Racial Bias in Motor Vehicle Searches: Theory and Evidence

John Knowles and Nicola Persico

University of Pennsylvania

Petra Todd

University of Pennsylvania and National Bureau of Economic Research

Police checking for illegal drugs are much more likely to search the vehicles of African-American motorists than those of white motorists. This paper develops a model of police and motorist behavior that suggests an empirical test for distinguishing whether this disparity is due to racial prejudice or to the police's objective to maximize arrests. When applied to vehicle search data from Maryland, our test results are consistent with the hypothesis of no racial prejudice against African-American motorists. However, if police have utility only for searches yielding large drug finds, then our analysis would suggest bias against white drivers. The model's prediction regarding nonrace characteristics is also largely supported by the data.

Earlier versions of this paper were circulated as CARESS Working Paper no. 99-06 (Center for Analytical Research in Economics and the Social Sciences, University of Pennsylvania) and as NBER Working Paper no. 7749. We thank the editor Steven Levitt and an anonymous referee for their recommendations. We are grateful to Ian Ayres, Gary Becker, Hanming Fang, John Ham, James Heckman, George Mailath, Charles Manski, Kevin Murphy, Andrew Postlewaite, Rafael Rob, Cecilia Rouse, Joel Waldfogel, Kenneth Wolpin, and participants at seminars at the University of Pennsylvania, New York University, Oxford, Hebrew University, Cornell, Penn State, Rochester, and the University of Chicago for helpful comments. We thank the Maryland ACLU for providing us with the data and information. Persico and Todd are grateful to the National Science Foundation for support under grants SBR-9905564 and SBR-9730688, respectively.

[*Journal of Political Economy*, 2001, vol. 109, no. 1]

I. Introduction

African-American motorists in the United States are much 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, African Americans constituted 63 percent of motorists searched by state police on Interstate 95 between January 1995 and January 1999 but only 18 percent of motorists on the road. 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 searches is that race is one of the criteria police officers use in deciding whether to search cars. 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" ("Clinton Order Targets Racial Profiling," Associated Press, June 9, 1999).

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 searches falls on the courts, which consider a variety of statistical evidence. The case for discrimination rests largely on the observation that the proportion of African Americans among the drivers searched by police far exceeds the proportion in the general population of drivers. This simple comparison is the basis of expert witness testimonies in several prominent legal cases.[1] A refined version of the test estimates the probability of being searched as a function of race and other observable characteristics thought to be related to criminal propensity. If race has no explanatory power in the regression, this is taken as evidence of no discrimination (see, e.g., expert witness testimony by John Donohue in the case *Chavez v. Illinois State Police* [1999]).

The drawback of this type of test for discrimination is that it requires data on the full set of characteristics that a police officer uses in deciding whether to search a motorist.[2] If some characteristics were missing from

[1] In the 1996 case *Wilkins v. Maryland State Police*, a statistician testified that "the disparities 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."

[2] A training manual issued by the Illinois State Police highlights some indicators of criminal activity. They include such characteristics as tinted windows, cell phones, leased vehicles, religious paraphernalia used to divert suspicion, and attorney business cards. Source: John Donohue in expert witness testimony in *ACLU v. State of Illinois*.

the data, then race could have explanatory power due to omitted-variable bias. If race were found to be insignificant, 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. Conditioning on those characteristics may lead to the wrong conclusion that race did not affect the search decision. Thus the validity of this type of test for discrimination hinges crucially on judgments about what constitutes a set of admissible conditioning variables and on whether the analyst has access to the full set of variables.[3]

Even if such tests find evidence of discrimination, they are not informative about the underlying motivation. Police may use race as a criterion in traffic stops because they are trying to maximize successful searches and race helps predict criminality or because they prefer stopping one racial group over another. We call the first type of situation *statistical discrimination*, using the terminology of Arrow (1973). An equilibrium exhibits statistical discrimination if police are not racially prejudiced and yet one race is searched more often than another.[4] In contrast, we say that officers are *racially prejudiced* if, ceteris paribus, they have a preference for searching motorists of a particular race. We model prejudice as a taste for discrimination, following Becker (1957). Prejudice is a property of the officer's utility function, whereas statistical discrimination is a property of equilibrium.

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

The idea of testing for discrimination by looking at differential outcomes is originally due to Becker (1957). It has previously appeared in the discrimination literature in a variety of settings: in mortgage lending, the setting of levels of bail, and the publication of academic articles. Subsection *B* discusses these applications.

Our model assumes that the police maximize the number of successful

[3] Similar considerations arise in audit pair studies in which auditees have to be matched on all relevant characteristics to ensure that disparate treatment is due to sex or race and not to some other unobservable characteristic (see, e.g., Heckman and Siegelman 1993).

[4] We give a formal definition of statistical discrimination in Sec. II.

searches, net of the cost of searching motorists.[5] One possible definition of a successful search is one in which a motorist is found to be carrying contraband. Alternatively, a search could be deemed successful by police only if it yields a sufficiently large drug find. In the empirical analysis, we consider several alternative definitions. Our model assumes that motorists take into account the probability of being searched in deciding whether to carry contraband. Police make their search decision on the basis of observable characteristics of the motorist, including race. Some of these characteristics may be informative about a driver's propensity to carry contraband. Prejudice is introduced as a difference in the cost to the police of searching drivers of different groups.

The key implication of the model is that if a police officer has the same cost of searching two subgroups of the population and if these two subgroups are searched at equilibrium, then the returns from searching will be equal across the subgroups. For example, suppose that searching one subgroup of motorists yielded a higher return. Then police would always search these motorists, who would in turn react by carrying contraband less often, until the returns to searching are equalized across groups. If the returns to searching are equal across *all* subgroups distinguishable by police, they must also be equal across *aggregations* of these subgroups, which is what we can distinguish in the data. Thus equality of the returns to searching can be tested without knowing all the characteristics observed by the police.

We test the implication that returns to searches are equal across groups using data on highway searches gathered by the Maryland State Police. The Maryland data attracted national attention when the American Civil Liberties Union (ACLU) asked that the police be found in contempt of court, using these data as evidence of racial bias on the part of the state police against black motorists in violation of the terms of a 1993 class action lawsuit settlement.[6] In our data, vehicles of African-American motorists are searched much more frequently than those of white motorists. However, the probability that a searched driver is found carrying any amount of contraband is very similar across races. Thus we cannot reject the hypothesis that the disparity in the probability of being searched is due purely to statistical discrimination and not to racial prejudice. When we look at the probability that a searched driver is carrying contraband in excess of a high threshold, this probability is

[5] An alternative model is that police maximize the number of arrests given a certain amount of resources (number of police officers, say). In terms of the implications tested in this paper, the two models are equivalent.

[6] The class action suit was filed by the ACLU on behalf of Robert L. Wilkins, an African-American attorney who was stopped and searched by the Maryland State Police. The ACLU contended that the data show a "continuing pattern of race discrimination in drug interdiction activities carried out along the I-95 corridor."

higher for African Americans. Under our model, this would imply a bias against white motorists.

A. *Legal Background*

The judicial standpoint on racial profiling is not clear-cut. The dominant view seems to be that race or ethnicity can be used as a factor in determining the likelihood that a person is engaging in or has committed a crime, as long as its use is reasonably related to law enforcement and is not a pretext for racial harassment (Kennedy 1997). In an early influential case (*United States v. Martinez-Fuerte,* 428 U.S. 543 [1976]), the Supreme Court found that it was permissible for police at border crossings to detain motorists of Mexican ancestry and to search their vehicles, because these cars were more likely to be transporting illegal aliens. The U.S. Court of Appeals adopted a similar position in *United States v. Weaver* (966 F.2d at 394, n.2 [1992]) in permitting airport searches of young black males prompted by suspicions that members of a black gang were trafficking in cocaine. Thus the courts have allowed race to be used in police decisions, provided that race is part of the suspect's description.

However, Kennedy (1997) argues that there is also a significant dissenting tradition. For example, in *United States v. Nicholas* (448, F.2d 622 [CA 8 1971]), the court found it impermissible to stop a black driver with out-of-state plates, maintaining that this did not constitute sufficient grounds for suspicion of criminality. In *United States v. Laymon* (730 F.Supp. 332 [D. Colo. 1990]), the court suppressed incriminating evidence found in a vehicle, arguing that the police officer did not have sufficient justification for searching the car and had used race as a factor in the decision to search.

Whether discrimination is deemed reasonable or not by the courts depends on assessments about the degree to which discrimination assists in apprehending criminals, the benefits of apprehending criminals, and the costs imposed on people erroneously searched or detained.[7] In evaluating the legality of racial disparities in law enforcement, the courts have clearly sought to determine the motivation for discriminating. Sometimes discrimination motivated by reasons of efficiency is considered permissible, whereas discrimination motivated by racial prejudice is never permissible. The standard regression-based test for discrimination described earlier is informative only about whether a racial disparity in car searches exists and is silent on the question of motivation.

[7] Also, there is some debate in the courts over whether "reasonableness" or "strict scrutiny" is the appropriate criterion to use in racial profiling cases (see Kennedy 1997).

In contrast, the test we propose distinguishes between different motivations for disparities in search behavior.

B. Related Literature

The theoretical model we develop belongs to the literature on optimal auditing. Early auditing models, such as Becker (1968) and Stigler (1970), examined citizens' incentives to misbehave under an exogenous probability of being audited. The more recent theoretical literature on optimal auditing, which mainly deals with income reporting and tax evasion, assumes that both parties, the auditor and the auditee, behave strategically (see Reinganum and Wilde 1986; Border and Sobel 1987; Scotchmer 1987).

In our model, if police are prejudiced, the equilibrium returns to searching members of the group that is discriminated against will be below average. This idea, that tastes for discrimination lead to lower profits for the discriminators, originated with Becker (1957) and is also discussed in Epstein (1992). One area in which the empirical link between profitability and discrimination has been investigated is mortgage lending, where studies compare the profitability of loans to minorities and nonminorities (see Berkovec et al. 1994; Van Order and Zorn 1995).[8] The profitability test has also been used by Smart and Waldfogel (1996) to test for bias in academic publications against authors outside top academic institutions by examining the citation frequencies of their published articles. Their analysis assumes that discrimination would reveal itself in a greater number of citations for persons subject to discrimination, with the quality of the journal and a measure of editorial treatment held constant.[9] Ayres and Waldfogel (1994) use a test similar in spirit to the one we develop here in investigating discrimination against minorities in the setting of levels of bail bonds. Under the assumption that unprejudiced judges would set the amount of bail for

[8] See also Ross (1996) for a criticism of the use of average mortgage default rates in testing for discrimination. Berkovec et al. (1998) find that default rates on loans granted to minorities are higher than those on loans granted to nonminorities, which is consistent with reverse discrimination. However, the relative loan performance of minorities vs. nonminorities is not found to be correlated with the degree of market competition, which leads the authors to conclude that the differences in loan performance do not reflect tastes for discrimination.

[9] Blank (1991) uses a different approach to detect discrimination against women in academic publications by comparing the probability of acceptance under double-blind vs. single-blind refereeing.

defendants so as to equalize the probability that the defendants flee, they find evidence of racial prejudice.[10]

II. The Model

We next develop the model of police and motorist behavior that underlies the empirical work in Section III. The model assumes a continuum of police officers and motorists. Let $r \in \{A, W\}$ denote the race of the motorist, which is observable by the police officer. Let c denote all characteristics other than race that are potentially used by the officer in the decision to search cars. The variable c may be unobserved or only partially observed by the econometrician. For expositional ease, we treat c as a one-dimensional variable, but the results extend to the case in which c is multidimensional. Let $F(c|W)$ and $F(c|A)$ denote the distribution of c in the white and African-American populations, respectively.

Each officer can choose to search motorists of any type (c, r). The police officer maximizes the total number of convictions minus a cost of searching cars. The marginal cost of searching a motorist of race r is denoted by t_r. We normalize the benefit of each arrest to equal one, so that the cost is scaled as a fraction of the benefit. To avoid trivial cases, assume $t_W, t_A \in (0, 1)$. Let G denote the event that the motorist searched is found guilty (i.e., found with drugs in the car).

We assume that motorists consider the probability of being searched in deciding whether to carry contraband. If they do not carry, their payoff is zero whether or not the car is searched. If they do carry, their payoff is $-j(c, r)$ if they are searched and $v(c, r)$ if not searched. We can interpret $v(c, r)$ as the expected value of carrying drugs and $j(c, r)$ as the expected cost of being convicted.[11] We assume that both $j(c, r)$ and $v(c, r)$ are positive.

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

$$\gamma(c, r)[-j(c, r)] + [1 - \gamma(c, r)]v(c, r). \tag{1}$$

Given $\gamma(c, r)$, the motorist chooses to carry contraband if this expression is greater than zero. When the expression is zero, motorists are willing

[10] In the absence of data on frequencies of flight, Ayres and Waldfogel (1994) use data on the rates bond dealers charge to defendants as an indication of the probability that defendants flee. Bond dealers charge significantly lower rates to minorities, which they interpret as evidence of discrimination against minorities at the bail bond setting stage.

[11] If there were discrimination in the court system leading to higher penalties for minority drivers found with contraband, this could be thought of as operating through $j(c, r)$. We do not test for this type of discrimination.

to randomize between carrying and not carrying. We denote the probability that a motorist of type c, r carries contraband by $P(G|c, r)$.[12]

The officer chooses the probability $\gamma(c, r)$ of searching each motorist of type c, r to solve

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

taking as given $P(G|c, r)$.[13] We can think of the term $P(G|c, r) - t_r$ as the expected profit from searching a motorist of type c, r. If $P(G|c, r) - t_r > 0$, then optimizing behavior implies $\gamma(c, r) = 1$; that is, always search motorists of type c, r. If $P(G|c, r) = t_r$, then the police officer is willing to randomize over whether or not to search type c, r.

Next, we introduce two definitions. First, a police officer is defined to be racially prejudiced if he or she exhibits a preference for searching motorists of one race. We model this preference as a difference in the cost of searching motorists.

DEFINITION 1. A police officer is racially prejudiced, or has a taste for discrimination, if $t_A \neq t_W$.

Next, we say that an equilibrium exhibits statistical discrimination if police officers have no taste for discrimination and yet the police officer chooses search probabilities that differ by race. Define the probability that a motorist of race r is searched as $\gamma(r) = \int \gamma(c, r)dF(c|r)$.

DEFINITION 2. Assume $t_A = t_W$. Then an outcome exhibits statistical discrimination if $\gamma(W) \neq \gamma(A)$.

An alternative definition of statistical discrimination would require that $\gamma(c, W) \neq \gamma(c, A)$ for some c; that is, blacks are searched at different rates than whites *with the same observable characteristics* c. This definition is more stringent than definition 2 in the sense that if $\gamma(c, W) \neq \gamma(c, A)$, then one expects that $\gamma(W) \neq \gamma(A)$. For our purposes, it is more convenient to use definition 2.

A. *Equilibrium*

We next construct an equilibrium in which police officers randomize over whether to search motorists, and motorists randomize over whether to carry contraband. The notion that motorists randomize is useful for describing the equilibrium in simple terms; in Section II*C*, we show that randomization is not crucial to the argument and present a simple

[12] We do not allow for the possibility of false accusation by police or planting of evidence. Donohue and Levitt (1998) analyze this issue in a context different from ours.

[13] In Sec. II*C*, we consider an alternative specification of the police officer's problem in which police maximize the number of arrests subject to a resource constraint. The testable implications are the same.

alternative interpretation that does not require that people be indifferent across actions and allows for unobservable heterogeneity within groups. We denote equilibrium values with an asterisk.

For a motorist to be willing to randomize, expression (1) must equal zero. Solving for γ yields

$$\gamma^*(c, r) = \frac{v(c, r)}{v(c, r) + j(c, r)}.$$

This ratio determines the police officer's search intensity. From this expression, we see that $\gamma^*(c, r)$ is between zero and one, so at equilibrium the officer randomizes over whether to search each type c, r.[14]

For a police officer to be willing to randomize, it must be that $P^*(G|c, r) = t_r$ for all c, r. At equilibrium, for all c,

$$P^*(G|c, A) = t_A,$$

$$P^*(G|c, W) = t_W,$$

$$\gamma^*(c, A) = \frac{v(c, A)}{v(c, A) + j(c, A)}, \tag{2}$$

and

$$\gamma^*(c, W) = \frac{v(c, W)}{v(c, W) + j(c, W)}. \tag{3}$$

Suppose that $t_A = t_W = t$; that is, police officers are not prejudiced. Then, for all c, guilt probabilities at equilibrium must be equal across races:

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

Notice that this does not imply $\gamma^*(c, W) = \gamma^*(c, A)$. The equilibrium search intensity may be higher for African Americans even in the absence of prejudice. This happens if

$$\frac{v(c, W)}{v(c, W) + j(c, W)} < \frac{v(c, A)}{v(c, A) + j(c, A)},$$

that is, if the expected value of carrying drugs is higher or the cost of being convicted lower for black motorists, after conditioning on observables c.

[14] Observe that in equilibrium, police would get the same payoff by setting $\gamma(c, r) = 0$, but $\gamma(c, r) = 0$ cannot be an equilibrium. If $\gamma(c, r)$ were equal to zero, then all motorists of type c, r would carry drugs, so it would not be rational for police to set $\gamma(c, r)$ to zero.

B. *Testing for Prejudice*

Equation (4) provides a test for prejudice ($t_W \neq t_A$) that is implementable even in the absence of data on c and on γ^*. It suffices to have data on the frequency of guilt by race conditional on being searched,

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

Using (4) to substitute for $P^*(G|c, r)$, we get

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

which is the implication that we test in the data.

In the model, there is nothing special about the characteristic "race." The analogue of (5) should hold for any observed characteristic. Equality of posterior frequencies of guilt should hold across any characteristic on which the police officer conditions his searching decision. In Section III, we test condition (5) using variables that describe the race and gender of the driver and characteristics of the vehicle.

When we compare the guilt probabilities across whites and African Americans, the empirical evidence is consistent with (5), which we interpret as the absence of racial prejudice against African Americans. At the same time, our data show that African Americans are searched more often than whites, that is, $\gamma^*(A) > \gamma^*(W)$. Statistical discrimination can arise for multiple reasons. It is possible that at equilibrium $\gamma^*(c, A) > \gamma^*(c, W)$; that is, African Americans are more likely to be searched than whites *with the same observable characteristics* c. If that is the case, race may proxy for other variables that are unobservable by the police officer and are correlated with both race and crime. Possible examples of such unobservables are the schooling level or the earnings potential of the motorist. Statistical discrimination may also arise from differences in the distribution of observed characteristics c among races (i.e., $F(c|W) \neq F(c|A)$). Distinguishing between these alternatives would require comprehensive data on c.

C. *Discussion of the Model*

Our model assumes that motorists respond to the probability of being searched. This assumption is key to obtaining a test for prejudice that can be applied without data on all the characteristics police use in the search decision. If motorists did not react to the probability of being searched, testing for prejudice would require data on c.

To see why, consider a model in which the probability that a motorist with characteristic c and race r carries drugs is a primitive (does not depend on the actions of police). In this model, the only optimizing

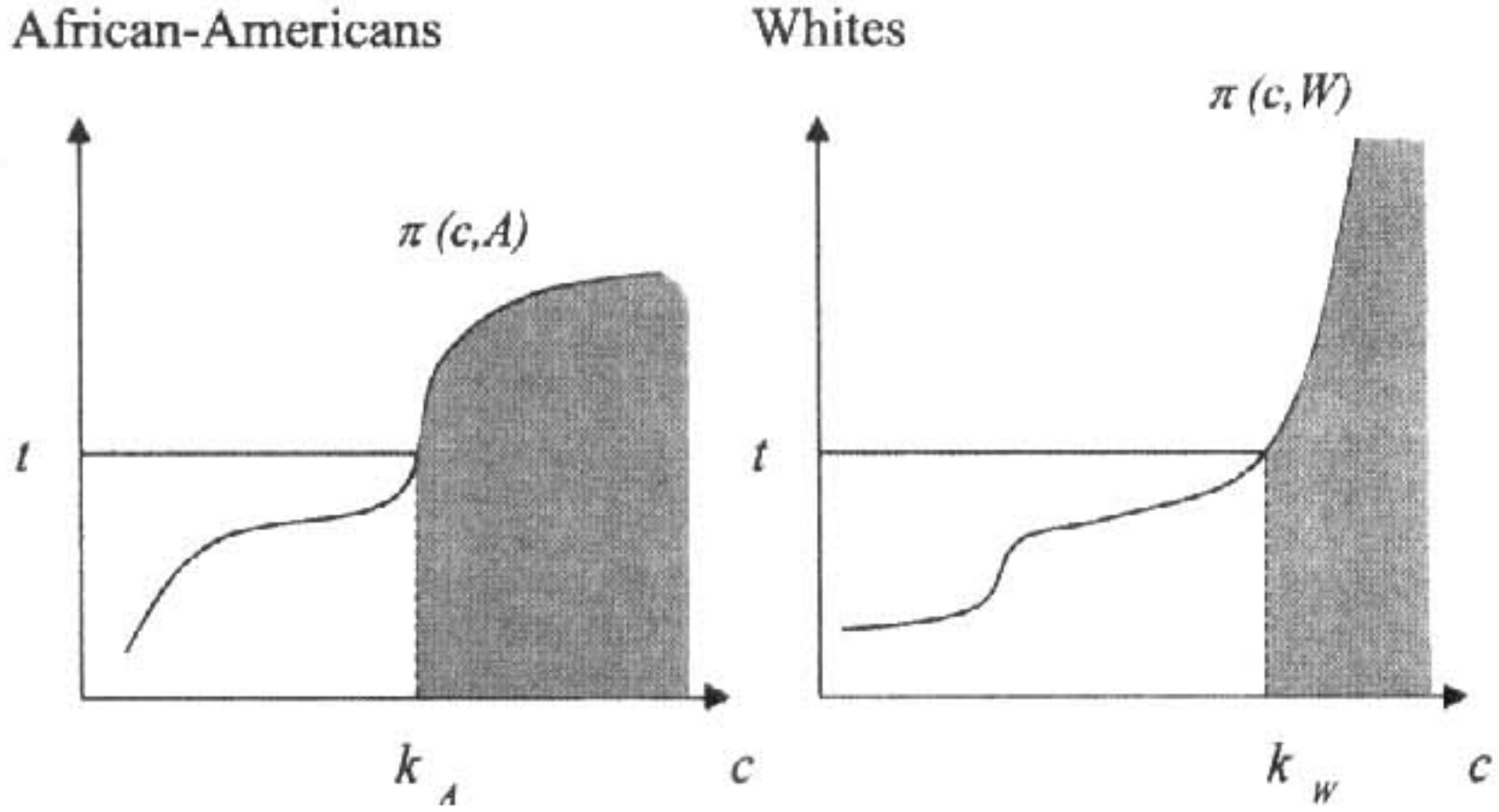


FIG. 1.—Model in which motorists do not respond to the probability of being searched

agents are the police. Let $\pi(c, r)$ denote the probability (now, exogenously given) that a type c, r carries drugs. If one supposes that $\pi(c, r)$ is increasing in c, it is optimal for police officers to choose two cutoffs, k_W and k_A, and search any motorist of race r with a c greater than k_r. In the absence of prejudice, police will choose k_W and k_A so that the probability that types k_W, W and k_A, A are guilty equals the marginal cost t of searching motorists, as shown in figure 1.

What we observe in the data is the average probability of guilt of types with c above k_r. Without data on c, we cannot identify the marginal motorists and so cannot test this model in the absence of assumptions on the shape of $\pi(c, r)$ and on the distribution of the unobservables. Thus we cannot estimate the equilibrium implications of a model in which motorists do not react to the probability of being searched, except under strong assumptions.

How does the endogenous response of motorists change figure 1? All motorists above the cutoff k_r, knowing that they are going to be searched for sure, would respond by not carrying drugs. All motorists below the cutoff would similarly react by choosing to carry drugs.[15] Then $\pi(c, r)$ would become *decreasing* in c, which would make suboptimal the policy of searching motorists above k_r. Therefore, when $\pi(c, r)$ is determined endogenously, the only equilibrium is for $\pi(c, r)$ to equal t, for all c.

[15] In this stylized model, without loss of generality, we restrict attention to the population that has some positive value from carrying drugs. We could easily add a fraction of motorists who police know would never want to carry drugs (even if they are never searched), and the police would therefore never search these motorists in equilibrium.

Now, let us return to the model in which motorists react to the probability of being searched. A number of stylized features of the model can be made more realistic without jeopardizing the test for racial prejudice.

First, it is unrealistic that at equilibrium motorists randomize over whether to carry drugs. The model developed earlier assumes that the characteristics c in the individuals' utility function are the same as those observed by the police officer (i.e., no private information). We can obtain a version of our model in which motorists have private information about their utility from carrying drugs, and they never randomize. This is achieved by adding a random variable X to the utility $v(c, r)$ of each motorist, as in a random utility model (McFadden 1984). The realization of X represents the idiosyncratic component of a motorist's propensity to carry drugs, which is private information. Given a certain probability of being searched, motorists with a high realized value of X strictly prefer to carry drugs and those with small values strictly prefer not to carry. Thus no motorist randomizes. But police cannot distinguish individuals inside each group, so whether the decision to carry drugs depends on X or on the roll of dice is observationally equivalent to the police. Therefore, as before, if police are not racially prejudiced, then in equilibrium expected returns to searching motorists must be the same across all groups searched (i.e., all groups searched are marginal). So, our test for racial prejudice continues to be valid in this environment.[16]

The "random utility" model accommodates the presence, in each group c, r, of some individuals who may not be deterred from carrying drugs except by very high search intensities. These are individuals with high values of X. We could think of these individuals as drug dealers, whose responsiveness to the probability of being searched would depend on the elasticity of the demand for drugs, which is generally thought to be quite inelastic. Conversely, the model also accommodates the presence, in each group c, r, of some individuals who may not derive any utility from carrying drugs ($v(c, r) + X \leq 0$). If these individuals make up a sufficiently large fraction of their group, the likelihood that a member of that group carries drugs could not reach t_r, and in equilibrium the police will not search that group (for that group $\gamma^*(c, r) =$

[16] As before, police would search with probability one (zero) motorists in groups with a high (low) probability of being guilty, which is inconsistent with equilibrium. For our test to fail, we would need to have a fraction of "crazy" criminals who are not deterred even if they know *for sure* that they are going to be caught. If these crazy criminals are sufficiently numerous in group c, r, then in equilibrium that group is searched with probability one and the success rate of police on that group is exactly equal to the fraction of crazy criminals. In such a model we have a divergence between the marginal and average rate of success of searches. In the absence of data on c, the only prediction from this model that we can test is that the success rate of police on race r must be greater than t_r. The argument is outlined more formally in Knowles, Persico, and Todd (1999).

0). Yet, it is still the case that the probability of being guilty is equal among all groups *that are searched in equilibrium.*

Our test for prejudice would also hold under the alternative specification that police maximize the number of arrests subject to a resource constraint on the total time spent searching in the random utility model. In this scenario, police would target its available resources on those groups in which the fraction of motorists carrying drugs is highest. If the expected return to the police were not equal across all groups that are searched with positive probability, the police could reallocate its (given) resources to achieve a higher number of arrests. Thus, even under this alternative specification, the equilibrium probability of being guilty should be constant among all groups searched.

Another stylized feature of our model that can be relaxed is that individuals do not get to choose their characteristics. In reality, some characteristics—such as tinted windows—can plausibly be viewed as endogenous. When characteristics can be purchased, some types with characteristics closely linked to criminal behavior will choose to purchase more innocuous characteristics. In equilibrium, police will take this into account when computing the probability of being guilty of motorists in a certain group. In an appendix available on request from the authors, we extend the model to incorporate endogenous characteristics and show that the key implication of the model is maintained.

Finally, our model abstracts from the issue of the thoroughness of searches. Suppose that it were the case that police search African-American motorists more thoroughly than whites, because of a lower "cost of thoroughness." As a result, searches of African Americans would not necessarily be more successful, because of the equilibrium reaction of motorists. In fact, we may expect searches of African Americans to be less successful since in equilibrium police equate the (lower) cost of searching thoroughly to the expected benefit from searching. Testing a model that takes into account thoroughness requires data on effort spent searching. In the absence of such data, we leave this question for future research.

III. Empirical Results

A. Data Description

We now apply the test described above to data that were collected as part of the settlement of a lawsuit filed in February 1993 by the ACLU. The lawsuit challenged as unconstitutional the Maryland State Police's alleged use of racial profiling as a basis for stopping, detaining, and searching motorists. As part of the settlement, the state agreed to maintain detailed records on motorist searches and to file quarterly reports

with the court and the ACLU. However, in November 1996, the Maryland ACLU asked a federal court to hold the state police agency in contempt of court on the basis of evidence that it continued a pattern of racial discrimination in its motor vehicle searches.

Our data set consists of 1,590 observations on all motor vehicle searches on a stretch of Interstate 95 in Maryland from January 1995 through January 1999.[17] The data provide information 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 police officer requested consent to search the car and, if consent to search was not granted, the probable cause that the police officer invoked to search the car.[18] In addition, we know whether dogs were used in the search and whether illegal drugs were found. If any drugs were recovered, we know what type and how much.[19] Finally, the data include the name (but not the race) of the police officer performing the search.

It is important to note that our data pertain only to motorists who were both stopped and searched; they do not include motorists who were stopped but not searched. While data on stops might reflect differences in driving habits between different race and sex groups, search data include only motorists whom police officers suspect to be carrying contraband. The total population searched is the relevant one for the denominator of the success rate $P(G|c, r)$.

One might be concerned that police require a legal pretext for stopping motorists before searching them. In practice, this constraint does not bind. The Supreme Court upheld the right of police to stop any motorist in violation of the law, even though it acknowledged that motor vehicle regulations are so restrictive and complex that this effectively gives the police the right to arbitrarily stop any motorist (*Whren and Brown v. US* in 1996).[20] Part of the controversy about racial profiling is that state troopers often use minor traffic violations, such as exceeding

[17] The searches were conducted in Baltimore, Cecil, Harford, Howard, and Prince George's counties. Because the data are available only 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.

[18] If the driver refuses to give consent, the search can still be performed if the police have probable cause. In our data, a search is always performed, and only a small fraction of motorists refuse to give consent (35, or 2 percent, with only eight found guilty). We take the fact that so few guilty motorists refuse consent to search as an indication that refusing consent does not appreciably decrease the probability of being searched.

[19] A few amounts were reported as "residue" or "trace amount," or in terms of small numbers of cigarettes. We classified these amounts as "less than 2 grams."

[20] Also, the vast majority of drivers along Interstate 95 commit speeding violations. According to expert witness testimony by John Lamberth in *Wilkins v. Maryland State Police* (Civil Action no. CCB-93-483, 1996), 98.1 percent of all cars on a stretch of the New Jersey Turnpike were clearly exceeding speed limits.

the speed limit by five miles per hour, as pretexts to stop and search the motorist, other occupants, and the vehicle.[21]

B. *Empirical Findings*

Our test for prejudice compares the probability of being found guilty across groups with different observed characteristics. The model developed in Section II has a strong implication: no matter what the set of characteristics, the guilty rate should be the same across groups. Therefore, we can test the model by testing the null hypothesis

$$\Pr(G = 1 | r, c) = \Pr(G = 1) \quad \text{for all } r, c,$$

where c is a set of characteristics and r is the race indicator variable. The probability could be estimated using a parametric procedure such as probit or logistic regression. With discrete regressors, a test for equality of guilt probabilities corresponds to a test of whether all the coefficients associated with the conditioning variables are equal to zero. This is a more stringent requirement than the conventional test of whether the coefficient on race is statistically different from zero. (Also, the conventional test is usually applied to the probability of being searched, not the probability of being guilty.)

A drawback of using a parametric estimation approach is that the test will generally be valid only if the systematic component and distribution of the error component of the model are correctly specified.[22] Instead, we use a simpler nonparametric test, the Pearson χ^2 test, which compares the proportion guilty within conditioning cells against the proportion that would be expected under the null hypothesis of no association between guilt and the conditioning characteristics. For example, the test statistic for testing the hypothesis of no association between guilt and race is given by

$$\sum_{r \in \mathcal{R}} \frac{(\hat{p}_r - \hat{p})^2}{\hat{p}_r} \sim \chi^2(R - 1),$$

where R is the cardinality of the set of race categories, $\mathcal{R}$, and $\hat{p}_r$ and $\hat{p}$ are conditional and unconditional estimated guilt proportions.

[21] In 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.

[22] However, a fully saturated parametric model with binary regressors would consistently estimate the conditional guilty probabilities even if the error distribution were misspecified.

TABLE 1
MEANS AND STANDARD ERRORS OF VARIABLES USED IN ANALYSIS

		By Race				By Sex	
	All Observations	African-American	Hispanic	White	Other	Female	Male
African-American	.63	1.00	.00	.00	.00	.64	.63
	(.01)	(.00)	(.00)	(.00)	(.00)	(.04)	(.01)
White	.29	.00	.00	1.00	.00	.35	.29
	(.01)	(.00)	(.00)	(.00)	(.00)	(.04)	(.02)
Hispanic	.06	.00	1.00	.00	.00	.00	.07
	(.01)	(.00)	(.00)	(.00)	(.00)	(.00)	(.01)
Female	.07	.07	.00	.09	.30	1.00	.00
	(.01)	(.01)	(.00)	(.01)	(.05)	(.00)	(.00)
Cocaine	.08	.10	.03	.04	.15	.09	.08
	(.01)	(.01)	(.02)	(.01)	(.08)	(.03)	(.01)
Marijuana	.23	.23	.04	.26	.20	.21	.23
	(.01)	(.01)	(.02)	(.02)	(.09)	(.04)	(.01)
Crack cocaine	.04	.05	.01	.01	.05	.06	.04
	(.01)	(.01)	(.01)	(.004)	(.05)	(.02)	(.01
Heroin	.02	.02	.03	.03	.05	.06	.02
	(.004)	(.004)	(.02)	(.01)	(.05)	(.02)	(.003)
Morphine	.001	.00	.00	.002	.00	.00	.001
	(.001)	(.00)	(.00)	(.002)	(.00)	(.00)	(.001)
Other drugs	.01	.00	.00	.00	.00	.01	.01
	(.003)	(.00)	(.00)	(.00)	(.00)	(.01)	(.003)
Paraphernalia	.01	.003	.010	.02	.00	.01	.007
	(.002)	(.002)	(.01)	(.01)	(.00)	(.01)	(.002)
Older vehicle (10 years or older)	.22	.20	.247	.28	.20	.18	.23
	(.011)	(.013)	(.044)	(.021)	(.092)	(.036)	(.011)
Luxury model	.08	.10	.09	.05	.05	.08	.08
	(.007)	(.01)	(.03)	(.05)	(.01)	(.025)	(.007)
Third-party vehicle	.18	.22	.25	.09	.20	.20	.18
	(.010)	(.013)	(.044)	(.014)	(.092)	(.027)	(.010)
Night (midnight to 6 A.M.)	.31	.35	.44	.23	.25	.35	.43
	(.01)	(.02)	(.05)	(.02)	(.10)	(.04)	(.01)
Number of observations	1,590	1,007	97	466	20	117	1,473

NOTE.—Standard errors of the means are shown in parentheses.

1. Descriptive Statistics

Table 1 summarizes the means and variances of variables we use in our empirical tests. Of the 1,590 total searches, 1,007 (63 percent) were performed on African Americans, 466 (29 percent) on whites, 97 (6 percent) on Hispanics, and the remaining on other race/ethnic groups. Female motorists were rarely searched: a total of 117 female motorists appear in the data compared with 1,473 men. Marijuana is the drug most commonly found (23 percent of all searches), and it is not uncommon for drivers to be carrying up to three different types of drugs

as well as drug paraphernalia. Among the drivers searched, about 8 percent drive cars generally considered to be in the class of luxury cars, with a higher proportion of African-American and Hispanic drivers driving luxury cars.[23] About a third of the searches occur during the hours of midnight to 6 A.M.

Figure 2 plots the time trend in the proportion of drivers searched who were African-American. The circle radius is proportional to the sample size. The figure reveals a downward trend over time in the proportion of African-American drivers searched (panel *a*) and an upward trend in the proportion of white drivers searched (panel *b*). There is no clear trend for female motorists (panel *c*). If police practices changed over time, then our test results could vary on the basis of the date of observation. However, our findings are robust across data subperiods, as discussed below.

Figure 3 displays the histogram of quantities of drugs found by race/ethnicity. Panel *a* shows the quantities for all drugs other than marijuana and panel *b* the quantities for marijuana only. Among drivers found with marijuana, whites and Hispanics are more likely to be carrying small quantities. Among drivers found with other "harder" kinds of drugs, African Americans and Hispanics are more likely to be carrying amounts in excess of 100 grams.

2. Test Results

To implement our test for detecting racial prejudice in police searches, we need to define what it means to be guilty. Under the definition guilty 1, we define motorists as guilty if they are found with any amount of marijuana, heroin, cocaine, crack, PCP, LSD, and methadone.[24] In Section III*B*3, we experiment with other, more lenient definitions of guilty.

Panel A of table 2 compares the probability of being found guilty by race. In this subsection we focus on column 1 (guilty 1). Although African-American motorists are much more likely to be searched by police, the proportion of guilty motorists among whites and African Americans whose cars are searched is nearly identical (0.32 vs. 0.34)—a result that is consistent with the hypothesis of no racial prejudice. For Hispanics, however, the guilty rate is 0.11, which is significantly lower than for African Americans or whites. Table 3 reports *p*-values for Pearson χ^2 tests of the null hypothesis of no association between guilt and various conditioning characteristics, including race. When all racial groups are used in the test, the null hypothesis is rejected at conven-

[23] We categorized the cars as luxury or not on the basis of a listing at the Internet site www.autobytel.com.

[24] A small number of individuals were found in possession of barbiturates (such as valium); we did not classify them as guilty.

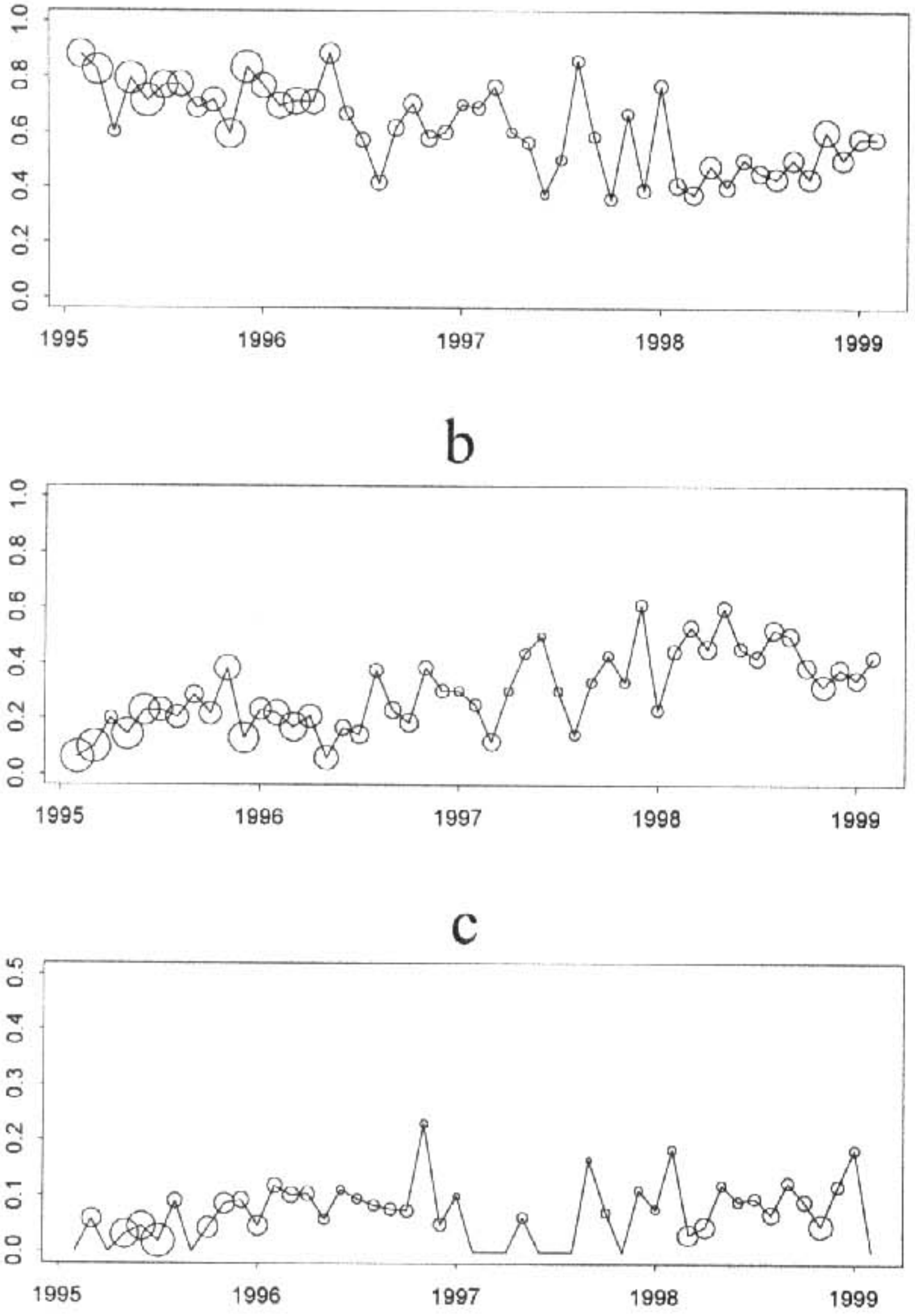
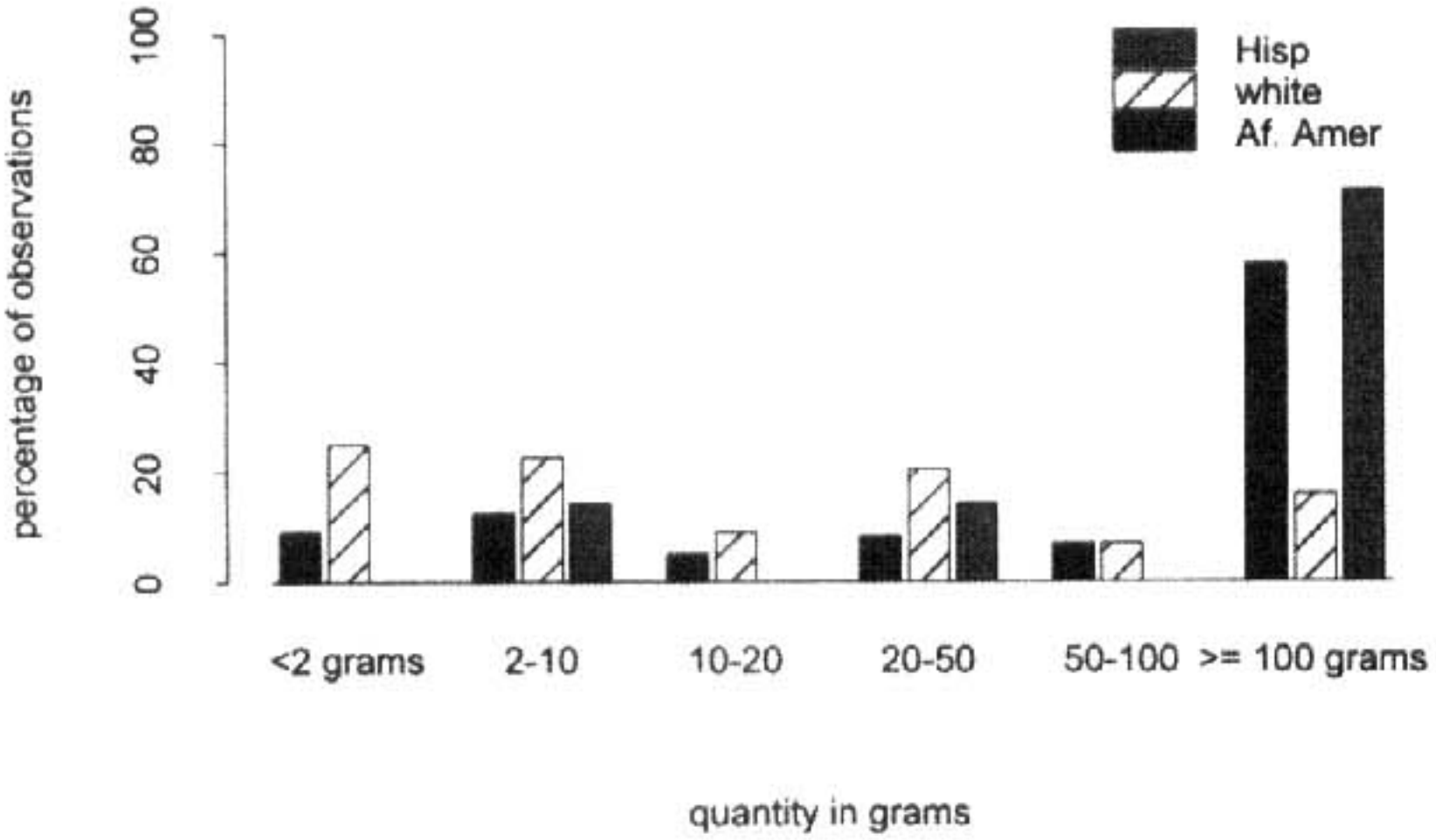


FIG. 2.—Proportion of drivers searched: *a*, black; *b*, white; *c*, female

a



b

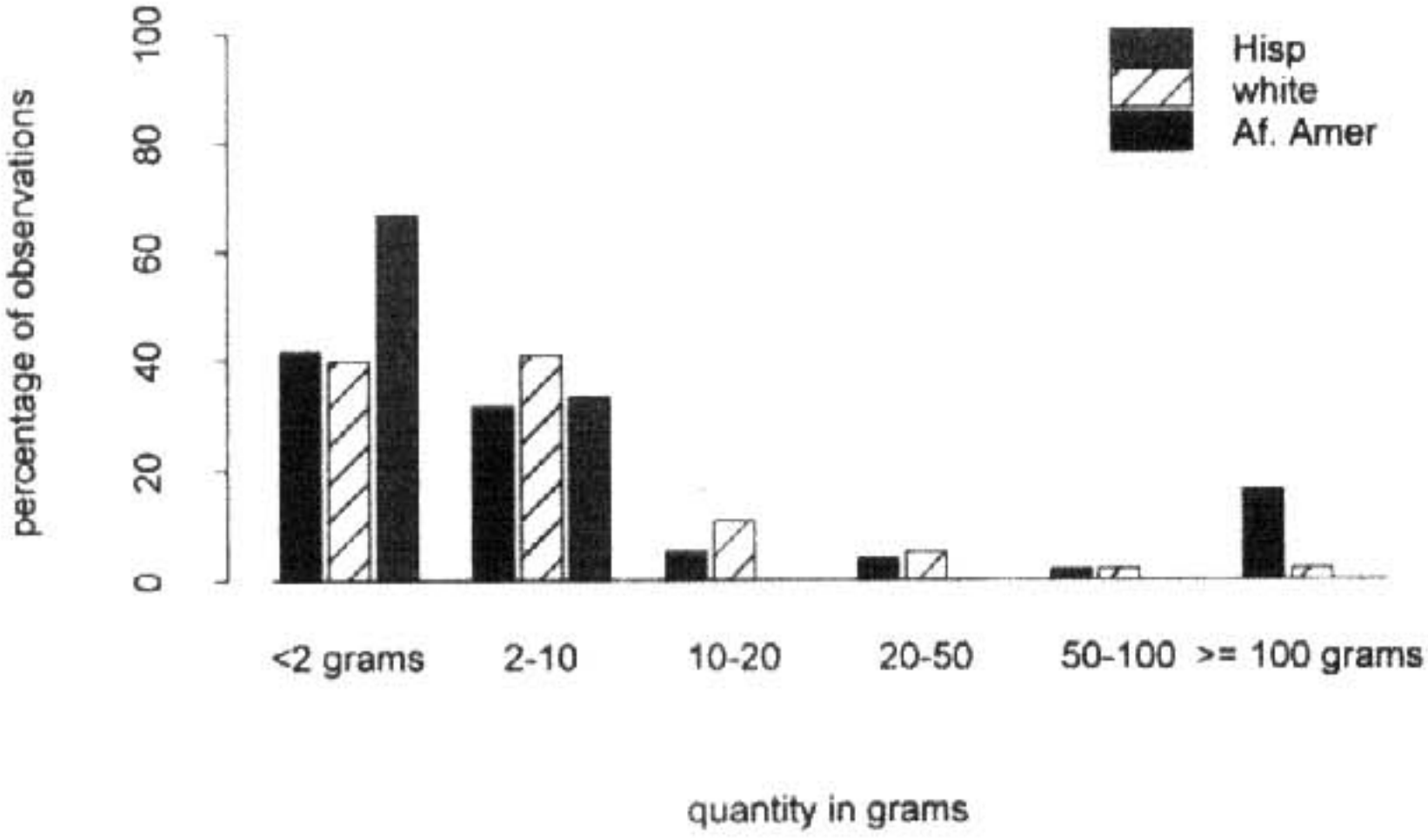


FIG. 3.—Quantities of drugs by race/ethnicity: *a*, hard drugs; *b*, marijuana

TABLE 2
PROPORTION OF VEHICLES SEARCHED FOUND TO BE CARRYING DRUGS

	Definition			
	Guilty 1 (1)	Guilty 2 (2)	Guilty 3 (3)	Guilty 4 (4)
	A. By Race/Ethnicity			
African-American	.34	.26	.16	.13
White	.32	.21	.07	.03
Hispanic	.11	.08	.08	.06
Other	.30	.30	.15	.10
	B. By Sex			
Male	.32	.23	.12	.09
Female	.36	.27	.17	.15
	C. By Time of Day			
Day	.32	.24	.13	.09
Night	.33	.23	.13	.10
	D. By Type of Car			
Luxury	.25	.19	.12	.10
Not luxury	.33	.24	.13	.09
Older car (≥10 years old)	.33	.16	.13	.08
Newer car (<10 years old)	.32	.15	.13	.10
Third-party vehicle	.29	.22	.19	.17
Own vehicle	.33	.14	.11	.08
	E. By Race/Ethnicity and Sex			
Male:				
African-American	.34	.25	.15	.12
White	.33	.21	.08	.03
Hispanic	.11	.08	.07	.06
Other	.32	.32	.16	.11
Female:[a]				
African-American	.44	.32	.23	.21
White	.22	.17	.07	.02

NOTE.—Guilty 1 includes drugs in any amount. Guilty 2 excludes less than 2 grams of marijuana. Guilty 3 excludes marijuana in any amount. Guilty 4 includes only felonious amounts.

[a] There are no Hispanic or Other females in our data.

tional significance levels. However, when only whites and African Americans are used in the test, the hypothesis is not rejected. Thus our findings suggest that police search behavior is not biased against African-American drivers. The lower guilty rates for Hispanics are suggestive of prejudice against this group.

Because our model implies equal guilt probabilities across all observable characteristics, in panels B–E of table 2, we compare the guilty rate for a variety of other conditioning variables, including the sex of the driver, time of day of the search, whether the car is a luxury model, whether the car is older than 10 years, and whether or not the driver is the owner of the car (third-party vehicle).

TABLE 3
p-VALUES FROM PEARSON χ^2 TESTS ON HYPOTHESIS THAT PROPORTION GUILTY IS EQUAL ACROSS VARIOUS GROUPINGS

Groups	Guilty 1	Guilty 2	Guilty 3	Guilty 4
Race:				
African-American, Hispanic, and white	<.001	<.001	<.001	<.001
African-American and white	.33	.05	<.001	<.001
African-American and white, males only	.75	.08	<.001	<.001
African-American and white, females only	.02	.13	.04	.01
Sex: male and female	.37	.52	.18	.06
Sex and race:				
African-American, Hispanic, and white, male and female	<.001	<.001	<.001	<.001
African-American and white, male or female	.10	.12	<.001	<.001
Luxury model	.06	.2	.77	.91
Older vehicle	.67	.53	.92	.19
Third-party vehicle	.28	<.001	<.001	<.001
Time of day: day/night	.69	.74	.71	.42

The proportions guilty by sex are similar, and the Pearson χ^2 test does not reject the null that the proportions do not differ by sex (see table 3, row 5). When we condition simultaneously on race and sex, however, we find that the proportions guilty for white and African-American males are very similar, but the proportions are somewhat lower for white women than for African-American women (0.22 vs. 0.44). The χ^2 test does not reject the null hypothesis of equal rates across racial groups (whites and African Americans) for males, but does reject it for females. These results could be interpreted as showing that police are biased against white women, but an alternative interpretation is that they perhaps derive some extra utility from searching cars of white women.[25]

Finally, tests of equality of guilty rates according to whether the car is a luxury model, whether the car is more than 10 years old, whether the driver owns the car (third-party vehicle), and time of day of the search all do not reject the null hypothesis that the guilty rates are equal across the different sets of conditioning characteristics.[26]

Our data were collected during three distinct time periods: first, when the police were being audited as part of a lawsuit; second, when the audit period was over but police were still gathering data as part of the settlement agreement; and third, after the ACLU filed a second lawsuit alleging that the police were still discriminating. During the audit period and after the filing of the second lawsuit, more searches were con-

[25] We are grateful to an anonymous referee for suggesting this interpretation.
[26] In these tests we include Hispanic motorists. The *p*-values generally increase when we exclude Hispanics.

ducted.[27] Tests combining all time periods might mask variation over time in police search behavior. To check the robustness of our findings, we performed identical tests after disaggregating the data by the three time periods: prior to May 1996, between June 1996 and December 1998, and after December 1998. Figure 4 plots the proportions guilty by race over time. The tests on disaggregated data are reported in Knowles et al. (1999) and generally yield the same conclusions as those performed on the full sample.

The fact that guilt probabilities are relatively stable over time suggests that the decline in the percentage of African Americans searched seen in figure 2 is not due to an increase in the relative utility cost of police of searching African Americans, because in equilibrium this would be reflected in a lower proportion guilty among African-American drivers searched. Rather, in the context of our model, the time trend suggests that the propensity of African-American drivers to carry drugs decreases over time relative to that of white drivers (perhaps because of increased economic opportunities for African Americans, greater penalties for this group, changing preferences for drug use, etc.).

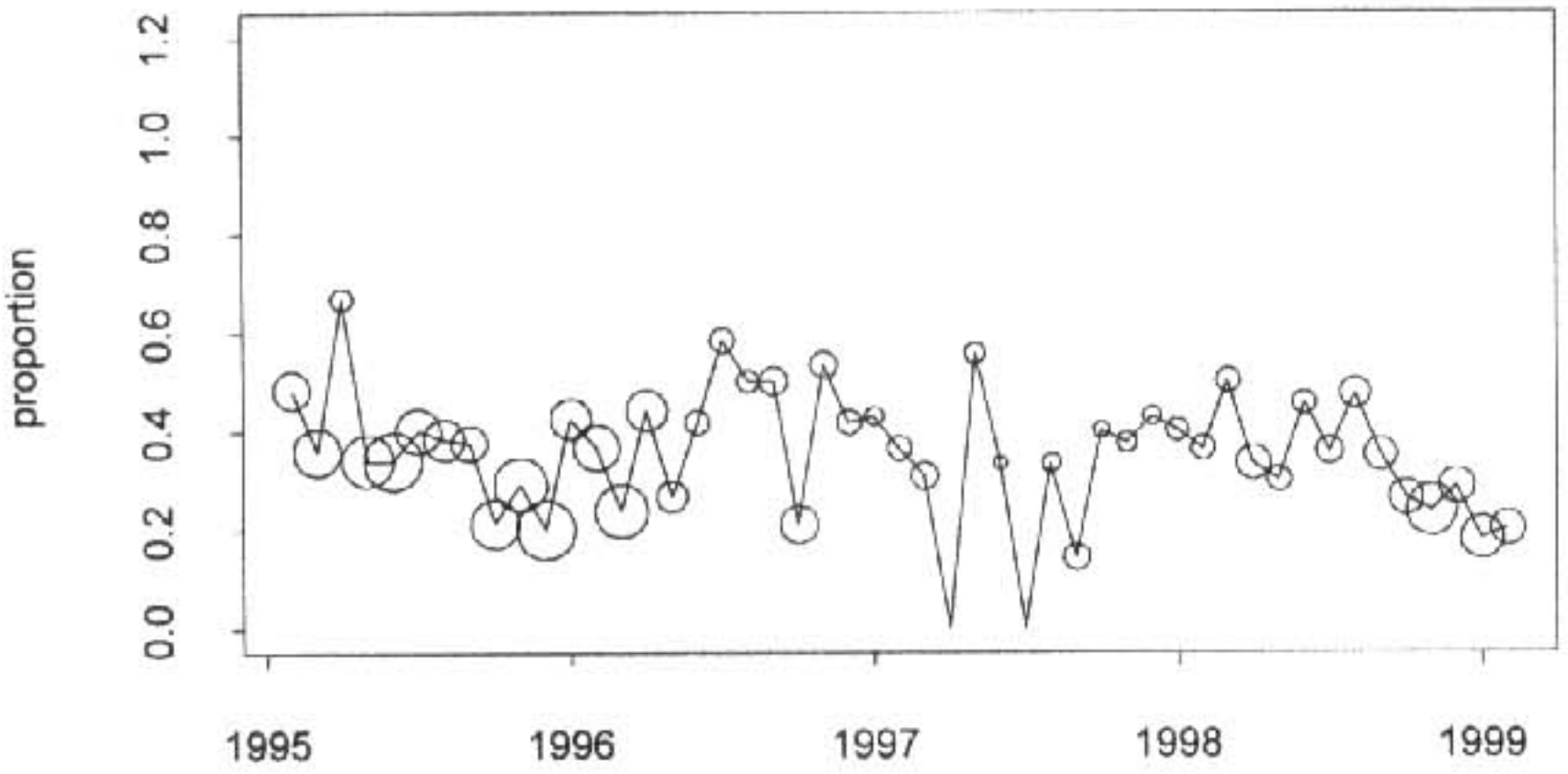
3. Other Models of Police Incentives

The tests above are based on the assumption that police value equally each successful search (one in which the driver is found guilty). It is possible, however, that police place more value on searches yielding a higher quantity of drugs. Apprehending a narcotics dealer may be considered more important than apprehending a person in possession of drugs intended for personal consumption. Therefore, a possible concern with the previous analysis is that extending it to account for drug quantities might uncover evidence of prejudice against African Americans. Alternatively, extending the analysis along these lines might help explain the apparent animus of police against Hispanics as simply reflecting the incentive to make large drug seizures. To address these issues, we now experiment with alternative definitions of a successful search.

We repeat the tests of the previous subsection using the following definitions of guilty. "Guilty 2" classifies as guilty only those motorists found with any amount of hard drugs (drugs other than marijuana) or with marijuana in excess of 2 grams. "Guilty 3" includes only motorists found with hard drugs. Finally, "guilty 4" includes only motorists in possession of sufficient amounts of drugs to constitute a felony under

[27] An ACLU NewsWire dated July 14, 1996, reported that "the Maryland State Police has reduced an 'elite' highway drug interdiction team after court-order statistics showed that most motorists searched by the team members along Interstate 95 are black" (www.aclu.org/news/w071496b.html).

a



b

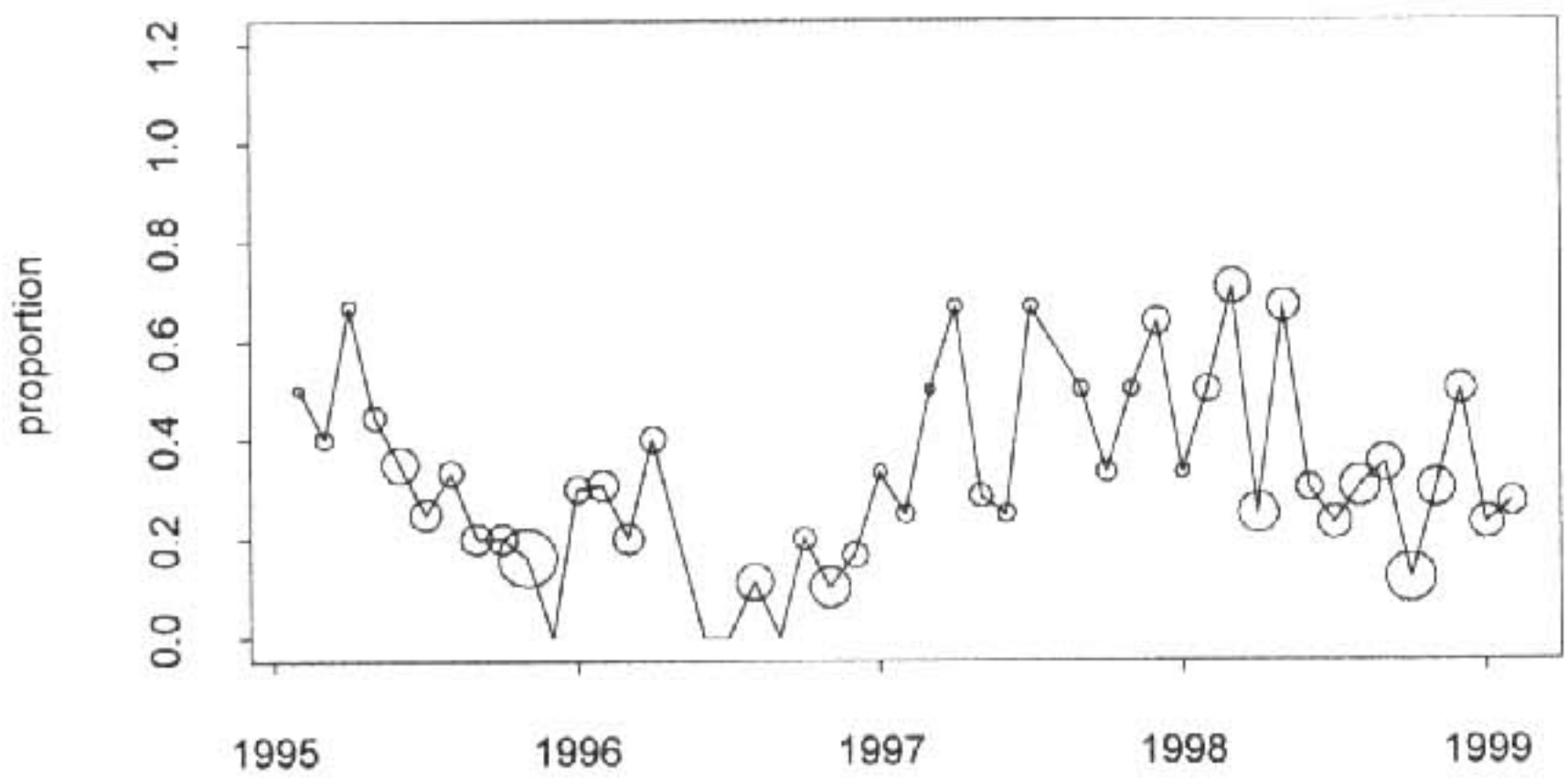


FIG. 4.—Proportion of drivers found with drugs: *a*, black; *b*, white

Maryland's drug laws.[28] Checking how the interpretation of our tests changes with these alternative definitions allows for the possibility that the police care more about finding motorists with larger quantities of drugs or with "harder" drugs.

Test results under alternative definitions of "guilty."—Definition guilty 2, which excludes individuals with less than 2 grams of marijuana, reproduces most of the results found earlier under definition guilty 1. The only changes are the following: (1) Under this definition, we reject the null hypothesis of no association between type of vehicle (luxury or not) and guilt probabilities: luxury cars are significantly less likely to yield seizures above 2 grams of marijuana than nonluxury cars (0.19 vs. 0.24). (2) We also reject the null for the variable "third-party vehicle" since third-party vehicles are more likely to yield seizures above 2 grams of marijuana.

Under the two most lenient definitions of guilt (guilty 3 and 4), results differ substantially and would imply that police behavior is biased against whites and Hispanics in favor of African Americans. The reason is that African Americans are significantly more likely than whites to be found guilty, under these definitions, when searched (0.16 vs. 0.7 for guilty 3 and 0.13 vs. 0.03 for guilty 4). This conclusion also holds when we restrict the population to one sex (males or females) and test for equality of guilty rates by race.[29] Under guilty 3, Hispanics have almost the same probability of being guilty as whites. However, the results for Hispanics should be interpreted cautiously since there are very few observations of guilty Hispanics under definitions guilty 3 and 4.

Tests of equality of guilty rates by type of vehicle (luxury or not), age of the vehicle, and time of day do not reject the null hypothesis of equality at conventional significance levels for all the definitions of guilt. However, tests involving a third-party vehicle reject the null under all definitions except guilty 1.

We place greater emphasis on the results obtained under the stricter definition 1 or 2 because we doubt that police are not rewarded for finding small quantities of controlled substances in vehicles. Under these definitions, our results are broadly consistent with no bias in police behavior along several dimensions of driver characteristics. The fact that searches of vehicles driven by white and African-American drivers are equally productive in terms of yielding positive quantities of drugs leads

[28] According to the Maryland Uniform Controlled Substance Act, it is a felony to bring into the state quantities of drugs in excess of the following amounts: 100 pounds of marijuana, 4 grams of opium or any opium derivative, 28 grams of cocaine, 1,000 dosage units of LSD, 28 grams of phencyclidine, 1,000 dosage units of methaqualone, and 4 grams of fentanyl (Source: *Digest of Criminal Laws*, available on the Internet at ns1.dpscs.state.md.us/pct/digest/n.htm).

[29] Under both definitions, proportions guilty are lower for white women than for African-American women, consistent with the interpretation given earlier.

us to conclude that police are not biased against African Americans. Under definitions guilty 3 and 4, whites who are searched are less likely to be guilty than African Americans. Hence, if we thought that police regard as unsuccessful all seizures that yield only marijuana or all seizures below the legal limits for a felony, then our model would suggest that police discriminate in favor of African Americans. This interpretation would be consistent with police's fear of future litigation and influence by public pressure against racial profiling.

The apparent prejudice against Hispanics persists under all definitions of guilt. We regard our results concerning Hispanics as only suggestive because our data set contains so few Hispanics. Further investigation is needed with a larger data set.

IV. 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 police troopers are needed to justify different types of tests. In this paper, we developed a simple equilibrium model of law enforcement via traffic searches and considered its implications for testing for racial prejudice in policing. Existing tests for discrimination typically regress an indicator for whether a motorist is searched on a number of characteristics and check whether race has any additional explanatory power. We discussed two disadvantages of these kinds of tests. First, their validity relies crucially on which variables are considered admissible (nondiscriminatory) variables that police can use in their search decision and on whether those variables are available in the data. Second, these tests are informative only on whether a disparity by race exists and not about the motivation for the disparity. The question of motivation plays a prominent role in racial profiling court cases.

Our equilibrium model of police and motorist behavior provides a test for whether racial disparities in motor vehicle searches reflect prejudice or instead are consistent with maximizing behavior by nonprejudiced police. The test is based on the success rates of police searches. It compares the probabilities that various subgroups of the population are found guilty of carrying contraband when searched. An attractive feature of our test is that it is valid even when the set of characteristics observed by the police is only partially observable by the econometrician.

Our model implies that at equilibrium, both races should have the same probability of carrying drugs, but one race may be searched more often than another. In fact, searching some groups more often than others may be *necessary* to sustain equality in the proportions guilty across groups. Sometimes, equality in the proportions guilty by race is used in court to argue that police are racist in searching a greater fraction of

cars of African-American motorists. The argument that infers racism from this evidence relies on two very strong assumptions: (1) that motorists of all races are equally likely to carry drugs and (2) that motorists do not react to the probability of being searched. Relaxing these assumptions, as we do in this paper, leads to a very different kind of test.

Our empirical results for the Maryland data showed that the probabilities of being found with drugs in any amount are equal across African Americans and whites, which is consistent with maximizing behavior by police who are not racially prejudiced. We also find equality of proportions guilty across several other dimensions of driver characteristics, which further supports the conclusion that police are trying to maximize the number of successful searches. When we look at the probability of being found with drugs in large quantities, this probability tends to be higher for African-American drivers, which would imply a bias against white motorists. Our finding of lower probabilities of guilt for Hispanics suggests that police may be biased against Hispanics, but the sample size for this group is small and further investigation is warranted.

Although this paper focuses on traffic searches, our analysis extends straightforwardly to some other similar settings; for example, our test could be applied to analyze 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 (see, e.g., *Anderson v. Cornejo*, 1999 [no. 97 C 7556]).

Statistical discrimination, even if not due to prejudice, may be considered unfair because innocent drivers experience different probabilities of being searched depending on their race. In ongoing research (Knowles et al. 1999), we address the issue of fairness and show that implementing color-blind search behavior does not necessarily entail a cost in terms of efficiency in interdiction.

References

Arrow, Kenneth J. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees. Princeton, N.J.: Princeton Univ. Press, 1973.

Ayres, Ian, and Waldfogel, Joel. "A Market Test for Race Discrimination in Bail Setting." *Stanford Law Rev.* 46 (May 1994): 987–1046.

Becker, Gary S. *The Economics of Discrimination*. Chicago: Univ. Chicago Press, 1957.

———. "Crime and Punishment: An Economic Approach." *J.P.E.* 76 (March/April 1968): 169–217.

Berkovec, James A.; Canner, Glenn B.; Gabriel, Stuart A.; and Hannan, Timothy H. "Race, Redlining, and Residential Mortgage Loan Performance." *J. Real Estate Finance and Econ.* 9 (November 1994): 263–94.

———. "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending." *Rev. Econ. and Statis.* 80 (May 1998): 241–50.

Blank, Rebecca M. "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from *The American Economic Review*." *A.E.R.* 81 (December 1991): 1041–67.

Border, Kim C., and Sobel, Joel. "Samurai Accountant: A Theory of Auditing and Plunder." *Rev. Econ. Studies* 54 (October 1987): 525–40.

Donohue, John J., III, and Levitt, Steven D. "The Impact of Race on Policing, Arrest Patterns, and Crime." Working Paper no. W6784. Cambridge, Mass.: NBER, November 1998.

Epstein, Richard A. *Forbidden Grounds: The Case against Employment Discrimination Laws*. Cambridge, Mass.: Harvard Univ. Press, 1992.

Heckman, James J., and Siegelman, Peter. "The Urban Institute Audit Studies: Their Methods and Findings." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael Fix and Raymond J. Struyk. Washington: Urban Inst. Press, 1993.

Kennedy, Randall. *Race, Crime, and the Law*. New York: Pantheon Books, 1997.

Knowles, John; Persico, Nicola; and Todd, Petra. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." Working Paper no. W7449. Cambridge, Mass.: NBER, December 1999.

McFadden, Daniel L. "Econometric Analysis of Qualitative Response Models." In *Handbook of Econometrics*, vol. 2, edited by Zvi Griliches and Michael D. Intriligator. New York: North-Holland, 1984.

Reinganum, Jennifer F., and Wilde, Louis L. "Equilibrium Verification and Reporting Policies in a Model of Tax Compliance." *Internat. Econ. Rev.* 27 (October 1986): 739–60.

Ross, Stephen L. "Mortgage Lending Discrimination and Racial Differences in Loan Default." *J. Housing Res.* 7, no. 1 (1996): 117–26.

Scotchmer, Suzanne. "Audit Classes and Tax Enforcement Policy." *A.E.R. Papers and Proc.* 77 (May 1987): 229–33.

Smart, Scott, and Waldfogel, Joel. "A Citation-Based Test for Discrimination at Economics and Finance Journals." Working Paper no. 5460. Cambridge, Mass.: NBER, February 1996.

Stigler, George J. "The Optimum Enforcement of Laws." *J.P.E.* 78 (May/June 1970): 526–36.

Van Order, Robert, and Zorn, Peter. "Testing for Discrimination: Combining Default and Rejection Data." In *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics*, edited by Anthony M. Yezer. Washington: American Bankers Assoc., 1995.