

Racial Bias in Motor Vehicle Searches: Theory and Evidence

John Knowles, Nicola Persico and Petra Todd¹

CARESS WORKING PAPER #99-06

August 19, 1999

¹Knowles, Persico and Todd are Assistant Professors in the Economics Department of the University of Pennsylvania. We thank George Mailath and Ken Wolpin for helpful discussions. We thank the Maryland ACLU for providing us with data and information. Please direct correspondence to any of the authors at 3718 Locust Walk, Philadelphia, PA 19104 or to jknowles@econ.sas.upenn.edu, persico@econ.sas.upenn.edu, or petra@athena.sas.upenn.edu.

Abstract

African American motorists in the United States are more likely than white motorists to have their cars searched by police checking for illegal drugs and other contraband. The courts are faced with the task of deciding on the basis of traffic-stop data whether police are basing their decisions to stop cars on the race of the driver. We develop a model of law enforcement for a population with two racial types who also differ along other dimensions relevant to criminal behavior. We discuss why a simple test commonly applied by the courts is inadequate when the econometrician observes only a subset of the characteristics observed by the policemen. Next, we show how to construct a test for whether differential treatment is motivated purely out of efficiency grounds, i.e. to maximize the number of arrests, or reflects racial prejudice. The test is valid even when the set of characteristics observed by the policemen are only partially observable by the econometrician. We apply the tests for discrimination to traffic stop data from Maryland. Finally, we present a simple analysis of the tradeoff between efficiency and fairness.

1 Introduction

African American motorists in the United States are more likely than white motorists to have their cars searched by police checking for illegal drugs and other contraband. In the state of Maryland, for example, 80% of motorists searched by state police on the I-95 highway between January 1995 and September 1996 were African-American, although only 18% of motorists on this road were African American.¹ While it is conceivable that African American motorists are more likely to commit the types of traffic offenses that police use as pretexts for vehicle checks, traffic studies and police testimony suggest that blacks and whites are not distinguishable by their driving habits. An alternative explanation for the racial disparity in traffic stops is racial discrimination: police officers make their decision to stop cars partly on the basis of race. This explanation, known as “racial profiling,” is the basis of several recent lawsuits against state governments. The issue has also attracted attention in political spheres, forcing the resignation of the New Jersey chief of police and provoking the U.S. President to describe racial profiling as a “morally indefensible, deeply corrosive practice.”²

Evidence of racial profiling is often interpreted as an indication of racist preferences on the part of the police. The task of deciding whether racism is a factor in police traffic stops falls on the courts, which consider various pieces of statistical evidence in reaching a decision. The case for discrimination rests largely on the observation that the proportion of African-Americans among the drivers stopped by police far exceeds the proportion in the general population of drivers. This simple comparison is the basis of expert witness testimonies in various legal cases.³ A refined version of the test performs

¹This information comes from Lamberth (1996), a study commissioned by the Maryland ACLU.

²“Clinton Order Targets Racial Profiling”, Associated Press, June 9, 1999.

³In the 1993 case of *Maryland v. Wilkins*, a statistician testified that “The disparities

the same comparison after conditioning on other observable characteristics of drivers thought to be related to criminal propensities. If race has no additional explanatory power in a regression after introducing the conditioning variables, then this is taken as evidence of no discrimination.⁴

The drawback to this approach to testing for discrimination is that it requires the analyst to have data on the full set of characteristics that a policeman uses in deciding whether to stop a motorist.⁵ If some characteristics are unobserved, then race may have explanatory power due to omitted variable bias. Another drawback of the test is that even if race has no explanatory power, there is still the possibility that police target individuals with certain characteristics because those characteristics are correlated with race and not because they are good predictors of criminality. Thus, the validity of the test hinges crucially on judgements about what constitutes a set of admissible conditioning variables and on whether the analyst has access to the full set of variables.

Even if regression-based tests conclude in favor of discrimination, they are not informative about the motivation for discrimination. Police may use race as a criterion in traffic stops for efficiency reasons (because they are trying to maximize arrests and race helps predict criminality) or because they prefer stopping one racial group over another. We call discrimination for efficiency reasons *statistical discrimination*, using the terminology of Arrow (1973). In contrast, we say that an officer is *prejudiced* (or racist) if, *cæteris paribus*, he has a preference for searching motorists of a particular race. We model are sufficiently great that, taken as a whole, they are consistent with and strongly support the assertion that the Maryland State Police are targeting the community of black motorists for stop, detention and investigation...".

⁴See Donohue (1999).

⁵A training manual issued by the Illinois State Police highlights some indicators of criminal activity. These include tinted windows, religious paraphernalia used to divert suspicion, and attorney's business cards. We report the complete list in the Appendix.

prejudice as a taste for discrimination, following Becker (1957). Prejudice is a property of the officer's utility function, while statistical discrimination is a property of equilibrium. An equilibrium exhibits statistical discrimination if, conditional on all other observables available to non-racist policemen, one race is searched more often than another.

In this paper, we propose a test for distinguishing between statistical discrimination and racist preferences, which is derived from a simple model of law enforcement via police searches. A key advantage of the test is that it is feasible even when only a subset of the variables used by the policeman in deciding whether to search a motorist are available. In fact, while more variables allow for a more powerful test, the test can be carried out when race is the only characteristic observed.

Our model belongs to the literature on optimal auditing. Early models, such as Becker (1968) and Stigler (1970) examined citizens' incentives to misbehave under an exogenous probability of being audited. More recent literature, mainly dealing with income reporting and tax evasion, assumes that both parties, the auditing and the auditee, behave strategically (see Reinganum and Wilde (1986), Border and Sobel (1987), and Scotchmer (1987)). To the best of our knowledge, ours is the first paper that attempts an empirical test of an optimal auditing model.

Our model assumes that the police force maximizes the number of arrests net of costs of searching motorists. Police take into account observable information about the driver that may or may not include race, while motorists take into account the probability of being searched in deciding whether to carry contraband in their car. Prejudice is introduced into the model as a difference in the perceived cost of searching different drivers. In equilibrium, police are most likely to search those motorists whose visible characteristics are most closely linked to crime, and the minimum level of suspicion that

results in a search may vary across racial groups. However, a key implication of the equilibrium model is that if a police officer has the same tastes for searching two subgroups of the population, such as two different racial groups, then the probability that a stopped driver is carrying contraband should be identical across those subgroups.

We apply this test to data on drug-related highway searches gathered by the Maryland State Police on the Maryland stretch of the I-95 highway. In our data, cars of African American motorists are searched much more frequently than those of white motorists. However, we cannot reject the hypothesis that this is due purely to efficiency considerations, i.e. statistical discrimination.⁶ In other words, it is possible to explain the black-white racial disparities observed in the Maryland traffic stop data without recourse to racist preferences. Of course, efficiency is not the only relevant criterion in evaluating economic outcomes. At the end of our paper, we consider the costs imposed by statistical discrimination and discuss the trade-offs between racial fairness and effectiveness of drug interdiction.

2 The Model

This section describes the model of police and motorist behavior that generates the implications that we will later test. Our model assumes that there are a continuum of policemen and motorists. Let $r \in \{A, W\}$ denote the race of the motorist, which we assume is observable by the policeman. Let c denote all observable characteristics other than race that are potentially used by the policeman in searching cars. The variable c may be unobserved or only partially observed by the econometrician. Although c is a multidimensional

⁶Although limitations in our data do not allow us to carry out a test of whether statistical discrimination is present, we are currently attempting to address this issue by enriching our data set.

variable, for expositional ease we treat it as a one dimensional variable. All the results in this section extend straightforwardly to the multidimensional case. Let $F(c|W)$ and $F(c|A)$ denotes the distribution of c in the white and African American populations, respectively.

Policemen search motorists, and we assume that each policeman can choose from an infinite supply of motorists of any type c, r . The policeman maximizes the total number of convictions minus a cost of searching cars. The marginal cost of searching a motorist of race r is t_r . We assume $t_W, t_A \in (0, 1)$. Let G denote the event where the motorist searched is found guilty. (In our data, G corresponds to being found with drugs in the car).

We assume that motorists consider the probability of being searched in deciding whether to carry contraband.⁷ If they do not carry contraband their payoff is zero whether or not the car is searched. Suppose they carry contraband; if they are searched their payoff is $-j(c, r)$, while if they are not searched the payoff is $u(c, r)$. We can interpret $u(c, r)$ as the expected propensity to carry drugs, and $j(c, r)$ as the expected cost of being convicted for a motorist with characteristics c and race r . We assume that $j(c, r), u(c, r) > 0$.

Denote by $\gamma(c, r)$ the probability that the policeman searches a motorist of type c, r . The expected payoff to a motorist of type c, r from carrying contraband is

$$\gamma(c, r)[-j(c, r)] + [1 - \gamma(c, r)]u(c, r). \quad (1)$$

The motorist decides to carry contraband if this expression is greater than zero. When the expression is zero the motorist is willing to randomize between carrying and not carrying. We denote the probability that a motorist of type c, r carries contraband by $P(G|c, r)$.

⁷An alternative assumption is that motorists do not react to the probability of being searched. In the Appendix we present that model, and present a testable implication allowing to distinguish between the two models.

The policeman's problem is to choose the probability of searching each motorist as a function of the observable characteristics, $\gamma(c, r)$. The policeman solves

$$\max_{\gamma(c,W), \gamma(c,A)} \sum_{r=W,A} \int [P(G|c, r) - t_r] \gamma(c, r) f(c|r) dc.$$

If $P(G|c, r) - t_r > 0$ then optimizing behavior implies $\gamma(c, r) = 1$, i.e. always search motorists of type c and r . If $P(G|c, r) = t_r$ then the policeman is willing to randomize over whether or not to search type c, r .⁸

Next, we introduce two definitions pertaining to racial bias in motor vehicle searchers. First, a policeman is defined to be racist if his utility function exhibits a preference for stopping one race. In our model, this is captured as a difference in the cost of searching motorists.

Definition 1 *We say that the policeman is prejudiced, or has a taste for discrimination, if $t_A \neq t_W$.*

Next, we say that an equilibrium exhibits statistical discrimination if, conditional on all characteristics observable by policemen with no tastes for discrimination, the probability of searching differs by race.

Definition 2 *We say that the equilibrium exhibits statistical discrimination if $t_A = t_W$ and $\gamma(c, W) \neq \gamma(c, A)$ for some c .*

It is important to observe that, while racism is defined as a characteristic of the policeman, statistical discrimination is a property of equilibrium that does not depend in any way on the policeman's characteristics. Also, the definition of statistical discrimination depends on what is observable. As we show below, the equilibrium may exhibit statistical discrimination if the

⁸Our model does not allow for police to have heterogeneous tastes for discrimination depending on the race of the policeman or for the possibility of false arrests. Donohue and Levitt (1998) incorporate such features into their model in a different context.

things on which the policeman can condition his search decision are not sufficiently detailed.

2.1 Equilibrium

We show that there is an equilibrium where motorists randomize over whether to carry contraband and policemen randomize over whether to search them.⁹

For a motorist of type c, r to be willing to randomize it must be that expression (1) equals zero.¹⁰ Solving for γ yields $\gamma(c, r) = u(c, r) / [u(c, r) + j(c, r)]$. This determines the policeman's searching intensity. Notice that $\gamma(c, r) \in (0, 1)$, so at equilibrium the policeman randomizes over whether to search each type c, r .

For a policeman to be willing to randomize it must be that $P(G|c, r) = t_r$ for all c, r .

⁹At the equilibrium, both the policemen and the motorists randomize. This may look unrealistic, since it requires agents to be indifferent across actions. There is a simple interpretation of these mixed strategies that does not require agents to actually flip coins. This interpretation goes under the name of "purification" of mixed strategies, and is presented in the Appendix.

¹⁰In our model, the probability that a motorist carries contraband is never one. The reason that this probability can never be one for any group c, r is that otherwise policemen would search these motorists always. In a more realistic model, it may not be possible to achieve such deterrence. To see this, consider a model where a finite number of policemen are spaced along a highway. Suppose that stopping and searching a vehicle takes a certain time, during which a policeman cannot stop other cars. In that case, even if it is common knowledge that motorist c, r carries contraband with probability one, there is a positive probability that he will not be stopped simply because all policemen along the highway are busy searching other motorists. Now, by making group c, r sufficiently rare in the population (so that policemen do not find it worth their while to wait for a member of that group to pass by) and by making $u(c, r)$ sufficiently high, one can manufacture the result that some group carry contraband with probability one at equilibrium.

The model implies that for all c ,

$$\begin{aligned} P(G|c, A) &= t_A \\ P(G|c, W) &= t_W \\ \gamma(c, A) &= \frac{u(c, A)}{[u(c, A) + j(c, A)]} \end{aligned} \tag{2}$$

$$\gamma(c, W) = \frac{u(c, W)}{[u(c, W) + j(c, W)]} \tag{3}$$

Suppose that $t_A = t_W = t$, i.e. policemen have no taste for discrimination. Then, for all c it must be

$$P(G|c, A) = t = P(G|c, W). \tag{4}$$

Notice that this does not imply that $\gamma(c, W) = \gamma(c, A)$. In other words, given c , the search intensity γ may be higher for African Americans even in the absence of prejudice. This happens if $\frac{u(c, W)}{[u(c, W) + j(c, W)]} < \frac{u(c, A)}{[u(c, A) + j(c, A)]}$, i.e. if the expected propensity to carry drugs is higher or the cost of being convicted is lower, for African American motorists after conditioning on all the observables c . This property may indicate that race proxies for some variable that is unobservable by the policeman and is correlated with both race and crime. Such an unobservable could be educational attainment.

2.2 Testing for Prejudice

It is easy to test for relationship (4) even in the absence of data on c and on γ . It suffices to have data on the posterior frequency of guilt by race, conditional on being searched,

$$D(r) = \int P(G|c, r) \frac{\gamma(c, r) f(c|r)}{\int \gamma(s, r) f(s|r) ds} dc.$$

Indeed, using (4) to substitute for $P(G|c, r)$ into $D(r)$ we get

$$D(W) = t = D(A), \tag{5}$$

which is the implication that will be tested with the data below.

Notice that there is nothing special about the characteristic “race” in the model. An implication analogous to (5) holds for male and female, and indeed the equality of posterior frequencies of guilty should hold true across any characteristic on which the policeman conditions his searching decision. Thus, in the empirical section of the paper we will test condition (5) using both race and sex.

3 Empirical Results

3.1 Data Description

The data we analyze were collected as part of the settlement of a federal lawsuit filed in February 1993 by the ACLU, challenging as unconstitutional the Maryland State Police’s alleged use of a “racial profile” as a basis for stopping, detaining and searching motorists. In the settlement, the state agreed to maintain detailed records of motorist stops, and to file quarterly reports with the Court and the ACLU. In November, 1996 the Maryland ACLU asked a federal court in Baltimore to hold the Maryland State Police agency in contempt of court and to impose a \$250,000 penalty, based upon evidence showing that state police violated a 1995 court decree by continuing a pattern of race discrimination in drug interdiction activities carried out along the Interstate 95 corridor. Because the data on traffic stops by race are only available for the time period after the initiation of the first ACLU lawsuit, the estimates we report cannot be construed as describing police behavior prior to the legal interventions.

Our dataset consists of 1590 observations on car searches on I-95 in Maryland from January, 1995 through January, 1999.¹¹ The data provide infor-

¹¹The searches were conducted in Baltimore, Cecil, Harford, Howard, and Prince

mation on the race and sex of the motorist as well as on the year, make and model of the vehicle, and the date, time and location of the search. We also know whether the policeman requested consent to search the car and, if consent to search was not requested, the probable cause that the policeman invoked to search the car. The probable-cause information provides insight into what types of characteristics are considered as grounds for initiating searches.¹² We also know whether canine units were used in the search and whether contraband (typically illegal drugs) was recovered. If any was recovered, we know what type, how much (in grams) and where it was found. There is also information on whether currency was recovered, how much and its location inside the car. Finally, the dataset includes the name of the officer.

It is important to recognize that our data does not pertain to the stopping decision, but rather to the decision to stop and search the car. We do not have data on motorists who were stopped, say because they were speeding, but who were not searched. While the decision to stop could depend on differences in driving habits between different race groups, the decision to search is more up to the policeman's discretion. The car searches at the heart of the racial profiling controversy are ones where the motorist involved has committed a minor traffic violation, such as changing lanes without signalling or exceeding the speed limit by less than 5 miles per hour. Highway studies show that the vast majority of drivers commit such violations.¹³ State troopers are not required to stop motorists for these kinds of infractions, so these stops are often used as a pretext to stop and search the motorist, other occupants and

George's Counties.

¹²For example, 'third-party-vehicle' (a vehicle not owned by the driver) is often listed as a grounds for requesting consent to search. The appendix lists other observables that police sometimes use in traffic stops.

¹³According to a study designed by John Lamberth, 98.1% of all cars on a stretch of the New Jersey Turnpike were clearly exceeding speed limits. See *New Jersey v. Soto*, 1996.

the vehicle.¹⁴ The decision to search is probably based on a great number of characteristics (nervous behavior, dress code, etc.), some of which we have no hope of getting reliable data on. Thus, our dataset affords a good application of the tests described above that are implementable with a partial set of observed characteristics.

Table 1 summarizes the means and variances of the subset of variables used in the empirical work. Of the 1590 total searches, 1007 or 63.4% were performed on African-Americans, 466 or 29.3% on Whites, 97 or 6.1% on Hispanics, and the remaining (1.3%) was performed on other race/ethnic groups. Female motorists were rarely searched: a total of 117 female motorists appear in the data, compared to 1473 men. Marijuana was the drug most commonly found during the searches (23% of the time). It is not uncommon for drivers to be carrying up to three different types of drugs as well as drug-related paraphernalia. Searches tend to be much more common at night: 43% of searches were performed during the hours of midnight-6am.

Figure 1 plots the proportion of drivers searched who are African American, white or female against time. The circle size is proportional to the inverse of the standard deviation of the estimates, with the smaller circles having the larger standard deviations. The figure reveals a downward trend

¹⁴In 1986 for instance, the Drug Enforcement Agency trained 27,000 police officers in 48 states in the use of pretext stops to find drugs in vehicles. According to the ACLU, the training materials in these and similar programs “implicitly” encourage the targeting of minority motorists. The practice of using discretionary stops as pretexts in this way was supported by the Supreme Court in *Whren and Brown vs. US* in 1996, which held that any traffic violation was a legal basis for stopping a motorist. The ACLU, arguing against the legality of pretextual stops, maintained that the traffic code is so detailed that every driver is in violation of at least one provision at any point in time, and so pretextual stops give police the right to stop anyone arbitrarily, and in particular would justify racist police practices. The court statement cites without comment the ACLU statement that in the decision to stop a motorist, race is a “decidedly impermissible factor”.

in the proportion of African American drivers searched and an upward trend in the proportion of white drivers searched.¹⁵ There is no clear trend for the proportion of female motorists. If there is variation over time in police practices, at least two issues arise: first whether the model’s predictions hold in all the subperiods of the data, and second, whether the model can help elucidate the cause of this variation. These issues are discussed later in the paper.

3.2 Test Results

To implement our test for detecting tastes for discrimination, we need to define what it means to be guilty. We classify as guilty anyone found carrying any amount of drugs of the following types: marijuana, heroine, cocaine, crack, PCP, LSD, and methadone. A small number of individuals were found in possession of barbituates (such as valium); these we did not classify as guilty.

Figure 2 plots the proportion of African American and white drivers found to be carrying drugs.¹⁶ As discussed above, our test for prejudice compares the probability of being found guilty conditional on a subset of observed characteristics. The model has a strong implication; namely, that no matter what the set of observable conditioning characteristics, the probability of being guilty should be the same across the groups defined by those characteristics.

We first describe test results obtained when we condition on characteristics nonparametrically through cell means. Then we will consider ways of carrying out the same test in a parametric setting.

¹⁵These trends are statistically significant in regressions of the proportions on a linear time trend.

¹⁶We only present the plots for white and black because sample sizes for the other groups are too small to disaggregate by month.

3.2.1 Nonparametric Tests

Tables 2a-2c compare the probability of being found guilty of carrying drugs, across cells, defined by race, sex and race*sex. Although African American motorists are much more likely to be searched by police, the proportion found guilty among whites and African Americans is nearly identical. This finding is consistent in our model with pure statistical discrimination. It is inconsistent with racial prejudice. Among Hispanics, though, the proportion found guilty is lower than among African-Americans or whites. Table 3 reports p-values from a Pearson chi-squared test of association between guilty and race. It rejects the null hypothesis of no association at conventional significance levels when all race groups are used in the test, but does not reject when the sample is restricted to African Americans and whites.

In Table 3b, we compare the probability of being found guilty for men and women. Again, the conditional probabilities are similar and a Pearson chi-squared test does not reject the null hypothesis of no association with sex. (See p-values in Table 4). Table 3c conditions more finely, this time on both sex and race. Again, the joint test rejects the null of no association if Hispanic males are included, but not when they are excluded.¹⁷

Thus, for the african-american sample our findings are consistent with no prejudice. At the same time, existing research suggest that $\int \gamma(c, A) dF(c|A) > \int \gamma(c, W) dF(c|W)$, i.e. African American are searched proportionately more often than whites. The disparity in the search intensity between blacks and whites may be due to statistical discrimination, i.e. $\gamma(c, A) \neq \gamma(c, W)$. Alternatively, the disparity may simply be evidence of a different distribution of characteristics c among races. In other words, it may be that $\gamma(c, A) = \gamma(c, W)$ but $F(c|W) \neq F(c|A)$, hence any difference in the search intensity between the two races is due to the distribution of observables c in

¹⁷There are no Hispanic women in our data, so the rejection is due to Hispanic males.

the two populations. However, for the Hispanic sample the lower probabilities of being guilty would imply under our model tastes for discrimination against Hispanics.¹⁸

Our data were collected during three distinct time periods, the first when the police were being audited as part of the law suit, the second when the audit period was over but the police were still gathering data as part of the settlement agreement, and the third after a second lawsuit alleging that the police were still discriminating had been filed. These distinct periods can be clearly seen in Figure 3, which gives a histogram of how many observations were collected in different time periods. During the audit period and after the filing of the second lawsuit, data collection was more intense.

We were concerned that police behavior could have been changing over these three time periods and that the tests of equality of guilt probabilities, which combine all time periods, mask variation over time. To address this concern, we perform the identical tests after disaggregating the data by the three time periods: prior to May, 1996, in between June, 1996 and December, 1998, and after December 1998. The tests performed on the disaggregated data yield the same conclusions as those performed on the full sample. They do not reject equality of the probabilities within each of the time periods when Hispanic males are excluded from the sample, but do reject when Hispanics are included. These results are reported in Appendix B. (See p-values in Table B.2)

3.2.2 Parametric Tests

As discussed in the introduction, a common approach to testing for discrimination is to ask whether race has a significant coefficient in a regression after

¹⁸However, a caveat is in order: when Hispanics are found with drugs, they appear to be found with very large quantities relative to other races. Unfortunately, because we have so few guilty Hispanics in our data, this feature is not statistically significant.

including other covariates. One such test compares

$$\Pr(G = 1|r, c) = \Pr(G = 1|c)$$

where c is a particular set of characteristics considered to be valid conditioning variables and r is the race indicator variable. The probability can be estimated using a parametric procedure such as probit or logistic regression. Under a probit assumption,

$$G = 1(c\gamma - v > 0)$$

where v is normally distributed.

A drawback of a parametric approach is that the test is only valid if the systematic component and the error component of the model are correctly specified.¹⁹ Accumulated Monte Carlo evidence suggests that probit estimates are relatively robust to misspecification in terms of the distribution of the error component, but can be quite sensitive to misspecification in the systematic component of the model.²⁰ With a small number of binary covariates, one can minimize the possibility of misspecification in the systematic component of the model by including all possible interactions between the elements of c . In practice, however, many covariates are often introduced in a linear fashion without higher order interactions, so a parametric test for the significance of coefficient on the race indicator variable can be highly sensitive to functional form misspecification.

Our theoretical model has a much stronger prediction. If the probit model is correctly specified, then the coefficients associated with all the conditioning

¹⁹An exception is the case where the error terms are misspecified but the regressors are all normally distributed.

²⁰Intuitively, this robustness is partly due to the fact that a normal error distribution has some properties in common to all cdf's, namely that $F_v(-\infty) = 0$ and $F_v(\infty) = 1$. If the true error distribution is symmetric, the normally distribution also gets the midpoint right, $F_v(0) = 0.5$. See discussion in Todd (1996).

variables should equal zero (since the conditional probability should be equal across all groups). Three versions of this test are shown in Table B.3 of the Appendix. The results for each model show that our conclusions from the non-parametric tests remain valid under the probit specification: only the indicator for Hispanic seems to have a statistically significant effect on the guilty rate, even when the model allows effects to vary over time.

The nonparametric test results presented above are not subject to the criticism that the probability model could be misspecified. However, a problem often encountered in using nonparametric conditioning methods is that the cell sizes become small as the number of covariates increases.²¹ Fortunately, our test requires only that a subset of the covariates be used, making a nonparametric test viable. But if additional covariates were available and the sample size were very large, a more powerful test for racial prejudice could be constructed by combining results from tests on multiple sets of conditioning variables.

3.3 Testing for statistical discrimination

If we had data on the entire vector c and we had a complete record of those motorists who are stopped but not searched, we would be able to recover the search intensity:

$$\gamma(c, r) = \frac{\# \text{ of motorists with characteristics } c, r \text{ that are searched}}{\# \text{ of motorists with characteristics } c, r \text{ that are stopped}}$$

Using expressions (2) and (3), we can recover $u(c, r) / [u(c, r) + j(c, r)]$ from $\gamma(c, r)$. This allows us to test whether for some (or all) c it is true that simultaneously $u(c, W) = u(c, A)$ and $j(c, W) = j(c, A)$. Remember that $u(c, r)$ and $j(c, r)$ capture a motorist's expected propensity to carry drugs, given his characteristics c and race r . If for all c we have that $u(c, r)$ or $j(c, r)$

²¹This is a version of the curse of dimensionality problem.

is not invariant to r for some c , then upon seeing that c the policeman expects different races to have different propensities to carry contraband. This may be because race is correlated with some trait that affects the propensity to carry drugs (say, education level) which the policeman cannot observe.

4 Efficiency vs. Fairness

If we take the viewpoint that the benefit of law enforcement is crime deterrence and that its cost is the cost of policing, then a non-prejudiced police force would implement the efficient search behavior given the behavior of motorists. This is because each police officer trades off the probability of arresting against the cost of performing searches. The fact that the equilibrium exhibits statistical discrimination is implied by efficiency in crime reduction, which according to this simple viewpoint, is not affected by considerations of racial disparities.

This outcome may be considered less than satisfactory in terms of fairness. Assume that a person who does not carry contraband suffers a cost when searched. Assume further that one race has a higher average propensity to carry contraband. According to our model, persons from this race will be searched more frequently, and will therefore suffer a higher expected cost of being searched.

We say that an outcome is fair when individuals of different races with the same observables other than race, c , experience the same expected cost of being stopped.

Definition 3 *An outcome is fair if for all c we have $\gamma(c, W) = \gamma(c, A)$.*

Under this definition, statistical discrimination is unfair. One way of implementing the fair outcome is simply to constrain the police force to stop

both races at the same rate, conditional on c .²² To explore the tradeoff between fairness and efficiency, it is useful to consider a stylized version of our model, which allows us to derive a simple expression for the inefficiency associated with the “fair” outcome.²³

Consider a simplified version of our model where $t_A = t_W$ and c is absent, so the only observable that the police can condition on is race. Assume further that the propensity to carry drugs is higher for African Americans than for whites, i.e. $\frac{u(c,W)}{[u(c,W)+j(c,W)]} < \frac{u(c,A)}{[u(c,A)+j(c,A)]}$. According to the previous analysis, African Americans and whites will be stopped at rates $\gamma^*(A)$ and $\gamma^*(W)$, with $\gamma^*(A) > \gamma^*(W)$. The benefit of law enforcement is t , the probability that any motorist carries contraband at equilibrium. The cost is

$$t [\gamma^*(A) \varphi(A) + \gamma^*(W) \varphi(W)]$$

where $\varphi(r)$ represents the fraction of people with race r in the entire population.

Consider now the behavior of police under the constraint that $\gamma(A) = \gamma(W)$. This constraint requires that both races be stopped at the same rate (this is as if policemen were color-blind when deciding whom to search). Denote with ** the equilibrium of this constrained game. Assume that police do not allow African Americans to carry drugs for sure, i.e. $t < \varphi(A)$.²⁴ Then at equilibrium whites do not carry drugs, while African Americans randomize between carrying drugs or not. To be willing to randomize it

²²One might also argue, however, that several variables potentially included in c should not be considered ‘fair’ conditioning variables. For example, age is a such a characteristic. It is known to be correlated with criminal behavior, which would lead to younger drivers being searched more frequently.

²³If monetary transfers are possible, then a straightforward way of recovering fairness without compromising efficiency is to avoid putting any constraint on police behavior, and offer side payments to members of the race that is searched more often.

²⁴If they carry drugs for sure, the expected probability of the average motorist being guilty conditional on being stopped is at least $\varphi(A)$.

must be that $\gamma^{**}(A) = \gamma^*(A)$. For the police to be willing to randomize it must be that $t = P^{**}(G|A)\varphi(A)$. The benefit of law enforcement is the average probability that a motorist carries contraband at equilibrium, $P^{**}(G|A)\varphi(A) = t$. The cost is

$$t[\gamma^*(A)\varphi(A) + \gamma^*(W)\varphi(W)].$$

Comparing the two equilibria we notice that the benefit is the same, but the number of searches necessary to achieve that benefit differs. It is $t[\gamma^*(A)\varphi(A) + \gamma^*(W)\varphi(W)]$ in the unconstrained equilibrium, and $t[\gamma^*(A)\varphi(A) + \gamma^*(A)\varphi(W)]$ in the fair equilibrium. Thus, the efficiency cost of fairness is

$$t\varphi(W)[\gamma^*(A) - \gamma^*(W)].$$

The quantities in this expression could in principle be measured: in order to compute the cost of fairness, we need only the proportion of motorists found guilty and the probabilities that police will stop motorists of different races at the unconstrained optimum, and the racial proportions of the population. These numbers would give us a rough estimate, based on the key simplifying assumption, which is that police use only race to decide whether to stop motorists.

The above efficiency analysis reduces the benefits of drug interdiction to the reduction of contraband on the road, and the cost of stopping to the cost of policing. In practice, the benefits and costs of drug interdiction are much more complex and some aspects may be difficult to quantify even abstracting from considerations of fairness. A realistic measure of the efficiency cost of fairness must then be confronted with the expected benefits of the fair outcome, which we do not attempt here.

More generally, the model of Section 2 tells us what would be the effects on the rates of guilt by race of raising the costs to police of stopping African American motorists. Since police equate the marginal benefit of stopping a

motorist to the marginal cost, an increase in the cost of stopping African American motorists should result in a rise in the equilibrium probabilities of carrying drugs. We saw in Figure 1 that the proportion of motorists searched who are African American seems to be falling over time. One may wonder whether this trend is due to an increase in the effective cost of stopping black motorists, say due to political pressure on the police. However, Figure 2 suggests that the proportion of guilty among either African American or white motorists searched has remained roughly constant over time, which is inconsistent with this hypothesis. It suggests that the fall in the proportion of African American motorists searched is due instead to other factors affecting either $u(c, r)$ and/or $j(c, r)$.

5 Summary and Conclusions

Given the key role of statistical testing in detecting discrimination, it is important to know what assumptions on the behavior of motorists and troopers are needed to justify the types of tests that are being applied. In this paper, we develop a simple equilibrium model of law enforcement via traffic stops and consider its implications for our ability to test for the presence and nature of discrimination. We argue that tests for whether any sort of discrimination occurs rely crucially on what sets of variables are considered to be valid, nondiscriminatory variables that the police can use in stopping cars and on whether those variables are available in the data.

However, our model shows that a test for whether discrimination, if it occurred, is motivated purely out of efficiency considerations (statistical discrimination) or for other reasons (prejudice) is less demanding in terms of data requirements and only requires comparing the probability of being found guilty of carrying contraband across various subgroups of the population. In our model one race may be searched more often than another although at

equilibrium both races have the same probability of carrying drugs. This is true even when there is no taste for discrimination, so it is ironic that equality in the proportions guilty by race has been used in court as evidence that police are racist.²⁵

Our model showed that the disparities observed in the Maryland data in terms of the proportions searched by race do not necessarily imply that Maryland state police were motivated by racism. Our empirical finding that the probabilities of being found with drugs are equal across the two race groups suggests that troopers were maximizing the effectiveness of their drug interdiction campaign, and that this was consistent with statistical discrimination on the basis of race. The results do not, however, imply statistical discrimination by our definition, because the police may also be conditioning their decisions on other observable motorist characteristics not in our data. Our findings for Hispanic males are consistent with a taste for discrimination against this group, although the sample size for this group is modest.

Statistical discrimination, although efficient from a crime-fighting perspective, may be considered unfair. This is because innocent drivers experience different probabilities of being searched depending on their race. Our estimates suggest that efficiency, not fairness, is currently the main consideration in drug interdiction. Achieving the fair outcome entails a cost in terms of efficiency. We derived a simple expression for the cost of fairness in terms of increased criminality, which could conceivably be estimated. Interestingly, we observe a reduction over time in the proportion of African American motorists stopped, which does not appear to be associated with a reduction

²⁵>From a memorandum prepared by ACLU lawyers: “MSP’s own data demonstrates that this racial distortion is unnecessary to successful drug interdiction. ...[Indeed,] MSP data shows that statewide, police find contraband on black and white motorists at equal rates. Blacks were found in possession of contraband in 28.4% of searches, whereas whites were found with contraband 28.8% of the time.” Mertens and Jeon (1996).

in efficiency. This suggests that the observed decline is not due to political pressure on the police.

Although in this paper we focus on traffic searches, our analysis applies more broadly to some other similar settings. Consider for example the behavior of security and customs agents in airports, where it is alleged that minorities and foreigners are unfairly targeted in baggage and passenger searches.²⁶ The principles of our model can also be extended to analyze other instances of discrimination, such as in employment or college admissions, where one would need to clearly define the nature of success, in analogy to that of guilt in the current paper.

References

- [1] Arrow, K. (1973) “The Theory of Discrimination.” In *Discrimination in Labor Markets*, O. Ashenfelter and A. Rees, eds., Princeton University Press, Princeton NJ.
- [2] Becker, G. (1957) *The economics of discrimination*, Chicago University Press.
- [3] Becker, G. (1968) “Crime and Punishment: An Economic Approach” *Journal of Political Economy*, 76, p. 169-217.
- [4] Border, Kim C; Sobel, Joel. (1987) “Samurai Accountant: A Theory of Auditing and Plunder” *Review of Economic Studies*. Vol. 54 (4). p 525-40.
- [5] Donohue, John (1999) Expert witness testimony in the case *Chavez v. Illinois State Police*.

²⁶See for instance *Anderson v. Cornejo*, 1999 (No. 97 C 7556).

- [6] Donohue, John and Levitt, Steven (1998): “The Impact of Race on Policing, Arrest Patterns, and Crime” NBER Working Paper # W6784.
- [7] Lambert, John (1996) Report in connection with the case *Wilkins v. Maryland State Police*, Civil Action No. CCB-93-483.
- [8] Mertens William J. and Deborah A. Jeon (1996) “Memorandum in support of plaintiff’s motion for enforcement of settlement agreement and for further relief in the United States District Court for the District of Maryland,” *Wilkins at al. v. Maryland State Police*. Civil Action No. CCB-93-468
- [9] Reinganum, J. and L. Wilde (1986) “Equilibrium Verification and Reporting Policies in a Model of Tax Compliance” *International Economic Review*, 27, p. 739-760.
- [10] Scotchmer, Suzanne (1987) “Audit Classes and Tax Enforcement Policy” *American Economic Review*. Vol. 77 (2). p 229-33.
- [11] Stigler, G (1970) “The optimum enforcement of laws” *Journal of Political Economy*, 78, p. 526-536.
- [12] Todd (1996) “Empirical Methods for Evaluating the Impact of Interventions in Education and Training” Ph.D. Dissertation, University of Chicago.

A Purification of Mixed Strategies

In this section we explain how the model's mixed-strategy equilibrium is consistent with a world in which motorists do not randomize over the decision to carry drugs. A simple example is sufficient to illustrate the idea. Consider an agent who randomizes between actions a and b with $\Pr(a) = p$. Utility maximization implies that the agent is indifferent between the two actions, $u(a) = u(b)$. To purify this mixed strategy, we can imagine that the utility from action a is really $u(a) + X$, where X is a random variable with the property that $\Pr(X > 0) = p$.²⁷ If this is the case, the agent chooses a with probability p , and is never indifferent between actions a and b . This perturbed model is equivalent in terms of outcomes to the mixed strategy, but the agent is never indifferent between the two actions and, as a consequence, never flips a coin.

In the context of our model, we can imagine of a class of individuals, all with characteristics c, r . Let i denote a generic individual in that class, and let i 's propensity to carry contraband be $u(c, r) + X_{c,r}^i$ where $X_{c,r}^i$ is a random variable independent across all motorists in that class. We can imagine that motorist i in class c, r knows his realized value $x_{c,r}^i$. Then, given a certain $\gamma(c, r)$ motorist i carries contraband if and only if $\gamma(c, r)[-1] + [1 - \gamma(c, r)][u(c, r) + x_{c,r}^i] > 0$. Equating the left hand side of this inequality to zero we find a cutoff value $\bar{x}_{c,r}$ such that a motorist carries contraband if and only if $x_{c,r} > \bar{x}_{c,r}$. This cutoff determines the fraction of individuals in class c, r who carry contraband. Suppose that this fraction is higher than in class c', r' . Then policemen will never search an individual in class c', r' , so that $\gamma(c', r') = 0$. But then the cutoff in that class goes down, and the equilibrating process equalizes the probabilities of carrying contraband across

²⁷Notice that X may have small support, so that the two models may be quite close in terms of primitives.

classes.

B A Model with Exogenous Guilt

If we had data on c an interesting exercise could be carried out. In the main body of the paper we have assumed that motorists consider the probability of being searched in deciding whether to carry illegal substances. Thus, the probability that a motorist is found guilty if searched has been endogenous. An alternative modeling assumption is that the probability that the motorist carries contraband is exogenous, in that it does not depend on the probability of being searched. This partial equilibrium approach transforms the problem into one of optimal sampling. Under this assumption, a policeman takes as given $P(G|c, r)$, the probability that a motorist with characteristic c and race r is found guilty if searched. If we assume that $P(G|c, r)$ is increasing in c then it is optimal for the policeman to choose two cutoffs k_W and k_A , so that if c exceeds these cutoffs a person is searched. To set the cutoffs optimally the policeman solves

$$\max_{k_W, k_A} \int_{k_W}^{\infty} [P(G|c, W) - t] f(c|W) dc + \int_{k_A}^{\infty} [P(G|c, A) - t] f(c|A) dc.$$

The first order conditions are

$$[P(G|k_W, W) - t] f(k_W|W) dc + [P(G|k_A, A) - t] f(k_A|A) = 0,$$

whence the implicit condition

$$P(G|k_W, W) = P(G|k_A, A) = t. \tag{6}$$

Knowledge of c would allow us to build the entire schedule $P(G|c, r)$ for all c greater than k_r . The exogenous model predicts that this schedule should be strictly increasing in c , while the endogenous model predicts that this schedule should be constant in c . Thus, knowledge of c would permit us to determine which model seems to fit the data better. Notice that condition (6) is expressed in terms of the probability of being guilty of the *marginal*

motorist searched. To achieve a reliable estimate of this relationship one would need a large number of observations.

C Observable Indicators of Criminal Activity

Citizen's band radios

Cellular telephones

Pre-paid phone cards

Tinted Windows

Radar Detectors

Perfumes

Duct tape

Pagers

Screws

Handles and knobs

Inability to completely roll down windows (which may hide concealment of drugs in the doors)

Religious paraphernalia used to divert suspicion

Police materials used in attempt to show support for law enforcement

High odometer mileage, particularly on late model vehicles

Switches and buttons, which may activate electronic compartment doors

Large amounts of cash

Attorney's business cards

Too little or too much luggage for stated length of trip

Signs of recent drug use

Hiding places

Weapons

Only one key in ignition or no trunk key

Maps from source cities or states, map turned to locations other than those mentioned by occupants

Leased vehicles: 'Leased vehicles are used frequently by drug traffickers. Many times these

vehicles are rented from airports. The person(s) authorized to drive the vehicle should

be noted on the least agreement.'

Third-party vehicles: 'Question when the owner of the vehicle is not at the scene of the stop or if

the occupants cannot tell you the owner's/lessee's name without looking at the registration or lease agreement themselves.

Cashier's checks

Bondo

Source: Expert witness testimony by Professor John Donohue in *ACLU v. State of Illinois*. List from Valkyrie police officer manual (p. 13-15).

Table 1
Means and Standard Deviations of Variables used in Analysis
(standard deviations in parentheses)

	All Obs	By Race				By Sex	
		Black	Hisp.	White	Other	Female	Male
Black	0.63 (0.01)	1.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.64 (0.04)	0.63 (0.01)
White	0.29 (0.01)	0.00 (0.00)	0.00 (0.00)	1.00 (0.00)	0.00 (0.00)	0.35 (0.04)	0.29 (0.02)
Hispanic	0.06 (0.01)	0.00 (0.00)	1.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.07 (0.01)
Female	0.07 (0.01)	0.07 (0.008)	0.00 (0.00)	0.09 (0.01)	0.22 (0.09)	1.00 (0.00)	0.00 (0.00)
Guilty	0.33 (0.01)	0.35 (0.02)	0.12 (0.03)	0.32 (0.02)	0.47 (0.11)	0.38 (0.05)	0.32 (0.01)
Cocaine	0.08 (0.01)	0.10 (0.01)	0.03 (0.02)	0.03 (0.01)	0.37 (0.11)	0.09 (0.03)	0.08 (0.007)
Marijuana	0.23 (0.01)	0.23 (0.01)	0.23 (0.01)	0.26 (0.02)	0.41 (0.11)	0.21 (0.04)	0.23 (0.01)
Crack Cocaine	0.04 (0.005)	0.05 (0.01)	0.01 (0.01)	0.01 (0.004)	0.22 (0.09)	0.06 (0.02)	0.04 (0.005)
Heroin	0.02 (0.003)	0.02 (0.004)	0.03 (0.02)	0.03 (0.01)	0.22 (0.09)	0.06 (0.02)	0.02 (0.004)
Morphine	0.001 (.001)	0.00 (0.00)	0.00 (0.00)	0.002 (0.002)	0.00 (0.00)	0.00 (0.00)	0.001 (0.001)
Other Drugs	0.01 (0.002)	0.00 (0.00)	0.00 (0.00)	0.01 (0.005)	0.00 (0.00)	0.01 (0.01)	0.02 (0.003)
Paraphernalia	0.01 (0.002)	0.003 (0.002)	0.010 (0.010)	0.02 (0.006)	0.00 (0.00)	0.01 (0.01)	0.01 (0.002)
Night (12am-6am) Number of Observations	0.43 (0.01)	0.46 (0.02)	0.44 (0.05)	0.35 (0.02)	0.51 (0.11)	0.47 (0.05)	0.43 (0.01)
	1582	1002	97	463	20	117	1465

Table 2a
Proportion of Vehicles Searched Found with Drugs
by Race/Ethnicity

	Not Guilty	Guilty
African American	0.66	0.34
White	0.68	0.32
Hispanic	0.87	0.11

Table 2b
Proportion of Vehicles Searched Found with Drugs
by Sex

	Not Guilty	Guilty
male	0.68	0.32
female	0.64	0.36

Table 2c
Proportion of Vehicles Searched Found with Drugs
by Race/Ethnicity and Sex

		Not Guilty	Guilty
male	African American	0.66	0.34
	White	0.67	0.33
	Hispanic	0.89	0.11
	Other	0.68	0.32
female	African American	0.56	0.44
	White	0.78	0.22
	Hispanic	*	*
	Other	100.00	*

Table 3
P-values on Pearson Chi-Squared Tests on
Hypothesis that Proportion Guilty is Equal Across Various Groups

Groups	χ^2	p-value
race (African American, Hispanic and white)	21.59	<0.001
race (African American, White)	0.97	0.33
sex (male, female)	0.82	0.37
sex and race (African American, Hispanic, white and male, female)	26.97	<0.001
sex and race (African American, white and male or female)	6.29	0.10

Table B.1a
Proportion of Vehicles Searched Found with Drugs
by Race/Ethnicity

	Period 1		Period 2		Period 3	
	Not Guilty	Guilty	Not Guilty	Guilty	Not Guilty	Guilty
African American	0.66	0.34	0.62	0.38	0.69	0.31
White	0.72	0.28	0.71	0.29	0.64	0.36
Hispanic	0.88	0.12	0.87	0.13	0.91	0.09

Table B.1b
Proportion of Vehicles Searched Found with Drugs
by Sex

	Period 1		Period 2		Period 3	
	Not Guilty	Guilty	Not Guilty	Guilty	Not Guilty	Guilty
male	0.68	0.32	0.67	0.33	0.69	0.31
female	0.63	0.37	0.64	0.36	0.65	0.35

Table B.1c
Proportion of Vehicles Searched Found with Drugs
by Race/Ethnicity and Sex

		Period 1		Period 2		Period 3	
		Not Guilty	Guilty	Not Guilty	Guilty	Not Guilty	Guilty
male	African American	0.66	0.34	0.63	0.37	0.71	0.29
	White	0.73	0.27	0.69	0.31	0.62	0.38
	Hispanic	0.88	0.12	0.87	0.13	0.91	0.09
	Other	0.70	0.30	0.50	0.50	0.80	0.20
female	African American	0.64	0.36	0.52	0.47	0.48	0.52
	White	0.56	0.44	0.00	100.00	0.79	0.21
	Hispanic	n/a	n/a	n/a	n/a	n/a	n/a
	Other	100.00	0.00	n/a	n/a	n/a	n/a

Table B.2
P-values on Pearson Chi-Squared Tests on
Hypothesis that Proportion Guilty is Equal Across Various Groups

Groups	Period 1		Period 2		Period 3	
	χ^2	p-value	χ^2	p-value	χ^2	p-value
race (African American, Hispanic and white)	8.68	0.01	8.90	0.01	9.72	0.01
race (African American, White)	2.41	0.12	2.54	0.11	1.18	0.28
sex (male, female)	0.7225	0.40	0.11	0.74	0.20	0.66
sex and race (African American, Hispanic, white and male, female)	9.9481	0.04	13.99	0.007	17.00	0.002
sex and race (African American, white and male or female)	3.66	0.30	7.47	0.06	8.27	0.04

Table B.3
Parameter Estimates for Probit Model of Conditional Probability of being ‘Guilty’
Models without Covariates
(p-values from Hypothesis Tests shown in footnote)

Variable	Model (1) ^(a)	Model (2) ^(b)	Model (3) ^(c)
Indicator for white	-0.46 (0.06)	-0.66 (0.13)	...
Indicator for black	(-0.38 (0.04)	-0.32 (0.07)	...
Indicator for Hispanic	-1.16 (0.16)	-1.20 (0.32)	...
Indicator for white * time	...	0.007 (0.004)	...
Indicator for black * time	...	-0.003 (0.004)	...
Indicator for Hispanic * time	...	0.002 (0.011)	...
indicator for white * period 1	-0.58 (0.11)
indicator for white * period 2	-0.53 (0.13)
indicator for white * period 3	-0.34 (0.09)
indicator for black * period 1	-0.39 (0.05)
indicator for black * period 2	-0.27 (0.09)
indicator for black * period 3	-0.45 (0.09)
indicator for Hispanic * period 1	-1.17 (0.28)
indicator for Hispanic * period 2	-1.13 (0.29)
indicator for Hispanic * period 3	-1.17 (0.28)

- (a) P-value from test of hypothesis white=black=Hispanic is 0.0001. P-value from test that white=black is 0.2523.
- (b) P-value from test of hypothesis black=white=Hispanic for both intercept and time trend is 0.0001. P-value from test that black=white for both intercept and time trend is 0.0530.
- (c) P-value from test of hypothesis that black=white=Hispanic within all time periods is 0.0007. P-value from test that black=white for all time periods is 0.2266.

Figure 3

Number of Cars Searched by Maryland Troopers on I-95

