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

Shamena Anwar† Hanming Fang‡

December 15, 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 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 none of the racial groups of troopers are racially prejudiced.

Keywords: Statistical Discrimination, Racial Prejudice, Racial Profiling.

JEL Classification Number: J7.

∗We are grateful to Joseph Altonji, Kate Antonovics, Richard Freeman, Nicola Persico and participants at the 2004 MacArthur Foundation Meeting on Social Interactions and Economic Outcomes in Stanford for helpful discussions. We especially thank the Co-Editor and three anonymous referees for many detailed and constructive comments on an earlier draft. We 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 06510-8264. 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 police practice of using a motorist’s race as one of the criteria in their motor vehicle search decisions. Racial profiling originated with the attempt to interdict the flow of drugs from Miami up Interstate 95 to the cities of the 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 have been 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 using race as a criterion in stop and search decisions. But many citizens, especially minorities, are skeptical of this claim: motor vehicle search decisions, by their very nature, are made in the midst of face-to-face interactions, and thus it is simply hard to imagine that troopers can block the race and ethnicity information that a motorist presents. 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 contraband, then the aggregate rate of stops and searches would be higher for black drivers even when race was 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 will not observe the true proportion of drivers who carry contraband. Ethnographic studies such as 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. Because we have no way of controlling for all of the legitimate factors that might cause minority drivers to be searched with higher probability than white motorists, it becomes very difficult to determine the true motivation behind racial profiling with the available data.

**Benchmarking vs. Outcome Tests.**  
There are two main statistical approaches that aim to test whether troopers impose disparate treatment on motorists of different races. The first approach is called “benchmarking test,” which typically compares the shares of racial or ethnic minorities in the population to their shares in the sample of motorists selected for discretionary stops and searches by police. There are at least two main problems with this approach. The first problem is called the denominator problem, which refers to the question of what should be the right benchmark to compare the stop and search rates. It ideally should be the racial or ethnic composition of drivers on the road, but such information is typically unavailable except in few instances of time and location. The second problem is the omitted-variable problem. If there exist certain characteristics whose distributions are correlated with motorists’ race or ethnicity and if such characteristics may be observed by police but not available to researchers, then benchmarking tests will not be completely informative of whether motorists’ race affected the search decision. Moreover, benchmarking tests are not informative about the underlying motivation even when they find evidence of discrimination.

The second approach is called the “outcome test,” whose idea originated in Becker (1957). In the context of motor vehicle searches, the outcome test is based on the following intuitive notion. If troopers are profiling minority motorists due to racial prejudice, they will search minorities even when the returns from searching them, i.e., the probabilities of successful searches against minorities, are smaller than those from searching 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

---

1. There are parallel and closely related approaches to test for disparate treatment in the literature on mortgage lending discrimination (see Ross and Yinger 2002 and Ross 2003 for comprehensive literature reviews). Paired-audit is a third frequently used method in the context of housing market, mortgage lending and car purchases (Ayres 2001).

2. A refined version of this test uses regression to estimate the probability of being searched as a function of race and other observable characteristics that may be related to propensity to commit crimes. Fridell (2004) provides a comprehensive review of different benchmarks in this approach.

3. Becker (1993a, 1993b) further elaborated on this idea. Also see Ayres (2001) for other applications of the outcome test idea.
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 (i.e., 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 will never be able to directly observe the search success rate of the marginal motorist. Instead we can only observe the average success rate of white and minority searches. Equality of marginal search success rates does not imply and is not implied by the equality of the average success rate – this has been called the infra-marginality problem in the literature. The infra-marginality problem is further magnified by the possible differences in the distributions of unobserved characteristics that may indicate guilt of carrying contraband for black and white motorists - another manifestation of the “omitted variable” problem. These problems severely limit the rigorous application of the outcome test idea, especially in situations where the decision or the outcome is dichotomous.4

Related Racial Profiling Literature. A seminal paper by Knowles, Persico and Todd (2001, KPT hereafter) provides the first solution to the infra-marginality problem of the outcome test.5 They develop a simple but elegant theoretical model about motorist and police behavior and show that in equilibrium the infra-marginality problem may not arise. 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 contraband. 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, and thus there is no difference between the marginal and the average search success rates. A nice feature of the KPT model is that it allows the motorists of different races to have different distributions of characteristics, as long as those characteristics are observable to the police (though they may not be observable to the researcher). Motorists with different characteristics may have different costs and

4See Ross and Yinger (1999 and 2002, Chapter 8) and Ayres (2002) for detailed discussions of the infra-marginality problem in the context of mortgage lending and police practices respectively. In fact, in the case of mortgage lending, Ross (2003) and Ross and Yinger (2002, Chapter 8) argue that the complete elimination of the omitted variable bias results in a test with no power.

5The ideas in our paper are inspired from reading KPT, from which we learned a great amount.
benefits from carrying contraband, but these differences only imply that in equilibrium troopers will search motorists with different characteristics at different rates, which in fact provides the necessary deterrence to ensure that all motorists will carry contraband with equal probabilities. Because the infra-marginality problem does not arise at all in the equilibrium of KPT’s simple model, they provide a solid theoretical basis for an empirical test based on the comparison of the average search success rates by the race of the motorists, a statistic typically available to researchers. 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 the I-95 in Maryland from January 1995 through January 1999, they find no evidence of racial prejudice against African-American motorists, but do find evidence of racial prejudice against Hispanics.

Dharmapala and Ross (2004) 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. If the probability of being a “dealer” is higher for minorities, then the average success rate against 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. Antonovics and Knight (2004) argued that KPT’s test may not be robust when its model is generalized to allow for trooper heterogeneity.

Motivations of This Paper. This paper proposes an alternative, albeit partial, solution to the “infra-marginality” and “omitted variables” problems. The motivation for our alternative test is our view that in some situations KPT model may not be the most suitable model to apply, as we will illustrate below with data from Florida Highway Patrol.

First, KPT’s model predicts 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. This is 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, thus avoiding the infra-marginality problem. This prediction, 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 police

---

6KPT recognized this issue in their footnote 16.
7However, Persico and Todd (2004) show that, if officers’ goal is to maximize search success rate, rather than total number of successes, KPT’s test can be generalized to allow for police heterogeneity.
behavior 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 ruling out the plausible scenario that a motorist’s actions when stopped is intimately related to whether or not he or she is carrying contraband. This is not just a minor quibble about details: once we allow the motorists’ actions when stopped to enter into the officers search decisions, the infra-marginality problem re-appears into the empirical analysis. An important feature of the model and test presented in our paper is that we allow for quite general differences, by motorists’ races, in the characteristics distributions that may be unobservable to researchers.

Second, KPT (and this field of research in general) assume that all troopers’ behavior is mono-

<table>
<thead>
<tr>
<th>Motorist’s Race</th>
<th>Trooper Race</th>
<th>All Troopers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
<td>Black</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>(6.68E-4)</td>
<td>(7.33E-4)</td>
</tr>
<tr>
<td>White</td>
<td>1.74</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>(1.30E-3)</td>
<td>(1.42E-3)</td>
</tr>
<tr>
<td>Black</td>
<td>1.61</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(1.46E-3)</td>
<td>(0.76E-3)</td>
</tr>
</tbody>
</table>

### Table 1: Search Rates and Average Search Success Rates against Motorists of Different Races.

**Note:** Standard errors of the means are shown in parentheses.
This assumption may not be valid. Most existing data sets on police behavior do not contain detailed information about the trooper characteristics for an examination of the monolithic trooper behavior assumption, thus it is assumed that all troopers, regardless of their race, have the same racial prejudice against minority motorists. 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) is not monolithic. Within the framework of KPT, an invalid monolithic trooper behavior assumption can lead to wrong conclusion on whether officers are racially prejudiced. 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 are not racially prejudiced. But this may seriously underestimate the harassment experienced by both white and minority motorists. This paper deviates from the KPT model and embrace the possibility that police behavior may vary by their races. It turns out that the variation of trooper behavior by their race provides the key additional information of our empirical test.

In the Florida data we analyze, the violation of the monolithic trooper behavior assumption is quite evident from Table 1. Panels A and B respectively show the search rate given stop and the average search success rate against motorists of different races. The column labeled “Trooper Race” are for given motorist/trooper race pairs; and the column labeled “All Troopers” aggregate all troopers. For example, the first row shows that, of the white motorists stopped by white, black and Hispanic troopers, respectively 0.96, 0.27 and 0.76 percent of them were searched; and in the aggregate 0.81 percent of all white drivers who are stopped ended up being searched. It can be seen that white troopers search motorists of all races more frequently than Hispanic troopers, followed by black troopers. Conversely, the average search success rate is highest, against

---

8A formal definition of monolithic behavior is given in Section 2.

9The Maryland data set KPT used has only very limited information about troopers (see KPT 2001 and Barnes and Gross 2002).

10We are grateful to an anonymous referee for clarifying this important point.

11Table 1 is our main table and we will discuss it more in Section 3.2. The formal test of monolithic behavior is presented in Section 3.2.3.

12The numbers in the column labeled “All Troopers” are calculated directly from the raw data; but the numbers in the columns labeled “Trooper Race” are calculated from artificial samples constructed from the raw data. See Section 2.4 for details about how we construct the artificial samples.
all motorists’ races, for black troopers followed in order by Hispanic and white troopers. KPT’s test only uses the information in Panel B of the column “All Troopers” and will conclude that police officers exhibit racial bias.\textsuperscript{13} But, when we admit the possibility that the unobservable characteristics among motorists of different races may differ (in a possibly arbitrary way), we will argue that even such strong disparities in search rates and average search success rate may not prove discrimination.

**Main Idea for our Alternative Test.** In this paper, we develop an alternative model of motorist and police behavior in which troopers are allowed to behave differently depending on their own race and the race of the motorists they interact with.\textsuperscript{14} 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. Instead, we allow for the average and marginal search success rates to differ; moreover, as we mentioned earlier we allow the distributions of unobservable characteristics to differ by motorists’ race in a general way. Our model follows the spirit of labor market statistical discrimination models (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 contraband, are more likely than innocent drivers to generate suspicious signals. A police officer incurs a cost of search $t(r_m; r_p)$ that depends on both his/her own race $r_p$ and the race of the motorist $r_m$. 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 exhibits non-monolithic behavior if the cost of searching motorists of a given race $r_m$ depend on the race of the trooper. Troopers are assumed to make their search decisions to maximize the number of successful searches (or arrests). The optimal decision of a race-$r_p$ police officer in deciding whether a race-$r_m$ motorist should be searched satisfies a threshold property: motorists should be searched if and only if their posterior probability of being guilty exceeds the search cost of race-$r_p$ officers 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 of any given race of motorists are independent of the race of the troopers conducting the search. Moreover, if none of the racial groups of troopers are racially prejudiced, then the

\textsuperscript{13} More discussion of KPT test on this data set is provided in Section 3.2.6.

\textsuperscript{14} We assume that race is the only characteristic of troopers that is likely to affect their search behavior. This is plausible assumption because we are examining if troopers search white and minority motorists differently, so the race of the trooper is the most likely characteristic to affect their search patterns.
ranking over the race of troopers of search rates and average search success rates against a given race of motorists should not depend on the race of the motorists. That is, if troopers of race \( r_p \) have a higher search rate (and lower average search success rate) against race-\( r_m \) motorists than troopers of race \( r'_p \), then race-\( r_p \) troopers should also have a higher search rate (and lower average search success 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 desirable feature of our model is that it has direct implications on the ranking of both the search rates and the average search success rates, thus it could potentially be refuted by the data we have available.

To summarize, the key idea of our empirical test is as follows. When there is no racial prejudice for all troopers, then the race of motorists should not affect the ranking of search rates and search success rates over officer races. It is also important to point out that our test can only detect relative racial prejudice of one officer race vis-à-vis another officer race and instead of the absolute racial prejudice of each officer race.\(^{15}\)

The implementation of our empirical tests relies on data sets that have race information on both the troopers and motorists.\(^{16}\) While such data has not been available for use in earlier empirical studies on racial profiling, we were able to obtain a data set 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 find 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.

**More Related Literature.** There is now a growing economics literature on the issue of empirical distinction between statistical discrimination and racial prejudice in motor vehicle searches. Hernandez-Murillo and Knowles (2004) use KPT framework and semi-parametric bounds to reject the official explanation that lower hit rates on minorities are due to higher rates of non-discretionary search using Missouri’s annual aggregate traffic-stop report for the year 2001. Dominitz and Knowles (2004) consider tests of racial prejudice when officers are assumed to minimize

\(^{15}\)More discussions on this issue is provided in Section 3.3 when we discuss the power of our test.

\(^{16}\)While our tests can in principle be implemented with only search data (by looking only at average search success rates), having data on all stops would be more desirable because we can then examine whether our model is refuted by the data besides providing supporting evidence from the search rates.
crime. Antonovics and Knight (2004) also use data with both motorist and officer information. As we do in our paper, they show that if officers of different races have the same search cost against motorists of a given race, then the search rate against these motorists should be independent of the officers’ race. They run a Probit regression using data from the Boston Police Department where the dependent variable is an indicator for whether a search took place for a given stop, and the explanatory variables include some observable characteristics of the driver and officer and a dummy variable indicating whether there is a racial mismatch between the officer and the driver. In their baseline regression, they find a positive coefficient on the “racial mismatch” variable, indicating that officers are more likely to conduct a search against motorists of races different from their own. They interpret this finding as evidence of racial prejudice. We argue in subsection 2.3.2 that their interpretation of the evidence may be misleading. It is also useful to point out that their data is from the Boston Police Department and consists mainly of stops and searches in local neighborhoods. There are two potential problems with such data. First, as Hernandez-Murillo and Knowles (2004) argued, many stops and searches conducted in local streets are in response to specific crime reports. In these situations, officers tend to have less discretion over who they search. Second, as argued by Donohue and Levitt (2001), for stops and searches conducted in local neighborhoods, it is much more likely that officers of different races may possess different amounts of information regarding a motorist, as residents in the neighborhood may be more willing to share information with officers with the same race as theirs. In contrast, our data consists only of stops and searches conducted on highways, and as a result the above two issues are less concerning.

The remainder of the paper is structured as follows. Section 2 presents and analyzes our model of trooper search behavior, and proposes 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.\textsuperscript{17} There is a continuum of troopers (interchangeably, police officers) and motorists

\textsuperscript{17}Borooah (2001) and Bjerk (2004) develop somewhat related models of policing behavior.
(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.\(^{18}\) Suppose that among motorists of race \( r_m \in \{ M, W \} \), a fraction \( \pi^{r_m} \in (0, 1) \) of them carry contraband.\(^{19}\)

The information that is available to an officer when he or she makes the search decision consists of the motorist’s race and many other characteristics pertaining to the motorist. Such characteristics may include, for example, the gender, age and residential 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 seriousness of the reason for the stop, etc.\(^{20}\) Note that while the police officer observes all the characteristics in the decision to search, a researcher will typically have access to only a small subset of them. We assume, however, that the police officer will use a single-dimensional index \( \theta \in [0, 1] \) that summarizes all of the information that these characteristics indicate about 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 \( f_g^{r_m} (\cdot) \); if a race \( r_m \) driver does not carry contraband, \( \theta \) would be randomly drawn from \( f_n^{r_m} (\cdot) \). [The subscripts \( g \) and \( n \) stand for “guilty” and “not guilty,” respectively.] Without loss of generality, we can assume that the two densities \( f_g^{r_m} \) and \( f_n^{r_m} \) satisfy the strict monotone likelihood ratio property (MLRP), i.e., for \( r_m \in \{ M, W \} \),

\[
\text{MLRP: } f_g^{r_m} (\theta) / f_n^{r_m} (\theta) \text{ is strictly increasing in } \theta.
\]

The MLRP property on the signal distributions essentially means that a higher index \( \theta \) is a signal that a driver is more likely to be guilty.\(^{21}\) To the extent that there may be obviously guilty drivers (for example, if illicit drugs are in plain view), we assume that:

\(^{18}\)In the empirical part of the paper, we will examine three racial or ethnic groups: whites, blacks, and Hispanics. For now, though, we group blacks and Hispanics together as minorities for ease of exposition.

\(^{19}\)For the purpose of deriving our empirical test, we will assume that \( \pi^{r_m} \) is exogenous. In Appendix A, we present an equilibrium model in which \( \pi^{r_m} \) is endogenously determined.

\(^{20}\)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.

\(^{21}\)For any one dimensional index \( \theta \), we can always reorder them according to their likelihood ratio \( f_g^{r_m} (\theta) / f_n^{r_m} (\theta) \) in an ascending order. Thus the MLRP assumption is with no loss of generality.
Unbounded Likelihood Ratio: $f_g^r(\theta)/f_n^r(\theta) \to +\infty$ as $\theta \to 1$.

The MLRP also implies that the cumulative distribution function $F_g^r(\cdot)$ first order stochastically dominates $F_n^r(\cdot)$, which implies that drivers who actually carry contraband are more likely to generate higher and thus more suspicious signals. We think this single dimensional index formulation summarizes the information that is available to troopers when they make their 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 contraband) minus a cost of searching cars. This is an important assumption because it requires that police officers always use any statistical information contained in the race of the motorist in their search decisions.\footnote{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 contraband. This assumption rules out the possibility that some officers ignore the race of a motorist even when it provides useful information. See Section 2.1 for more discussion for this key assumption.}

Let $t(r_m; r_p)$ be the cost of a police officer with race $r_p$ searching a motorist with race $r_m$, where $r_p, r_m \in \{M, W\}$. We normalize the benefit of each arrest (or successful drug find) to equal one, and scale the search cost 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, and thus we can directly confront the possibility that police officers may not be monolithic in their search behavior.

We now introduce some definitions. First, a police officer of race $r_p$ is defined to be \textit{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.\footnote{Strictly speaking, we should have a broad interpretation of the search cost $t(r_m; r_p)$. For example, the cost of decoding the demeanor may be smaller if $r_m = r_p$. We are not able to distinguish such cost differences from racial prejudice.} \footnote{We are interpreting racial prejudice as “consequential animus” in the terminology of Ayres (2001, Chapter 3). In other contexts such as mortgage lending, racial prejudice may be manifested as “association animus,” i.e., a lender may be prejudiced against borrowers of a given race by not willing to engage in transactions with them. We believe that “consequential animus” is appropriate interpretation of racial prejudice in motor vehicle searches. We thank an anonymous referee for bringing this distinction to our attention.}
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 police do 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 has the same search cost against all groups of motorists, but that search costs depend on the race of the trooper.

2.1 Discussion of the Model

Assumption on the Signal Distributions. Our model allows 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. As we explained in the introduction, black and white drivers may exhibit different characteristics in their encounters with highway troopers, and thus imposing $f_{g^{M}}$ and $f_{n^{M}}$ to be equal to $f_{g^{W}}$ and $f_{n^{W}}$, respectively would be very strong and may be empirically implausible. Indeed, it is possible for example that minority drivers not carrying contraband might tend to be more nervous during a stop than whites. Also note that, since $\theta$ is most likely not observable by researchers, we do not want to impose parametric distributional assumptions. While sharper tests may be designed if we were to impose more parametric distributional restrictions on $f_{g^{r_m}}$ and $f_{n^{r_m}}$, the desirable feature of our test is its robustness.

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 for the possibility that minority drivers will present a signal that is drawn from one distribution when they are stopped by a minority trooper and another signal that is drawn from a different distribution when they are stopped by a white trooper. This would be a suspicious assumption, for example, if the stops and searches occur on local streets. As argued in Donohue and Levitt (2001), a black community may be more willing to cooperate with a black officer, and 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 officers have about black motorists on local streets may not extend to the highways. Thus as long as white and black troopers observe the same list of characteristics and summarize them in the same way, our assumption will be valid.

One may also argue that minority drivers might be more nervous with white officers than they are with minority officers, regardless of whether or not they are carrying contraband. But as long as white officers properly take this fact into account, they should put a lower weight on the observed nervousness from a black motorist when they formulate the signal index \( \theta \). Thus this argument does not necessarily invalidate our assumption that \( f_{\theta}^m \) and \( f_{\theta}^{\bar{m}} \) do not depend on the race of the police officers \( r_p \).

**Assumption on the Officers’ Objective Function.** We assume that officers maximize the total number of convictions minus a cost of searching cars. We allow the officers to exploit all statistically valid racial inferences in making their search decisions. Most highway patrol departments prohibited using race as the primary cause for police-citizen contact. For example, California Highway Patrol prohibits racial profiling which it defined as occurring “when a police officer initiates a traffic or investigative contact based primarily on the race/ethnicity of the individual.”

Thus our assumption may not be at odds with the official policies on racial profiling. Another possible way that our model incorporates such official denouncement against racial profiling is in the search cost \( t(\cdot, \cdot) \), to the extent that the search cost includes the possibility of being disciplined for engaging in racial profiling. Our assumption is problematic if officers respond differently to the official denouncement of using race/ethnicity information as the primary cause of stops and searches: if some troopers of a given race or some race of officers are not willing to engage in statistical discrimination, then our test becomes invalid. Unfortunately our data do not allow us to examine such possibilities because each officer in our sample only conduct a handful of searches. Thus we have to maintain the assumption that within a trooper race all troopers are monolithic.

In our model, we assume that the officers do not have search capacity constraint and thus they judge each stopped vehicle individually to determine whether it is worth a costly search. If officers have a search capacity constraint they may then target their searches only to those most suspicious ones. In reality, however, such concerns may not be important: in our data, an officer has on

---

\[25\] See California Highway Patrol Public Contact Demographic Data Summary (P. 1).

---
average less than 7 searches in a span of almost two years. An officer may also care about the
quality of the contraband found, but unfortunately we do not have such information in our data
set.26

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 contraband $\pi^{r_m} \in (0,1)$ does not
depend on the race of the troopers searching them. 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 that this is the case, see Tables 4 and 5). In Section 2.4 we
describe a resampling procedure to deal with this problem empirically.

2.2 Theoretical Implications

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:

$$\Pr (G|r_m, \theta) = \frac{\pi^{r_m} f_{\theta}^{r_m} (\theta)}{\pi^{r_m} f_{\theta}^{r_m} (\theta) + (1 - \pi^{r_m}) f_{\theta}^{r_m} (\theta)}.$$  

It immediately follows from the MLRP that $\Pr (G|r_m, \theta)$ is monotonically increasing in $\theta$. From the
unbounded likelihood ratio assumption, we know that $\Pr (G|r_m, \theta) \to 1$ as $\theta \to 1$.

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, which is normalized to 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:

\[\text{26The Maryland data set used by KPT does contain the quantity of drug found in the searches (see Knowles, Persico and Todd 2001).}\]
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 a successful search for the marginal motorist is equal to the cost of search. Any infra-marginal motorist will 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. We define the equilibrium search rate of race-\(r_p\) police officers against race-\(r_m\) motorists as \(\gamma (r_m; r_p)\), which is given by
\[
\gamma (r_m; r_p) = \pi^{r_m} \left[1 - F_g^{r_m} (\theta^* (r_m; r_p))\right] + (1 - \pi^{r_m}) \left[1 - F_n^{r_m} (\theta^* (r_m; r_p))\right].
\]

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
\[
S (r_m; r_p) = \frac{\pi^{r_m} \left[1 - F_g^{r_m} (\theta^* (r_m; r_p))\right]}{\pi^{r_m} \left[1 - F_g^{r_m} (\theta^* (r_m; r_p))\right] + (1 - \pi^{r_m}) \left[1 - F_n^{r_m} (\theta^* (r_m; r_p))\right]}.
\]

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 \(\theta^* (M; r_p) \neq \theta^* (W; r_p)\).

Officers will choose to use statistical discrimination if the distribution of the signal \(\theta\) among white and minority motorists is different. When these distributions differ and \(t (M; r_p) = t (W; r_p)\) (as assumed), Proposition 1 implies that the race-\(r_p\) police will choose search criteria \(\theta^* (M; r_p)\) and \(\theta^* (W; r_p)\) so that the marginal search success rates against white and minority motorists are both equal to the search cost. This typically implies that \(\theta^* (M; r_p) \neq \theta^* (W; r_p)\). One reason why the distribution of the signal \(\theta\) might be different across motorists of different races is that one group might be more likely to carry contraband. For example, if minority drivers are more likely to carry contraband (\(\pi^W < \pi^M\)), then it will be optimal for a non-prejudiced officer to search relatively
more minority drivers (assume everything else is the same for white and minority drivers), and thus they will set $\theta^* (M; r_p) < \theta^* (W; r_p)$. Another reason why the distribution of $\theta$ might be different for whites and minorities is that $f_{g,m}^m (\theta)$ and $f_{n,m}^m (\theta)$ can differ between motorist races.

Now we derive some simple implications of the model that will serve as the basis of our empirical test. First, note that if police officers are monolithic, then the cost of searching any given race of motorists is the same, regardless of the race of the officer. That is, $t(W; W) = t(W; M)$ and $t(M; W) = t(M; M)$. If we assume that white and minority troopers face the same population of white motorists and the same population of minority motorists, then Proposition 1 implies that both races of officers will use the same search criterion against a given race of motorists, so that $\theta^* (W; W) = \theta^* (W; M)$ and $\theta^* (M; W) = \theta^* (M; M)$. 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 it immediately follows from Definition 1 that the ranking of $t(r_m; M)$ and $t(r_m; W)$ does not depend on the motorist’s race $r_m$, regardless of whether or not troopers are monolithic. We can illustrate the implication of this using an example where white troopers find searching both minority and white motorists more costly than minority troopers do. More formally this can be written as $t(M; M) = t(W; M) < t(M; W) = t(W; W)$. Because the search threshold given in Proposition 1 is monotonically increasing in $t(r_m; r_p)$ and both white and minority troopers face the same population of white and minority motorists, this implies that $\theta^* (M; M) < \theta^* (M; W)$ and $\theta^* (W; M) < \theta^* (W; W)$. Because the equilibrium search rate given in formula (1) is monotonically decreasing in $\theta^* (r_m; r_p)$, we immediately have that $\gamma (M; M) > \gamma (M; W)$ and $\gamma (W; M) > \gamma (W; W)$, so that race-M

---

27In Section 2.4 we describe a resampling procedure to empirically deal with data sets in which this assumption may be invalid in the raw data.

28Consider, for illustrative purposes, the case that $t(W; M) < t(W; W)$. Since race-M officers are assumed not to be racially prejudiced, we have $t(W; M) = t(M; M)$. Similarly since race-W officers are not racially prejudiced, we have $t(W; W) = t(M; W)$. Thus it must be the case $t(M; M) < t(M; W)$. Thus $t(r_m; M) < t(r_m; W)$ for all $r_m$. Similar arguments show that if $t(W; M) > t(W; W)$, then we must have $t(M; M) > t(M; W)$; and if $t(W; M) = t(W; W)$ then we must have $t(M; M) = t(M; W)$. Thus the ranking of $t(r_m; M)$ and $t(r_m; W)$ does not depend on the motorist’s race $r_m$.

29Note that the relationship $t(M; r_p) = t(W; r_p)$ does not imply that $\theta^* (M; r_p) = \theta^* (W; r_p)$, because troopers can be engaged in statistical discrimination.
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 rates of motorists than race-\( W \) officers. Finally, if \( t(M; M) = t(W; M) = t(M; W) = t(W; W) \), then race-\( M \) officers’ search rates will be equal to those of race-\( W \) officers for both races of motorists.

We can also show that if none of the police officers are racially prejudiced, then the rank order of average search success rates between white and minority troopers for any race of motorists should also be independent of the motorists’ race. Recall the previous example where white troopers had a higher overall search cost than minority troopers. We showed this would imply that \( \theta^*(M; M) < \theta^*(M; W) \) and \( \theta^*(W; M) < \theta^*(W; W) \). The average search success rate with a search criterion \( \theta^* \) against race-\( r_m \) motorist is simply

\[
\frac{\pi^{r_m} [1 - F_g^{r_m}(\theta^*)]}{\pi^{r_m} [1 - F_g^{r_m}(\theta^*)] + (1 - \pi^{r_m}) [1 - F_n^{r_m}(\theta^*)]},
\]

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 above discussion is summarized in the following proposition:

**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 \( S(r_m; W) \).

In our model if race-\( r_p \) troopers are not racially prejudiced, we know that race-\( r_p \) troopers’ *marginal* search success rate against white motorists will be equal to that against minority motorists.

---

30To see this, note that it will be strictly increasing in \( \theta^* \) if and only if

\[
H(\theta^*) = \frac{1 - F_g^{r_m}(\theta^*)}{1 - F_n^{r_m}(\theta^*)}
\]

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

\[
H'(\theta^*) = \frac{-f_g^{r_m}(\theta^*) [1 - F_g^{r_m}(\theta^*)] + [1 - F_g^{r_m}(\theta^*)] f_n^{r_m}(\theta^*)}{[1 - F_n^{r_m}(\theta^*)]^2}
\]

\[
= \frac{-f_g^{r_m}(\theta^*) \int_0^1 f_n^{r_m}(\theta) d\theta + f_n^{r_m}(\theta^*) \int_0^1 f_g^{r_m}(\theta) d\theta}{[1 - F_n^{r_m}(\theta^*)]^2}
\]

\[
= \int_0^1 \left[ f_n^{r_m}(\theta^*) f_g^{r_m}(\theta) - f_n^{r_m}(\theta^*) f_n^{r_m}(\theta) \right] d\theta
\]

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

\[
\frac{f_g^{r_m}(\theta)}{f_n^{r_m}(\theta)} > \frac{f_g^{r_m}(\theta^*)}{f_n^{r_m}(\theta^*)}
\]

thus the integrand in the numerator is always positive. Thus \( H'(\theta^*) > 0 \).
But because in our model the marginal motorist’s guilty probability is smaller than that of the infra-
marginal motorists, we can not conclude that race-$r_p$ troopers’ average search success rate against
white motorists will be equal to that against minority motorists. This is in stark contrast to
KPT’s model where there is no distinction between marginal and average motorists. Nonetheless,
Proposition 3 provides robust testable implications of our model based on rank orders of observable
statistics – the search rates and the average search success rates.\textsuperscript{31}

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
troopers exhibit racial prejudice. Without further assumptions, it is not possible to determine
which group of troopers are racially prejudiced.

\section{Empirical Tests}

\subsection{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 $\theta$ observed by
troopers in making their search decisions. Under the null hypothesis that police officers exhibit
monolithic behavior, then, for any race of drivers, the search rates and average search success rates
against drivers of that race should be independent of the race of the troopers that conduct the
searches. That is, under the null hypothesis of monolithic trooper behavior, we must have, for all
$r_m \in \{M, W\}$,

\begin{align}
\gamma (r_m; M) &= \gamma (r_m; W), \\
S (r_m; M) &= S (r_m; W).
\end{align}

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. To illustrate why this is true we need to show that when the null hypothesis is not true
we will never satisfy equality (3) and (4). Without loss of generality, suppose that troopers are not
monolithic in their search behavior against white motorists ($r_m = W$). That is, $t(W; W) \neq t(W; M)$.

\textsuperscript{31}Proposition 3 provides testable implications on the rank orders of both search rate and average search success
rates. In this regard, our test is in agreement to Ross and Yinger (2002, Chapter 8) in the context of mortgage lending
discrimination, where they emphasize the inextricable link between loan approval decisions and loan performance (see
also Ross 1997).
If \( t(W; W) > t(W; M) \), then, because both white and minority troopers face the same population of white motorists, we know 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. This then simultaneously implies that \( \gamma(W; W) < \gamma(W; M) \) and that \( S(W; W) > S(W; M) \), following from the proof in footnote 30. Thus the test using either (3) and (4) has an asymptotic power of one.

Moreover, the relationship between search rates and average search success rates suggests that, in principle, our model can be refuted. According to our model, whenever \( \gamma(W; W) < \gamma(W; M) \), this must be because \( \theta^* (W; W) > \theta^* (W; M) \) which directly implies that \( S(W; W) > S(W; M) \). Thus if the rank order between the search rates between racial groups of troopers for a given race of motorists is not exactly the opposite of the rank order between the average search success rates, then we know that at least some of the conditions of our model are not satisfied.\(^{32}\)

### 2.3.2 Test for Racial Prejudice

Proposition 3 suggests a test for whether some racial groups of troopers exhibit racial prejudice in their search behavior. Under the null hypothesis that none of the racial groups of 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 ranking of \( \gamma (r_m; M) \) and \( \gamma (r_m; W) \), or the ranking of \( S(r_m; M) \) and \( S(r_m; W) \), depends on the race of the motorists \( r_m \).

**Difference From Test of Antonovics and Knight (2004).** Now we relate our test of racial prejudice to the test proposed in Antonovics and Knight (2004). As we described in the introduction, they use evidence that police officers are more likely to conduct a search if the race of the officer differs from the race of the driver as evidence of racial prejudice. First, it is useful to point out that their test is different from our rank order test proposed above. Consider the following simple example. Suppose that \( r_m, r_p \in \{ W, M \} \) and let the search rates be as follows: \( \gamma (M; M) = .05, \gamma (W; M) = .10, \gamma (M; W) = .20 \) and \( \gamma (W; W) = .15 \). That is, minority officers are more likely to search white motorists than minority motorists, and white officers are more likely

\(^{32}\)Of course, if the search rates between racial groups of troopers for a given race of motorists are equal, then the average search success rates between racial groups of troopers for a given race of motorists must also be equal.
to search minority motorists than white motorists. Thus officers in this example are more likely to conduct a search if the race of the motorist is different from their own, causing Antonovics and Knight’s test to conclude that racial prejudice is occurring. However, such patterns of search rates satisfy our rank independence condition, that is, \( \gamma(r_m; W) > \gamma(r_m; M) \) for \( r_m \in \{W, M\} \), and thus our test would not consider this as evidence of racial prejudice. If we allow for arbitrary differences, including higher moments, in the signal distributions between white and minority motorists [determined by \((\pi^W, f_g^W, f_n^W)\) and \((\pi^B, f_g^B, f_n^B)\) respectively], a positive coefficient on “racial mismatch” can be consistent with the hypothesis that both racial groups of officers are not prejudiced, even though they must behave non-monolithically. We would like to emphasize, however, that we do not mean to say that our test proves no racial prejudice: our conclusion is simply that no racial prejudice could not be ruled out by the data without making stronger, and non-verifiable, distributional assumptions on the signal distribution. A second difference between Antonovics and Knight’s (2004) test and ours is that we use both search rate and average search success rates in our test, while theirs uses only search rates. Using both pieces of information permits us to potentially refute our behavioral model on which our test is based. We think this is an additional strength of our test (see Ross and Yinger 2002, Chapter 8 for related discussion in the context of mortgage lending).

### 2.4 A Resampling Procedure

As we mentioned in Section 2.1, our model assumes that the fraction of race-\( r_m \) motorists carrying contraband \( \pi^{r_m} \in (0, 1) \) does not depend on the race of the troopers searching them. Our raw data, summarized in Tables 4 and 5, indicated that white and minority troopers are systematically assigned to patrol in different locations or time of day. Here we propose an empirical method that can resolve this problem even when the raw data does not satisfy this condition. For illustration purposes, suppose that there are two troop stations 1 and 2, each with 100 officers. Suppose that in troop station 1, 80 officers are white and 20 are minorities; in station 2, 60 officers are white and 40 are minorities. Thus, on average 70 percent of the troopers are white and 30 percent 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 select 47 out of the 80 white officers. Similarly, we keep all the white officers (60 of them) in station...
2, but randomly 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 resampling method can effectively ensure 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.

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).\footnote{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.} 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 1,469
troopers. Florida State Highway Patrol troopers are assigned to one of 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 2 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 of the motorists, male motorists account for at least 67 percent of the stopped motorists for all race categories. Among all the motorists that were stopped, 48 percent were in the 16-30 age group, 33.6 percent were in the 31-45 age group and 18.3 percent were 46 and older. 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 between 6am and 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 searches 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 the 16-30 age group, 31.7 percent were in the 31-45 age group and only 9.9 percent were in the 46 and older age group. Vehicles with in-state plates account 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 3 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 13.7, 10, and 76.3 percent of the troopers respectively. 89 percent of the troopers are male. The second and the third columns show 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 they are 11.4 and 9.5 percent. Female troopers conducted 9.3 percent of all stops and 6.9 percent of all searches. 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 of the other racial groups to consider them.
Figure 1: Troop Station Coverage Map
<table>
<thead>
<tr>
<th>Motorists' Characteristics</th>
<th>Stops</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>By Motorist Sex</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
</tr>
<tr>
<td>Black</td>
<td>.162 (.368)</td>
<td>.327 (.470)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.173 (.378)</td>
<td>.225 (.417)</td>
</tr>
<tr>
<td>White</td>
<td>.665 (.472)</td>
<td>.319 (.466)</td>
</tr>
<tr>
<td>Female</td>
<td>.304 (.460)</td>
<td>1.00 (.00)</td>
</tr>
<tr>
<td>Male</td>
<td>.696 (.460)</td>
<td>0.00 (.00)</td>
</tr>
<tr>
<td>Age: 16-30</td>
<td>.481 (.500)</td>
<td>.325 (.468)</td>
</tr>
<tr>
<td></td>
<td>.336 (.472)</td>
<td>.295 (.456)</td>
</tr>
<tr>
<td></td>
<td>.183 (.386)</td>
<td>.269 (.444)</td>
</tr>
<tr>
<td>License Plate:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-state</td>
<td>.899 (.302)</td>
<td>.310 (.462)</td>
</tr>
<tr>
<td>Out-of-state</td>
<td>.101 (.302)</td>
<td>.252 (.434)</td>
</tr>
<tr>
<td>Time: Day (6am-6pm)</td>
<td>.697 (.459)</td>
<td>.316 (.465)</td>
</tr>
<tr>
<td></td>
<td>.303 (.459)</td>
<td>.275 (.447)</td>
</tr>
<tr>
<td>Contraband Seized:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>.792 (.406)</td>
<td>.155 (.362)</td>
</tr>
<tr>
<td>Drugs</td>
<td>.151 (.358)</td>
<td>.137 (.344)</td>
</tr>
<tr>
<td>Paraphernalia</td>
<td>.015 (.122)</td>
<td>.156 (.364)</td>
</tr>
<tr>
<td>Currency</td>
<td>.003 (.051)</td>
<td>.174 (.388)</td>
</tr>
<tr>
<td>Vehicles</td>
<td>.010 (.100)</td>
<td>.154 (.363)</td>
</tr>
<tr>
<td>Alcohol/Tobacco</td>
<td>.021 (.142)</td>
<td>.151 (.359)</td>
</tr>
<tr>
<td>Weapons</td>
<td>.006 (.078)</td>
<td>.055 (.229)</td>
</tr>
<tr>
<td>Other</td>
<td>.003 (.049)</td>
<td>.318 (.477)</td>
</tr>
</tbody>
</table>

Number of Observations: 906,339 275,527 630,812 8,976 1,364 7,612

Table 2: Means of Variables Related to Motorists.

Note: Standard errors of the means are shown in parentheses.
<table>
<thead>
<tr>
<th>Troopers' Characteristics</th>
<th>Troopers</th>
<th>Stops</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td></td>
<td>By Trooper Sex</td>
<td>By Trooper Sex</td>
</tr>
<tr>
<td>Black</td>
<td>.137</td>
<td>.160</td>
<td>.115</td>
</tr>
<tr>
<td></td>
<td>(.344)</td>
<td>(.366)</td>
<td>(.319)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.100</td>
<td>.114</td>
<td>.070</td>
</tr>
<tr>
<td></td>
<td>(.300)</td>
<td>(.318)</td>
<td>(.256)</td>
</tr>
<tr>
<td>White</td>
<td>.763</td>
<td>.726</td>
<td>.092</td>
</tr>
<tr>
<td></td>
<td>(.425)</td>
<td>(.446)</td>
<td>(.289)</td>
</tr>
<tr>
<td>Female</td>
<td>.106</td>
<td>.093</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>(.307)</td>
<td>(.291)</td>
<td>(.00)</td>
</tr>
<tr>
<td>Male</td>
<td>.894</td>
<td>.907</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(.307)</td>
<td>(.291)</td>
<td>(.00)</td>
</tr>
<tr>
<td>Ranks:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Captain</td>
<td>.022</td>
<td>.002</td>
<td>.239</td>
</tr>
<tr>
<td></td>
<td>(.148)</td>
<td>(.041)</td>
<td>(.426)</td>
</tr>
<tr>
<td>Lieutenant</td>
<td>.070</td>
<td>.013</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td>(.255)</td>
<td>(.112)</td>
<td>(.151)</td>
</tr>
<tr>
<td>Sergeant</td>
<td>.145</td>
<td>.062</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td>(.352)</td>
<td>(.241)</td>
<td>(.226)</td>
</tr>
<tr>
<td>Corporal</td>
<td>.147</td>
<td>.112</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td>(.354)</td>
<td>(.316)</td>
<td>(.252)</td>
</tr>
<tr>
<td>LEO</td>
<td>.602</td>
<td>.810</td>
<td>.101</td>
</tr>
<tr>
<td></td>
<td>(.490)</td>
<td>(.392)</td>
<td>(.301)</td>
</tr>
</tbody>
</table>

Table 3: Means of Variables Related to Troopers.

**Note:** Standard errors of the means are shown in parentheses.
Table 4: 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-value</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 Examining 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 $\pi_{rm}, f_{g}^{rm}(\theta), f_{n}^{rm}(\theta)$ and $\theta$ are all unobservable. The best we can do is to examine the distribution of observable motorist characteristics faced by troopers of different races. Table 4 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) relative to Troop A and H (covering counties in the Florida Panhandle). Indeed, Table 5 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 2.1.\textsuperscript{35} By construction, in the artificial data we created with the resampling method, troopers of a given race are assigned to different troops with the same probabilities. The Pearson’s $\chi^2$ test also reveal that in the artificial sample troopers of

\textsuperscript{35}One 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.
different races are assigned to night shifts with the same probability. Thus we can maintain our hypothesis that the distribution of the observable characteristics of the stopped motorists faced by troopers are the same in the artificial sample. We report our test results below using data from the artificial samples.

3.2.3 Test for Monolithic Trooper Behavior

We now implement our test for the hypothesis that troopers of different races exhibit monolithic behavior. Our main empirical results are presented in Table 1 in the introduction. Panel A shows two facts: first, regardless of motorists’ race, white officers search the highest percentage of the motorists they stop, and black officers search the lowest percentage; second, for all officers’ races, the percentage of black motorists searched is higher than Hispanic motorists, which in turn is higher than white motorists. Table 1 also 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, the 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

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

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, $R$. The $p$-value for a given motorist race gives the significance level above which we can reject the null hypothesis that the three search rates corresponding to that row are equal, which is the prediction under the null hypothesis of monolithic behavior. Note that all the $p$-values are less than 0.001, indicating strong evidence against monolithic trooper behavior.

Panel B presents the average search success rate for given motorist/trooper race pairs. The first finding from Panel B is exactly converse to the first finding from Panel A: for any given motorist race, black officers’ average search success rate is higher than that of Hispanic officers, which in turn is higher than that of white officers. The $p$-value in each row is 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

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

where $S(\hat{r}_m; r_p)$ is the estimated average search success rate of race-$r_p$ officers against race-$r_m$ motorists, and $S(\hat{r}_m)$ is the estimated average search success rate against race-$r_m$ motorists unconditional on the race of the officers. All $p$-values are less than .001, again indicating strong evidence against monolithic trooper behavior.

The second finding from Panel B is that, for all officers the average search success rate is highest against white motorists, followed in order by black and Hispanic motorists. Though this finding is not directly related to our test for monolithic behavior, it provides strong support for our modelling assumption that the distributions of unobservable characteristics for motorists of different races may be very different, not only in means but also in higher moments. For example, Panel A shows that black officers search about the same percentage of white and Hispanic motorists (0.27 vs. 0.28), but their average search success rate against white motorists are much higher than that for Hispanic motorists (39.4 vs. 21.0).

### 3.2.4 Test for Racial Prejudice

We have so far provided strong evidence that troopers do not exhibit monolithic search criteria when deciding whether to search motorists of a given race. Now we describe the results from our test for racial prejudice as described in subsection 2.3.2. Under the null hypothesis that none of the racial groups of troopers are racially prejudiced, we argued that the rank order over the search rates $\gamma (r_m; W), \gamma (r_m; B)$ and $\gamma (r_m; H)$, and the rank order over the average search success rates $S(r_m; W), S(r_m; B)$ and $S(r_m; H)$, should both be independent of $r_m$. From the estimated mean search rates and average search success rates in Table 1, we have for all $r_m \in \{W, B, H\}$,

$$\gamma (\hat{r}_m; W) \succ \gamma (\hat{r}_m; H) \succ \gamma (\hat{r}_m; B),$$

$$S(\hat{r}_m; W) < S(\hat{r}_m; H) < S(\hat{r}_m; B).$$

We can use simple $Z$-statistics to formally test that

$$\gamma (r_m; W) \succ \gamma (r_m; H) \succ \gamma (r_m; B), \quad (5)$$

$$S(r_m; W) < S(r_m; H) < S(r_m; B). \quad (6)$$
For example, let the null hypothesis be \( \gamma (r_m; W) = \gamma (r_m; H) \). We can test it against the one-sided alternative hypothesis \( \gamma (r_m; W) > \gamma (r_m; H) \) by using

\[
Z = \frac{\gamma (r_m; W) - \gamma (r_m; H)}{\sqrt{\frac{\text{SVar}_W}{n_W} + \frac{\text{SVar}_H}{n_H}}}
\]

where \( n_W \) and \( n_H \) are the number of stops conducted by white and Hispanic officers respectively against race-\( r_m \) motorists, and \( \text{SVar}_W \) and \( \text{SVar}_H \) are respectively the sample variances of the search dummy variables in the samples of stops against race-\( r_m \) motorists conducted by white and Hispanic officers. By the Central Limit Theorem (due to our large sample size), \( Z \) has a standard normal distribution under the null hypothesis. The null will be rejected in favor of the alternative at significance level \( \alpha \) if \( Z \geq z_\alpha \) where \( \Phi (z_\alpha) = 1 - \alpha \). When \( r_m = W \), the value of the \( Z \)-statistic is 27.4 under the null, thus we can reject it in favor of the alternative \( \gamma (W; W) > \gamma (W; H) \) at significance level close to 0. Similarly, for the test of the null hypothesis \( \gamma (W; H) = \gamma (W; B) \) against \( \gamma (W; H) > \gamma (W; B) \), we obtain a \( Z \)-statistic of 65, thus again rejecting the null in favor of the alternative. Implementing this test to other races of motorists, we find that the evidence supports inequality (5).

We can use an analogous \( Z \)-test to formally test inequality (6) by using

\[
Z' = \frac{S (r_m; W) - S (r_m; H)}{\sqrt{\frac{\text{SVar}_W'}{n_W'} + \frac{\text{SVar}_H'}{n_H'}}} \sim N (0, 1), \tag{7}
\]

where \( n_W' \) and \( n_H' \) are the number of searches against race-\( r_m \) motorists conducted by white and Hispanic officers respectively, and \( \text{SVar}_W' \) and \( \text{SVar}_H' \) are respectively the sample variances of the search success dummy variables in the sample of searches against race-\( r_m \) motorists 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 −324.1 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 \( Z \)-statistic of −254). Implementing this test to other races of motorists, we find that the evidence supports inequality (6).

To summarize, we cannot 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.

### 3.2.5 Other Implications from the Tests

It is interesting to note some additional implications from the tests we conducted above. First of all, inequality (5) implies that the search criterion used by troopers against race-$r_m$ motorists have the ranking

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

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

$$t(r_m; W) < t(r_m; H) < t(r_m; B).$$

That is, white troopers seem to have smaller costs of searching motorists of any race, followed by Hispanic troopers. Black troopers have the highest search costs.

Second, as we mentioned at the end of subsection 2.3.1, our model is refuted if, for each $r_m$, the rank order of the search rates against race-$r_m$ motorists $\gamma(r_m; W), \gamma(r_m; B)$ and $\gamma(r_m; H)$ is not exactly the opposite of the rank order of the corresponding average search success rates $S(r_m; W), S(r_m; B)$ and $S(r_m; H)$. As we showed above, the statistical evidence in our data does not refute our model.

### 3.2.6 Replicating KPT’s Test

It is useful 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. Last column in Panel B of Table 1 shows the average search success rate for different races of the motorists in the raw data, and Table 6 shows the $p$-values from Pearson’s $\chi^2$ test on the hypothesis that the search rates and average search success rates 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.
### Table 6: p-Values from Pearson’s $\chi^2$ Tests on the Hypothesis that Average Search Success Rate are Equal Across Various Groupings: KPT Test.

<table>
<thead>
<tr>
<th>Groupings</th>
<th>Average Search Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>White, Black, Hispanic</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>White, Black</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>White, Hispanic</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Black, Hispanic</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

#### 3.3 Discussion

**Power of the Test** Two features of our empirical test for racial prejudice is worth remarking. First, it only speaks to “relative bias” among troopers. When we do find evidence of racial prejudice, we only know that at least some race of officers are racially prejudiced, but would not be able to determine the race of the prejudiced officers. Second, the power of our test is not one, even when the sample size goes to infinity. To illustrate, let $r_m, r_p \in \{M, W\}$ for simplicity. 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, and 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 a type-II error. In general, our test has an asymptotic power of zero if $[t(W; W) - t(W; M)] [t(M; W) - t(M; M)] > 0$;\(^{36}\) and has an asymptotic power of one if $[t(W; W) - t(W; M)] [t(M; W) - t(M; M)] < 0$.

The low power of our test may be considered a weakness of our test. If we do find evidence against the null hypothesis, however, we can be quite confident that at least one racial group of troopers is racially prejudiced. If we were willing to assume that the signal distributions $f_g^{r_m}$ and $f_n^{r_m}$ do not depend on $r_m$, then one can derive more powerful tests for racial prejudice. Our test can be considered as the robust implication from a plausible behavioral model without imposing unverifiable distributional assumptions.\(^{37}\)

\(^{36}\)This will be true either $t(W; W) > t(W; M)$ and $t(M; W) > t(M; M)$, or $t(W; W) < t(W; M)$ and $t(M; W) < t(M; M)$.

\(^{37}\)In this regard, our position is similar to Manski (1995) who preached the tolerance of ambiguity in empirical
Finally it is worth pointing out that our test is able to detect racial prejudice when we apply it to the Boston data analyzed in Antonovics and Knight (2004). Their table 1 indicates that in their data, black officers’ search rate is higher than white officers’ against white motorists; but white officers’ search rate is higher than black officers’ against black motorists. This is a violation of the rank order independence for the search rate, indicating at least one race group of the officers are racially prejudiced.38

Racial Bias in Stops. Our paper only focuses on the officers’ search decisions. But the trooper must first stop the motorist prior to a search. In our analysis, we took the sample of cars that are stopped as our population and focus solely on determining racial prejudice in troopers’ search decisions. Our data does not allow us the possibility of racial prejudice in highway stops is beyond the scope of this paper. However, it is possible that the racial prejudice of police officers are reflected in their stop decisions as well as (or instead of) their search decisions. Because our model allows for general differences in the unobservable distributions among motorists of different races, the presence (or lack thereof) of racial prejudice at the stop level should not affect our conclusions about additional racial prejudice in the search decisions. Investigating racial bias in stops is clearly an important topic for future research, when suitable data sets that include random samples of drivers on the road become available.

4 Conclusion

Black and Hispanic motorists in the United States are much more likely than white motorists to be searched by highway troopers. Is this apparent racial disparity driven by racist preferences by the troopers, or by motives of effectiveness in interdicting drugs? Our paper presents a simple but plausible model of police search behavior, and we define racial prejudice, statistical discrimination and monolithic trooper behavior within the confines of our model. We then exploit the theoretical predictions from this model 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 troopers’ desire for effective policing (statistical discrimination) or by their racial prejudice (racism)? Relative to the seminal research in Knowles, Persico and Todd (2001), our model allows

38Their paper did not present information about average search success rates which, as we remarked earlier, could have been used to potentially refute our model.
troopers of different races to behave differently, thus allowing us to examine non-monolithic trooper behavior; moreover, our model does not yield, and the subsequent empirical test does not rely on, the convenient, but in our view unrealistic, implication that all drivers of the same race carry contraband with the same probability regardless of characteristics other than race, which is the vital prediction underlying their tests. We also propose a resampling method to deal with raw data sets where one of the major assumptions underlying our model and empirical tests is violated. 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 none of the racial groups of troopers are racially prejudiced. Finally we would like to emphasize that our test for racial prejudice is relatively conservative in that we may not always conclude there is racial prejudice when it is actually present. Although our test is a low-power one, which implies a high probability of type-II error will occur, the positive side of this is that when we do find evidence of racial prejudice it is rather conclusive.

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 as \( \pi^{r_m} \in (0,1) \). 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 like reducing crimes and the “war on drugs,” one may want to know how any changes in trooper behavior may affect the motorists’ drug carrying decisions.\(^{39}\) 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 \( r'_p \).\(^{40}\) Suppose

\(^{39}\)See Persico (2002) for an analysis on how racially blind search policies may affect the total crimes committed by motorists.

\(^{40}\)Because we are only considering one race group of motorists, we will omit \( r_m \) from the subsequent notation.
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 contraband $v \in \mathbb{R}_+$ 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 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

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

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

$$\Pr(G|\theta) = t'_p.$$  

Having more than one racial groups of motorists will not change any of the results below.
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

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

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-\( r_p \) and 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, whose utility cost is \( v \), from not carrying contraband is simply the inconvenience cost of being searched by mistaken troopers:

\[
\left\{ \alpha \left[ 1 - F_g(\theta_p^*) \right] + (1 - \alpha) \left[ 1 - F_g(\theta_{p'}^*) \right] \right\} c_n.
\]

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

\[
v^* (\theta_p^*, \theta_{p'}^*) = \left[ \alpha F_g(\theta_p^*) + (1 - \alpha) F_g(\theta_{p'}^*) \right] b - \left\{ \alpha \left[ 1 - F_g(\theta_p^*) \right] + (1 - \alpha) \left[ 1 - F_g(\theta_{p'}^*) \right] \right\} c_g + \left\{ \alpha \left[ 1 - F_n(\theta_p^*) \right] + (1 - \alpha) \left[ 1 - F_n(\theta_{p'}^*) \right] \right\} c_n.
\]

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

An equilibrium of the model is a triple \( (\pi, \theta_p^*, \theta_{p'}^*) \) such that:

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

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 (i.e., \( v \in \mathbb{R}_+ \)), the equilibrium is \textit{unique}. Suppose that there are two equilibria in which the proportion of guilty motorists are \( \pi \) and \( \tilde{\pi} \) with
\( \pi > \hat{\pi} \). Observe from (A1) that \( v^*(0,0) = c_n - c_g < 0 \) and

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

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

By the MLRP, we know that there exists \( \left( \theta^*_p, \theta^*_p' \right) \in [0,1]^2 \) such that \( v^* (\theta^*_p, \theta^*_p') \) is strictly increasing in both \( \theta^*_p \) and \( \theta^*_p' \) when \( (\theta^*_p, \theta^*_p') > \left( \hat{\theta}^*_p, \hat{\theta}^*_p' \right) \). Since \( v^*(0,0) < 0 \) and the support of \( G \) is non-negative, we have \( G (v^*(0,0)) = 0 \). Moreover, \( G (v^*(\theta^*_p, \theta^*_p')) \) will be zero for all \( (\theta^*_p, \theta^*_p') \leq \left( \hat{\theta}^*_p, \hat{\theta}^*_p' \right) \). Thus any \( (\theta^*_p, \theta^*_p') \leq \left( \hat{\theta}^*_p, \hat{\theta}^*_p' \right) \) cannot be part of the equilibrium (because if \( \pi = 0 \), the optimal thresholds should be 1 from the troopers’ best response). Thus in both equilibria of the model, we must have \( (\theta^*_p, \theta^*_p') > \left( \hat{\theta}^*_p, \hat{\theta}^*_p' \right) \) and \( \left( \hat{\theta}^*_p, \hat{\theta}^*_p' \right) > \left( \hat{\theta}^*_p, \hat{\theta}^*_p' \right) \). That is, both equilibria lie in the region where \( v^* (\cdot, \cdot) \) is strictly increasing in both arguments. If \( \pi > \hat{\pi} \), equilibrium conditions (A2) and (A3) imply that \( \theta^*_p < \hat{\theta}^*_p \) and \( \theta^*_p' < \hat{\theta}^*_p' \), therefore \( 0 < v^* (\theta^*_p, \theta^*_p') < v^* (\hat{\theta}^*_p, \hat{\theta}^*_p') \). But then it implies that \( \hat{\pi} > \pi \), a contradiction.

References


