Author Query Sheet

Manuscript Information	
Journal Acronym	RJQY
Volume and issue	24:4
Author name	PERSICO AND TODD
Manuscript No. (if applicable)	271625

AUTHOR: The following queries have arisen during the editing of your manuscript. Please answer the queries by making the necessary corrections on the CATS online corrections form. Once you have added all your corrections, please press the SUBMIT button.

QUERY NO.	QUERY DETAILS
1	Numbered sections are mentioned. For clarity, please specify the section name instead.
2	“increasing the **(psychic)** benefit” -- Is this correct?
3	Some keywords have been added. Please check.

JUSTICE QUARTERLY VOLUME 24 NUMBER 4 (DECEMBER 2007)

The Hit Rates Test for Racial Bias in Motor-Vehicle Searches

Nicola Persico and Petra E. Todd

This paper exposits a rational choice model of police–motorist interactions originally developed in Knowles, Persico, and Todd (Knowles, J., Persico, N., & Todd, P. [2001]. Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, *109*, 203–299). It discusses how the model provides a test for detecting bias in police enforcement based on the success rate of searches and presents empirical evidence on the model's implications based on policing data from Maryland.

AQ3 *Keywords* hit rates test; motor-vehicle searches; racial bias

Introduction

This paper exposits a rational choice model of police–motorist interactions and addresses a number of concerns raised in a recent article by Engel (2007). The policing model was originally developed in Knowles, Persico, and Todd (2001) and generalized in Persico and Todd (2005, 2006). The model assumes that police make decisions about which vehicles to search for drugs/contraband, taking into account the benefits of searching different kinds of motorists, and motorists make decisions about whether to carry drugs/contraband, taking into account the risk of getting searched. We discuss the model's key assumptions in nontechnical terms and its implications. The most significant testable implication is that unbiased police should equalize hit rates across observable categories of drivers. We present empirical evidence in support of this implication.

Nicola Persico is a Professor of Economics, and a Professor of Law and Society, at New York University. His research interests include policing, and the law and economics of discrimination. Petra E. Todd is a Professor of Economics at University of Pennsylvania. Her research interests include policing and evaluating the effects of program interventions. Correspondence to: Nicola Persico, Department of Economics, NYU, 19 W. 4th Street, New York, NY 10012, USA. Email nicola@nicolapersico.com

Routledge Taylor & Francis Group
ISSN 0741-8825 print/1745-9109 online/07/040741-17

DOI: 10.1080/07418820701717201

Model

The rational choice policing model of Knowles et al. (2001) describes the behavior of police and of motorists. It assumes that there are many motorists, each of whom makes a dichotomous choice: whether or not to transport drugs or other contraband. These motorists are outwardly distinguishable only by race, A (African American) and W (White), but they are inwardly heterogeneous in their propensity to carry drugs: some motorists (the majority, perhaps) are not intent on carrying drugs; others may be willing to carry if their likelihood of being searched does not exceed some threshold.[1]

There are many police officers, each choosing which type of motorists to search, A or W. All searches have the same intensity, and each officer has exactly one search to allocate. We assume that all police officers seek to find drugs, and we allow for the possibility that some police intend to discriminate.

We assume that an unbiased police officer will choose whom to search in the pursuit of successful searches only, where a successful search is one that yields a find. To this end, unbiased officers will focus their searches on whichever group presents the highest likelihood of success. A biased police officer, on the other hand, also takes some pleasure from the mere act of searching a given type of motorist, say a motorist of type A. Operationally, our definition means that the biased officer will favor searching A's even when this group presents a somewhat lower likelihood of success relative to W's. The extent to which a biased officer is willing to trade off a lower likelihood of success against the pleasure of searching A's measures the intensity of the bias.[2,3]

The reader familiar with Game Theory will have noticed that our simple model is in fact a game. In what follows, we shall explore the workings of this simple game; more formally, we shall characterize the Nash Equilibrium. The aim is to develop an intuition for the features of the data that are relevant when checking for police bias.

It is important to note that this model is a much-simplified version of reality, because it assumes away many realistic details. Some of these so-called "frictions" may change the implications of the model, while others may be incorporated without altering the key implications. We describe below the robustness properties of the model when some of its assumptions are changed.

1. We assume two races here only for exposition. The framework is generalizable to more categories, based on race or other characteristics, such as age or gender.

2. Besides lowering the cost of searching A's, bias could also be modeled as increasing the (psychic) benefit to the officer of finding a guilty member of group A. This is immaterial for our analysis. Our conclusions are unchanged if we admit this additional channel for bias. AQ2

3. A comment is in order. The reader might argue that, in using race to choose whom to search, the police are behaving impermissibly, even if they are not motivated by bias. A more realistic labeling of the variables in our model should be "individuals who share a characteristic (or set of characteristics) that is more frequently found in the African American (resp., White) population."

Intuition for the Workings of the Model and Derivation of the Hit Rates Test

To develop some intuition for the implications of this model, it may be helpful to introduce an analogy between searching motorists for drugs and sampling from an urn. Let us imagine that a racial group corresponds to an urn, which is filled with red balls representing drug-carrying motorists, and green balls representing honest motorists. The fraction of red to green balls in an urn represents the crime rate within that group. The unbiased police officer is deciding from which of the two urns to draw in hopes of drawing a red ball. If the proportions of red to green balls are as depicted in Figure 1, they would choose to draw from the W urn, because the proportion of red balls is greater in that urn.

Our immediate goal is to predict what type of police and motorist behavior is "stable" in this environment. By stable, we mean that no motorists or police officer wants to change their behavior. Technically, we are looking for the Nash Equilibrium of this game.[4] Let us start by imagining that all police officers only searched from the W urn. This is the situation depicted in Figure 2, where the arrow on top of urn W indicates that police search efforts are focused solely on that group (there is no arrow over urn A-A). Given that the police choose this particular enforcement pattern, motorists in group A would not be deterred from carrying drugs relative to group W, and we would expect proportionally more motorists choosing to carry drugs in group A than in group W. This is reflected in Figure 2, where the proportion of red balls is greater in urn A-A than in urn W.

Is the situation depicted in Figure 2 a Nash equilibrium? No, because it is not stable: the police officers would make their searches more effective by changing their search behavior. Just as sampling from urn A-A maximizes the chances of drawing a red ball, police officers who are motivated to find drugs will want to allocate some of their search efforts to group A.

Let us then follow the natural adjustment process, and examine whether shifting some search intensity towards group A, as depicted in Figure 3a and b,

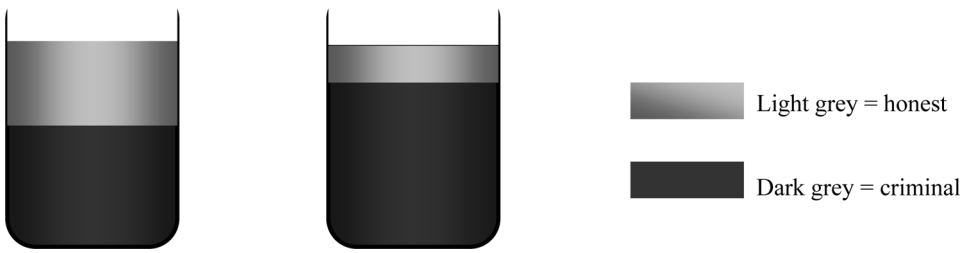

Figure 1 Urn model.

4. Formally, a game is in a Nash equilibrium when no player of the game finds it profitable to change their action, given the behavior of the other players. The Nash equilibrium is the standard solution concept used in game theory to predict how a game will be played by rational agents.

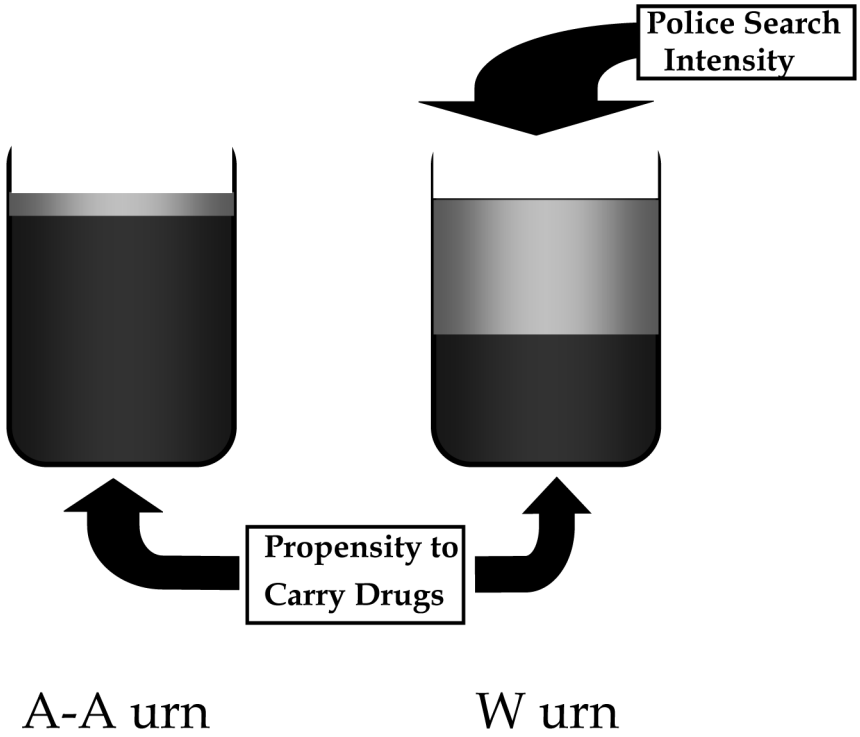

Figure 2 Disequilibrium.

puts the game in equilibrium. As the police progressively shift to searching more members of group A, the difference in crime rate between group A and W falls. This is because A motorists realize that they are being searched with increasing intensity, which discourages them from carrying drugs. Thus, the proportion of red balls is closer to equal across urns in Figure 3a than in Figure 2, and it becomes perfectly equal in Figure 3b.

In Figure 3b, the two urns have exactly the same proportion of red and green balls. At this point, a person who wants to maximize the chance of drawing a red ball would be indifferent as to which urn to sample. Analogously, at the level of police search intensity represented by the thick arrows in Figure 3b, a police officer has exactly the same chance of finding drugs when searching from either A or W motorists. For this reason, an unbiased police officer would be indifferent between groups. This is a stable situation, meaning that the game is at a Nash equilibrium. Observe that, for the game to be in equilibrium, the probability of finding drugs has to be equalized across the two groups. If the probability were lower in one of the groups, then unbiased police officers would want to reallocate search efforts away from that group, and so we would not have an equilibrium.

Consider now the case in which all police officers are to some degree biased against members of group A, in the specific sense that they derive additional pleasure from searching a member of the A group. This means that, if the two

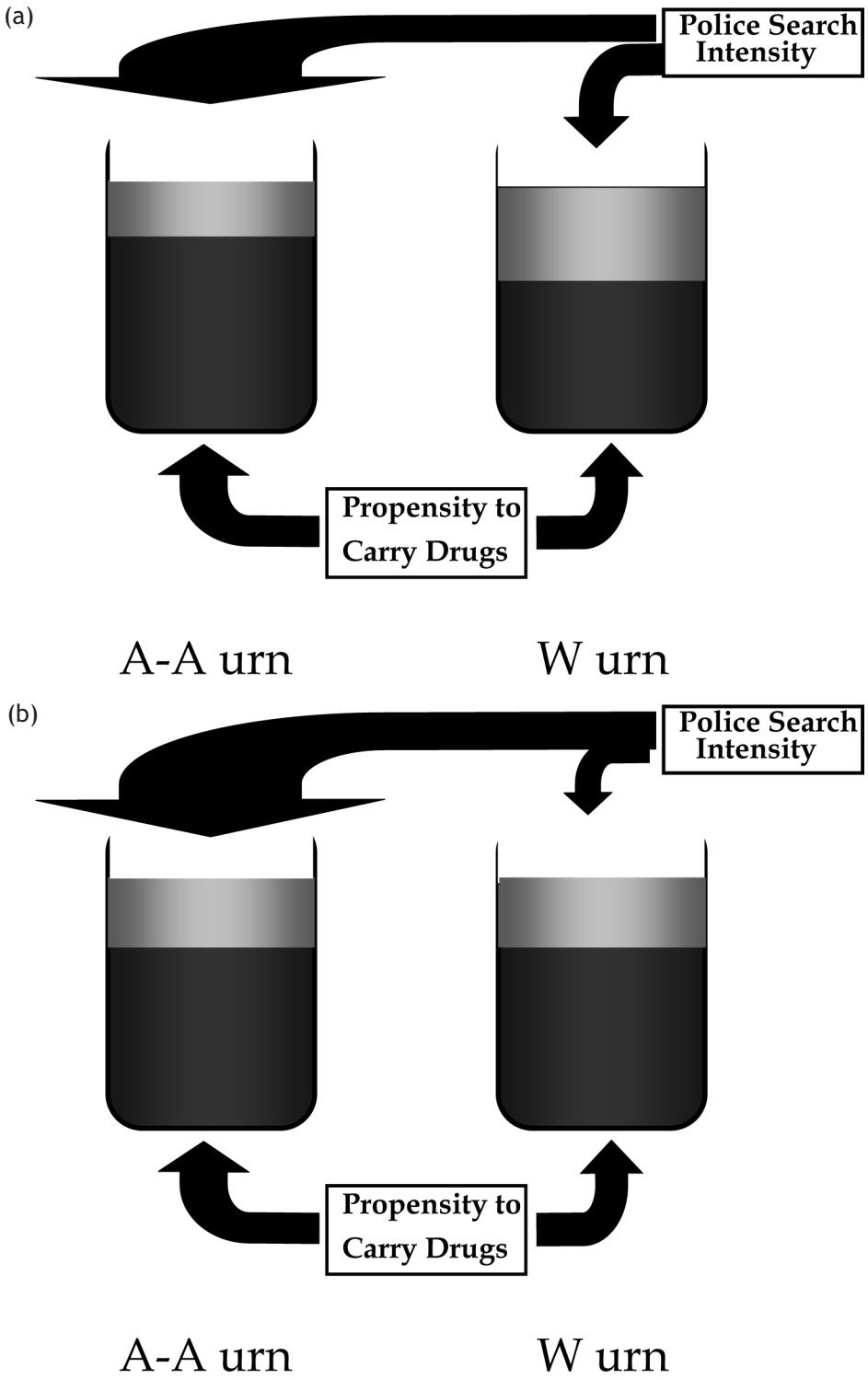

Figure 3 (a) Towards equilibrium. (b) Equilibrium with unbiased police.

groups had the same crime rate, all officers would strictly prefer to search A motorists over W motorists. In this case, the situation depicted in Figure 3b is no longer stable. We would expect officers to switch to searching group A motorists which, in turn, induces a reduction in the crime rate in group A, as motorists respond to the higher search intensity. Eventually, though, as officers increasingly shift towards searching A motorists, the higher success rate from searching group W motorists will exactly offset the extra pleasure that police derive from searching group A motorists, and police will again become indifferent between searching group A or W motorists. To summarize, when the police are biased against A, Nash equilibrium is achieved only when the crime rate among A motorists is lower than among W motorists.

The simple model exposited above has thus yielded a sharp implication. If a group is discriminated against, then searches of that group (the "hit rate") should be less successful. If, conversely, police are equally likely to find drugs upon searching a member of either group, i.e., the hit rates are equalized, then disparities in search frequencies across groups, while possible,[5] are not the result of police bias. This is a useful finding because it relates the existence of police bias, which is difficult and subjective to document, with disparities in the success rate ("hit rates") of police searches, which are relatively straightforward to document.

Robustness

It is important to emphasize, as Engel (2007) does, that the hit rates test is properly applied to that subset of searches which is discretionary. Also, the test is informative only about bias in searches, not in stops. That being said, the hit rates test has been shown to be valid in environments much more general than that presented in Section 2. We now describe several dimensions along which the model can be generalized, and explain why the hit rates test is preserved. In the process, we address some of the concerns raised in Engel (2007). AQ1

Richness of Observable Characteristics and "Subgroup Validity"

We have assumed that the police can only distinguish two groups in the population, A and W. In reality, the police can observe more than just race. Police can likely classify individuals in groups characterized by a combination of characteristics: race, sex, age, length of hair, etc. In terms of the model, this means that instead of being able to draw from two urns only, the police are able to pick whom to search from many urns. An urn might represent, for example, the

5. In light of our model, search disparities (in some direction) are almost inevitable in equilibrium; it is very unlikely that the police, be they biased or not, would search both groups with the same intensity. The only way that this could happen is if the difference in the average propensities to carry drugs between the two groups is exactly offset by the police bias.

group of White female motorists in their thirties with short hair who drive minivans. In this more sophisticated many-urns model, the same logic that was developed in Section 3 applies. If the police are unbiased, the hit rate should be equalized across all urns searched. If not, then the police could benefit from reallocating searches to urns with higher hit rates. By the same token, if the police are biased against African Americans, we should observe lower hit rates on all urns that contain African Americans. Thus, the hit rate test is still valid for detecting police bias even when the set of observable characteristics is richer.

With regards to the concern about "subgroup validity" raised in Engel (2007), note that we do not need to assume that police use a given characteristic in the same way across groups. For instance, it may well happen that in equilibrium, women are searched more intensely if they have short hair, and the opposite is true for men. (This would be the case if, all other things equal, women with short hair were more inclined to carry drugs than women with long hair, but the opposite was true of men.) Then, in equilibrium the police would be using short hair as a suspicious indicator for women and long hair as an indicator for men. The hit rates test does not require or imply that the police use clues in the same way across racial groups. The concern about "subgroup validity" does not arise in the model.

Omitted Variables and the Inframarginality Concern

Characteristics that are unobservable both to the police and to the researcher obviously do not pose a problem for our analysis: they are incorporated in the "average propensity to commit a crime within a group." What could potentially pose a problem are characteristics that are observable to the police but not to the researcher. Suppose, for example, that the police used vehicle ownership as an indicator of drug-carrying propensity. If vehicle ownership is not recorded in our data, then we have a problem of omitted variable. This type of omission is especially relevant, because an experienced officer likely uses a variety of subtle clues in deciding whether to search a motorist.

It has been shown, however, that the previous analysis is robust to the presence of omitted variables that take the form of characteristics observed by the officer but not recorded in the data. That is, the hit rates test continues to be a valid test of racial bias. To see why, consider how omitted variables can be represented within the model. Suppose, for example, we know the race of the driver but we do not know whether the driver owns the car. In the parlance of the earlier section, we are unable to distinguish between the "White driver, own vehicle" urn and the "White driver, third-party vehicle" urn. One can think of omitted variable (car ownership, in this case) as the confounding of two or more urns. This confounding is potentially troublesome, because the police are able to distinguish the two urns and may treat them differently. The police might, for example, search third-party vehicles more intensely than other cars,

but the data will not reveal this disparity. Alternatively, the police may treat the groups the same, but the outcomes (hit rates) might be different—this is the so-called inframarginality problem.

Fortunately, under the null hypothesis that the police are unbiased, in equilibrium the hit rate from the third-party and own-vehicle urns ought to be the same (otherwise, the police would switch to searching more motorists from the urn with the higher hit rate). Under the null hypothesis that the police are unbiased, being unable to distinguish the two urns is not a problem for the hit rates test, since their hit rate should be the same. It is therefore valid to compare the A and W hit rates even if we realize that, in so doing, we are aggregating hit rates of vehicle owners and non-owners.

Engel (2007) raises the concern that the inframarginality problem might be more likely to arise due to the presence of behavioral cues during the stop. But note that the logic described above still applies even if behavioral characteristics are substituted for car ownership. For example, if the police thought that disrespectful motorists were more likely to carry drugs, these motorists would be searched more intensely. This will lead motorists who carry drugs to be more respectful, and motorists who expect that they will be disrespectful to avoid carrying drugs. Thus, our model does not require that police enforcement strategy be insensitive to the motorist's behavior during the stop—in fact, the hit rates test is not subject to the inframarginality problem, precisely because the police react to behavioral cues.

Non-monolithic Officers and Officer Motivation

Engel (2007) also raises a concern that "non-monolithic officers" might invalidate the hit rates test. The term "non-monolithic officer" refers to the possibility that officers might differ in whether they are biased and/or in their degree of bias. Persico and Todd (2006) address the easier case in which officers may differ in the degree of their bias, under the assumption that all officers are biased against the same group. In that case, they show that the hit rates test remains valid.

The more difficult case that requires some discussion is the case in which some officers are biased *against* a group of motorists and other officers have a bias *in favor* of that group. One can get some intuition for the workings of this variant of the model if we start from an equilibrium where all the police are unbiased, and so the hit rates are equalized. Let us replace a small fraction of the police with officers who are extremely biased and search African Americans regardless of the hit rate. While these biased officers will only search the A group, in equilibrium this will lead the remaining (unbiased) officers to increase the fraction of Whites that they search, so as to perfectly compensate for the imbalance created by the presence of the biased officers. Once this adjustment is done, the crime rates, the hit rates, aggregate police behavior, and the position of the motorists, are all brought back to their initial level. In effect,

the presence of the biased officers is immaterial in terms of aggregate outcomes for all the players, and the hit rate test reflects that—the hit rates remain equalized across races. In other words, the presence of a small fraction of biased police officers who only search the A group does not affect outcomes in the aggregate, because other officers are a countervailing force.

If the fraction of biased officers is large enough, though, the remaining unbiased officers may not be able to fully compensate even if all switch to only searching Whites. In this case, Black motorists will be oversearched relative to the unbiased equilibrium. Note that in this case, the hit rates for White motorists will be higher than for Black motorists, and so the hit rate test picks up the fact that the police force as a whole behaves in a biased fashion. Thus, the hit rates test is informative in the presence of non-monolithic police force, namely, it picks up aggregate bias to the extent that it is reflected in equilibrium outcomes. In other words, the presence of a non-monolithic police force per se does not invalidate the hit rates test.

Regarding officer motivation, Engel (2007) correctly emphasizes that the validity of the hit rates test relies on the assumption that officers maximize number of drug finds. If the benefit to the police is increasing in the magnitude of the drugs find, then the test must be amended. This point was discussed in AQ1 some detail in Knowles et al. (2001); we will return to this topic in Section 6.2.

Incomplete Information, and Reliance on Equilibrium Condition

The assumptions of rationality made in the model, including the motorists' ability to infer the probabilities of being searched and the police's ability to infer the probability of a successful search, may strike some readers (including Engel, 2007) as extreme. Intuitively, it seems plausible that most agents in the population might not be able to articulate the decision-theoretic foundations for their behavior, and probably some agents may not even behave like rational agents. The general argument in favor or against rational agents and equilibrium behavior is a broad one, and it is unlikely to be settled here. Suffice it to say that there are many environments in which the Nash equilibrium predicts behavior very well.[6] The relevant question for us is whether *in our environment* there are enough agents who respond to incentives in a way that is sufficiently close to the ideal rational agent. What we can do is to look at the equilibrium predictions of the model and compare them with data. If the data fit well the predictions of the model, then that gives greater confidence that the set of assumptions, including that of rationality and knowledge of probabilities, are plausible assumptions. In Section 5, we present evidence to this effect.

6. Nash equilibrium, for example, has been found to accurately describe the decision of which side to serve on in tennis (Walker & Wooders, 2001) or which side to kick the ball in soccer penalty kicks (Chiappori, Levitt, & Groseclose, 2002; Palacios-Huerta, 2003).

Other Dimensions in Which the Model Can Be Made More Realistic

Number and effectiveness of searches

It is not important to the model that each police officer can search only once. The analysis generalizes straightforwardly to the case of multiple searches per officer. Perhaps less obviously, the hit rates test also remains valid if officers have different degrees of effectiveness in detecting the presence of drugs on one race relative to another. Say, for example, that officers were on average less able to detect drugs on A's than on W's, owing perhaps to their lack of familiarity with the cultural cues among the A's. Ceteris paribus, this will decrease the hit rate of a search conducted on a member of the A group. The lower relative success rates on the A group will drive officers to search fewer A's and more W's to the point where the hit rate equality is again equalized between the two groups.[7] If the police are unbiased, this hit rate equalization will represent an equilibrium point. So, again, we find that equal hit rates are associated with unbiased police—the hit rates test remains valid.

Motorists from all groups need not carry drugs with equal probability

In the simple model described above, there were only two groups, and so motorists from the two groups carried drugs at the same rate in equilibrium. We may generalize the model to allow for groups of motorists, such as older women, who are likely to exhibit a low propensity to carry drugs. In equilibrium, these groups will need to be searched very lightly, if at all, to bring their crime rates in line with that of more criminally inclined groups. If their propensity to carry drugs is low enough, such groups will not need to be searched at all, and still their crime rate will be below that of the groups that are searched. So, our model does not predict that all groups on the road should have the same crime rate. It only predicts that those groups that are searched—a small minority of all motorists—should have the same crime rate (if the police are unbiased).

Quantity of drugs transported

We have assumed that the motorists' decision is dichotomous—whether or not to carry drugs. In reality, motorists choose how much drugs to carry. In terms of the theoretical development of the model, this generalization does not pose particular problems. We can think of the decision to commit a crime as a continuous variable (how much drug to carry), and we can let motorists choose their preferred quantity, taking into account the probability of being searched

7. Hit rates equalization will require the crime rate to be higher among the A's given the primitives in this thought experiment.

as well as the penalty schedule that is associated with being discovered with different quantities of drugs. If we assume that the police still maximize the probability of a "hit," and so that the police receive the same benefit (in terms of psychic and other benefits) from big and small hits, then in equilibrium hit rates must be equalized across groups if and only if the police have no bias. Thus, the hit rates test remains valid. If police perceive a benefit from finding drugs that is increasing in the size of the find, then the hit rates test must be amended to take into account the weights placed on drug finds of different quantities. See below for more discussion on this point.

Maryland Case Study

To illustrate how the methodology could be applied, we report the results of statistical analyses carried out on data collected by the police as a result of the settlement agreement in *Wilkins v. Maryland State Police* (Maryland I)[8] The settlement entailed the payment of money to the plaintiff, the formulation of a statement by the police renouncing racial profiling, and the collection of the data presented here. The NAACP used these data in a followup case, alleging that Maryland was still discriminating (Maryland II).[9] The tables reported in this section are based on the analysis originally performed in Knowles et al. (2001), which is based on data from 1,590 vehicular searches performed between January 1995 and 1999 on a stretch of I-95 in Maryland.

A first look at the data reveals a familiar pattern of disparate impact. Table 1 reveals that, of those searched, 63 percent were African Americans, unquestionably a much higher percentage than the fraction of Black motorists on the road.[10] We also note that men are disproportionately more likely to be searched than women (93 percent of those searched are men). Other features of the data, while of lesser interest because they are not directly related to protected categories, are still instructive about police behavior. For example, older vehicles represent 22 percent of all searches, luxury vehicles 8 percent, and third-party vehicles 18 percent, and 31 percent of searches were made at night. The variable "Guilty" refers to the fraction of searches that resulted in a find of any drugs or paraphernalia. Many searches resulted in marijuana finds (23 percent), and 8 percent resulted in cocaine finds.

The first two rows in Table 2 report the hit rates by race. Of the White motorists searched, 32 percent were found with some illegal drugs. Of the African American motorists searched, 34 percent were found with some illegal drugs. Despite the wide disparity in search rates, hit rates are very close. In fact, a Pearson chi-square test cannot reject the hypothesis that the two hit rates are the same.[11] According to our analysis, this suggests that the police are

8. *Wilkins v. Maryland State Police*, Civil Case No CCB-93-468 (D. Md. 1993).
9. *NAACP v. Maryland*, 72 F.Supp.2d (D. Md. 1999).
10. The number in parentheses represents the standard deviation.
11. The p value is 0.33.

Table 1 Summary statistics: proportion (*SD*)

African American	.63
	(.01)
White	.29
	(.01)
Hispanic	.06
	(.01)
Female	.07
	(.01)
Guilty	.33
	(.01)
Cocaine	.08
	(.01)
Marijuana	.23
	(.01)
Crack cocaine	.04
	(.005)
Heroine	.02
	(.003)
Morphine	.001
	(.001)
Other drugs	.01
	(.002)
Paraphernalia	.01
	(.002)
Night (12 am–6 am)	.43
	(.01)
No. of observations	1582

not racially biased against African Americans.[12] Although they search African Americans at a higher rate than Whites, the data are consistent with the underlying motivation being to maximize successful searches rather than racial bias.

The next two rows examine disparities in hit rates by gender of the driver. Again, despite the wide disparity in search rates, hit rates are very close by gender, and the chi-square test does not reject the hypothesis that the two hit rates are the same.[13] According to our analysis, this suggests that the police

12. For Hispanics, however, the hit rate (not reported) is lower, possibly indicating a bias against Hispanics. An alternative explanation for the Hispanics finding could be that Hispanics are more likely to be "mules," i.e., to transport high-value shipments of drugs not for personal use. To address this issue, one can perform more sophisticated analyses based on the quantities of drugs found. See Knowles et al. (2001).

13. The *p* value is 0.37.

Table 2 Hit rates: proportion of motorists found to be carrying drugs

By race	African American	0.34
	White	0.32
	Hispanic	0.11
By gender	Male	0.32
	Female	0.36
By time of day	Day	0.32
	Night	0.33
By characteristic of car	Luxury car	0.25
	Non-luxury car	0.33
	Older car	0.32
	Newer car	0.33
	Third party	0.29
	Own	0.33

have no intent to discriminate against men, even though they search them at a relatively higher rate.[14] The remaining rows compare hit rates along other dimensions. Hit rates are similar across these dimensions, too.[15]

Discussion: Model and Evidence

In our analysis, we have assumed that the MSP faithfully recorded all instances of searches and finds. We must acknowledge the possibility that the police might misreport or underreport certain incidents. If, for example, the police decided to let go a young motorist who was found with a minimal amount of marijuana, the police might not record the search, or might record the outcome as "no drugs found" when in reality some drugs were found. If the police behaved in this fashion, and if these under-reportings were skewed by race, then the hit rate computed from our data would not accurately reflect the true hit rate. We have no direct way of disproving this possibility. With regard to the Maryland data, we note that the police knew that the data collection was mandated by court order and that the scrutiny did not focus on disparities in search rates rather than disparities in hit rates.

14. For Hispanics, however, the hit rate (not reported) is lower, possibly indicating a bias against Hispanics. An alternative explanation for the Hispanics finding could be that Hispanics are more likely to be "mules," i.e., to transport high-value shipments of drugs not for personal use. To address this issue, one can perform more sophisticated, though somewhat more subjective, analyses based on the quantities of drugs found. See Knowles et al. (2001).

15. The only time that the Pearson test rejects equality of hit rates is for luxury cars, which are somewhat less likely to result in a find. According to our model, this indicates some preference on the part of the police for searching luxury cars. It is interesting, however, that the disparity disappears when we focus on "large" drug finds. See Knowles et al. (2001).

Having raised this cautionary point concerning data quality, we proceed to discuss the predictions of the model in light of some accumulated empirical evidence. We see this as a fruitful way of checking the validity of some maintained assumptions. If the model was grossly misspecified—if it missed something very important—then we would see large deviations between the model's predictions and the evidence.

A Competing Hypothesis: Unresponsive Motorists

One of the most important rationality assumptions in our model is that motorists react to the probability of being searched by carrying drugs less often. The degree to which this assumption is valid is of course an empirical question. The fact that hit rates are equalized in the Maryland data suggests to us that our model of police-motorist behavior approximates the actual behavior, at least on the stretch of highway 95 and time period covered by the Maryland data. It is remarkable that such a stark prediction as hit rates equalization is observed in the data, particularly in light of the wide disparity in search rates.

There are several different types of frictions that could be introduced into the model and that would lead to a breakdown of its hit equalization prediction.[16] The fact that hit rate equalization appears to be so pervasive in the Maryland data suggests that such frictions did not play a significant role. When applying our analysis to other data sets, however, we can expect that the neat equalization of hit rates we observe may not be so pervasive. In those cases, it may be necessary to augment the simple model we presented to account for these frictions.[17]

Police Incentives to Go After Large Finds

Another reason why equalization of hit rates could be considered a remarkable finding is that the model assumes that police maximize successful searches, while in reality it seems likely that police receive greater rewards from finding greater quantities of drugs. When we amend along those lines the model of Section 2, what we should look at to test for bias are not hit rates, but rather some other function of drug finds that takes into account the quantity of drugs recovered. If, for example, we knew that the police received a benefit that was exactly proportional to the weight of marijuana found in a car, we would look for equalization of the expected weight (in grams) of the marijuana found in searches of African Americans and Whites.

16. See Anwar and Fang (2006) for an example of such frictions.
17. The equalization of the hit rates is not special to the Maryland case, however: It also happens in other environments. See Persico and Todd (2006).

In practice, we do not know exactly how the size of the find factors in the reward accruing to the officer. The Maryland data did allow us to make some preliminary inquiries into this question. When we restrict attention to more significant drug finds, and consider as "hits" only those searches that exceed a certain magnitude, we find that the hit rate is higher for African Americans than it is for Whites.[18] So, if the police only considered a search successful if it yielded large quantities of drugs, then our data would suggest that the Maryland police behavior is biased, if at all, in favor of African Americans. That is, searches of African Americans yield on average larger drug finds, indicating that police should allocate more search effort to this group. This finding does not seem to us the most plausible interpretation of the data. Instead, the data are consistent with the simpler interpretation that the police are actually equalizing hits, defined as any drug recovered. This simple interpretation is also consistent with the equalization of hit rates across other (non-race) dimensions.

Differential Search Intensity

Another aspect of reality that the model does not do justice to is the question of differential intensity in searches. If the police search one group more intensely than another, comparing hit rates may be less meaningful than our analysis might suggest. In reality, we do not know if there is a systematic correlation between the intensity of searches and race, and few datasets contain information on search intensity. It is difficult to gauge how restrictive is the model's assumption, but the question of search intensity is an important area for future research.

Costless Focusing of Search Activities

Finally, the model assumes that the police can costlessly direct searches toward any subgroup of the motoring population. This assumption implied that any subgroup whose "hit rate" was higher than the others could, and would, be subject to such intense focus of police interdiction that members of that group would decrease their criminal activity until the "hit rate" in that group was brought back to the population average. In reality, the police might find it difficult, or impractical, to direct a lot of searches on a subgroup, especially if that subgroup is small.

Imagine, for example, that a rare subgroup of the population is highly likely to be involved in crime. To fix ideas, let us imagine these are first-generation

18. See Knowles et al. (2001, table 2). As the definition of a hit is made more stringent, so that only very large drug finds are counted as hits, the disparity between the hit rates of African Americans and Whites becomes statistically significant (Knowles et al. 2001, table 3).

Sicilians.[19] Imagine that there is only one officer stationed on the road. This officer can stop many drivers (it does not take long to stop someone), but can only search a few of them (a search takes a long time). Moreover, while the officer is busy searching one car, other drivers who pass by cannot be searched. Now, it will be extremely costly for the police to devote himself solely to the task of searching Sicilians, because that would mean that most of the times the officer will not search anyone, waiting for the off chance of meeting a Sicilian. Instead, the practical thing to do is for the officer to search a broader variety of motorists. But this means that Sicilians cannot be searched with probability 1, because sometimes as they are driving by, the officer will be occupied with searching some other motorist.

This thought experiment suggests that it might theoretically be difficult for the police to focus their interdiction efforts on narrowly defined subgroups. We do not know how empirically relevant this potential consideration might be.

Conclusion

In this paper, we have given an intuitive exposition of a model originally developed in the economics literature. The model yields a test for police bias based on statistical data on the success rate of searches. The test finds bias against a protected group if the success rate of searches on that group is lower than on another group. Within a rational choice model of crime and policing, this test flags police behavior if it is driven by bias against a protected category. This test therefore operationalizes the concept of "intent to discriminate" which is central to the legal approach to selective enforcement. We presented some evidence based on data from Maryland State Police that supports the validity of the model.

In the process of expositing the model, we have addressed a number of concerns raised by Engel (2007). To the extent that some of the concerns have been shown to be unfounded, we acknowledge that they are so only within the assumptions of the model. If the model is not a good fit for the data, then the model's predictions will not be well founded.

References

Anwar, S., & Fang, H. (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review, 96,* 127–151.

Chiappori, P.A., Levitt, S., & Groseclose, T. (2002). Testing mixed strategy equilibria when players are heterogeneous: the case of penalty kicks in soccer, *American Economic Review, 92,* 1138–1151.

Engel (2007), A critique of the "outcome test" in racial profiling research. *Justice Quarterly, 24,* this issue.

19. Full disclosure: the first author is a first-generation Sicilian.

Knowles, J., Persico, N., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy, 109,* 203–299.

Palacios-Huerta, I. (2003) Professionals play minimax, *Review of Economic Studies, 70,* 395–415.

Persico, N., & Todd, P. E. (2005) Passenger profiling, imperfect screening, and airport security. *Papers and Proceedings of the American Economic Association, 95,* 127–131.

Persico, N., & Todd, P. E. (2006). Generalising the hit rates test for racial bias in law enforcement, with an application to vehicle searches in Wichita. *The Economic Journal, 116,* F351–F367.

Walker, M., & Wooders, J. (2001). "Minimax play at Wimbledon," *American Economic Review, 91,* 1521–1538.