<td>.114 .070 .930</td>
<td>.095 .021 .979</td>
</tr>
<tr>
<td>White</td>
<td>.754</td>
<td>.726 .092 .908</td>
<td>.859 .076 .924</td>
</tr>
<tr>
<td>Female</td>
<td>.111</td>
<td>.093 1.00 0.00</td>
<td>.069 1.00 0.00</td>
</tr>
<tr>
<td>Male</td>
<td>.889</td>
<td>.907 0.00 1.00</td>
<td>.931 0.00 1.00</td>
</tr>
<tr>
<td>Ranks:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Captain</td>
<td>.008</td>
<td>.002</td>
<td>.0021</td>
</tr>
<tr>
<td>Lt.</td>
<td>.042</td>
<td>.013</td>
<td>.0067</td>
</tr>
<tr>
<td>Sgt.</td>
<td>.136</td>
<td>.062</td>
<td>.0530</td>
</tr>
<tr>
<td>Corp.</td>
<td>.151</td>
<td>.112</td>
<td>.0709</td>
</tr>
<tr>
<td>LEO</td>
<td>.662</td>
<td>.810</td>
<td>.866</td>
</tr>
</tbody>
</table>

Table 2: Means of Variables Related to Troopers
### Table 3: Distribution of Characteristics of Stopped Motorists, by Trooper Race in the Raw Data

<table>
<thead>
<tr>
<th>Motorist’s Race</th>
<th>Motorist’s Characteristics</th>
<th>White Troopers</th>
<th>Black Troopers</th>
<th>Hispanic Troopers</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>White Male</td>
<td>Night stops</td>
<td>.679</td>
<td>.684</td>
<td>.701</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Age: 16-30</td>
<td>.471</td>
<td>.460</td>
<td>.445</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Age: 31-45</td>
<td>.325</td>
<td>.341</td>
<td>.349</td>
<td>0.02</td>
</tr>
<tr>
<td>Black Male</td>
<td>Night stops</td>
<td>.671</td>
<td>.667</td>
<td>.686</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Age: 16-30</td>
<td>.514</td>
<td>.514</td>
<td>.507</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Age: 31-45</td>
<td>.340</td>
<td>.344</td>
<td>.356</td>
<td>0.03</td>
</tr>
<tr>
<td>Hispanic Male</td>
<td>Night stops</td>
<td>.783</td>
<td>.774</td>
<td>.761</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Age: 16-30</td>
<td>.516</td>
<td>.497</td>
<td>.494</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Age: 31-45</td>
<td>.350</td>
<td>.363</td>
<td>.355</td>
<td>0.01</td>
</tr>
</tbody>
</table>

3.2.2 Examine the Assumption that Troopers Face the Same Population of Motorists

Before we conduct our tests of monolithic behavior and racial prejudice, we first examine whether a crucial assumption of our test, that all troopers face the same population of motorists, are satisfied in the raw data (before resampling). This assumption, of course, is not directly testable, because both $\pi^m$ and $\theta$ are unobservable. The best we can do is to examine the distribution of observable motorist characteristics faced by troopers of different races. Table 3 shows the proportions of stopped motorists with given characteristics faced by troopers of different races. The characteristics of motorists reported in the table include race, gender, age, and time of the stops. For each row, we also report in the last column the $p$-values for Pearson $\chi^2$ tests of the null hypothesis that the proportions of stopped motorists with the characteristics specific to that row are the same for all three race groups of the troopers. As one can see, the hypothesis that troopers of different races face the same population of motorists can be statistically rejected in the raw data, even though the differences are numerically quite small. One may suspect that the reason
that troopers of different races are stopping motorists with different characteristics is that Black, Hispanic and White troopers are assigned to different troops. For example, Hispanic troopers are likely to have an over-representation in Troop E (covering Miami in Dade County) than Troop A and E (covering counties in Florida Panhandle). Indeed, Table 4 shows that the allocations of troopers of different races to different troops and time of the assignment do not seem random in the raw data. For this reason, we think it is important to conduct the resampling methods we described in Subsection 2.2. As expected, in the artificial data we created with the resampling method we can not statistically reject the hypothesis that the distribution of the observable characteristics of the stopped motorists faced by troopers are the same. We report our test results below with the artificial samples from resampling.

### 3.2.3 Test of Monolithic Behavior

We now implement our test for the hypothesis that troopers of different races exhibit monolithic behavior. Table 5 is the main table. In Panel A, we show the search rate given stop for motorist/trooper race pairs. For example, the first row shows that white, black and Hispanic troopers respectively search 0.96, 0.27 and 0.76 percent of the white motorists they stop. The last column shows the p-value from the Pearson’s $\chi^2$ test under the null hypothesis that troopers of all races search white motorists with equal probability. Specifically, Pearson’s $\chi^2$ test statistic under the null hypothesis all troopers with race in $R$ search race-$r_m$ motorists with equal probability is given by

$$\sum_{r_p \in R} \frac{(\gamma(r_m; r_p) - \gamma(r_m))^2}{\gamma(r_m; r_p)} \sim \chi^2(R - 1),$$

where $\gamma(r_m; r_p)$ is the estimated search probability of race-$r_p$ officers against race-$r_m$ motorists, $\gamma(r_m)$ is the estimated search probability against race-$r_m$ motorists unconditional on the race of the officer, and $R$ is the cardinality of the set of troopers’ race categories, $R$.

---

16One may argue that all of the stops occurred on Florida highways, and the drug flow in Florida tends to go from Miami (a city in the southern tip of Florida) to cities in the northeastern United States; that is, drug couriers are moving throughout Florida (except for possibly the panhandle). Thus troopers stationed in different areas are likely to face similar population of drivers; and the differences in the stopped motorists’ characteristics reflect the differences in stop behavior of the troopers of different races, rather than the differences in the driver population. It is plausible, but in this paper we take the stopped motorists population as given.
<table>
<thead>
<tr>
<th>Troop</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.930</td>
<td>.054</td>
<td>.016</td>
</tr>
<tr>
<td>B</td>
<td>.889</td>
<td>.081</td>
<td>.030</td>
</tr>
<tr>
<td>C</td>
<td>.816</td>
<td>.116</td>
<td>.068</td>
</tr>
<tr>
<td>D</td>
<td>.793</td>
<td>.117</td>
<td>.090</td>
</tr>
<tr>
<td>E</td>
<td>.412</td>
<td>.236</td>
<td>.352</td>
</tr>
<tr>
<td>F</td>
<td>.880</td>
<td>.056</td>
<td>.063</td>
</tr>
<tr>
<td>G</td>
<td>.833</td>
<td>.135</td>
<td>.032</td>
</tr>
<tr>
<td>H</td>
<td>.886</td>
<td>.114</td>
<td>0.00</td>
</tr>
<tr>
<td>K</td>
<td>.698</td>
<td>.147</td>
<td>.155</td>
</tr>
<tr>
<td>L</td>
<td>.603</td>
<td>.298</td>
<td>.099</td>
</tr>
<tr>
<td>% Night Stops</td>
<td>.300</td>
<td>.283</td>
<td>.349</td>
</tr>
</tbody>
</table>

Table 4: Proportion of Troopers with Different Races by Troop and Time Assignment in the Raw Data
Panel B presents the average search success rate for given motorist/trooper race pairs. The last column in each row shows the $p$-value from the Pearson’s $\chi^2$ test under the null hypothesis that troopers of all races have the same average search success rate against motorists of race in that specific row. Again the Pearson’s $\chi^2$ test statistics under the null hypothesis that all troopers with race in $R$ have the same average search success rate against race-$r_m$ motorists is given by

$$\sum_{r_p \in R} \frac{\left( \widehat{S}(r_m; r_p) - \widehat{S}(r_m) \right)^2}{\widehat{S}(r_m; r_p)} \sim \chi^2 (R - 1),$$

where $\widehat{S}(r_m; r_p)$ is the estimated average search success rate of race-$r_p$ officers against race-$r_m$ motorists, $\widehat{S}(r_m)$ is the estimated average search success rate against race-$r_m$ motorists unconditional on the race of the officer.

As we argued in subsection 2.1.1, under the null hypothesis that troopers exhibit monolithic behavior, $\gamma(r_m; r_p) = \gamma(r_m)$ and $\widehat{S}(r_m; r_p) = \widehat{S}(r_m)$ for all $r_p$, thus Pearson’s $\chi^2$ test statistic should be small under the null. The $p$-values in Table 5 show that we can soundly reject the null hypothesis.

Moreover, the estimated search rates in Panel A have the ranking that, for any race of motorists $r_m \in \{W, B, H\}$,

$$\gamma(\overline{r_m}; W) > \gamma(\overline{r_m}; H) > \gamma(\overline{r_m}; B).$$

This of course implies that the search criterion used by troopers against race-$r_m$ motorists have the ranking

$$\theta^*(\overline{r_m}; W) < \theta^*(\overline{r_m}; H) < \theta^*(\overline{r_m}; B).$$

In the light of Proposition 1, this implies a ranking over the search costs: for any $r_m$,

$$t(\overline{r_m}; W) < t(\overline{r_m}; H) < t(\overline{r_m}; B);$$

that is, white troopers seem to have smaller costs of searching motorists of any race, followed by Hispanic troopers and black troopers has the highest search cost.

As we mentioned at the end of subsection 2.1.1, if the search criteria used by troopers against race-$r_m$ motorists satisfy $\theta^*(r_m; W) < \theta^*(r_m; H) < \theta^*(r_m; B)$, then our model is refuted unless the average search success rates satisfy the same ranking, i.e.,

$$\widehat{S}(r_m; W) < \widehat{S}(r_m; H) < \widehat{S}(r_m; B).$$

(5)
Because we have a large sample, we can use a simple $Z$-statistics to test the null hypothesis $S(r_m; W) = S(r_m; H)$ against the one-sided alternative hypothesis $S(r_m; W) < S(r_m; H)$:

$$Z = \frac{S(r_m; W) - S(r_m; H)}{\sqrt{\frac{\text{SVar}_W}{n_W} + \frac{\text{SVar}_H}{n_H}}} \sim N(0, 1),$$  \hspace{1cm} (6)

where $n_W$ and $n_H$ are the number of searches conducted by white and Hispanic officers respectively, and $\text{SVar}_W$ and $\text{SVar}_H$ are the sample variance of the search success dummy variables in the sample of searches conducted by white and Hispanic officers. The null will be rejected in favor of the alternative at significance level $\alpha$ if $Z \leq -z_\alpha$ where $\Phi(z_\alpha) = 1 - \alpha$. For example when we consider white motorists, we obtain a $Z$-statistic of $-3.24$ for white and Hispanic officers, thus we are able to reject the null in favor of the alternative $S(W; W) < S(W; H)$ at a significance level essentially equal to 0. Likewise, we can reject the null $S(W; H) = S(W; B)$ in favor of the alternative $S(W; H) < S(W; B)$ at significance level close to 0 (with a $Z$-statistics of $-2.54$). In fact the rankings over the average search success rates in (5) hold for all races of the motorists. Thus our model is not statistically refuted by the data.

### 3.2.4 Test for Racial Prejudice

We have so far provided strong evidence that the troopers do not exhibit monolithic search criteria when deciding whether to search motorists of a given race, and we have shown that the data suggests that, for any given race, white troopers tend to use the weakest standard, followed in order by Hispanic and black troopers. As we argued in subsection 2.1.2, under the null hypothesis that none of racial groups of troopers are racially prejudiced, then the ranking over the search rates $\gamma(r_m; W), \gamma(r_m; B)$ and $\gamma(r_m; H)$ should be independent of $r_m$. This clearly holds when we look at the estimated mean search rates in Table 5: for all $r_m \in \{W, B, H\}$,

$$\gamma(r_m; W) > \gamma(r_m; H) > \gamma(r_m; B).$$

One can formally test the null hypothesis $\gamma(r_m; W) = \gamma(r_m; H) = \gamma(r_m; B)$ against the alternative $\gamma(r_m; W) > \gamma(r_m; H) > \gamma(r_m; B)$ using a $Z$-test similar to (6). For example, the null $\gamma(W; W) = \gamma(W; H)$ is rejected against the alternative $\gamma(W; W) > \gamma(W; H)$ at significance level close to 0 ($Z$-statistics is $27.4$). Similarly, for the test of null $\gamma(W; H) = \gamma(W; B)$ against $\gamma(W; H) > \gamma(W; B)$, we obtain a $Z$-statistic of 65, thus rejecting the null in favor of the alternative. To summarize, the
<table>
<thead>
<tr>
<th>RACE</th>
<th>Motorist’s Race</th>
<th>Trooper Race</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
<td>Black</td>
<td>Hispanic</td>
</tr>
<tr>
<td>Panel A: Search Rate Given Stop (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0.96 (6.68E-4)</td>
<td>0.27 (7.73E-4)</td>
<td>0.76 (9.26E-4)</td>
</tr>
<tr>
<td>Black</td>
<td>1.74 (1.30E-3)</td>
<td>0.35 (1.42E-3)</td>
<td>1.21 (2.28E-3)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.61 (1.46E-3)</td>
<td>0.28 (0.76E-3)</td>
<td>0.99 (3.03E-3)</td>
</tr>
<tr>
<td>Panel B: Average Search Success Rate (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>24.3 (9.43E-3)</td>
<td>39.4 (5.57E-2)</td>
<td>26.0 (2.28E-2)</td>
</tr>
<tr>
<td>Black</td>
<td>19.9 (1.26E-2)</td>
<td>26.0 (5.32E-2)</td>
<td>20.8 (2.67E-2)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>8.5 (9.78E-3)</td>
<td>21.0 (4.55E-2)</td>
<td>14.3 (6.63E-2)</td>
</tr>
</tbody>
</table>

Table 5: Search Rates and Average Search Success Rates for by Races of Motorists and Troopers in the Artificial Data Sets.

Note: Standard errors of the means are shown in parentheses.
evidence supports the hypothesis that $\gamma(r_m; W) > \gamma(r_m; H) > \gamma(r_m; B)$ for all $r_m$. Thus we can not reject the null hypothesis that troopers are not racially prejudiced. Of course, we would like to emphasize caution in interpreting our finding: while we do not find definitive evidence of racial prejudice, it is still possible that some or all groups of troopers are racially prejudiced. If the latter is true, then we have committed a type-II error as a result of the weak test. Analogously our data supports the hypothesis that for all $r_m, S(r_m; W) < S(r_m; H) < S(r_m; B)$. Thus our model is not statistically refuted.

### 3.2.5 Replicating KPT Test

Finally, we would like to contrast our findings with those from KPT’s test. Recall that KPT’s test relies on the prediction from their model that, under the null hypothesis of no racial prejudice, the average search success rates should be independent of the motorists’ race. Table 6 shows the search rate and average search success rate for different races of the motorists in the raw data;\(^{17}\) and Table 7 shows the $p$-value from Pearson’s $\chi^2$ test on the hypothesis that the average search rate and search rate are equal across various race groupings. Their test immediately implies that the troopers show racial prejudice against black and Hispanic motorists, especially the Hispanics. However, as we argued, this conclusion is only valid if their model of motorist and trooper behavior is true.

---

\(^{17}\)While KPT’s model does make predictions of the average search rate, their test does not utilize such information. In fact, they do not have the search rate information in their application to the Maryland data since their data consist of searches only. We include the search rate information in the tables as simple descriptive statistics of the data.
4 Conclusion

The validity of any test to detect and determine the motivations behind racial profiling relies heavily on whether the theoretical model used to justify the test generates realistic predictions about the motorist and trooper behavior. We contribute to this important literature by presenting a simple but realistic model of trooper behavior to design empirical tests that address the following two questions. Are police monolithic in their search behavior? Is racial profiling in motor vehicle searches motivated by the troopers’ desire for effective policing (statistical discrimination) or by their racial prejudice (racism)? Our tests require data sets with race information about both the motorists and troopers. When applied to vehicle stop and search data from Florida, our tests can soundly reject the null hypothesis that troopers of different races are monolithic in their search behavior; but fail to reject the null hypothesis that no racial groups of troopers are racially prejudiced.

A Appendix: A Model with Endogenous Drug Carrying Decisions.

In Section 2 we assumed that the proportion of motorists in race group \( r_m \) is exogenously given at \( \pi^m \in (0, 1) \). For the purpose of testing for monolithic behavior and racial prejudice, this partial equilibrium approach suffices. However, for other purposes such as public policy considerations such as reducing crimes and “war on drugs” in particular, one may want to know how any changes
in trooper behavior may affect the motorists’ drug carrying decisions.\textsuperscript{18} One needs an equilibrium model to address such questions. In this appendix, we propose a simple model. We show that closing our partial equilibrium model in Section 2 is easy; moreover, such an equilibrium model has nice equilibrium uniqueness properties under reasonable conditions. This is in contrast to the labor market statistical discrimination models where multiple equilibria naturally arise and are the driving force for statistical discrimination (see, among others, Coate and Loury 1993).

Consider a single motorist race group $r_m$; and two trooper racial groups, $r_p$ and $r'_p$.\textsuperscript{19} Suppose that in the trooper population, a fraction $\alpha$ is of race $r_p$ and the remainder fraction $1 - \alpha$ is of race $r'_p$. Suppose that Nature draws for each driver a utility cost of carrying contrabands $v \in \mathbb{R}_+$ from CDF $G$ with a continuous density. The utility cost $v$ represents feelings of fear experienced by a driver from the act of carrying contraband. If a driver carries contraband and is not caught, he/she derives a benefit of $b > 0$. If a guilty driver is searched and thus arrested, he/she experiences an additional cost (over and above $v$) of $c_g$. If a driver does not carry contraband, he/she does not incur the utility cost of $v$. But the inconvenience experienced by an innocent driver when he/she is searched is denoted by $c_n$. Naturally we assume that $c_g > c_n$. We assume that a driver’s realization of $v$ is his or her private information; $b, c_g$ and $c_n$ are constants known to all drivers and police officers. Each driver decides whether to carry contraband.

As before, we normalize the benefit of each arrest to the police officer to be one; and for notational simplicity, the cost of search for a race-$r_p$ trooper is written as $t_p \in (0, 1)$ and that for a race-$r'_p$ trooper is $t'_p \in (0, 1)$. As in Section 2, troopers observe noisy but informative signals regarding whether or not a driver is carrying contraband: if a driver is guilty, the signal $\theta \in [0, 1]$ is drawn from PDF $f_g(\cdot)$; if the driver is not guilty, then $\theta$ is drawn from PDF $f_n(\cdot)$. As before $f_g/f_n$ is strictly increasing in $\theta$. Let $F_g$ and $F_n$ denote the corresponding CDFs of $f_g$ and $f_n$. We assume that a trooper wants to maximize the total number of convictions minus the cost of searching cars.

We first suppose that a proportion $\pi$ of drivers choose to carry contraband and analyze the optimal search behavior of the troopers. Let $\Pr(G|\theta)$ denote the posterior probability that a driver

\textsuperscript{18}See, e.g., Persico (2002) for an analysis on how racial blind search policy may affect the total crimes committed by motorists.

\textsuperscript{19}Because we are only considering one race group of motorists, we will omit $r_m$ from the subsequent notation. Having more than one racial groups of motorists will not change any of the results below.
with signal $\theta$ is guilty of carrying illicit drugs, which is given by

$$\Pr(G|\theta, \pi) = \frac{\pi f_g(\theta)}{\pi f_g(\theta) + (1 - \pi) f_n(\theta)}.$$  

A race-$r_p$ trooper will decide to search a driver with signal $\theta$ if and only if

$$\Pr(G|\theta, \pi) - t_p \geq 0;$$

which, from the MLRP, is equivalent to $\theta \geq \theta^*_p(\pi)$ where $\theta^*_p(\pi) \in [0, 1]$ is the unique solution to

$$\Pr(G|\theta, \pi) = t_p.$$

Obviously $\theta^*_p(\pi)$ is strictly decreasing in $\pi$. Similarly, race-$r'_p$ troopers will search a motorist if and only if the motorist’s signal $\theta$ exceeds $\theta^*_{p'}(\pi)$ where $\theta^*_{p'}(\pi)$ solves

$$\Pr(G|\theta, \pi) = t_{p'}.$$

Now suppose that race-$r_p$ and race-$r'_p$ troopers use search criteria of $\theta^*_p$ and $\theta^*_{p'}$ respectively. The expected payoff of a driver with utility cost $v$ from carrying contraband is given by

$$\text{Term 1} = \alpha F_g(\theta^*_p) + (1 - \alpha) F_g(\theta^*_{p'}) b - \left\{ \alpha \left[ 1 - F_g(\theta^*_p) \right] + (1 - \alpha) \left[ 1 - F_g(\theta^*_{p'}) \right] \right\} c_g - v$$

where Term 1 is the probability of not being caught multiplied by the benefit from drugs if the motorist is not caught. Note that a fraction $\alpha$ of the troopers are of race-$r_p$ who use a search criterion of $\theta^*_p$, and $1 - \alpha$ of the troopers use $\theta^*_{p'}$, thus the expected probability of not being caught is $\alpha F_g(\theta^*_p) + (1 - \alpha) F_g(\theta^*_{p'})$. Term 2 is the expected probability of being caught multiplied by the cost of being caught with illicit drugs. Of course, the driver suffers a disutility $v$ whenever he or she carries drugs.

The expected payoff of a driver with utility cost $v$ from not carrying contraband is simply the inconvenience cost of being searched by troopers in mistake:

$$- \left\{ \alpha \left[ 1 - F_n(\theta^*_p) \right] + (1 - \alpha) \left[ 1 - F_n(\theta^*_{p'}) \right] \right\} c_n.$$  

Thus a driver with utility cost realization $v$ will decide to carry illicit drugs if and only if $v \leq v^*(\theta^*_p, \theta^*_{p'})$ where

$$v^*(\theta^*_p, \theta^*_{p'}) = \left[ \alpha F_g(\theta^*_p) + (1 - \alpha) F_g(\theta^*_{p'}) \right] b - \left\{ \alpha \left[ 1 - F_g(\theta^*_p) \right] + (1 - \alpha) \left[ 1 - F_g(\theta^*_{p'}) \right] \right\} c_g$$

$$+ \left\{ \alpha \left[ 1 - F_n(\theta^*_p) \right] + (1 - \alpha) \left[ 1 - F_n(\theta^*_{p'}) \right] \right\} c_n.$$  

(A1)
Thus if the troopers follow search criteria $\theta_p^*$ and $\theta_p^0$ respectively, the proportion of drivers who will choose to carry contraband is given by $G\left(v^*\left(\theta_p^*,\theta_p^0\right)\right)$.

An equilibrium of the model is a triple $(\pi, \theta_p^*, \theta_p^0)$ such that:

\[
\Pr\left(G|\theta_p^*, \pi\right) = t_p \quad \text{(A2)}
\]
\[
\Pr\left(G|\theta_p^0, \pi\right) = t_p' \quad \text{(A3)}
\]
\[
G\left(v^*\left(\theta_p^*, \theta_p^0\right)\right) = \pi \quad \text{(A4)}
\]

The existence of equilibrium follows directly from Brouwer’s Fixed Point Theorem. Now we show that in fact for any CDF $G$ with non-negative support, the equilibrium is unique. Suppose that there are two equilibria in which the proportion of guilty motorists are $\pi$ and $\tilde{\pi}$ with $\pi > \tilde{\pi}$. From equilibrium conditions (A2) and (A3) we know that $\theta_p^* < \tilde{\theta}_p^*$ and $\theta_p^0 < \tilde{\theta}_p^0$. Observe from (A1) that $v^*(0,0) = c_n - c_g < 0$ and

\[
\frac{\partial v^*\left(\theta_p^*, \theta_p^0\right)}{\partial \theta_p^*} = \alpha c_n f_n\left(\theta_p^*\right) \left[\frac{f_g\left(\theta_p^*\right) b + c_g}{f_n\left(\theta_p^*\right) c_n} - 1\right],
\]
\[
\frac{\partial v^*\left(\theta_p^*, \theta_p^0\right)}{\partial \theta_p^0} = \alpha c_n f_n\left(\theta_p^*\right) \left[\frac{f_g\left(\theta_p^0\right) b + c_g}{f_n\left(\theta_p^0\right) c_n} - 1\right].
\]

By MLRP, $v^*\left(\theta_p^*, \theta_p^0\right)$ is decreasing in both $\theta_p^*$ and $\theta_p^0$ when they are small. Therefore when $(\theta_p^*, \theta_p^0)$ are small, $G\left(v^*\left(\theta_p^*, \theta_p^0\right)\right) = G(0) = 0$; and when $(\theta_p^*, \theta_p^0)$ is larger than a threshold, $v^*\left(\theta_p^*, \theta_p^0\right)$ are strictly increasing in them. Therefore we have $0 < v^*\left(\theta_p^*, \theta_p^0\right) < v^*\left(\tilde{\theta}_p^*, \tilde{\theta}_p^0\right)$. But then it implies that $\tilde{\pi} > \pi$, a contradiction.

References


