

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

By SHAMENA ANWAR AND HANMING FANG\*

*We propose a simple model of trooper behavior to design empirical tests for whether troopers of different races are monolithic in their search behavior, and whether they exhibit relative racial prejudice in motor vehicle searches. Our test of relative racial prejudice provides a partial solution to the well-known infra-marginality and omitted-variables problems associated with outcome tests. When applied to a unique dataset from Florida, our tests soundly reject the hypothesis that troopers of different races are monolithic in their search behavior, but the tests fail to reject the hypothesis that troopers of different races do not exhibit relative racial prejudice. (JEL J7)*

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 (Robin S. Engel et al., 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 troopers 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, John Knowles et al., 2001; Engel et al., 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 not a factor in troopers’ decision-making. 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 dis-*

\* Anwar: Department of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520-8268 (e-mail: shamena.anwar@yale.edu); Fang: Department of Economics, Yale University, P.O. Box 208264, New Haven, CT 06520-8264 and National Bureau of Economic Research (e-mail: hanming.fang@yale.edu). 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 four anonymous referees for many detailed and constructive comments on an earlier draft. We thank Richard Zeller at the 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 for financial support from a Yale University Junior Faculty Fellowship. All remaining errors are our own.

*crimination* in the terminology of Kenneth J. 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. Furthermore, ethnographic studies such as those of Lawrence W. Sherman (1980) and Eric C. Riksheim and Steven M. 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 drivers, it becomes very difficult to determine the true motivation behind racial profiling with the available data.

One prominent approach that has been used to distinguish between racial prejudice and statistical discrimination is the "outcome test," which originated in Gary S. Becker (1957).<sup>1</sup> 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 against the *marginal* minority motorist (i.e., the last minority motorist deemed suspicious enough to be searched) will be lower than the success rate against 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 against the marginal minority motorist be equal to that against 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 search success rates against marginal motorists. This is due to the fact that we cannot identify the marginal motorist, since accomplishing this would require having complete information on all of the variables that troopers use in determining the suspicion level of motorists. Because of this *omitted-variables* problem, we can observe only the average success rate of searches against white and minority motorists, and not the marginal success rate. Since the equality of marginal search success rates does not imply, and is not implied by, the equality of the average search success rates, we cannot determine the relationship between the marginal search success rates of white and minority motorists by looking at average success rates. In past literature, this has been referred to as the "infra-marginality" problem. These problems severely limit the rigorous application of the outcome test idea, especially in situations where the decision or the outcome is dichotomous.<sup>2</sup>

A seminal paper by Knowles et al. (2001, KPT hereafter) provides the first solution to the infra-marginality problem associated with the outcome test. 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 factors that are observable to troopers but may or may not be available to researchers. Troopers decide whether or not to search motorists while motorists decide whether or not to carry contraband. In this "matching pennies"-like model they show that if troopers are not racially prejudiced, all motorists, if they are searched at all, must in equilibrium carry contraband with equal probability regardless of their race and other characteristics. Thus, in their model there is *no difference* between the marginal and the average

<sup>1</sup> Becker (1993a, 1993b) further elaborated on this idea and Ian Ayres (2001) presented several interesting applications.

<sup>2</sup> See Stephen L. Ross and John Yinger (1999 and 2002, chap. 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, chap. 8) argue that the complete elimination of the omitted variable bias results in a test with no power.

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 benefits from carrying contraband, but these differences imply only 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 dataset of 1,590 searches on a stretch of 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 Hispanic motorists.

While KPT's model provides a good starting point to distinguish between racial prejudice and statistical discrimination empirically, there are a couple of drawbacks to their theoretical model which cast doubt on the validity of their empirical test. First, KPT's model predicts that all motorists of a given race, if they are ever searched, will carry contraband with equal probability regardless of their other characteristics that may be observed by the police. 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 about the presence of contraband when a trooper decides whether to search. This implication of police 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 (see, e.g., Sherman, 1980; 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 are intimately related to whether 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 reappears in the empirical analysis. Our main contribution to the racial profiling literature in this paper is that we develop a more realistic model of trooper behavior that allows officers to use information they gather about motorists during traffic stops when they make their search decisions; and we exploit the theoretical implications of our model to propose an alternative empirical test to detect racial prejudice *in the presence of* potential infra-marginality and omitted-variables problems associated with outcome tests.

The second issue we have concerning KPT (and this field of research in general) is that they implicitly assume that all troopers' behavior is *monolithic*. This assumption may not be valid.<sup>3</sup> Most existing datasets on police behavior do not contain detailed information about the trooper characteristics, and thus it is assumed that all troopers, regardless of their race, have the same racial prejudice against minority motorists.<sup>4</sup> John J. Donohue and Steven D. 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 a wrong conclusion about 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

<sup>3</sup> A formal definition of monolithic behavior is given in Section II.

<sup>4</sup> The Maryland dataset KPT used has only very limited information about troopers (see KPT, 2001; Samuel R. Gross and Katherine Y. Barnes, 2002).

TABLE 1—SEARCH RATES AND AVERAGE SEARCH SUCCESS RATES AGAINST MOTORISTS OF DIFFERENT RACES

Motorists' race	Trooper race				All troopers
	White	Black	Hispanic	<i>p</i> -value	
Panel A: Search rate given stop (percent)					
White	0.96 (6.68E-4)	0.27 (7.73E-4)	0.76 (9.26E-4)	<0.001	0.81 (0.090)
Black	1.74 (1.30E-3)	0.35 (1.42E-3)	1.21 (2.28E-3)	<0.001	1.35 (0.115)
Hispanic	1.61 (1.46E-3)	0.28 (0.76E-3)	0.99 (3.03E-3)	<0.001	1.34 (0.115)
Panel B: Average search success rate (percent)					
White	24.3 (9.43E-3)	39.4 (5.57E-2)	26.0 (2.28E-2)	<0.001	25.1 (0.434)
Black	19.9 (1.26E-2)	26.0 (5.32E-2)	20.8 (2.67E-2)	<0.001	20.9 (0.407)
Hispanic	8.5 (9.78E-3)	21.0 (4.55E-2)	14.3 (6.63E-2)	<0.001	11.5 (0.319)

Note: Standard errors of the means are shown in parentheses.

the police as a whole are not racially prejudiced. But this may seriously underestimate the harassment experienced by both white and minority motorists. Our paper deviates from the KPT model and embraces the possibility that police behavior may vary by their racial group, which is our second main contribution to the existing literature. As we will show later, the variation of trooper behavior by their race will provide the key additional information that allows us to develop our empirical test.<sup>5</sup> We are able to relax the typical assumption that requires troopers to behave monolithically because we have a *unique* dataset of highway stops and searches conducted by the Florida Highway Patrol which contains information on the race of the trooper making the stop *and* information about the motorist who is stopped.

The information we present in Table 1 further illustrates why it is unrealistic to impose the assumption that all troopers behave monolithically in our Florida dataset.<sup>6,7</sup> Panels A and B,

respectively, show the search rate given stop and the average search success rate against motorists for the different combinations of motorist and trooper racial groups. The first row of panel A 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. Monolithic search behavior requires all troopers to search white motorists at the same rate, which clearly is not the case. Thus, any empirical test that relies on this assumption will not be valid for our dataset. The “All troopers” column in panel B of Table 1 also contains the information used in KPT’s test.<sup>8</sup> Because the average search success rates against both black and Hispanic motorists fall below those against white motorists, KPT would conclude that troopers have racial prejudice toward black and Hispanic motorists. 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 rates may not prove racial prejudice.

Our model of trooper search behavior follows

<sup>5</sup> We are grateful to an anonymous referee for clarifying this important point.

<sup>6</sup> Table 1 is our main table and we will discuss it more in Section IV. The formal test of monolithic behavior is presented in Section IVD.

<sup>7</sup> The 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 the reweighted samples constructed from the raw data. See Sec-

tion IIIC for details about how we construct the reweighted samples.

<sup>8</sup> More discussion of the KPT test on this dataset is provided in Section IVG.

the spirit of labor market statistical discrimination models (see, e.g., Stephen Coate and Glenn C. 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 depends on the race of the motorist.<sup>9</sup> The police force exhibits nonmonolithic behavior if the cost of searching motorists of a given race  $r_m$  depends 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 against 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 is racially prejudiced, then the 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. The key idea of our empirical test is as

follows: when there is no racial prejudice, the race of motorists should not affect the *ranking* of search rates and search success rates *over officer races*.

An additional desirable feature of our model is that it has direct implications for the ranking of both the search rates and the average search success rates, and thus our model could potentially be refuted by the data we have available. It is also important to point out, though, that our test can detect only what we term to be *relative* racial prejudice, and not *absolute* racial prejudice. This is because when the ranking of search rates and search success rates over officer races depends on the race of the motorists, we know that at least one of the racial groups of officers is using racial prejudice, but we cannot identify which group it is. Thus, all we can conclude is that one group of troopers is more racially prejudiced relative to another group of troopers, instead of an absolute conclusion which would identify which groups of troopers were racially prejudiced.<sup>10</sup>

The implementation of our empirical tests relies on datasets that have race information on both troopers and motorists.<sup>11</sup> While such data have not been available for use in earlier empirical studies on racial profiling, we are able to obtain a dataset 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 who 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 hypothesis that troopers of different races do not exhibit relative racial prejudice.

The remainder of the paper is structured as follows. Section I provides some additional discussion of the related literature. Section II presents and analyzes our model of trooper search

<sup>9</sup> We assume that race is the only characteristic of troopers that is likely to affect their search behavior. This is a 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.

<sup>10</sup> More discussions on this issue are provided in Section IIIB when we discuss the power of our test.

<sup>11</sup> 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 is more desirable because we can then examine whether our model is refuted by the data in addition to providing supporting evidence from the search rates.

behavior. Section III proposes empirical tests based on the theoretical predictions of the model. Section IV describes the dataset from the Florida Highway Patrol, presents our test results, and contrasts our results with those using KPT's test. Section V concludes. In the Appendix, we present a simple equilibrium model of drug-carrying behavior to show that our focus on trooper behavior in Section II is not problematic.

### I. Related Literature

Dhammika Dharmapala and Ross (2004) and Kate L. Antonovics and Brian G. Knight (2004) also discussed the possible shortcomings of the KPT model.<sup>12</sup> 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.<sup>13</sup> Under those circumstances, KPT's test fails because the infra-marginality and omitted variables problems reemerge. More specifically, the equilibrium of the KPT model under those circumstances may involve a group of motorists carrying drugs with probability one (being a "dealer") 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.

Antonovics and Knight (2004) argue that KPT's test may not be robust when its model is generalized to allow for trooper heterogeneity.<sup>14</sup> They also propose using 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 Section IIIB that their interpretation of the evidence may be misleading. It is also useful to point out that their data are from the Boston Police Department and consist mainly of stops and searches in *local* neighborhoods. There are two potential problems with such data. First, as Hernandez-Murillo and Knowles (2004) argue, 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 about motorists, as residents in the neighborhood may be more willing to share information with officers of their same race. In contrast, our data consist only of stops and searches conducted on highways, and as a result the two issues above are less concerning.

We would also like to point out that, besides the "outcome test" approach, a large field of literature has used a different statistical test, known as the "benchmarking test," to determine whether troopers impose disparate treatment on motorists of different races.<sup>15,16</sup> The bench-

<sup>12</sup> Ruben 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 nondiscretionary search using Missouri's annual aggregate traffic-stop report for the year 2001. Jeff Dominitz and Knowles (2005) consider tests of racial prejudice when officers are assumed to minimize crime.

<sup>13</sup> KPT recognized this issue in their footnote 16.

<sup>14</sup> However, Nicola Persico and Petra Todd (2005) 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.

<sup>15</sup> A refined version of this test uses regressions 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. Lorie A. Fridell (2004) provides a comprehensive review of different benchmarks in this approach.

<sup>16</sup> There are parallel and closely related approaches to test for disparate treatment in the literature on mortgage

marking test 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. The main drawback of the benchmark test is that it cannot determine if racial disparities arise out of racial prejudice or statistical discrimination. Furthermore, the benchmark test suffers from two main problems. 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. The second problem is the *omitted-variables* 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 are not available to researchers, benchmarking tests will not be completely informative about whether motorists' race affected the search decision.

## II. The Model

We now present a simple model of trooper search behavior that underlines the empirical work in Section IV.<sup>17</sup> There is a continuum of troopers (interchangeably, police officers) and motorists (interchangeably, drivers). Let  $r_m$  and  $r_p \in \{M, W\}$  denote the race of the motorists and the troopers, respectively, where  $M$  stands for minorities and  $W$  for whites.<sup>18</sup> Suppose that among motorists of race  $r_m \in \{M, W\}$ , a fraction  $\pi^{r_m} \in (0, 1)$  of them carry contraband.<sup>19</sup>

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 char-

acteristics 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.<sup>20</sup> 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*  $\theta \in [0, 1]$  which summarizes all of the information that these characteristics indicate about the likelihood that a driver may be carrying contraband.<sup>21</sup> 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  $r_m \in \{M, W\}$ ,

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

The MLRP property on the signal distributions essentially means that a higher index  $\theta$  is a signal that a driver is more likely to be guilty.<sup>22</sup> To the extent that there may be obvi-

lending discrimination (see Ross and Yinger, 2002; 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).

<sup>17</sup> Vani K. Borooah (2001) and David Bjerk (2004) develop somewhat related models of policing behavior.

<sup>18</sup> 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.

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

<sup>20</sup> The questions the trooper will ask the motorist are typically focused on where the motorist is headed and the purpose of the 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.

<sup>21</sup> It is useful to think that troopers aggregate their observed variables into the index  $\theta$  by assigning them different weights. The weight troopers assign to a particular variable, however, can be different for motorists of different races. For example, Hispanic motorists in Florida tend to have more limited English skills than whites. Thus the weight on English skills in the formation of  $\theta$  should differ for Hispanic and white drivers.

<sup>22</sup> 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.

ously guilty drivers (for example, if illicit drugs are in plain view), we assume that:

Unbounded Likelihood Ratio:  $f_g^{r_m}(\theta)/f_n^{r_m}(\theta) \rightarrow +\infty$  as  $\theta \rightarrow 1$ .

The MLRP also implies that the cumulative distribution function  $F_g^{r_m}(\cdot)$  first order stochastically dominates  $F_n^{r_m}(\cdot)$ , 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.<sup>23</sup>

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 said to be *racially*

*prejudiced* if he or she exhibits a preference for searching motorists of one race. Following KPT, we model this preference in the cost of searching motorists.<sup>24,25</sup>

**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 nonmonolithic police force does not necessarily imply that some racial group of troopers is 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.

### A. 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,

<sup>24</sup> 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.

<sup>25</sup> We are interpreting racial prejudice as "consequential animus" in the terminology of Ayres (2001, chap. 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 being willing to engage in transactions with them. We believe that "consequential animus" is an appropriate interpretation of racial prejudice in motor vehicle searches. We thank an anonymous referee for bringing this distinction to our attention.

<sup>23</sup> 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 IIB for more discussion on this key assumption.



$\Pr(G|r_m, \theta)$ , is obtained via Bayes's rule:

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

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

The decision problem faced by a police officer of race  $r_p$  when dealing with a motorist with race  $r_m$  and signal  $\theta$  is thus as follows:

$$(1) \quad \max\{\Pr(G|r_m, \theta) - t(r_m; r_p); 0\}$$

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:

**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 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

$$(2) \quad \gamma(r_m; r_p) = \pi^{r_m} [1 - F_g^{r_m}(\theta^*(r_m; r_p))] + (1 - \pi^{r_m}) [1 - F_n^{r_m}(\theta^*(r_m; r_p))].$$

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

$$(3) \quad S(r_m; r_p) = \frac{\pi^{r_m} [1 - F_g^{r_m}(\theta^*(r_m; r_p))]}{\left\{ \pi^{r_m} [1 - F_g^{r_m}(\theta^*(r_m; r_p))] + (1 - \pi^{r_m}) [1 - F_n^{r_m}(\theta^*(r_m; r_p))] \right\}}.$$

We say that race- $r_p$  police officers exhibit *statistical discrimination* if they have no taste for discrimination and yet they use different search criteria 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 nonprejudiced 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  $f_g^{r_m}(\theta)$  and  $f_n^{r_m}(\theta)$  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 pop-

ulation of minority motorists, then Proposition 1 implies that both races of officers will use the same search criterion against a given race of motorists,<sup>26</sup> so that  $\theta^*(W; W) = \theta^*(W; M)$  and  $\theta^*(M; W) = \theta^*(M; M)$ . Thus, following from the formula for the search rate (2) and average search success rate (3), 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 is 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.<sup>27</sup> 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  $t(M; M) = t(W; M) < t(M; W) = t(W; W)$ .<sup>28</sup> 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 (2) 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* 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 is 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  $\theta^*$ .<sup>29</sup> 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 discussion above 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  $S(r_m; W)$ .*<sup>30</sup>

<sup>29</sup> To 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, after some simplification,

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

From MLRP, we know that, for all  $\theta > \theta^*$ ,  $f_n^{r_m}(\theta)/f_n^{r_m}(\theta^*) > f_g^{r_m}(\theta)/f_g^{r_m}(\theta^*)$ , thus the integrand in the numerator is always positive. Hence,  $H'(\theta^*) > 0$ .

<sup>30</sup> The last statement in Proposition 3 holds, regardless of whether troopers are racially prejudiced.

<sup>26</sup> In Section IIIC we describe a resampling procedure to deal empirically with datasets in which this assumption may be invalid in the raw data.

<sup>27</sup> Consider, 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  $t(M; M) < t(M; W)$  and therefore  $t(r_m; M) < t(r_m; W)$  for all  $r_m$ . Similar arguments show that if  $t(W; M) > t(W; W)$ , we must have  $t(M; M) > t(M; W)$ ; and if  $t(W; M) = t(W; W)$ , 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  $r_m$ .

<sup>28</sup> Note 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.

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 their marginal search success rate against minority motorists. But because in our model the marginal motorist's guilt probability is smaller than that of the infra-marginal motorists, we cannot conclude that race- $r_p$  troopers' *average* search success rate against white motorists will be equal to their average search success rate 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.<sup>31</sup>

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

### B. 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 an 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 a very strong assumption and may be empirically implausible. Indeed, it is possible, for example, that minority drivers not carrying contraband might 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. We maintain, however, 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. 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 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  $f_g^{r_m}$  and  $f_n^{r_m}$  do not depend on the race of the police officers  $r_p$ .

*Assumptions on the Officers' Optimization Problem.*—We assume in officers' optimization problem (1) that they maximize the total number of convictions minus a cost of searching cars. We also assume that officers exploit all statistically valid racial inferences in making their search decisions. Our assumption that troopers will always use the race of motorists as a factor in deciding whether or not to search is not necessarily at odds with the official policies

<sup>31</sup> 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 with Ross and Yinger (2002, chap. 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).

on racial profiling. Most highway patrol departments prohibit using race as the primary cause for police-citizen contact, but do not rule out using it as one of many factors. For example, the California Highway Patrol prohibits racial profiling, which it defines as occurring “when a police officer initiates a traffic or investigative contact based primarily on the race/ethnicity of the individual.”<sup>32</sup> Federal courts have ruled that race cannot be the only basis for search and seizure, but it can be one among other factors (see, for example, *Whren v. United States* (1996) and *United States v. Waldon* (2000)).

Finally, in officers’ optimization problem (1), we assume that they do not have search capacity constraints and thus they judge each stopped vehicle individually to determine whether it is worth a search. But if officers did have a search capacity constraint, they would choose to search only the most suspicious motorists. In reality, however, capacity constraints are not likely to be important: in our data, an officer has, on average, fewer than seven searches in a span of almost two years. One might also think that an officer may care about the quality of the contraband found and that this should be reflected in his objective function, but unfortunately we do not have such information in our dataset and thus cannot include this in our model.<sup>33</sup>

*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 assume 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 and different times of the day (indeed, our raw data indicate that this is the case—see Tables 4 and 5 on pp. 144 and 145). In Section IIIC we describe a resampling procedure to deal with this problem empirically.

<sup>32</sup> See California Highway Patrol Public Contact Demographic Data Summary (p. 1).

<sup>33</sup> The Maryland dataset used by KPT does contain the quantity of drugs found in the searches (see Knowles et al., 2001).

### III. Empirical Tests

#### A. 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 who conduct the searches. That is, under the null hypothesis of monolithic trooper behavior, we must have, for all  $r_m \in \{M, W\}$ ,

$$(4) \quad \gamma(r_m; M) = \gamma(r_m; W),$$

$$(5) \quad S(r_m; M) = S(r_m; W).$$

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

It is worth pointing out that both equalities (4) and (5) 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 (4) and (5). Without loss of generality, suppose that troopers are not monolithic in their search behavior against white motorists ( $r_m = W$ ). That is,  $t(W; W) \neq t(W; M)$ . 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 29. Thus, the test using either (4) or (5) 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.<sup>34</sup>

### B. 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 has 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  $r_m$ .

*Power of the Test.*—Two features of our empirical test for racial prejudice are worth discussing in further detail. First, our test will indicate only whether or not there is a “relative bias” among troopers. This is because when we do find evidence of racial prejudice, we know that at least one racial group of officers is racially prejudiced, but we cannot determine which one. Second, the power of our test is not one, even when the sample size goes to infinity. To illustrate, suppose 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 is 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. Therefore, the null hypothesis is not rejected even if it is false, and as a result we commit a type-II error. In general, our test has an asymptotic power of zero if  $[t(W; W) - t(W; M)][t(M; W) - t(M; M)] > 0$ ,<sup>35</sup> and it 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. On the other hand, if we do find evidence against the null hypothesis, 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 the robust implication from a plausible behavioral model that does not impose strong and unverifiable distributional assumptions.<sup>36</sup>

Finally, it is worth pointing out that even though our test has low power, it 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 search rates, which indicates that at least one racial group of the officers is racially prejudiced.<sup>37</sup>

*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:  $\gamma(M; M) = 0.05$ ,  $\gamma(W; M) = 0.10$ ,  $\gamma(M; W) = 0.20$ , and  $\gamma(W; W) = 0.15$ . That is, minority officers are more likely to search white motorists than minority motorists, and white

<sup>34</sup> Of course, if the search rates between racial groups of troopers for a given race of motorists are equal, the average search success rates between racial groups of troopers for a given race of motorists must also be equal.

<sup>35</sup> This will be true either when  $t(W; W) > t(W; M)$  and  $t(M; W) > t(M; M)$ , or when  $t(W; W) < t(W; M)$  and  $t(M; W) < t(M; M)$ . That is, our test will fail to detect relative

racial prejudice if the troopers of some race have smaller (or larger) costs of searching drivers of any race than troopers of other races.

<sup>36</sup> In this regard, our position is similar to Charles F. Manski (1995) who preached the tolerance of ambiguity in empirical research in social sciences.

<sup>37</sup> Their paper did not present information about average search success rates which, as we remarked earlier, could have been used potentially to refute our model.

officers are more likely 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.<sup>38</sup> 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  $(\pi^M, f_g^M, f_n^M)$  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 nonmonolithically. 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 nonverifiable distributional assumptions on the signal distribution.

A second difference between Antonovics and Knight's test and ours is that we use both search rate and average search success rates in our test, while their test uses only search rates. Using both pieces of information permits us potentially to 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, chap. 8, for related discussion in the context of mortgage lending).

### C. A Resampling Procedure

As we mentioned in Section IIB, 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 on pp. 144 and 145, indicate that white and mi-

nority troopers are systematically assigned to patrol in different locations and at different times of the day, and thus might face different populations of motorists. We will not explain an empirical method that can possibly resolve this problem so that we can use our empirical test even when the raw data do not satisfy this condition. For illustration purposes, suppose that there are two troop stations, denoted 1 and 2, each with 100 officers. Suppose that in 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 reweighted 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 a reweighted sample of 107 white officers and 46 minority officers. Among the 153 officers in the artificial sample, (roughly) 70 percent of them are white 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 reweighted sample. To alleviate the sampling error, we use independent resampling to create a list of such reweighted data sets.<sup>39</sup>

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.

<sup>38</sup> When we ran Probit regressions as specified in Table 6 of Antonovics and Knight (2004) on our Florida dataset, the coefficient on the "racial mismatch" variable is positive and significant (the point estimate is about 0.1 with a robust standard error of 0.013). Thus their test, in contrast to ours, would have concluded racial prejudice.

<sup>39</sup> As pointed out by a referee, our resampling procedure is reminiscent of the "reweighting" method proposed in John DiNardo et al. (1996) in the context of decomposing the effects of institutional and labor market factors on changes in the U.S. wage distributions. The main difference from our procedure is that we use independent multiple resampling to alleviate sampling error.

## IV. Empirical Results

### A. Data Description

We now apply the tests described above to data from the Florida State Highway Patrol. The Florida data are composed of two parts. The first is the *traffic* dataset 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 dataset, 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).<sup>40</sup> The *unique* feature of our dataset is the second part, which is the *personnel* data which contains information on each of the troopers who conducted the stops and searches in the traffic dataset, including 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 datasets. The merged dataset thus provides information about the demographics of the trooper who made each stop and search. After eliminating cases in which there was missing information on the demographics of the trooper who conducted the stop, we end up with 906,339 stops and 8,976 searches conducted by a total of 1,469 troopers.<sup>41</sup> 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.

### B. 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

<sup>40</sup> 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.

<sup>41</sup> We also eliminated cases where the race of the motorist and trooper was neither white, black, nor Hispanic, since there are not enough observations of the other racial groups to consider them.

were carried out against white motorists, 17.3 percent against Hispanic motorists, and 16.2 percent against black motorists. In all race categories of motorists, male motorists account for at least 67 percent of the stopped motorists. Among all the motorists who 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 had in-state license plates, and close to 70 percent of the stops were conducted in the daytime (defined to be between 6 a.m. and 6 p.m.).

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 on black motorists. 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 who 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). Of the total, 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, Hispanics, 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.

### C. 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

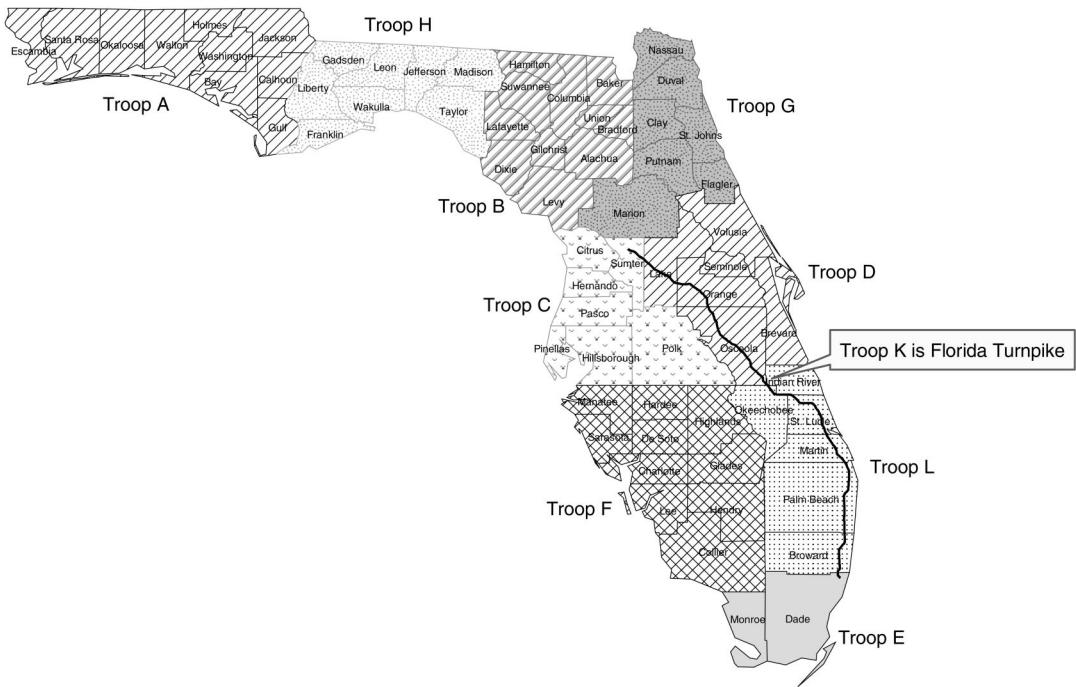


FIGURE 1. TROOP STATION COVERAGE MAP

whether a crucial assumption of our test, that all troopers face the same population of motorists, is satisfied in the raw data (before resampling). This assumption, of course, is not directly testable, because  $\pi^{r,m}$ ,  $f_g^{r,m}(\theta)$ ,  $f_n^{r,m}(\theta)$ , and  $\theta$  are all unobservable. The best we can do is 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 racial groups of 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 overrepresentation 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 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 Section IIB.<sup>42</sup> By construction, in the reweighted data we created with the resampling method, troopers of a given race are assigned to different troops with the same probabilities. The Pearson  $\chi^2$  test also

<sup>42</sup> 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 a similar population of drivers, and the differences in the stopped motorists' characteristics reflect the differences in stop behavior of the troopers of different races, rather than the differences in the driver population. It is plausible, but in this paper we take the stopped motorists population as given.



TABLE 2—MEANS OF VARIABLES RELATED TO MOTORISTS

Motorists' characteristics	Stops			Searches		
	All stops	By motorist sex		All searches	By motorist sex	
		Female	Male		Female	Male
Black	0.162 (0.368)	0.327 (0.470)	0.673 (0.470)	0.221 (0.415)	0.146 (0.354)	0.851 (0.354)
Hispanic	0.173 (0.378)	0.225 (0.417)	0.775 (0.471)	0.234 (0.423)	0.098 (0.296)	0.902 (0.296)
White	0.665 (0.472)	0.319 (0.466)	0.681 (0.466)	0.546 (0.498)	0.178 (0.382)	0.822 (0.382)
Female	0.304 (0.460)	1.00 (0.00)	0.00 (0.00)	0.152 (0.359)	1.00 (0.00)	0.00 (0.00)
Male	0.696 (0.460)	0.00 (0.00)	1.00 (0.00)	0.848 (0.359)	0.00 (0.00)	1.00 (0.00)
Age						
16–30	0.481 (0.500)	0.325 (0.468)	0.675 (0.468)	0.584 (0.493)	0.149 (0.356)	0.851 (0.356)
31–45	0.336 (0.472)	0.295 (0.456)	0.705 (0.456)	0.317 (0.465)	0.162 (0.368)	0.838 (0.368)
46+	0.183 (0.386)	0.269 (0.444)	0.731 (0.444)	0.099 (0.299)	0.136 (0.343)	0.864 (0.343)
License plate						
In-state	0.899 (0.302)	0.310 (0.462)	0.690 (0.462)	0.857 (0.350)	0.155 (0.362)	0.845 (0.362)
Out-of-state	0.101 (0.302)	0.252 (0.434)	0.748 (0.434)	0.143 (0.350)	0.132 (0.338)	0.868 (0.338)
Time						
Day (6am–6pm)	0.697 (0.459)	0.316 (0.465)	0.684 (0.465)	0.475 (0.499)	0.161 (0.367)	0.839 (0.367)
Night	0.303 (0.459)	0.275 (0.447)	0.725 (0.447)	0.525 (0.499)	0.144 (0.351)	0.856 (0.351)
Contraband seized						
None				0.792 (0.406)	0.155 (0.362)	0.845 (0.362)
Drugs				0.151 (0.358)	0.137 (0.344)	0.863 (0.344)
Paraphernalia				0.015 (0.122)	0.156 (0.364)	0.844 (0.364)
Currency				0.003 (0.051)	0.174 (0.388)	0.826 (0.388)
Vehicles				0.010 (0.100)	0.154 (0.363)	0.846 (0.363)
Alcohol/tobacco				0.021 (0.142)	0.151 (0.359)	0.849 (0.359)
Weapons				0.006 (0.078)	0.055 (0.229)	0.945 (0.229)
Other				0.003 (0.049)	0.318 (0.477)	0.682 (0.477)
Number of observations	906,339	275,527	630,812	8,976	1,364	7,612

Note: Standard errors of the means are shown in parentheses.

reveals that in the reweighted sample, troopers of 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 reweighted sample. We report our test results below using data from the reweighted samples.

D. Empirical Result for the Test of Monolithic Trooper 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.

We now implement our test for determining whether troopers of different races exhibit mono-

lithic behavior in their search decisions. Recall that we said that if troopers are monolithic they will all search a given race of motorists at the same rate. Thus, we need to test whether the search rates that are given in panel A of Table 1 differ among trooper racial groups for a given group of motorists. To accomplish this, we compute the  $p$ -values from the Pearson  $\chi^2$  test under the null hypothesis that troopers of all races search race- $r_m$  motorists with equal probability. These  $p$ -values are shown in Table 1. Specifically, the Pearson  $\chi^2$  test statistic, under the null hypothesis that all troopers with race in  $\mathcal{R}$  search race- $r_m$  motorists with equal probability, is given by

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

where  $\widehat{\gamma}(r_m; r_p)$  is the estimated search probability of race- $r_p$  officers against race- $r_m$  motor-

TABLE 3—MEANS OF VARIABLES RELATED TO TROOPERS

Troopers' characteristics	Troopers		Stops		Searches		
	All troopers	All stops	By trooper sex		All searches	By trooper sex	
			Female	Male		Female	Male
Black	0.137 (0.344)	0.160 (0.366)	0.115 (0.319)	0.885 (0.319)	0.046 (0.208)	0.044 (0.206)	0.956 (0.206)
Hispanic	0.100 (0.300)	0.114 (0.318)	0.070 (0.256)	0.930 (0.256)	0.095 (0.293)	0.025 (0.155)	0.975 (0.155)
White	0.763 (0.425)	0.726 (0.446)	0.092 (0.289)	0.908 (0.289)	0.859 (0.348)	0.076 (0.265)	0.924 (0.265)
Female	0.106 (0.307)	0.093 (0.291)	1.00 (0.00)	0.00 (0.00)	0.069 (0.254)	1.00 (0.00)	0.00 (0.00)
Male	0.894 (0.307)	0.907 (0.291)	0.00 (0.00)	1.00 (0.00)	0.931 (0.254)	0.00 (0.00)	1.00 (0.00)
Rank							
Captain	0.022 (0.148)	0.002 (0.041)	0.239 (0.426)	0.761 (0.426)	0.002 (0.046)	0.474 (0.513)	0.526 (0.513)
Lieutenant	0.070 (0.255)	0.013 (0.112)	0.023 (0.151)	0.977 (0.151)	0.007 (0.081)	0.000 (0.000)	1.000 (0.000)
Sergeant	0.145 (0.352)	0.062 (0.241)	0.054 (0.226)	0.946 (0.26)	0.053 (0.224)	0.052 (0.223)	0.948 (0.223)
Corporal	0.147 (0.354)	0.112 (0.316)	0.068 (0.252)	0.932 (0.252)	0.071 (0.257)	0.030 (0.170)	0.970 (0.170)
LEO	0.602 (0.490)	0.810 (0.392)	0.101 (0.301)	0.899 (0.301)	0.866 (0.341)	0.073 (0.261)	0.927 (0.261)

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

Motorists' race	Motorists' characteristics	White troopers	Black troopers	Hispanic troopers	p-value
White	Male	0.679	0.684	0.701	<0.001
	Night stops	0.288	0.272	0.318	<0.001
	Age: 16–30	0.471	0.460	0.445	<0.001
	Age: 31–45	0.325	0.341	0.349	0.02
Black	Male	0.671	0.667	0.686	<0.001
	Night stops	0.332	0.308	0.354	<0.001
	Age: 16–30	0.514	0.514	0.507	0.001
	Age: 31–45	0.340	0.344	0.356	0.03
Hispanic	Male	0.783	0.774	0.761	<0.001
	Night stops	0.322	0.288	0.393	<0.001
	Age: 16–30	0.516	0.497	0.494	<0.001
	Age: 31–45	0.350	0.363	0.355	0.01

ists,  $\widehat{\gamma}(r_m)$  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  $\mathcal{R}$ . 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. Because all the  $p$ -values are less than 0.001,

this provides 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. To test for monolithic behavior,

TABLE 5—PROPORTION OF TROOPERS WITH DIFFERENT RACES BY TROOP AND TIME ASSIGNMENT IN THE RAW DATA

Troop	Troopers' race		
	White	Black	Hispanic
A	0.930 (0.256)	0.054 (0.227)	0.016 (0.124)
B	0.889 (0.316)	0.081 (0.274)	0.030 (0.172)
C	0.816 (0.389)	0.116 (0.321)	0.068 (0.253)
D	0.793 (0.406)	0.117 (0.322)	0.090 (0.287)
E	0.412 (0.494)	0.236 (0.426)	0.352 (0.479)
F	0.880 (0.326)	0.056 (0.231)	0.063 (0.245)
G	0.833 (0.374)	0.135 (0.343)	0.032 (0.176)
H	0.886 (0.320)	0.114 (0.320)	00.00 (0.00)
K	0.698 (0.461)	0.147 (0.355)	0.155 (0.364)
L	0.603 (0.491)	0.298 (0.459)	0.099 (0.300)
% Night stops	0.283 (0.172)	0.284 (0.192)	0.349 (0.179)

Note: Standard errors of the means are shown in parentheses.

we need to see whether all racial groups of troopers have the same average search success rate against a given racial group of motorists. The  $p$ -value in each row is from the Pearson  $\chi^2$  test under the null hypothesis that troopers of all races have the same average search success rate against motorists of the race in that specific row. Again, the Pearson  $\chi^2$  test statistic, under the null hypothesis that all troopers with race in  $\mathcal{R}$  have the same average search success rate against race- $r_m$  motorists, is given by

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

where  $\widehat{S}(r_m; r_p)$  is the estimated average search success rate of race- $r_p$  officers against race- $r_m$  motorists, and  $\widehat{S}(r_m)$  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 0.001, which again provides 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 motor-

ists 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 percent versus 0.28 percent), but their average search success rate against white motorists is much higher than that for Hispanic motorists (39.4 percent versus 21.0 percent).

### E. Empirical Result for the Test of Racial Prejudice

Thus far, we have 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 Section IIIB. Under the null hypothesis that none of the racial groups of troopers is 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 know that, for all

$$r_m \in \{W, B, H\}, \widehat{\gamma}(r_m; W) > \widehat{\gamma}(r_m; H) > \widehat{\gamma}(r_m; B), \text{ and } \widehat{S}(r_m; W) < \widehat{S}(r_m; H) < \widehat{S}(r_m; B).$$

We can use simple Z-statistic to formally test that

$$(6) \quad \gamma(r_m; W) > \gamma(r_m; H) > \gamma(r_m; B),$$

$$(7) \quad S(r_m; W) < S(r_m; H) < S(r_m; B).$$

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

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

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 (6).

Analogously we can formally test inequality (7) by using a  $Z$ -test

$$(8) \quad Z' = \frac{\widehat{S(r_m; W)} - \widehat{S(r_m; H)}}{\sqrt{\frac{\text{SVar}'_W}{n'_W} + \frac{\text{SVar}'_H}{n'_H}}} \sim N(0, 1)$$

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 (7).

To summarize, we cannot reject the null hypothesis that troopers do not exhibit relative racial prejudice. 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.

#### F. Other Empirical Implications

It is interesting to note some additional implications from the tests we conducted. First of all, inequality (6) 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 Section IIIA, 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.

#### G. Replicating KPT's Test

It is useful to contrast our findings with those of 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. The 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 the Pearson  $\chi^2$  test on the hypothesis that the 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 Hispanics. As we argued, however, this conclusion is valid only if their model of motorist and trooper behavior is true. We would like to emphasize, though, that our test does not necessarily refute the presence of racial prejudice. Our results are simply that, without strong (and possibly untenable) assumptions, we cannot confidently prove the presence of relative, let alone absolute, racial prejudice.

TABLE 6— $p$ -VALUES FROM PEARSON'S  $\chi^2$  TESTS ON THE HYPOTHESIS THAT AVERAGE SEARCH SUCCESS RATES ARE EQUAL ACROSS VARIOUS GROUPINGS: KPT TEST

Groupings	Average search success rate
White, Black, Hispanic	<0.001
White, Black	<0.001
White, Hispanic	<0.001
Black, Hispanic	<0.001

## V. 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 et al. (2001), our model allows troopers of different races to behave differently, thus enabling us to examine nonmonolithic 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 datasets where one of the major assumptions underlying our model and empirical tests is violated. Our tests require datasets with race information about both the motorists and troopers. When applied to vehicle stop-and-search data from Florida, our tests soundly reject the hypothesis that troopers of different races are monolithic in their search behavior, but fail to reject the hypothesis that troopers of different races do not exhibit relative racial prejudice. 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 that a type-II error will occur, the positive side of this is that when we do find evidence of racial prejudice, it is rather conclusive.

Our paper focuses solely 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 only on determining racial prejudice in troopers' search decisions. Given data limitations, examining the possibility of racial prejudice in highway stops is beyond the scope of this paper. It is possible, however, that the racial prejudice of police officers is 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 datasets that include random samples of drivers on the road become available.

Finally, our proposed test of racial prejudice in motor vehicle searches in the presence of infra-marginality and omitted variable problems may also be applicable to detect racial prejudice in mortgage lending. For example, the analog of troopers of different races in the context of mortgage lending can be banks operating in the same metropolitan area, the analog of search rates by driver races can be mortgage denial rates by applicant races, and the analog of search success rates can be mortgage default rates. The test developed in this paper suggests that the comparison of mortgage denial rates and default rates for banks operating in the same metropolitan area can potentially reveal (relative) racial prejudice of the banks.

#### APPENDIX: A MODEL WITH ENDOGENOUS DRUG CARRYING DECISIONS

In Section II, we assumed that the proportion of motorists in race group  $r_m$  is exogenously given as  $\pi^m \in (0, 1)$ . For the purpose of testing for monolithic behavior and racial prejudice, this partial equilibrium approach suffices. For other purposes, however, 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.<sup>43</sup> 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 II 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  $r'_p$ .<sup>44</sup> Suppose that in the trooper population a fraction  $\alpha$  is of race  $r_p$  and the remainder fraction  $1 - \alpha$  is of race  $r'_p$ . Suppose that Nature draws for each

<sup>43</sup> See Persico (2002) for an analysis on how racially blind search policies may affect the total crimes committed by motorists.

<sup>44</sup> Because we are considering only one race group of motorists, we will omit  $r_m$  from the subsequent notation. Having more than one racial group of motorists will not change any of the results below.

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 II, 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

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

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

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

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

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

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

$$\begin{aligned} & \overbrace{[\alpha F_g(\theta_p^*) + (1 - \alpha)F_g(\theta_{p'}^*)]b}^{\text{Term 1}} \\ & - \overbrace{\{\alpha[1 - F_g(\theta_p^*)] + (1 - \alpha)[1 - F_g(\theta_{p'}^*)]\}c_g}^{\text{Term 2}} - v \end{aligned}$$

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:

$$-\{\alpha[1 - F_n(\theta_p^*)] + (1 - \alpha)[1 - F_n(\theta_{p'}^*)]\}c_n.$$

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

$$\begin{aligned} \text{(A1)} \quad v^*(\theta_p^*, \theta_{p'}^*) &= [\alpha F_g(\theta_p^*) \\ &+ (1 - \alpha)F_g(\theta_{p'}^*)] \\ &\times b - \{\alpha[1 - F_g(\theta_p^*)] \\ &+ (1 - \alpha)[1 - F_g(\theta_{p'}^*)]\}c_g \\ &+ \{\alpha[1 - F_n(\theta_p^*)] \\ &+ (1 - \alpha)[1 - F_n(\theta_{p'}^*)]\}c_n. \end{aligned}$$

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  $G(v^*(\theta_p^*, \theta_{p'}^*))$ .

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

$$\text{(A2)} \quad \Pr(G|\theta_p^*, \pi) = t_p$$

$$\text{(A3)} \quad \Pr(G|\theta_{p'}^*, \pi) = t_{p'}$$

$$\text{(A4)} \quad G(v^*(\theta_p^*, \theta_{p'}^*)) = \pi.$$

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 *unique*. Suppose that there are two equilibria in which the proportion of guilty motorists are  $\pi$  and  $\tilde{\pi}$  with  $\pi > \tilde{\pi}$ . Observe from (A1) that  $v^*(0, 0) = c_n - c_g < 0$  and

$$\frac{\partial v^*(\theta_p^*, \theta_{p'}^*)}{\partial \theta_p^*} =$$

$$\alpha c_n f_n(\theta_p^*) \left[ \frac{f_g(\theta_p^*)}{f_n(\theta_p^*)} \frac{b + c_g}{c_n} - 1 \right],$$

$$\frac{\partial v^*(\theta_p^*, \theta_{p'}^*)}{\partial \theta_{p'}^*} =$$

$$\alpha c_n f_n(\theta_{p'}^*) \left[ \frac{f_g(\theta_{p'}^*)}{f_n(\theta_{p'}^*)} \frac{b + c_g}{c_n} - 1 \right].$$

By the MLRP, we know that there exists  $(\hat{\theta}_p^*, \hat{\theta}_{p'}^*) \in [0, 1]^2$  such that  $v^*(\theta_p^*, \theta_{p'}^*)$  is strictly increasing in both  $\theta_p^*$  and  $\theta_{p'}^*$ , when  $(\theta_p^*, \theta_{p'}^*) > (\hat{\theta}_p^*, \hat{\theta}_{p'}^*)$ . Since  $v^*(0, 0) < 0$  and the support of  $G$  is nonnegative, we have  $G(v^*(0, 0)) = 0$ . Moreover,  $G(v^*(\theta_p^*, \theta_{p'}^*))$  will be zero for all  $(\theta_p^*, \theta_{p'}^*) \leq (\hat{\theta}_p^*, \hat{\theta}_{p'}^*)$ . Thus, any  $(\theta_p^*, \theta_{p'}^*) \leq (\hat{\theta}_p^*, \hat{\theta}_{p'}^*)$  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  $(\theta_p^*, \theta_{p'}^*) > (\hat{\theta}_p^*, \hat{\theta}_{p'}^*)$  and  $(\tilde{\theta}_p^*, \tilde{\theta}_{p'}^*) > (\hat{\theta}_p^*, \hat{\theta}_{p'}^*)$ . That is, both equilibria lie in the region where  $v^*(\cdot, \cdot)$  is strictly increasing in both arguments. If  $\pi > \tilde{\pi}$ , 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

- Antonovics, Kate L. and Knight, Brian G.** "A New Look at Racial Profiling: Evidence from the Boston Police Department." National Bureau of Economic Research, Inc., NBER Working Papers: No. 10634, 2004.
- Arrow, Kenneth J.** "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in labor markets*. Princeton: Princeton University Press, 1973, pp. 3–33.
- Ayres, Ian.** *Pervasive prejudice? Unconventional evidence of race and gender discrimination*. Chicago: University of Chicago Press, 2001.
- Ayres, Ian.** "Outcome Tests of Racial Disparities in Police Practices." *Justice Research and Policy*, 2002, 4(Special Issue), pp. 131–42.
- Becker, Gary S.** *The economics of discrimination*. Chicago: University of Chicago Press, 1957.
- Becker, Gary S.** "The Evidence against Blacks Doesn't Prove Bias." *Business Week*, April 19, 1993a. <http://bwarchive.businessweek.com/index.jsp>.
- Becker, Gary S.** "Nobel Lecture: The Economic Way of Looking at Behavior." *Journal of Political Economy*, 1993b, 101(3), pp. 385–409.
- Bjerk, David.** "Racial Profiling, Statistical Discrimination, and the Effect of a Colorblind Policy on the Crime Rate." McMaster University, Department of Economics Working Papers: No. 2004-11, 2004.
- Borooah, Vani K.** "Racial Bias in Police Stops and Searches: An Economic Analysis." *European Journal of Political Economy*, 2001, 17(1), pp. 17–37.
- Coate, Stephen and Loury, Glenn C.** "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review*, 1993, 83(5), pp. 1220–40.
- Dharmapala, Dhammika and Ross, Stephen L.** "Racial Bias in Motor Vehicle Searches: Additional Theory and Evidence." *Contributions to Economic Analysis & Policy*, 2004, 3(1), pp. 1–21.
- DiNardo, John; Fortin, Nicole M. and Lemieux, Thomas.** "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semi-parametric Approach." *Econometrica*, 1996, 64(5), pp. 1001–44.
- Dominitz, Jeff and Knowles, John.** "Crime Minimization and Racial Bias: What Can We Learn from Police Search Data?" University of Pennsylvania, Penn Institute for Economic Research Working Paper: No. 05-019, 2005.
- Donohue, John J., III and Levitt, Steven D.** "The Impact of Race on Policing and Arrests." *Journal of Law and Economics*, 2001, 44(2), pp. 367–94.
- Engel, Robin S.; Calnon, Jennifer M. and Bernard, Thomas J.** "Theory and Racial Profiling: Shortcomings and Future Directions in Research." *Justice Quarterly*, 2002, 19(2), pp. 249–73.
- Fridell, Lorie A.** *By the numbers: A guide for analyzing race data from vehicle stops*. Washington, DC: Police Executive Research Forum, 2004.
- Gross, Samuel R. and Barnes, Katherine Y.** "Road Work: Racial Profiling and Drug Interdiction on the Highway." *Michigan Law Review*, 2002, 101(3), pp. 653–754.
- Hernandez-Murillo, Ruben and Knowles, John.** "Racial Profiling or Racist Policing? Bounds Tests in Aggregate Data." *International Economic Review*, 2004, 45(3), pp. 959–89.
- Knowles, John; Persico, Nicola and Todd, Petra.** "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy*, 2001, 109(1), pp. 203–29.
- Manski, Charles F.** *Identification problems in the social sciences*. Cambridge, MA: Harvard University Press, 1995.
- National Research Council of the National Academies, Committee on National Statistics.** *Measuring racial discrimination: Panel on methods for assessing discrimination*. Washington, DC: National Academies Press, 2004.
- Persico, Nicola.** "Racial Profiling, Fairness, and Effectiveness of Policing." *American Economic Review*, 2002, 92(5), pp. 1472–97.
- Persico, Nicola G. and Todd, Petra.** "Using Hit Rates to Test for Racial Bias in Law Enforcement: Vehicle Searches in Wichita." University of Pennsylvania, Penn Institute for Economic Research Working Paper: No. 05-004, 2005.



- Ramirez, Deborah; McDevitt, Jack and Farrell, Amy.** *A resource guide on racial profiling data collection systems; Promising practices and lessons learned.* Washington, DC: U.S. Department of Justice, 2000.
- Riksheim, Eric C. and Chermak, Steven M.** "Causes of Police Behavior Revisited." *Journal of Criminal Justice*, 1993, 21(4), pp. 353–82.
- Ross, Stephen L.** "Mortgage Lending Discrimination and Racial Differences in Loan Default: A Simulation Approach." *Journal of Housing Research*, 1997, 8(2), pp. 277–97.
- Ross, Stephen L.** "What Is Known about Testing for Discrimination: Lessons Learned by Comparing across Different Markets." University of Connecticut, Department of Economics Working Papers: No. 2003-21, 2003.
- Ross, Stephen L. and Yinger, John.** "The Default Approach to Studying Mortgage Discrimination: A Rebuttal," in Margery A. Turner and Felicity Skidmore, eds., *Mortgage lending discrimination: A review of existing evidence.* Washington, DC: Urban Institute, 1999, chap. 5.
- Ross, Stephen and Yinger, John.** *The color of credit: Mortgage discrimination, research methodology, and fair-lending enforcement.* Cambridge, MA: MIT Press, 2002.
- Sherman, Lawrence W.** "Causes of Police Behavior: The Current State of Quantitative Research." *Journal of Research in Crime and Delinquency*, 1980, 17(1), pp. 69–100.