On The Observational Implications of Taste-Based Discrimination in Racial Profiling

William A. Brock, Jane Cooley, Steven N. Durlauf and Salvador Navarro

February 11, 2010

Revised December 20, 2010

Department of Economics, University of Wisconsin at Madison. This work was supported by the National Science Foundation grant SES-0518274 and the University of Wisconsin Vilas Trust, University of Wisconsin Graduate School, Institute for Research on Poverty, and Laurits Christensen Chair in Economics. We thank Hon Ho Kwok, Xiangrong Yu and Yu Zhu for excellent research assistance. We thank Elie Tamer and an anonymous referee for very helpful suggestions. Daniel Nagin, Alexis Piquero, David Weisburd, Sarit Weisburd, and especially Angela Duckworth have helped us with the criminological and psychological literatures. This paper was prepared for the festschrift volume in honor of Charles Manski and is written in his honor.

1. Introduction

Minorities are generally subject to higher rates of policing, such as automobile or pedestrian stops, than whites. For instance, Ridgeway (2007) reports that 87 percent of stops of the New York police department in 2006 were nonwhite, and 51 percent of these were black. These differentials are often referred to as racial profiling. Racial profiling has generated and continues to generate an active scholarly literature because differential treatment by police officers (often measured as stop or search rates) can stem from a range of factors, not all of which correspond to what is commonly understood as discrimination. A central difficulty in the study of racial bias stems from the fact that the police officer is likely to have more information about the potential guilt of suspects than the analyst. This makes it difficult to attribute racial disparities in police treatment of potential criminals to discrimination. From a policy perspective, this distinction is likely to be very important, as racial disparities that result from optimal (or productive) policing may be less of a concern than disparities that result from discrimination. In this paper, we examine the challenges to identifying discrimination in police behavior in observational data.

For our purposes, we equate discrimination with Becker's (1957) notion of taste-based discrimination (or racial prejudice). As originally argued in Knowles, Persico and Todd (2001), which we subsequently denote as KPT, and recently reviewed in Persico (2009) differential treatment of blacks and whites does not imply the presence of taste-based discrimination. The objective functions of police could be race neutral yet produce differential equilibrium search rates by race.[1] This has spurred a literature on the potential to detect taste-based discrimination using outcome data (such as arrest rates), a prominent example of which is a recent Rand study (Ridgeway, 2007), of pedestrian stops in New York City.

---

[1]This latter differential may be attributed to statistical discrimination. Our focus on taste-based discrimination does not imply a justification for statistical discrimination; in fact the ethics of statistical discrimination have been challenged by Harcourt (2004) and Sklansky (2008); see Risse and Zeckhauser (2004) for a defense. However, all of these positions recognize an ethical distinction between taste-based discrimination and statistical discrimination and hence the importance of empirically distinguishing them.

Our goal in this paper is to describe conditions under which taste-based discrimination in police stops and searches can be uncovered in observational data. One important contribution relative to the existing literature is that we consider the robustness of various tests for discrimination to heterogeneity across groups of potential criminals and officers.

With respect to potential criminals, we are concerned with the possibility that police stops involve information that is not available to the econometrician. KPT is also concerned with heterogeneity across potential criminals that may be observed to police officers but not the econometrician. If the payoffs to carrying contraband vary with characteristics correlated with race, such as income, then we might observe higher search rates for nonwhites even in the absence of taste-based discrimination. However, KPT shows that when police use this information to help maximize search success rates in their context, individuals with the "guilty characteristic" have the incentive to carry contraband less and those without the characteristic to carry contraband more. In equilibrium, the probability of carrying contraband equalizes across groups and so the heterogeneity provides no useful information with which to identify criminals. Differentials in the search success rates across race only occur in their model when police discriminate. We show that in a more general model where the police officer's utility depends on the intensity of search rates, the problem of unobservables resurfaces because search success rates no longer equalize across groups in the absence of discrimination.

Like us, Anwar and Fang (2006) (henceforth denoted as AF) and Bjerk (2007) are also concerned with unobservables that could prevent attributing differential success rates across racial groups to taste-based discrimination. However, they focus on a particular type of unobservable that signals the guilt of the potential criminal to the officer. Our analysis permits unobservables to play a more general role in that we do not restrict them to just act as a signal to the officer. In addition, unobservables can affect equilibrium guilt probabilities and enter into officers' preferences. Thus, their whole distribution matters for detecting bias, as highlighted in Heckman (1998).

Beyond heterogeneity in potential criminals, we also contribute to the literature by highlighting the role of heterogeneity across individual police officers in taste for discrimination. In our view, this heterogeneity is essential for capturing Becker's original thinking, in which the key determinant of degrees of discrimination involves the behavior of the marginal employer in a labor market. In particular, we consider the possibility that officers have differential tastes both for searching guilty criminals and for searching potential criminals regardless of guilt.

AF is also concerned with police officer heterogeneity, but a particular type of heterogeneity where the cost to searching criminals of different races varies by the race of the officer. They exploit this heterogeneity to derive testable implications about relative racial prejudice using a ranking condition that compares searches and success rates of criminals of different races across officers of different races. We discuss how our formulation links to AF's and argue that their rank condition is sensitive to model assumptions in a fashion similar to KPT. We further illustrate a set of assumptions through which we can exploit police officer heterogeneity to identify relative bias in our model.[2]

The literature notes reasons beyond unobservables and police officer heterogeneity why a test based on search success rates may fail. Dominitz and Knowles (2006) show that the condition of equal search success rates, while consistent with KPT's model of arrest rate maximization, would generally not hold in a model where the police officer objective was to minimize crime. Thus, like us, they are concerned with the sensitivity of the result of equal search success rates to the specification of the police officer objective function. However, we consider the opposite possibility, that equal search success rates can be consistent with crime minimization. We show that equal search success rates neither necessarily distinguish between arrest maximization and crime minimization as objective functions nor rule out the possibility of racial prejudice.

In terms of substantive conclusions, we make four claims. First, tests based on the cross groups comparisons of either the success rate of searches or search rates are not informative about the presence or absence of taste-based discrimination when one weakens the assumptions made by KPT and Persico (2009). Second, the presence or

---

[2]Antonovics and Knight (2009) also consider the potential for exploiting police officer heterogeneity as a means of identifying taste-based discrimination. Unlike AF, they focus on search rates (rather than search success rates) in a context where like KPT the unobservable (to the econometrician) does not act as a signal of guilt.

absence of taste-based discrimination does not necessarily differentiate between crime minimization and a decentralized determination of stop and search strategies except for relatively special assumptions. Third, even if a police department possesses superior information to an econometrician, it is possible for discrimination to go undetected when there is unobserved taste heterogeneity across police officers. A policy mandating officers achieve either equal search or success rates across observable groups would not prevent biased officers from exercising their bias. Fourth, by placing restrictions on unobserved heterogeneity it is possible to develop testable implications for detecting bias even in our more general setting. Together, these results suggest that model-specific claims about taste-based discrimination and racial profiling should be interpreted with caution; in our conclusions we discuss constructive ways to proceed given this.

It is worth observing that racial profiling studies require a somewhat different view of the process by which taste-based discrimination affects outcomes than Becker's original model. In Becker's analysis, which focused on labor markets, equilibrium wage gaps between blacks and whites is determined by the behavior of the marginal discriminator in a labor market; see Charles and Guryan (2008) for a recent test of this proposition. From this vantage point, it is possible for a market to contain prejudiced firms even though the wage gap between blacks and whites is zero; this may happen when potential discriminators employ all white workforces. An analogous argument does not directly apply to the study of police stops and searches. Each potential discriminator separately interacts with a common population and, ceteris paribus, would always prefer to sample blacks if the option arises. A segregating firm with an all white workforce has no incentive, at equal wage rates, to substitute a black worker for a white one.

Section 2 of this paper outlines a general model of police stops and equilibrium guilt rates. We use this abstract vantage point to argue that tests of taste-based discrimination based on simple criteria such as equality of hit rates or search rates are not informative about taste-based discrimination without very particular assumptions on preference structures[3]. The need for such assumptions does not render previous exercises

[3]We do not claim a general observational equivalence result nor do we argue that field experiments would not potentially be able to reveal taste-based discrimination in racial profiling.

uninteresting but rather indicates the importance of performing sensitivity analyses. Section 3 studies a generalization of the KPT model, one that follows aspects of its formulation in Persico (2009) focusing on predictions deriving from search success rates. Here we discuss AF (2006) in some detail. Section 4 considers how biased officers can survive police department policies to detect bias in the presence of unobservable officer heterogeneity. Section 5 considers empirical predictions of the absence of taste-based discrimination in our model using assumptions on unobservables. We show how different restrictions can lead to both point and partial identification results. Section 6 concludes.

Section 2. Basic model

i. decision problems for potential criminals and police officers

We consider the behavior of two populations: a population of civilians who make choices as to whether or not to engage in an illegal activity (for concreteness, we think of the activity as carrying contraband) and a population of police officers each of whom chooses a strategy for stopping individuals who will be searched.

Each member $j = 1,...,N$ of the civilian population is a member of some group. The characteristics of a group are described as a triple, (r, c_o, c_u), where r denotes the race of the group members, c_o denotes the characteristics of individuals in the group that are observable to both the econometrician and the officer, and c_u denotes the characteristics of individuals in the group that are observable to the officer but not to the econometrician. The number of groups is assumed to be finite. We will denote n_{r,c_o,c_u} as the number of members of a group. For example, an officer may follow a stop strategy that depends on a driver's race, car type and driving style, the last of which is unavailable to the analyst.[4] In addition, we assume that individual specific heterogeneity among members of a group is described by □, which is taken as independent across individuals.

We employ groups to define what it means for an officer to regard two individuals as indistinguishable. We assume that individual heterogeneity is not observable to the police, so group level characteristics summarize what police officers know about an individual. This is without loss of generality as one can always define a finer partition of individuals across groups in which the observable part of the individual heterogeneity is treated as a group attribute. The primary identification challenge is the presence of the characteristics that the police officer observes that the researcher does not (□ as opposed to □).

For individual j, a choice is made between committing a crime, for the profiling case this typically amounts to the possession of contraband, C, and not committing a crime, i.e. not carrying contraband, NC. We assume that all uncertainty associated with the payoff to this choice involves the possibility of being searched or not being searched (S or NS). Therefore, the individual's expected utility from carrying contraband/commission of a crime equals

$$EU_C\left(r,c_o,c_u,s_{r,c_o,c_u},\varepsilon_j\right)=s_{r,c_o,c_u}U_{C,S}\left(r,c_o,c_u,\varepsilon_j\right)+\left(1-s_{r,c_o,c_u}\right)U_{C,NS}\left(r,c_o,c_u,\varepsilon_j\right), \quad (1)$$

where s_{r,c_o,c_u} is the probability of being searched if one is a member of the group. As noted above, ε_j is not observed by an officer, hence an individual uses the probability of a member of his group being searched to assess his own probability of being searched. We assume that civilians have rational expectations in the sense that their subjective probabilities concerning searches will correspond to the equilibrium search probabilities of the complete model.

[4]Grogger and Ridgeway (2006) argues that the race of a driver is not known at night if one is focusing on car stops; their approach means group definitions are time dependent as some variables only periodically appear in the officer's information set; we do not pursue this generalization of our information structure.

Normalizing the utility of no crime commission to zero, an individual will carry contraband if equation (1) is positive. From the perspective of a given police officer, since ε_j is unobservable, the solutions of the individual-level crime choice problems will produce group-specific crime rates

$$\pi_{r,c_o,c_u} = \Pr\left(EU_C\left(r,c_o,c_u,s_{r,c_o,c_u},\varepsilon_j\right) > 0 \middle| r,c_o,c_u,s_{r,c_o,c_u}\right), \qquad (2)$$

which summarize the information available to the police on relative guilt probabilities between individuals. Police are assumed to have rational expectations in the sense that they employ these probabilities in their search calculations.

The crime choices made by civilians are paralleled in the search choices made by the police. Officer $i = 1,...,M$ chooses group-specific search rates s_{i,r,c_o,c_u}. We are not explicit here about the constraints that search rates must satisfy. In most of our analysis, search rates will be have to be consistent with a fixed number of stops per day; we will consider additional restrictions placed by the police department. We allow for heterogeneity in search rates across officers. We distinguish between officer types with observable characteristics $p_{o,i}$ and unobservable characteristics $p_{u,i}$ with respect to what is observable to the econometrician.

The expected utility for a police officer with characteristics $p_{o,i}$, $p_{u,i}$ of applying a search rate s_{r,c_o,c_u} to a group characterized by $\left(r,c_o,c_u\right)$ is

$$\pi_{r,c_o,c_u} V_C\left(p_{o,i},p_{u,i},r,c_o,c_u,s_{i,r,c_o,c_u}\right) + \left(1-\pi_{r,c_o,c_u}\right) V_{NC}\left(p_{o,i},p_{u,i},r,c_o,c_u,s_{i,r,c_o,c_u}\right). \qquad (3)$$

Here $V_C\left(p_{o,i},p_{u,i},r,c_o,c_u,s_{i,r,c_o,c_u}\right)$ denotes the utility of searching a guilty member of the group and $V_{NC}\left(p_{o,i},p_{u,i},r,c_o,c_u,s_{i,r,c_o,c_u}\right)$ denotes the utility of searching an innocent member. We include the individual search rate s_{i,r,c_o,c_u} in the utility function as it allows for the possibility that the utility to a given search depends on the intensity of search the officer applies to the group. Notice the presence of this term does not imply taste-based discrimination per se. For example, it may be interpreted as a cost parameter in the sense that different choices of search allocations across groups involve the use of different locations at which to search civilians.

Let a_{i,r,c_o,c_u} denote the number of members of a group searched by officer i, i.e. $a_{i,r,c_o,c_u} = s_{i,r,c_o,c_u} n_{r,c_o,c_u}$. Aggregating across police determines the probability of being searched via

$$s_{r,c_o,c_u} = \int s_{i,r,c_o,c_u} di = \frac{\int a_{i,r,c_o,c_u} di}{n_{r,c_o,c_u}}. \tag{4}$$

Combining (4) with (2), guilt probabilities for each group can be thought of as a function

$$\pi_{r,c_o,c_u} = \pi\left(r, c_o, c_u, s_{r,c_o,c_u}\right). \tag{5}$$

This closes the description of the crime choices of civilians and the search choices of police officers, once one is explicit about the restrictions on search rates that an officer must fulfill.

We do not prove the existence of equilibrium for the joint crime and search choices of civilians and police. Establishing sufficient regularity or smoothness conditions on preferences and search constraints in order to ensure a Nash or a Stackelberg equilibrium is not relevant to our goals in this paper. We will prove existence for the version of this model we study in the next section, so existence is by no means vacuous.

ii. defining taste-based discrimination

Our formulation of the individual officer decision problems provides one natural definition for the absence of discrimination in the preferences of a given police officer, namely irrelevance of race in an officer's utility function given other factors. Formally, the absence of a taste for race-based discrimination on the part of officer i requires that

$$V_C\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right) = V_C\left(p_{o,i}, p_{u,i}, r', c_o, c_u, s_{i,r',c_o,c_u}\right) \quad \text{if } s_{i,r,c_o,c_u} = s_{i,r',c_o,c_u} \tag{6}$$

and

$$V_{NC}\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right) = V_{NC}\left(p_{o,i}, p_{u,i}, r', c_o, c_u, s_{i,r',c_o,c_u}\right) \quad \text{if } s_{i,r,c_o,c_u} = s_{i,r',c_o,c_u} \tag{7}$$

Notice that we do *not* require that $\pi_{r,c_o,c_u} = \pi_{r',c_o,c_u}$. Officers may prefer to search one race or another because of differences in guilt probabilities; these appear in the expected utility, equation (3), from a given group-specific search rate.

One apparently odd feature of this definition of no taste-based discrimination is that it only imposes an equal utility requirement when equal effort is applied to both groups. Our reasoning for this parallels the incorporation of search effort in the payoff function. If utility were convex in effort, this would create the possibility of ruling out taste-based discrimination against blacks if an officer spends all effort on blacks, because he would, under the counterfactual, derive equal utility from applying all effort to whites. We will impose a form of concavity later on, so this possibility is moot for us.[5]

Further, we think our definition of the absence of taste-based discrimination is appropriate. In our view, the defining property of the absence of taste-based discrimination is that a police officer should experience the same utility from a given action applied to groups that are identical except for race. This definition follows the

---

[5]Ayres (2005) implicitly criticizes this definition of the absence of taste-based discrimination on another grounds, namely that the variables that comprise c_o, c_u may be correlated with race and so mask disparate treatments by race. His argument may have ethical and legal merit, but it is not germane to our objective which is to understand the observable implications of race based discrimination. We return to this issue below.

spirit of Levin and Robbins (1983) who focus on conditional exchangeability as a notion of lack of discrimination; similar reasoning appears in Heckman and Siegelman (1993). Search rates are simply one more variable that has to be held constant to engage in race-based comparisons of utility. If it were the case that increasing returns implied that there are multiple equilibrium search configurations on the part of officers, for no taste discrimination to exist, we would require that group/search rate pairs exhibit conditional exchangeability with respect to race. Put differently, taste-based discrimination means that a preference for stopping blacks is the source of discrepancies in treatment, as opposed to a feature of the search technology, for example, which implies differential treatment for some group occurs under race neutral preferences.

One could consider empirically implementing the model we just described as a discrete binary choice in which the officer decides whether or not to stop and search each civilian he encounters. By doing this, one could in principle recover preferences and test the no discrimination hypothesis directly. The main problem with the binary choice implementation is that it requires the analyst to recover the probability that an officer searches an individual from the civilian population as a whole. This in turns requires the analyst to know not only the distribution of characteristics of the individuals who are actually searched but also the distribution of characteristics of individuals who are not searched. This latter distribution is not directly observed in data sets on police stops and searches.[6]

One of the important insights in the reformulation of the problem in KPT and in Persico (2009) is that it is possible to test an empirical implication of no discrimination directly, sidestepping the need to recover the distribution of the unobservables. In fact, under the assumptions of their model, in equilibrium there is no selection on unobservables (c_u), as guilt probabilities equalize across groups. Bjerk (2007) shows that when selection on unobservables does occur, KPT's finding of equal guilt probabilities, which applies to the population, would not apply to the selected sample of potential

[6]This, in part, motivates different benchmarks used in studies, such as racial composition of area based on CPS data, as a means of detecting racial profiling. See Ridgeway (2007).

criminals who are actually searched.[7] In order to avoid this selection on unobservables problem, we follow a similar approach to KPT in that we focus on implications that can be tested by looking only at the treated sample. However, a significant difference is that the issue of unobservables remains an important part of our analysis.

Persico (2009) uses a specific version of the officer's expected utility calculation (3) in that he assumes that $V_C\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right) = \beta_r$, which means that officers only (potentially) care about the race of a guilty individual and that differences in the weights assigned to races do not covary with either officer or other group characteristics. He further assumes that $V_{NC}\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right) = 0$ so that there is no utility from searches per se. Taste-based discrimination is defined as $\beta_r \neq \beta_{r'}$ for any $r \neq r'$. If taste-based discrimination is absent, officers will only search members of the race with the highest guilt probability, π_r and effort across races is allocated to equalize guilt rates. If the ratio of guilt probabilities between races is not equal to 1, this is evidence that the taste for capturing criminals varies by race.

This is essentially the same logic employed in the earlier work by KPT, except that in KPT $V_C\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right) - V_{NC}\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right) = \beta$ and racial prejudice enters through differential costs to discrimination. In terms of functional form this is analogous to utility from searching someone who is not carrying contraband. They also focus on a general equilibrium setting in which a mixed-stop equilibrium occurs, i.e., one where members of all groups are searched and thus the guilt probabilities are equal across all subgroups in the absence of differential costs to search across races. KPT find that the guilt probabilities in their traffic stop data are approximately equal across blacks and whites. This leads them to conclude that the data are consistent with no discrimination against blacks. As our formulation demonstrates, the KPT and Persico (2009) analyses are based on a special case of an abstract officer decision problem. In our subsequent analysis, we work with a particular officer decision problem of which the KPT and Persico (2009) analyses are special cases. Although we call our decision

---

[7]Dharmapala and Ross (2004) also argue that if some subset of the population is never searched, then the result of equal guilt probabilities in KPT does not hold.

problem general, it is still based on parametric assumptions on the officer objective functions.

We are unaware of any work that attempts to test for taste-based discrimination in police stops and searches without restrictions on police preferences. In the absence of any assumptions on the objective functions of either potential criminals or the police, we conjecture that there are no testable implications for taste-based discrimination if one focuses on data on searches and their outcomes. A heuristic justification for this view is the following. Suppose that $V_{NC}\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right)=0$ and that $V_C\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right)$ does not depend upon race. Suppose that data are available on two objects. First, the average search rate for individuals with race r and observable characteristics c_o by officers with observable characteristics $p_{o,i}$:

$$s_{p_{o,i},r,c_o} = \int s_{p_{o,i},p_{u,i},r,c_o,c_u} dF_{p_{u,i},c_u|p_{o,i},r,c_o} . \tag{8}$$

Second, the number of guilty individuals with race r and observable characteristics c_o who are searched,

$$g_{r,c_o} = n_{r,c_o} \int \pi\left(r, c_o, c_u, s_{r,c_o,c_u}\right) s_{r,c_o,c_u} dF_{c_u|r,c_o} . \tag{9}$$

Since one is free to choose any distribution functions $F_{p_{u,i},c_u|p_{o,i},r,c_o}$ and $F_{c_u|r,c_o}$ and any utility function that does not depend on race, $V_C\left(p_{o,i}, p_{u,i}, c_o, c_u, s_{i,r,c_o,c_u}\right)$, it is difficult to see how the abstract search and guilt model can place any restrictions on these data.

For example, in the spirit of AF, suppose that one tests for the absence of taste-based discrimination by determining whether the ranking of search rates for two groups with identical values of c_o but different races, r and r', is preserved across two officers i, i' with observable characteristics $p_{o,i}$ and $p_{o,i'}$. Clearly, one could construct a pair of functions $V_C\left(p_{o,i}, p_{u,i}, c_o, c_u, s_{i,r,c_o,c_u}\right)$ and $V_C\left(p_{o,i'}, p_{u,i'}, c_o, c_u, s_{i',r,c_o,c_u}\right)$ together with associated distribution functions $F_{p_u,c_u|p_{o,i},r,c_o}$, $F_{p_u,c_u|p_{o,i'},r,c_o}$ and $F_{c_u|r,c_o}$ to produce any ranking across the two officers. The problem is that solely restricting $V_C\left(p_{o,i}, p_{u,i}, c_o, c_u, s_{i,r,c_o,c_u}\right)$ so that it does not depend on race leaves too many degrees of freedom in its possible forms.

This example illustrates why the literature on taste-based discrimination in racial profiling has not proceeded nonparametrically. The example, should not, in our judgment, be used to dismiss the KPT and Persico (2009) work, let alone the broader literature that followed. Rather, it suggests the importance of understanding how functional form assumptions influence the evaluation of taste-based discrimination. We therefore study the empirical implications of taste-based discrimination in a model that nests essential aspects of KPT and Persico (2009) as special cases.

Section 3. A generalized search and guilt model

In order to understand how the introduction of alternative preference structures can affect claims of no discrimination while at the same time see how some relaxation of the stringent KPT assumptions can still allow of observational implications of taste-based discrimination, we extend KPT using a device of Persico (2009). We emphasize that an important contribution of his paper is to set up a framework for search allocations across groups given a time constraint. Our contribution involves relaxing aspects of his (and KPT's) functional form assumptions as well as some of the informational assumptions implicitly assumed in their analyses. Relative to our abstract formulation, we assume that

$$V_C\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right) = \left(\beta_{i,r,c_o,c_u} + b_{i,r,c_o,c_u}\right) s_{i,r,c_o,c_u}^{\alpha-1} \tag{10}$$

and

$$V_{NC}\left(p_{o,i}, p_{u,i}, r, c_o, c_u, s_{i,r,c_o,c_u}\right) = b_{i,r,c_o,c_u} s_{i,r,c_o,c_u}^{\alpha-1}. \tag{11}$$

From this formulation, one can interpret $\beta_{i,r,c_o,c_u} s_{i,r,c_o,c_u}^{\alpha-1}$ as the (extra) utility of searching a guilty person in distinction to $b_{i,r,c_o,c_u} s_{i,r,c_o,c_u}^{\alpha-1}$, the utility of the act of searching itself. Since the total number of searches of members of a group by officer i is $n_{r,c_o,c_u} s_{i,r,c_o,c_u}$, the overall expected utility from a given allocation of searches across groups by officer i is

$$\int \left(\beta_{i,r,c_o,c_u}\pi_{r,c_o,c_u} + b_{i,r,c_o,c_u}\right) n_{r,c_o,c_u} s_{i,r,c_o,c_u}^{\alpha} dc_u dc_o dr. \tag{12}$$

For our specification, the term β_{i,r,c_o,c_u} corresponds to Persico's (2009) measure of taste-based discrimination as it characterizes how payoffs for searching the guilty depend on the group of which an individual is a member. In addition to this potential source of taste-based discrimination, we introduce the parameter b_{i,r,c_o,c_u} in order to allow for payoffs to depend on the act of stopping a civilian regardless of his guilt. This factor is related to what we would argue is the most clearly unambiguous morally objectionable aspect of profiling, the disproportionate searching of innocent blacks (Durlauf (2006)). In order to fully understand the relationship between the searching of innocent blacks and taste-based discrimination we need to allow for the possibility that there is utility from their being searched, not just that they are searched as a consequence of a biased desire to find guilty blacks.

This formulation embeds the Persico (2009) preference structure as a special case in which $\alpha = 1$; we are also interested in cases where $\alpha < 1$.[8] It also embeds the KPT preferences, with a slightly different set of assumptions on the source of taste-based discrimination, as described in the general problem above. One value of $\alpha < 1$ is that it allows for interior solutions to the individual officers' allocations of searches across races, as will be seen below. It is important to recognize that our formulation implies a

[8]Persico (2009) focuses on a setting where individuals maximize total search effort rather than search rates. While the implications for group size differ across the two models, the implications for relative search effort are similar. We prefer the search rate model because it has the property that in a no-discrimination equilibrium where individuals were simply searched at random, the search rates would be equal across races.

substantive preference difference relative to KPT and Persico (2009). We assume that officers have a preference against concentrating their searches on a single group, which renders the search rates for each officer determinate. This is not the case for Persico (2009) or KPT in the presence of time constraints, as individual officers, in equilibrium, are indifferent in their search allocations across groups.[9]

We defend our preference structure at two levels. First, from the vantage point of functional forms, our structure nests KPT and Persico (2009) smoothly in the sense (made precise below) that as $\alpha \to 1$ from below, the behavior of officers in our model converges to the appropriate analog of theirs. Hence we can investigate robustness of their findings with respect to a class of parametric specifications.

Second, there are substantive reasons why $\alpha < 1$ might better describe officer payoffs than $\alpha = 1$. One reason concerns the costs of searching different groups. It may be the case that these costs are convex in the number of searches of a given group (and hence convex in the search rate); $\alpha < 1$ proxies for this possibility. Convexity in costs, in turn, implies concavity in payoffs from group-specific search rates, as implied by $\alpha < 1$. Convexity of costs may occur if higher search rates for a given group, relative to their population fraction, require additional actions on the part of an officer, for example by choosing different locations from which to search.[10] Convexity in costs, in turn, implies concavity in payoffs from the search intensity. Alternatively, one can imagine that litigation-wary police departments impose costs on officers who deviate from equal search strategies.[11] Sanga (2009) provides some evidence that the litigation affects police behavior. He replicates the KPT analysis and extends it to include other highways in Maryland that were not the focus of the original racial profiling lawsuit in the state, which initiated the collection of police stops data. The finding of equal hit rates across

---

[9]To be precise, KPT does not constrain the number of officer stops to meet some fixed time, so α would be irrelevant for their context; if they did implement a time constraint, their analysis would map to the case where $\alpha = 1$.

[10]This would likely be the case if the groups used to determine group-specific search rates are formed based on a finely detailed set of characteristics.

[11]Of course, these conceptual arguments do not justify our particular functional form, which is chosen for analytical convenience and because it nests the KPT and Persico preference structures.

blacks and whites is not robust to the inclusion of these other highways, even after controlling for location fixed effects. Furthermore, he finds more evidence of disparities in White and Hispanic hit rates. We believe that a reasonable interpretation of this finding could be that the equal hit rates found in the original KPT analysis were a direct result of the litigation. .

An alternative view of convex costs to group-specific search is that they are psychological, i.e. the mental effort involved in identifying groups is increasing in the complexity of their description. Hence a random search involves less mental effort than a search that is trying to oversample blacks; and a search that looks for violations that serve as pretexts for traffic stops and searches (which is technically required for a stop to be legal) requires less effort than a search that requires concentration to identify both traffic violations and the race of the driver. This view of attention as requiring effort, with attention on multiple attributes or objects requiring greater effort, is standard in psychology; Kahneman (1973) is the classic reference.[12] A recent example of psychological research of this type is Warm, Parasuraman, and Matthews (2008) who review evidence on the mental energy costs of vigilance, which they defined as "the ability of organisms to maintain their focus of attention…over periods of time." (p. 433). Vigilance seems an appropriate way to think about police who are observing motorists and making decisions on who to search.

One feature of the function s^{α}_{i,r,c_o,c_u}, $\alpha < 1$, is that it implies that a police offer would experience lower stop costs if he were to focus on cruder group designations than (r, c_o, c_u). This is also consistent with existing psychological evidence and in fact provides a basis for understanding why individuals stereotype, see Macrae, Milne, and Bodenhausen (1994). The authors specifically describe stereotypes as "energy saving devices" (p.37) which is consistent with our functional form assumption as cruder group descriptions are isomorphic to greater stereotyping.

That said, our objective is *not* to argue that one value of α is more defensible than another, but rather to argue that there do not exist strong *a priori* reasons to privilege

[12]Cognitive costs to attention have been proposed in other economic contexts, see Banerjee and Mullainathan (2008) for an example.

$\alpha = 1$ over $\alpha < 1$. As suggested in the previous section, without functional form assumptions there is little to be said about taste-based discrimination and racial profiling. Hence it is important to understand how different functional forms affect claims of bias. In the case of α, this functional form assumption maps to a substantive assumption on the costs of search and hence is interpretable in terms of the underlying search technology.

Applying the general definition of no taste-based discrimination, equations (6) and (7), to the preferences embedded in equation (12), the absence of taste-based discrimination on the part of officer i requires that

$$\beta_{i,r,c_o,c_u} = \beta_{i,r',c_o,c_u} = \beta_{i,c_o,c_u} \tag{13}$$

and

$$b_{i,r,c_o,c_u} = b_{i,r',c_o,c_u} = b_{i,c_o,c_u} \tag{14}$$

We thus distinguish between discrimination with respect to the utility of searching a guilty driver versus an innocent driver. This general view of the way that taste-based discrimination can enter individual decisions will of course matter in interpreting existing tests of the no discrimination hypothesis as well as formulating new ones.

We describe optimal police behavior and the existence of equilibrium for this model. For simplicity, officers are assumed to have a fixed number of searches T which they allocate across groups,[13]

$$\int n_{r,c_o,c_u} s_{i,r,c_o,c_u} dc_u dc_o dr \leq T. \tag{15}$$

[13] Alternatively one can think of officers as having a technology to transform time τ_{i,r,c_o,c_u} into stops say $n_{r,c_o,c_u} s_{i,r,c_o,c_u} = \delta_i \tau_{i,r,c_o,c_u}$. Then, if officers face a time constraint $\int \tau_{i,r,c_o,c_u} dc_u dc_o dr \leq t$, the constraints may be rewritten as $\int n_{r,c_o,c_u} s_{i,r,c_o,c_u} dc_u dc_o dr \leq \frac{t}{\delta_i} \equiv T_i$ which is the same as (15) except it is i specific.

The Lagrangian for the officer's problem is therefore

$$\int\left(\left(\beta_{i,r,c_o,c_u}\pi_{r,c_o,c_u}+b_{i,r,c_o,c_u}\right)n_{r,c_o,c_u}s_{i,r,c_o,c_u}^{\alpha}-\mu_{BR}n_{r,c_o,c_u}s_{i,r,c_o,c_u}\right)dc_u dc_o dr+\mu_{BR}T. \qquad (16)$$

Under the assumption that individual officers are small relative to the population of police and the groups under scrutiny, each officer takes aggregate search effort (s_{r,c_o,c_u}), and hence π_{r,c_o,c_u}, as given and independent of his own effort choice. The subscript on the Lagrange multiplier μ_{BR} refers to this being the multiplier for a best response problem. The first order conditions of equation (16) are

$$\alpha\left(\beta_{i,r,c_o,c_u}\pi_{r,c_o,c_u}+b_{i,r,c_o,c_u}\right)s_{i,r,c_o,c_u}^{\alpha-1}=\mu_{BR}\ \forall r,c_o,c_u\,. \qquad (17)$$

The optimal search rate of a police officer can be directly computed from the first order conditions for our maximization problem and yields

$$s_{i,r,c_o,c_u}=\frac{\left(\beta_{i,r,c_o,c_u}\pi_{r,c_o,c_u}+b_{i,r,c_o,c_u}\right)^{\frac{1}{1-\alpha}}T}{\int\left(\beta_{i,r',c_o',c_u'}\pi_{r',c_o',c_u'}+b_{i,r',c_o',c_u'}\right)^{\frac{1}{1-\alpha}}n_{r',c_o',c_u'}dc_u'dc_o'dr'}. \qquad (18)$$

Aggregating across police gives group-specific search rates,[14]

---

[14]Notice that we are implicitly assuming that individuals are sampled without replacement in this calculation. That is, we are assuming that adding the individual search rates generates the appropriate aggregate search rate since

$$s_{r,c_o,c_u}=\frac{a_{r,c_o,c_u}}{n_{r,c_o,c_u}}=\frac{\sum_i a_{i,r,c_o,c_u}}{n_{r,c_o,c_u}}=\sum_i s_{i,r,c_o,c_u}$$

If individuals can be searched more than once, we would simply redefine the size of the group accordingly.

$$s_{r,c_o,c_u} = \int \frac{\left(\beta_{i,r,c_o,c_u}\pi_{r,c_o,c_u} + b_{i,r,c_o,c_u}\right)^{\frac{1}{1-\alpha}} MT}{\int \left(\beta_{i,r',c_o',c_u'}\pi_{r',c_o',c_u'} + b_{i,r',c_o',c_u'}\right)^{\frac{1}{1-\alpha}} n_{r',c_o',c_u'} dc_u' dc_o' dr'} dF_{\beta_i,b_i|r,c_o,c_u}, \tag{19}$$

where $dF_{\beta_i,b_i|r,c_o,c_u}$ denotes the distribution of taste parameters across police conditional on the characteristics of the group. Since the guilt probabilities for each group are a function of aggregate search probabilities and the equilibrium search probabilities are determined by search rates of the police, the equilibrium search rate is implicitly defined by

$$s_{r,c_o,c_u} = \int \frac{\left(\beta_{i,r,c_o,c_u}\pi\left(r,c_o,c_u,s_{r,c_o,c_u}\right) + b_{i,r,c_o,c_u}\right)^{\frac{1}{1-\alpha}} MT}{\int \left(\beta_{i,r',c_o',c_u'}\pi\left(r',c_o',c_u',s_{r',c_o',c_u'}\right) + b_{i,r',c_o',c_u'}\right)^{\frac{1}{1-\alpha}} n_{r',c_o',c_u'} dc_u' dc_o' dr'} dF_{\beta_i,b_i|r,c_o,c_u}. \tag{20}$$

Given the nonlinearity in the individual equilibrium stop choices, it is evident from equation (20) that even if the mean or median officer is unbiased, this does not imply that the presence of bias fails to affect aggregate group specific search rates.

Let $g = 1,...,G$ index the groups that can be formed from all the combinations of $\left(r,c_o,c_u\right)$ in the population. Assume the total number of groups, G, is finite. Equation (20) defines a mapping of the space

$$\Omega = \left(s_1 \in [0, MT/n_1]\right) \times \left(s_2 \in [0, MT/n_2]\right) \times ... \times \left(s_G \in [0, MT/n_G]\right), \tag{21}$$

into itself. This is a fixed point problem. The following proposition establishes existence under mild conditions. We therefore state

Proposition 1. Existence of Nash equilibrium

Assume $\pi\left(r,c_o,c_u,s_{r,c_o,c_u}\right)$ is continuous in $s_{r,c_o,c_u}\in[0,MT/n_{r,c_o,c_u}]$. Then a fixed point exists for equation (20). Therefore, a Nash equilibrium exists in aggregate officer stop choices across groups.

Proof: The claim is immediate from Brouwer's fixed point theorem since G has been assumed to be finite. □

i. guilt rates and taste-based discrimination

We now consider the empirical restrictions that our generalized model places on the equilibrium guilt rates across groups. It is evident from equation (18) that, as one takes the limit $\alpha\to 1$ from below, officer i's available searches T become entirely concentrated on a single group, determined by the maximum value of $\beta_{i,r,c_o,c_u}\pi_{r,c_o,c_u}+b_{i,r,c_o,c_u}$ across groups. This is the type of behavior found by KPT when $\beta_{i,r,c_o,c_u}=\beta$ and $b_{i,r,c_o,c_u}=b_r$. For KPT this cannot be an equilibrium, as individuals who are not being policed now have an incentive to carry contraband with probability 1 and those who are being policed with probability 1 should stop carrying contraband. This forms the basis of their conclusion that aggregate policing must be distributed so that guilt probabilities are equalized. Under the KPT or Persico (2009) preferences police allocate relative search efforts until guilt rates are equalized, unless there are groups whose guilt rates when they are never searched are lower than the guilt rates of every group that is searched. We ignore this type of corner solution.

We can now analyze how our model restricts guilt probabilities across groups. Equation (17) requires

$$\left(\beta_{i,r,c_o,c_u}\pi_{r,c_o,c_u}+b_{i,r,c_o,c_u}\right)s_{i,r,c_o,c_u}^{\alpha-1}=\left(\beta_{i,r',c_o',c_u'}\pi_{r',c_o',c_u'}+b_{i,r',c_o',c_u'}\right)s_{i,r',c_o',c_u'}^{\alpha-1}. \qquad (22)$$

For expositional purposes and to facilitate comparison with KPT and Persico (2009), we assume that officers have homogeneous nondiscriminatory preferences, which means that

$$\beta_{i,r,c_o,c_u} = \beta_{c_o,c_u} \tag{23}$$

and

$$b_{i,r,c_o,c_u} = b_{c_o,c_u}\,. \tag{24}$$

We introduce this level of parameter homogeneity as it implies that in equilibrium, each officer chooses an identical cross group distribution of search.[15] The average and individual group search rates thus coincide for each group. We assume homogenous preferences for most of this section. We return to officer heterogeneity in Subsection (iv).

Under this preference homogeneity, the first order condition for individual officers therefore requires that the equilibrium aggregate searches obey

$$\left(\beta_{c_o,c_u}\pi\left(r,c_o,c_u,s_{r,c_o,c_u}\right)+b_{c_o,c_u}\right)s_{r,c_o,c_u}^{\alpha-1} = \left(\beta_{c_o,c_u}\pi\left(r',c_o,c_u,s_{r',c_o,c_u}\right)+b_{c_o,c_u}\right)s_{r',c_o,c_u}^{\alpha-1}\,. \tag{25}$$

Equation (25) provides a way of describing how taste-based discrimination determines equilibrium guilt rates for each group. For our model, this depends on the informational content of race with respect to guilt, conditional on the other group characteristics. Let s denote a fixed level of aggregate policing. Suppose that

$$\pi\left(r,c_o,c_u,s\right) = \pi\left(r',c_o,c_u,s\right). \tag{26}$$

This condition means that given the group characteristics $\left(c_o,c_u\right)$ and equal policing rates s, the probability of guilt is independent of whether the group consists of members of race r or race r'. In other words, race has no marginal predictive value for guilt or innocence once one has accounted for the other elements of the officer's information set.

---

[15]This follows from uniqueness of the optimal search solutions.

Combining (25) and (26), it is immediate that unique equal search rates are applied to the groups and that the two groups have unique equilibrium guilt probabilities, i.e., the realized group-specific guilt and search rates are

$$\pi_{r,c_o,c_u} = \pi_{r',c_o,c_u} \text{ and } s_{r.c_o,c_u} = s_{r',c_o,c_u}. \tag{27}$$

This mimics the KPT finding which equates no discrimination with equal guilt probabilities and thus identifies a dimension along which their test for no discrimination is robust. Our equilibrium condition differs from KPT in that it also implies equal search rates across groups; we return to this below.

The robustness of the equal guilt probability requirement breaks down when we consider divisions of the population other than the groups we have defined. To fully observe a group, it is necessary to observe the entire (r, c_o, c_u) triple. For this reason, the racial profiling literature has focused on the idea of the hit rate for a given population partition. The hit rate is defined as the fraction of searches of members of a partition who are guilty. When one is working with the "true" groups as defined by (r, c_o, c_u) the hit rate equals the guilt rate for the group. For groups that are defined only by observables, the hit rate is influenced by differential search rates applied to subsets of the population comprising the group as well as by the different guilt probabilities for these subsets. For example, the hit rate for race r is the ratio of total guilty persons among those searched of race r, $n_r \int \pi\left(r, c_o, c_u, s_{r,c_o,c_u}\right) s_{r,c_o,c_u} dF_{c_o,c_u|r}$ to the total number of searched people of race r, $n_r \int s_{r,c_o,c_u} dF_{c_o,c_u|r}$. That is, the hit rate for race r, h_r, is defined as

$$h_r = \frac{\int \pi\left(r, c_o, c_u, s_{r,c_o,c_u}\right) s_{r,c_o,c_u} dF_{c_o,c_u|r}}{\int s_{r,c_o,c_u} dF_{c_o,c_u|r}}. \tag{28}$$

Note that $dF_{c_o,c_u|r}$, the conditional density of (c_o, c_u) given race r, is the appropriate conditional probability density in this calculation as the hit rate in this case asks what percentage of members of race r who are searched are guilty. Notice that s_{r,c_o,c_u} is an equilibrium condition that is determined by $\left(r,c_o,c_u\right)$ and varies across the values of $dF_{c_o,c_u|r}$.

Given (20), our functional forms restrict the race-level hit rate to the following

$$h_r = \frac{\int \pi\left(r,c_o,c_u,s_{r,c_o,c_u}\right)\left(\beta_{c_o,c_u}\pi\left(r,c_o,c_u,s_{r,c_o,c_u}\right)+b_{c_o,c_u}\right)^{\frac{1}{1-\alpha}} MTdF_{c_o,c_u|r}}{\int\left(\beta_{\hat{c}_o,\hat{c}_u}\pi\left(\hat{r},\hat{c}_o,\hat{c}_u,s_{\hat{r},\hat{c}_o,\hat{c}_u}\right)+b_{\hat{c}_o,\hat{c}_u}\right)^{\frac{1}{1-\alpha}} n_{\hat{r},\hat{c}_o,\hat{c}_u}d\hat{c}_u d\hat{c}_o d\hat{r}\int s_{r,c_o,c_u}dF_{c_o,c_u|r}} \quad (29)$$

It is evident that, unless $dF_{c_o,c_u|r} = dF_{c_o,c_u|r'}$, it is possible that

$$h_r \neq h_{r'}. \quad (30)$$

Similarly, if one considers observables pairs $\left(r,c_o\right)$ and $\left(r',c_o\right)$, it is possible that

$$h_{r,c_o} \neq h_{r',c_o} \quad (31)$$

unless $dF_{c_u|r,c_o} = dF_{c_u|r',c_o}$.[16]

This illustrates an essential restriction in the KPT framework relative to our framework. The KPT loss functions require that all groups have equal guilt probabilities if taste-based discrimination is absent regardless of any group characteristics. This follows as a general equilibrium result. For instance, if police suspect that men are more likely to carry contraband than women and therefore in the aggregate search them more, women have a relative incentive to carry contraband compared to men. Thus, guilt rates should equalize across observable characteristics, including race, unless police officers derive higher utility from catching criminals of a particular race.

[16]The same argument applies if one conditions on equal stop rates as well as $\left(r,c_o\right)$.

Put differently, KPT avoids the problem of unobservables because their loss functions for officers have the property that race-correlated unobservables do not differentially affect the marginal payoffs to searches. Hence in equilibrium, $\pi\left(r, c_o, c_u, s_{r,c_o,c_u}\right)$ is constant and factors out of (28). Our specification does not have this feature. Indeed it is clear from (28) that equal hit rates across race or across observable characteristics generally do not hold without this feature, so the link between (29) and (31) is not an artifice of our functional forms. What matters in our specification is that we allow preference weights to depend on group characteristics other than race and that we allow for preferences for equal group allocations, other things equal ($\alpha < 1$ in our functional form). Hence the presence of unmeasured group characteristics matters for us in a way that it does not for KPT.

AF and Bjerk (2007) also show that unobservables (c_u) can matter in the context of KPT's loss function if these unobservables are a direct result of the choice to commit the crime. For instance, a person who is carrying contraband may appear more anxious when stopped by police than a person not carrying contraband. It was the choice to commit the crime itself that produced the signal to the officer, and is outside the control of the individual. Our results show that this problem does not depend on their specifications of the source of the unobservables.

The finding that conditional probability dependences between unobservables and race can render the identification of discrimination problematic is a standard problem in discrimination analyses. Heckman and Siegelman (1993), Heckman (1998), and Bohnholz and Heckman (2005) have pioneered this recognition by showing how differences in higher moments in $dF_{c_u|r,c_o}$ and $dF_{c_u|r',c_o}$ can produce spurious evidence of discrimination in contexts ranging from success in loan applications to access to organ transplants. Our findings follow this same line of reasoning. KPT also recognize this problem. This motivates their focus on racial disparities in the productivity of searches (hit rates) rather than the searches themselves, which may be a result of statistical discrimination. The additional challenge that we highlight is that under more general loss functions the hit rates also generally fail to solve the problem of unobservables.

There is a second dimension along which our model produces different results from KPT and Persico (2009). Since our condition (27) also requires equal search rates, it fails to match a key idea in KPT, namely that unequal search rates are consistent with the absence of taste-based discrimination. Put differently, KPT is explicitly interested in the case where race is informative about guilt probabilities at equal policing levels. We therefore now follow KPT and consider the case where race is informative about guilt probabilities, in the sense that

$$\pi\left(r,c_o,c_u,s\right) > \pi\left(r',c_o,c_u,s\right). \tag{32}$$

In this case, unless $\alpha = 1$, equal guilt probabilities require taste-based discrimination (i.e. (23) and/or (24) do not hold). This is immediate from the equilibrium condition (22). Furthermore in the absence of taste-based discrimination, the equilibrium search rates implied by (22) require

$$\pi_{r,c_o,c_u} > \pi_{r',c_o,c_u} \text{ and } s_{r,c_o,c_u} > s_{r',c_o,c_u}. \tag{33}$$

From this vantage point, equal guilt probabilities can represent evidence *in favor of* taste-based discrimination. For example, suppose that the utility of searching guilty parties is independent of race, i.e. (23) holds. Equal guilt probabilities would necessarily imply that $b_{r,c_o,c_u} > b_{r',c_o,c_u}$ if groups are perfectly observable to the econometrician, i.e. there are no group level unobservables. Hence equal guilt rates would imply that officers possess a differential taste for searching various races regardless of guilt. This difference from KPT derives specifically from allowing officer utility to directly depend on the distribution of searches across groups. This is a plausible assumption as it may reflect, for example, returns to scale in the search process. It follows that hit rates for coarser population subdivisions are not necessarily equal either even if taste-based discrimination is absent.

One response to our argumentation is that KPT finding of equal hit rates is a knife edge equilibrium under our alternative modeling assumptions. However, in light of findings such as Sanga (2009), equal hit probabilities are not universally observed and may well result from litigation, in which officers are obliged to constrain the manifestation of their taste for discrimination. And of course, under convex search costs, equal hit rates generally imply discrimination, hence the only question is why prejudiced officers do not oversample certain groups beyond the levels producing equality. The ability to point to an observable feature of their stopping strategy such as equal hit rates as a defense is an example of why this may occur.[17]

Our findings may be summarized in the following two propositions.

Proposition 2. Lack of information about taste-based discrimination from equal hit rates

> Under our generalized stop model, the observation of equal hit rates between groups characterized by (r, c_o) and (r', c_o) is compatible with either the presence or the absence of taste-based discrimination on the part of police.

and

Proposition 3. Lack of information on taste-based discrimination from unequal hit rates

> Under our generalized stop model, the observation of unequal hit rates between groups characterized by (r, c_o) and (r', c_o) is compatible with either the presence or the absence of taste-based discrimination on the part of police.

Together, these propositions illustrate that without strong preference assumptions, guilt probabilities are not informative about bias in police searches even when one allows for additional control variables (which are by definition observable). Nothing above

---

[17] See also our discussion in section 4.

suggests that KPT's or Persico's (2009) methodological analyses are incorrect. Their substantive conclusion that observed hit rates are equal in the absence of taste-based discrimination holds in their model specification. Rather, our analysis illustrates which assumptions are important for their claims to be valid.

It is worth noting that the well-known inframarginality problem is a special case of our non-identification results in Propositions 2 and 3. The inframarginality problem arises when officers would like to search more of a given group, but are prevented from doing so by their time constraints. In this case, the average hit rate does not equal the marginal hit rate, making it difficult to detect discrimination using averages. Endogenizing criminal response, as in the case of KPT, helps alleviate the inframarginality problem because individuals in a group that is more likely to be searched respond by decreasing the probability of carrying contraband, which decreases the utility of searching that group. In our context, we additionally incorporate the potential diminishing marginal utility from searching a given group, which also helps alleviate the inframarginality problem. Importantly, the identification problems we discuss above extend beyond the inframarginality concern.

Some additional insight may be derived by imposing further assumptions. First, we assume a constant elasticity (with respect to searches) functional form assumption on group specific guilt probabilities, such that guilt probabilities do not directly depend on race, i.e.

$$\pi\left(r, c_o, c_u, s_{r,c_o,c_u}\right) = \pi\left(c_o, c_u\right) s_{r,c_o,c_u}^{-\gamma}, \tag{34}$$

Second, we assume that

$$b_{i,r,c_o,c_u} = 0,\ \forall i, r, c_o, c_u. \tag{35}$$

These assumptions allow for a closed form solution to aggregate search rates in each group:

$$s_{r,c_o,c_u} = \frac{\left(\beta_{c_o,c_u}\pi\left(c_o,c_u\right)\right)^{\frac{1}{(1-\alpha+\gamma)}}}{\int \left(\beta_{c_o',c_u'}\pi\left(c_o',c_u'\right)\right)^{\frac{1}{(1-\alpha+\gamma)}} n_{c_o',c_u'}dc_u'dc_o'} MT. \tag{36}$$

We may now draw several implications. Take two groups $\left(r,c_o,c_u\right)$ and $\left(r',c_o,c_u\right)$. From (36), we see that aggregate policing is the same across races with the same characteristics even if police are biased against certain characteristics other than race. Second, hit rates $h_{r,c_o,c_u} = h_{r',c_o',c_u'}$ are equated for all groups with the same characteristics independent of race. Third, the measured hit rates

$$h_{r,c_o} = \frac{\int \pi\left(r,c_o,c_u,s_{r,c_o,c_u}\right)s_{r,c_o,c_u}dF_{c_u|r,c_o}}{\int s_{r,c_o,c_u}dF_{c_u|r,c_o}} \tag{37}$$

for races with the same observable characteristics (c_o) do not have to be equated across r. We verify this last claim in detail.

To see why the absence of taste-based discrimination does not equalize h_{r,c_o} and h_{r',c_o}, first note that the numerator of h_{r,c_o}, as seen in equation (37), has the closed form solution

$$\int \pi\left(c_{o,},c_u\right)s_{r,c_o,c_u}^{1-\gamma}dF_{c_u|r,c_o} =$$
$$\int \pi\left(c_o,c_u\right)\left(\frac{\left(\beta_{c_o,c_u}\pi\left(c_o,c_u\right)\right)^{\frac{1}{(1-\alpha+\gamma)}} MT}{\int \left(\beta_{c_o',c_u'}\pi(c_o',c_u')\right)^{\frac{1}{(1-\alpha+\gamma)}} n_{c_o',c_u'}dc_u'dc_o'}\right)^{1-\gamma} dF_{c_u|r,c_o} \tag{38}$$

This expression provides the first reason why measured hit rates can be race dependent: the presence of $\pi(c_o,c_u)$ in the numerator of the measured hit rate. While we have assumed that guilt rates and preference are identical across individuals of different races but with the same characteristics, integration against $dF_{c_u|r,c_o}$ renders (38) race dependent.

Similar arguments apply to $\int s_{r,c_o,c_u} dF_{c_u|r,c_o}$, the denominator of (37), as it obeys

$$\int s_{r,c_o,c_u} dF_{c_u|r,c_o} = \int \frac{\left(\beta_{c_o,c_u}\pi\left(c_o,c_u\right)\right)^{\frac{1}{(1-\alpha+\gamma)}} MT}{\int \left(\beta_{c_o',c_u'}\pi(c_o',c_u')\right)^{\frac{1}{(1-\alpha+\gamma)}} n_{c_o',c_u'} dc_u' dc_o'} dF_{c_u|r,c_o} \quad (39)$$

The presence of $\pi\left(c_{o,},c_u\right)$ and integration against $dF_{c_u|r,c_o}$ in the RHS of this expression also can render (39) race-dependent.

ii. alternative loss functions and coordinated police activity

So far, we have considered a decentralized environment in which police officers noncooperatively choose effort levels when utility depends on both overall searches and searches of the guilty. This ignores the role of the police administration in influencing police behavior. A natural objective function of the administration is to allocate police searches so as to minimize overall crime. We call this the socially optimal allocation.[18] Dominitz and Knowles (2006) also consider how a preference for crime minimization affects the KPT test for racial profiling by studying the effort allocation of a single representative police officer who can be interpreted as a police chief. We consider whether the Dominitz and Knowles (2006) equilibrium search rates under coordination are different from our generalized model of noncooperative search decisions.

For our environment crime minimization occurs when aggregate police search rates are set via the optimization problem

---

[18]Justifications may be drawn from Persico (2002), Harcourt (2004) and Durlauf (2006) as to why a crime minimization criterion is a more appropriate policy objective than arrest maximization; these same arguments would also imply that crime minimization should trump officer utility from stops. Both Harcourt and Durlauf as well as Sklansky (2008, ch. 7) argue that there are ethical constraints that should impinge on the crime minimization solution when assessing social optima, but these are factors that lie outside of our model and so are ignored.

$$\min_{s_{r,c_o,c_u}} \int \left(\left(\pi\left(r, c_o, c_u, s_{r,c_o,c_u}\right) - \mu_S s_{r,c_o,c_u} \right) n_{c_o,c_u,r} dc_u dc_o dr \right) + \mu_S MT, \tag{40}$$

where the subscript on μ_S, the Lagrange multiplier for the aggregate police resource constraint, refers to this being a social problem. We maintain the constant elasticity crime function but allow for race-based differences in guilt, i.e. (34) is modified to

$$\pi\left(r, c_o, c_u, s_{r,c_o,c_u}\right) = \pi\left(r, c_o, c_u\right) s_{r,c_o,c_u}^{-\gamma} \tag{41}$$

to allow for race to have intrinsic predictive value in group-specific guilt. The first order condition for (40), under assumption (41) requires that the marginal product of policing between groups be equalized. This means

$$\gamma\pi\left(r, c_o, c_u\right) s_{r,c_o,c_u}^{-\gamma-1} = \gamma\pi\left(r', c_o', c_u'\right) s_{r',c_o',c_u'}^{-\gamma-1}, \tag{42}$$

so that the guilt rates across groups are related to search rates by

$$\frac{\pi\left(r, c_o, c_u\right) s_{r,c_o,c_u}^{-\gamma}}{\pi\left(r', c_o', c_u'\right) s_{r',c_o',c_u'}^{-\gamma}} = \frac{s_{r,c_o,c_u}}{s_{r',c_o',c_u'}}. \tag{43}$$

We contrast this with the equilibrium guilt probabilities when officers choose search rates noncooperatively. Suppose one assumes the Persico (2009) nondiscriminatory preferences in the sense that $\beta_{r,c_o,c_u} \equiv \beta, b_{r,c_o,c_u} \equiv 0$. Following (17), the ratio of equilibrium guilt rates in the noncooperative model is

$$\frac{\pi\left(r, c_o, c_u\right) s_{r,c_o,c_u}^{-\gamma}}{\pi\left(r', c_o', c_u'\right) s_{r',c_o',c_u'}^{-\gamma}} = \frac{s_{r,c_o,c_u}^{1-\alpha}}{s_{r',c_o',c_u'}^{1-\alpha}}. \tag{44}$$

As $\alpha \to 0$, the ratios of guilt probabilities of the social planner and noncooperative equilibria coincide as the search rates coincide. Recall that α measures the extent to which more intensive search of a group diminishes the payoffs from stops of guilty drivers; as we have assumed that utility of stops of innocent drivers is 0. Sufficient concavity of this payoff, i.e. sufficiently small α can render the planner and noncooperative equilibria arbitrarily close to one another. Therefore, in the absence of discrimination, there is no necessary difference in the implications of noncooperative and social planning views of search rate determination.

Proposition 4. Lack of information about objective functions from hit rates

> For the generalized choice model, hit rates across observed groups do not necessarily distinguish between a coordinated distribution of searches that minimizes crime and a noncooperative equilibrium in which police searches generate utility based on arrest maximization and search rate intensity.

To be clear, the observational equivalence of the planner and noncooperative equilibria that we state requires (41). But this also suggests that distinguishing between the types of equilibria is difficult when (41) is a good approximation of the guilt rate process and when marginal increases in search rates are extremely costly to an officer.

iii. lack of information from effort differences

We now discuss what information can be revealed on taste-based discrimination from differences in police effort across groups without restrictions on unobservables. We continue to assume (35) and that police officers are homogeneous (23). We also assume that

$$\pi\left(r,c_o,c_u,s_{r,c_o,c_u}\right)=\pi\left(r,c_o,c_u\right)s_{r,c_o,c_u}^{-\gamma} \tag{45}$$

holds. In analogy to (36), the unique group specific search rate is

$$s_{r,c_o,c_u} = \frac{\left(\beta_{c_o,c_u}\pi\left(r,c_o,c_u\right)\right)^{\frac{1}{(1-\alpha+\gamma)}} MT}{\int \left(\beta_{c'_o,c'_u}\pi\left(r',c'_o,c'_u\right)\right)^{\frac{1}{(1-\alpha+\gamma)}} n_{r',c'_o,c'_u} dc'_u dc'_o dr'}. \qquad (46)$$

From (46), the equilibrium ratio of policing on group $\left(r,c_o\right)$ to group $\left(r',c_o\right)$ is

$$\frac{s_{r,c_o}}{s_{r',c_o}} = \frac{\int\left(\beta_{c_o,c_u}\pi\left(r,c_o,c_u\right)\right)^{\frac{1}{(1-\alpha+\gamma)}} dF_{c_u|r,c_o}}{\int\left(\beta_{c_o,c_u}\pi\left(r',c_o,c_u\right)\right)^{\frac{1}{(1-\alpha+\gamma)}} dF_{c_u|r',c_o}} \qquad (47)$$

when taste-based discrimination is absent. We may now state a strong observational equivalence result.

Proposition 5. Lack of information about taste-based discrimination from differences in search rates applied to different races

Let $\frac{s_{r,c_o}}{s_{r',c_o}}$ be any observed ratio of policing on race r and race r' where the observed characteristics c_o are the same. Then, there exist values of $\beta_{c_o,c_u}, dF_{c_u|r,c_o}, dF_{c_u|r',c_o}$, $\pi\left(r,c_o,c_u\right)$, and $\pi\left(r',c_o,c_u\right)$ such that (47) holds.

As an extreme example suppose $\beta_{c_o,c_u}\pi\left(r,c_o,c_u\right)$ and $\beta_{c_o,c_u}\pi\left(r',c_o,c_u\right)$ each take only two values one of which is zero and the other is positive and identical for both. Suppose the two probability densities over which integration occurs in the variant of (46) that is germane to our assumptions only place positive mass on these two values. For this case, the ratio of search rates obeys $\frac{s_{r,c_o}}{s_{r',c_o}} = \frac{\Pr\left(r,c_o\right)}{\Pr\left(r',c_o\right)}$ where the $\Pr\left(\cdot,\cdot\right)$ terms are the conditional probabilities of the positive value. Since these probabilities can be anywhere between zero and one, we can choose them so their ratio is any positive number or zero or infinity. For more general cases, it is immediately evident that we have enough "degrees of freedom" to implement (47) when we are free to choose $\beta_{c_o,c_u}, dF_{c_u|r,c_o}, dF_{c_u|r',c_o}$, $\pi(r,c_o,c_u)$, and $\pi(r',c_o,c_u)$ such that (46) holds.

While this result is mathematically trivial it is important for identification issues. Even though the police above are not racially biased and the crime propensity functions of the two races are exactly the same, the conditional distributions of unobservables can be such that any observed ratio of policing across the races can be *observed*. Turning to actual policy debates the main public policy concern is "excessive" policing of blacks given the same observed characteristics as whites. As Proposition 5 demonstrates, one cannot conclude from the observation of "excessive" policing of blacks relative to whites that the police are biased. This of course is an essential insight of KPT; our proposition demonstrates that their finding holds across alternative specifications.

iv. lack of information from officer comparisons

As illustrated above, the key difference between our analysis and KPT is that unobservables affect guilt probabilities in our setting, thus making it difficult to detect taste-based discrimination. AF introduces a role for unobservables in the KPT setting by focusing on a particular type of unobservable that signals a potential criminal's guilt to the police officer. As opposed to KPT where in equilibrium guilt rates are equalized across unobservable characteristics, this is no longer the case. In the AF setting, these unobservables lead to differences across observed (to the econometrician) hit rates because they are behaviors observed by the officer that are beyond the individual's control. Officers use this unobservable to predict the guilt of the potential criminal and the criminal cannot eliminate the signal so that guilt probabilities do not equalize across the unobservable in equilibrium. An example AF use is that an individual who is carrying contraband may appear more nervous or agitated during a police stop which signals guilt to the officer.

Given these unobservables, AF draws a conclusion parallel to our Proposition 2, namely, equal hit rates across observed groups is no longer an implication of the absence of taste-based discrimination. Notice, however, that our setting does not restrict the role of the unobservables to that of a signal of guilt. We permit unobservables to matter in determining police officer preferences and do not specify the distribution of unobservables.

To obtain testable implications, AF assumes that the unobservable is single-dimensional and that it satisfies a monotone likelihood ratio property, so that effectively higher values of the unobservable mean that the individual is more likely to be guilty. AF also generalizes KPT's model by introducing officer heterogeneity, such that the cost to searching criminals of different races varies by the race of the officer.

Under these restrictions, AF's main empirical implication is that the relative rankings of search rates (and resulting hit rates) are preserved across officers if the officers are racially unbiased. In particular, the AF rank test (AF (2006, page 131, pages 145 and 146, equations (6) and (7)) in our notation is

$$\begin{aligned} s_{i,r,c_o} > s_{i',r,c_o} \text{ and } h_{i,r,c_o} < h_{i',r,c_o} \Rightarrow \\ s_{i,r',c_o} > s_{i',r',c_o} \text{ and } h_{i,r',c_o} < h_{i',r',c_o}. \end{aligned} \tag{48}$$

In words, if officer i searches a potential criminal of race r at a higher rate with a lower hit rate than officer i', the same must also be true in their relative behaviors toward race r' if one officer is not relatively more racially-biased than another. AF is particularly concerned with heterogeneity across black, white and Hispanic officers. We allow the heterogeneity to be individual-officer specific, but our argument could clearly be applied to their level of aggregation. In what follows, we consider whether a similar rank condition holds in our framework under equivalent assumptions on the unobservables.

Mapping the AF analysis into our framework, searches of a group are still determined by the values of the triple (r, c_o, c_u).[19] Similar to AF, we make the following assumptions. We restrict the role of the unobservables as follows. First, we assume that

[19] We do not need to take a stance on the source of unobservables. As stated above, unobservables retain meaning in our model because we assume a less restrictive functional form than KPT.

c_u is a scalar random variable. We further restrict its role so that the main implication of AF's monotone likelihood assumption holds, namely that π_{r,c_o,c_u} is monotonically increasing in c_u. Second, we mimic AF's officers' preferences by imposing $\beta_{i,r,c_o,c_u} = 1$ and $b_{i,r,c_o,c_u} = b_i$. if officers are unbiased, while maintaining the assumption of weakly diminishing marginal utility from searches, i.e. $\alpha < 1$.

Under these assumptions, if the relative search rates for two officers i, i' satisfy

$$\frac{s_{i,r,c_o}}{s_{i',r,c_o}} = \frac{\int \left(\pi_{r,c_o,c_u} + b_i\right)^{\frac{1}{(1-\alpha)}} dF_{c_u|r,c_o}}{\int \left(\pi_{r,c_o,c_u} + b_{i'}\right)^{\frac{1}{(1-\alpha)}} dF_{c_u|r,c_o}} > 1, \tag{49}$$

then it does not necessarily follow that

$$\frac{s_{i,r',c_o}}{s_{i',r',c_o}} = \frac{\int \left(\pi_{r',c_o,c_u} + b_i\right)^{\frac{1}{(1-\alpha)}} dF_{c_u|r',c_o}}{\int \left(\pi_{r',c_o,c_u} + b_{i'}\right)^{\frac{1}{(1-\alpha)}} dF_{c_u|r',c_o}} > 1, \tag{50}$$

even if $\frac{h_{i,r,c_o}}{h_{i',r,c_o}} < 1$ and $\pi_{r,c_o,c_u} = \pi_{c_o,c_u}$. Since the distribution of the unobservables depends on race, condition (49) does not provide enough of a restriction to determine whether (50) holds.

Thus, AF's rank condition (48) does not necessarily hold even in the absence of discrimination. This is precisely Heckman's argument about the effects of the conditional density of unobservables confounding tests for taste-based discrimination. The key is that unobservables play a different role than in AF's analysis. Suppose we start with a situation where (49) and (50) hold. We can change the distribution of unobservables, while maintaining π_{r,c_o,c_u} is monotonically increasing in c_u and violate the rank condition.

Rank condition (48) may also fail even if we replace $\pi_{r,c_o,c_u} = \pi_{c_o,c_u}$ with $dF_{c_u|r,c_o} = dF_{c_u|c_o}$. Only when the distributions do not depend on race ($dF_{c_u|r,c_o} = dF_{c_u|c_o}$) and the guilt probabilities are independent of race ($\pi_{r,c_o,c_u} = \pi_{c_o,c_u}$), does AF's condition necessarily follow since nothing depends on race anymore.

Section 4. Uncovering biased officers

In this section we consider the problem of a police administration (generically referred to as the police chief) who, intent on removing biased officers from the police force, follows two alternative commonly suggested policies.[20] In the first case, he eliminates racial profiling by making sure officers' search rates are the same across groups that differ only by race. In the second policy he makes sure officers' hit rates are the same across groups that differ only by race. In particular we want to understand whether biased officers could survive if either policy is implemented. Since throughout we assume that the police chief has access to the same information as an officer regarding the criminals, i.e. to (r, c_o, c_u), the distinction between c_o and c_u is not relevant. We simplify notation by letting $c = (c_o, c_u)$.

i. equal search rates

Suppose that a police chief implements a policy where officers who do not have equal search rates across groups that differ only by race are fired. We first consider what the chief would find if officers do not take the policy into account when choosing who to search. We assume that $b_{i,r,c} = 0$. From (18) it follows that the ratio of search rates for two groups differing only by race is

[20]In this section, we do not explicitly prove existence of an equilibrium for the constrained problems we present. Instead, we simply ask whether whenever an equilibrium exists biased officers can adjust their behavior and still act on their bias.

$$\frac{s_{i,r,c}}{s_{i,r',c}} = \left(\frac{\beta_{i,r,c}\pi_{r,c}}{\beta_{i,r',c}\pi_{r',c}} \right)^{\frac{1}{1-\alpha}}. \tag{51}$$

Biased officers can survive if the extent of their bias is such that this ratio equals one. Unbiased officers would not survive unless guilt rates were equal across races.

This conclusion ignores strategic behavior on the part of prejudiced officers. If officers take the policy into account when choosing their search rates, the question of interest is whether biased officers adjust their behavior to equalize search rates, while still acting on their bias. Officers choose search rates conditional on them being equal across groups that differ only by race so they face the constraint

$$s_{i,r,c} = s_{i,r',c} = s_{i,c}. \tag{52}$$

The officer chooses search rates $s_{i,r,c}$ to maximize his utility $\int \beta_{i,r,c}\pi_{r,c}n_{r,c}s_{i,r,c}^{\alpha}dcdr$ subject to both (52) and the time constraint $\int n_{r,c}s_{i,r,c}dcdr \le T$.

Replacing (52) into the time constraint we can rewrite it as

$$\int s_{i,c} \int n_{r,c} drdc = \int n_c s_{i,c} dc \le T. \tag{53}$$

If we also replace (52) into the utility function, we have the modified objective function

$$\int \left(s_{i,c}^{\alpha} \int \beta_{i,r,c}\pi_{r,c}n_{r,c}dr \right) dc. \tag{54}$$

Equations (53) and (54) produce the Lagrangian for the officer’s problem

$$\int \left(s_{i,c}^{\alpha} \int \beta_{i,r,c}\pi_{r,c}n_{r,c}dr - \mu_{ES}n_c s_{i,c} \right) dc + \mu_{ES}T, \tag{55}$$

where the subscript on μ_{ES}, the Lagrange multiplier for the officer's time constraint under (52), refers to this being the equal search rates problem . The solution to (55) produces an optimal search rate given by

$$s_{i,r,c} \equiv s_{i,c} = \frac{\left(\frac{\int \beta_{i,r,c} \pi_{r,c} n_{r,c} dr}{n_c} \right)^{\frac{1}{1-\alpha}}}{\int \left(\frac{\int \beta_{i,r',c'} \pi_{r',c'} n_{r',c'} dr'}{n_{c'}^{\alpha}} \right)^{\frac{1}{1-\alpha}} dc'} T. \tag{56}$$

By construction any officer, whether he is racially biased or not, can survive the policy followed by the police chief.

To illustrate how racial bias is reflected in the differential behavior between biased and unbiased officers, assume that tastes do not depend on the characteristics c of the potential criminal, so that $\beta_{i,r,c} = \beta_{i,r}$ for every officer. Assume further that officer i is biased, i.e. $\beta_{i,r} > \beta_{i,r'}$, and officer i' is not biased, i.e., $\beta_{i',r} = \beta_{i',r'}$. Then the optimal search rates in equation (56) are such that officer i focuses his searches more intensely on c-groups with higher proportions of guilty criminals of race r relative to officer i'. As a result, the total search rate of potential criminals of race r is higher for officer i relative to the unbiased officer.

This example illustrates that officers who equalize search rates for groups that differ only by race can still act on their biased preferences. The margin of discrimination in this context appears in the allocation of searches across c-groups. For example if black males disproportionately wear baggy-pants relative to white-males, an officer who wants to discriminate against black males in this setting searches baggy-pant-wearing males at higher rates than non-baggy-pant-wearing males. We return to this issue in Section 5.

ii. equal hit rates

Suppose that the police chief now imposes the objective of equal hit rates across groups that differ only by race. Since the hit rate and the guilt probability are the same at the (r,c) level, this means that officers need to satisfy

$$\pi_{r,c} = \pi_c \;\; \forall r. \tag{57}$$

As before, we first consider what the police chief would find if officers do not respond to such a policy. From our discussion in Section 3, guilt probabilities are generally not equalized, even for unbiased officers. Hence the majority of officers would not survive the policy.

Now, suppose that we allow officers to respond to the introduction of the policy. In this case, the officer chooses search rates $s_{i,r,c}$ to maximize his utility $\int \beta_{i,r,c}\pi_{r,c}n_{r,c}s_{i,r,c}^{\alpha}dcdr$ subject to both (56) and the time constraint $\int n_{r,c}s_{i,r,c}dcdr \leq T$. Replacing (56) into the utility function, we obtain $\int \left(\pi_c \int \beta_{i,r,c}n_{r,c}s_{i,r,c}^{\alpha}dr\right)dc$. The Lagrangian describing the officer's problem is then

$$\int \pi_c \int \left(\beta_{i,r,c}n_{r,c}s_{i,r,c}^{\alpha} - \mu_{EH}n_{r,c}s_{i,r,c}\right)drdc + \mu_{EH}T, \tag{58}$$

where μ_{EH} is the Lagrange multiplier for the equal hit rates problem.

The solution to this problem is given by the optimal search rate,

$$s_{i,r,c} = \frac{\left(\pi_c\beta_{i,r,c}\right)^{\frac{1}{1-\alpha}}}{\int \left(\pi_{c'}\beta_{i,r',c'}\right)^{\frac{1}{1-\alpha}} n_{r',c'}dr'dc'}T. \tag{59}$$

By construction the hit rates for groups that differ only by race are equal and so both biased and unbiased officers survive this policy. Furthermore, hit rates are also equal across individuals.

To understand how biased officers act on their bias in this context, take the ratio of search rates for two groups differing only by race

$$\frac{s_{i,r,c}}{s_{i,r',c}} = \left(\frac{\beta_{i,r,c}}{\beta_{i,r',c}}\right)^{\frac{1}{1-\alpha}}. \tag{60}$$

Unbiased officers also equalize search rates for groups that differ only by race. Biased officers on the other hand, search groups proportionally to their bias. If we further look at search rates at the race level we get

$$\frac{s_{i,r}}{s_{i,r'}} = \frac{\int \left(\pi_c \beta_{i,r,c}\right)^{\frac{1}{1-\alpha}} dF_{c|r}}{\int \left(\pi_{\hat{c}} \beta_{i,r',\hat{c}}\right)^{\frac{1}{1-\alpha}} dF_{\hat{c}|r'}}, \tag{61}$$

and neither biased nor unbiased officers equalize search rates (or hit rates) purely based on race if the distribution of c depends on race.

iii. equal search and hit rates

Finally, we consider the case where the police chief imposes the objective of both equal search rates and hit rates across groups that differ only by race. The officer now chooses to maximize his utility $\int \beta_{i,r,c} \pi_{r,c} n_{r,c} s_{i,r,c}^{\alpha} dcdr$ subject to (52), (57) and the time constraint $\int n_{r,c} s_{i,r,c} dcdr \leq T$. In parallel to earlier reasoning substitute (52) into the time constraint to get (53). Substituting (52) and (57) into the utility function produces

$$\int \left(s_{i,c}^{\alpha} \pi_c \int \beta_{i,r,c} n_{r,c} dr \right) dc. \tag{62}$$

The officer chooses search rates to maximize the Lagrangian

$$\int \left(s_{i,c}^{\alpha} \pi_c \int \beta_{i,r,c} n_{r,c} dr - \mu_{EHS} n_c s_{i,c} \right) dc + \mu_{EHS} T, \tag{63}$$

where μ_{EHS} is the Lagrange multiplier for the equal hit and search rates problem. The solution is

$$s_{i,r,c} \equiv s_{i,c} = \frac{\left(\frac{\pi_c \int \beta_{i,r,c} n_{r,c} dr}{n_c} \right)^{\frac{1}{1-\alpha}}}{\int \left(\frac{\pi_{c'} \int \beta_{i,r',c'} n_{r',c'} dr'}{n_{c'}^{\alpha}} \right)^{\frac{1}{1-\alpha}} dc'} T. \tag{64}$$

By construction both biased and unbiased officers survive the policy.

The ratio of search rates for officers i and i' depends only on the relative proportion of individuals of a given race in each c-group, i.e.

$$\frac{s_{i,r,c}}{s_{i',r,c}} = \frac{\left(\int \beta_{i,r,c} n_{r,c} dr \right)^{\frac{1}{1-\alpha}}}{\left(\int \beta_{i',r,c} n_{r,c} dr \right)^{\frac{1}{1-\alpha}}}. \tag{65}$$

As before, bias is reflected in higher search rates for groups containing more members of the particular race against which the officer is biased. In contrast to Subsection (i), where officers only equalized search rates, disparities in guilt probabilities across races do not play a role.

These examples above illustrate that when police chiefs act to eliminate racial bias imposing common tests used in the literature, racially biased officers can survive and continue to act on their bias. In comparison to our results in Section 3 and 4, these results do not depend on the existence of unobservables. For example, if police chiefs use equal hit rates to detect bias, KPT’s test may hold even in the presence of bias because police officers adjust their behavior in response to the policy. Notice that this is not the consequence of unobservables since the police chief has access to the same information regarding the criminal as officers. Instead, they follow because of the ability of officers to use variables that are correlated with race in their search allocations in order to indirectly sample different races differentially.

Section 5. Testable implications from assumptions on unobservables

In the previous sections, we illustrate the difficulties in obtaining information about the bias of police officers. We now develop examples of testable implications about bias in our model, based on additional assumptions on the nature of unobservable heterogeneity. These are not exhaustive. Throughout, we assume (35), i.e. that officers only care about arresting guilty criminals.

i. information from observable group search rates with homogeneous officers

Our first example focuses on restrictions placed on search rates for observed groups (r, c_o). Assume that officer preferences are homogeneous in the sense of (23). Suppose further that race has no intrinsic value in predicting guilt probabilities, so that (34) holds. Finally, assume that unobservables are such that

$$\forall c_o,\ \pi(c_o, c_u) \text{ and } \beta_{c_o, c_u} \text{ are both monotonically increasing in } c_u. \qquad (66)$$

For this case, one can relate assumptions about the conditional distribution of unobservables to the search rates across races.

Proposition 6. First order stochastic dominance of unobservables and search rates

(a) If $dF_{c_u|r,c_o}$ first order stochastically dominates $dF_{c_u|r',c_o}$ then the ratio of the rate at which members of race r are searched relative to race r' is greater than 1 even though the observed characteristics are the same

$$\frac{\int s_{r,c_o,c_u} dF_{c_u|r,c_o}}{\int s_{r',c_o,c_u} dF_{c_u|r',c_o}} > 1. \tag{67}$$

(b) If $dF_{c_u|r,c_o} = dF_{c_u|r',c_o}$, then the ratio in (66) is unity.

Proof: From (39) calculate the ratio $\frac{\int s_{r,c_o,c_u} dF_{c_u|r,c_o}}{\int s_{r',c_o,c_u} dF_{c_u|r',c_o}}$; note that the term $\int \left(\beta_{c_o',c_u'} \pi(c_o', c_u')\right)^{\frac{1}{(1-\alpha+\gamma)}} n_{c_o',c_u'} dc_u' dc_o'$ cancels out. Application of first order stochastic dominance to the remaining terms in the numerator and denominator immediately gives (a). Part (b) follows by definition. □

This proposition is of interest in understanding how assumptions about unobservables matter in interpreting racial profiling claims. The Henry Louis Gates affair is a prominent example of alleged racial profiling.[21] For high c_o individuals like Gates, policy makers might assume roughly the same conditional distributions $dF_{c_u|r,c_o}$ and $dF_{c_u|r',c_o}$. Note that Proposition 6.b implies that hit rates should be the same across races for Gate's level of c_o. It is this latter belief that renders differential guilt and differential policing conditional on high levels of c_o to be prima facie evidence of discrimination by police against blacks.

What conditions could lead to a low relative hit rate ratio for blacks relative to whites when police are not racially biased? Here is an example. Police are biased against young males who wear baggy pants and whose belts hang far below the waistline. Pants style is not observed by the econometrician but is observed by the officers. Suppose that the observed c_o is high, as might be the case for male college students. Suppose further that young black college males like wearing baggy pants. Finally,

[21] See Harcourt (2009).

suppose previous studies have shown that the average observed crime rate for black male college students is the same as for white male college students. Given this, it is reasonable to assume that conditional guilt rates for male college students are independent of race. Since baggy-pants-biased police like to search baggy-pants-wearing males we get a high ratio of policing of black males relative to white males. This example reinforces the conclusions of KPT, Persico (2009) and others that the policy problems associated with racial bias versus other types of police bias are subtle.

ii. information from comparisons of search rates across heterogeneous officers

Instead of placing restrictions on guilt probability and the distribution of unobservables across races, we now consider what information can be learned from comparisons across heterogeneous officers. We begin by assuming that the officer's preferences do not vary over unobservable characteristics, i.e. $\beta_{i,r,c_o,c_u} = \beta_{i,r,c_o}$. The relative search rates for two officers i, i' satisfy

$$\frac{s_{i,r,c_o}}{s_{i,r',c_o}} = \frac{\int \left(\beta_{i,r,c_o}\pi_{r,c_o,c_u}\right)^{\frac{1}{(1-\alpha)}} dF_{c_u|r,c_o}}{\int \left(\beta_{i,r',c_o}\pi_{r',c_o,c_u}\right)^{\frac{1}{(1-\alpha)}} dF_{c_u|r',c_o}} = \left(\frac{\beta_{i,r,c_o}}{\beta_{i,r',c_o}}\right)^{\frac{1}{(1-\alpha)}} \frac{\int \left(\pi_{r,c_o,c_u}\right)^{\frac{1}{(1-\alpha)}} dF_{c_u|r,c_o}}{\int \left(\pi_{r',c_o,c_u}\right)^{\frac{1}{(1-\alpha)}} dF_{c_u|r',c_o}}. \qquad (68)$$

Taking the ratio of relative search rates across officers, we have

$$\frac{\dfrac{s_{i,r,c_o}}{s_{i,r',c_o}}}{\dfrac{s_{i',r,c_o}}{s_{i',r',c_o}}} = \left(\frac{\dfrac{\beta_{i,r,c_o}}{\beta_{i,r',c_o}}}{\dfrac{\beta_{i',r,c_o}}{\beta_{i',r',c_o}}}\right)^{\frac{1}{(1-\alpha)}}. \qquad (69)$$

If this ratio is greater than 1, officer i exhibits relatively greater bias against race r (versus r') than officer i'. The hit rates provide no additional information in this case since by definition

$$h_{i,r,c_o} = \frac{\int \pi_{r,c_o,c_u} \left(\beta_{i,r,c_o} \pi_{r,c_o,c_u}\right)^{\frac{1}{1-\alpha}} dF_{c_u|r,c_o}}{\int \left(\beta_{i,r,c_o} \pi_{r,c_o,c_u}\right)^{\frac{1}{1-\alpha}} dF_{c_u|r,c_o}} = \frac{\int \pi_{r,c_o,c_u} \left(\pi_{r,c_o,c_u}\right)^{\frac{1}{1-\alpha}} dF_{c_u|r,c_o}}{\int \left(\pi_{r,c_o,c_u}\right)^{\frac{1}{1-\alpha}} dF_{c_u|r,c_o}}, \tag{70}$$

and so they do not depend on officer bias.

iii. multiple officer types

We conclude our analysis by considering how one can uncover the fraction of prejudiced officers in a population. This subsection illustrates how this might be done by further restricting the role of unobservables. This leads to partial identification results in the spirit of Manski. For tractability, we simplify the problem by assuming that officers form groups based only on race and that race can take only two values: r and r'. Since we focus on discrimination against race r we impose the normalization that $\beta_{r'} = 1$ across all officers. We further simplify the problem by assuming that there are only two types of officers: "type" *I* officers for whom $\beta_r \neq 1$ (i.e. biased officers) and "type" *II* officers for whom $\beta_r = 1$ (i.e. unbiased officers).

Under our assumptions, the search rates for type *I* officers are

$$s_r^I = \frac{\left(\beta_r \pi_r\right)^{\frac{1}{1-\alpha}} T}{\left(\beta_r \pi_r\right)^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}}, \tag{71}$$

and

$$s_{r'}^I = \frac{\pi_{r'}^{\frac{1}{1-\alpha}} T}{\left(\beta_r \pi_r\right)^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}}. \tag{72}$$

while the search rates for unbiased (type *II*) officers are

$$s_r^{II} = \frac{\pi_r^{\frac{1}{1-\alpha}} T}{\pi_r^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}}, \tag{73}$$

and

$$s_{r'}^{II} = \frac{\pi_{r'}^{\frac{1}{1-\alpha}} T}{\pi_r^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}}. \tag{74}$$

If we let p denote the (unknown) fraction of type I officers in the population, the observed search rates for each race are

$$s_r = \frac{(\beta_r \pi_r)^{\frac{1}{1-\alpha}} MT}{(\beta_r \pi_r)^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}} p + \frac{\pi_r^{\frac{1}{1-\alpha}} MT}{\pi_r^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}} (1-p), \tag{75}$$

and

$$s_{r'} = \frac{\pi_{r'}^{\frac{1}{1-\alpha}} MT}{(\beta_r \pi_r)^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}} p + \frac{\pi_{r'}^{\frac{1}{1-\alpha}} MT}{\pi_r^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}} (1-p). \tag{76}$$

Notice that, given our simplifying assumptions, the guilt probabilities are known to the econometrician since they equal the observed hit rates. It then follows that equations (75) and (76) give us a system of 2 equations with 3 unknowns: p, α and β_r. From this system one can find identification regions for the 3 parameters. For example, for each point in a grid of values $\alpha < 1$, equations (75) and (76) imply values for p and

β_r. If either $p \notin [0,1]$ or $\beta_r < 0$, then the parameter vector is rejected. By proceeding over all points in the grid, the identified region may be recovered.

In some situations interest may be centered on the proportion of racially biased officers p. If this is the case, a lower bound on p can be obtained by considering the extreme scenario in which β_r is so large that type *I* officers focus their search exclusively on race r, so that their search rates are $\frac{T}{n_r}$. In this case the observed search rates become

$$s_r = \frac{MT}{n_r} p + \frac{\pi_r^{\frac{1}{1-\alpha}} MT}{\pi_r^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}} (1-p) \tag{77}$$

and

$$s_{r'} = \frac{\pi_{r'}^{\frac{1}{1-\alpha}} MT}{\pi_r^{\frac{1}{1-\alpha}} n_r + \pi_{r'}^{\frac{1}{1-\alpha}} n_{r'}} (1-p). \tag{78}$$

Adding (77) and (78) and rearranging terms.

$$p = \frac{1 - s_r - s_{r'}}{\frac{MT}{n_r} - 1} \tag{79}$$

Equation (79) provides a lower bound on p since it is derived assuming the largest possible search rate for members of race r by the biased officers. Clearly the bound is not sharp since the assumption that prejudiced officers search at rate $\frac{T}{n_r}$ may not be consistent with the data. One easy way to sharpen the bound is to replace $\frac{T}{n_r}$ with the highest search rate in the sample of officers. We leave more sophisticated ways of sharpening the bound, which rely on the empirical distribution of search rates to future work.

Section 6. Summary and conclusions

In this paper, we explore how various facets of the stop and guilt data can shed light on taste-based discrimination in racial profiling. Working with parametric models that nest some of those most prominent in the literature, we argue that mapping claims about the absence of taste-based discrimination to relationships between searches and guilt is sensitive to functional form assumptions and to assumptions about the nature of unobserved group level characteristics.

Importantly, nothing we have argued suggests that KPT's, Persico (2009)'s or AF's methodological analyses are incorrect, nor is our discussion intended to diminish the value of their studies. Rather, our analysis clarifies how apparently innocuous assumptions in these models can lead to vastly different claims about the presence or absence of taste-based discrimination. In particular, we highlight how the general presence of unobservables affects the robustness of their tests. Heckman (2000, 2005) emphasizes that properties such as identification are characteristics of models and therefore are always reliant on assumptions. We demonstrate how this is true in the racial profiling context.

Overall, our results may appear to be discouraging for the potential to detect discrimination in data on police stops and searches. One way forward is to start with the KPT and Persico (2009) assumptions and generate a more general model that relaxes some of these assumptions, while still maintaining testable implications on the presence or absence of taste-based discrimination. Our paper provides some examples of how this may be done.

A systematic development of models of that span the plausible assumptions that distinguish different models of police stop and search process would produce a model space. A researcher may then evaluate evidence of taste-based discrimination given the model space, as opposed to the standard procedure, which amounts to evaluating evidence based on a single model. Model averaging methods, for example, provide a way to evaluate the evidence from these models. [22] As our results indicate, this will probably require that each of the alternative models place strong assumptions on the nature and effects of the unobserved heterogeneity that distinguishes groups observed by an econometrician versus an officer. But the model averaging approach provides a strategy for developing robust evidence of discrimination or its absence, i.e. evidence that is not contingent on particular sets of assumptions.

A different route is to reformulate tests of taste-based discrimination in a partial identification context; we give an example in Section 5.iii on how this may be done, but the general idea has yet to be systematically explored. A partial identification approach will allow for weaker assumptions on the model, particularly on unobservable heterogeneity. Brock and Durlauf (2007) show this is possible for social interactions. Theoretical work on partial identification in abstract profiling environments has been done by Brock (2006) and Manski (2006a), which is an appropriate conclusion to this paper.

---

[22]See Brock, Durlauf and West (2003) for a conceptual defense of model averaging and Cohen-Cole, Durlauf, Fagan, and Nagin (2009) for an example in criminology.

Bibliography

Anwar, S. and H. Fang, (2006), “An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence, *American Economic Review*, 96, 1, 127-151

Antonovics, K. and B. Knight, (2009), “A New Look at Racial Profiling: Evidence from the Boston Police Department, *Review of Economics and Statistics*, 91, 1, 163-177.

Ayres, I., (2005), “Three Tests for Measuring Unjustified Disparate Impacts in Organ Transplantation: The Problem of “Included Variable” Bias, *Perspectives in Biology and Medicine*, 48, 1, S68-S87.

Banerjee, A. and S. Mullainathan, (2008), “Limited Attention and Income Distribution,” *American Economic Review*, 98, 2, 489-493.

Becker, G. (1957), *The Economics of Discrimination*, Chicago; University of Chicago Press.

Bjerk, D., (2007), “Racial Profiling, Statistical Discrimination, and the Effect of a Colorblind Policy on the Crime Rate,” *Journal of Public Economic Theory*, 9. 3, 521-545, 06.

Bohnholz, R. and J. Heckman, (2005), “Measuring the Disparate Impacts and Extending Disparate Impact Doctrine to Organ Transplantation,” *Perspectives in Biology and Medicine,* 48, 1, S95-S122.

Brock, W., (2006) “Profiling Problems With Partially Identified Structure,” *Economic Journal*, 116, 515, F427-F440.

Brock. W. and S. Durlauf, (2007), “Identification of Binary Choice Models with Social Interactions,” *Journal of Econometrics*, 140, 1, 52-75.

Brock, W., S. Durlauf, and K. West (2003), “Policy Evaluation in Uncertain Economic Environments (with discussion),” *Brookings Papers on Economic Activity*, 1, 235-322.

Charles, K. and J. Guryan, (2008), “Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination" *Journal of Political Economy*, 116, 5, 773-809.

Cohen-Cole, E., S. Durlauf, J. Fagan, and D. Nagin, (2009), “Model Uncertainty and the Deterrent Effect of Capital Punishment,” *American Law and Economics Review*, forthcoming.

Dharmapala, D. and S. Ross, (2004) "Racial Bias in Motor Vehicle Searches: Additional Theory and Evidence,” *Contributions to Economic Analysis & Policy*: 3, 1, article 12.

Dominitz, J. and J. Knowles, (2006), “Crime Minimisation and Racial Bias: What Can Be Learned from Police Search Data,” *Economic Journal*, 116, F368-F384.

Durlauf, S., (2006), “Assessing Racial Profiling,” *Economic Journal*, 116, 515, F402-F426.

Grogger, J. and G. Ridgeway, (2006), “Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness,” *Journal of the American Statistical Association*, 101, 475, 878-887.

Harcourt, B., (2004), “Rethinking Racial Profiling: A Critique of the Economics, Civil Liberties, and Consitutional Literature, and of Criminal Profiling More Generally,” *University of Chicago Law Review*, 71, 4, 1275-1381.

Harcourt, G., (2009), “Henry Louis Gates and Racial Profiling: What’s the Problem?,” *University of Chicago John M. Olin Law & Economics Working Paper no.* 482.

Heckman, J. (1998), “Detecting Discrimination,” *Journal of Economic Perspectives* 12, 2, 101-116.

Heckman, J., (2000), “Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective,” *Quarterly Journal of Economics*, 115, 45-97.

Heckman, J., (2005), “The Scientific Model of Causality,” *Sociological Methodology*, 35, 1-98.

Heckman, J. and P. Siegelman, (1993), ‘The Urban Institute Audit Studies: Their Methods and Findings’ in *Clear and Convincing Evidence*, M. Fix and R. Struyk, eds. Washington D.C.: Urban Institute Press.

Knowles, J., N. Persico, and P. Todd, (2001), “Racial Bias in Motor Vehicle Searches: Theory and Evidence,” *Journal of Political Economy*, 109, 1, 203-229.

Kahneman, D., (1973), *Attention and Effort*, Englewood Cliffs: Prentice Hall.

Levin, B. and H. Robbins, (1983), “Urn Models for Regression Analysis, with Applications to Employment Discrimination Studies,” *Law and Contemporary Problems*, 46, 247-267.

MaCrae, C. N., A. Miller, and G. Bodenhausen, (1994), “Stereotypes as Energy-Saving Devices: A Peek Inside the Cognitive Toolbox,” *Journal of Personality and Social Psychology*, 66, 1, 37-47.

Manski, C., (1993), “Identification of Endogenous Social Effects: The Reflection Problem.” *Review of Economic Studies*, 60, 3, 531-542.

Manski, C., (2006a), “Interpreting the Predictions of Prediction Markets,” *Economics Letters*, 91, 3, 425-429

Manski, C., (2006b), “Search Profiling With Partial Knowledge of Deterrence,” *Economic Journal*, 116, 515, F385-F401.

Persico, N., (2002), “Racial Profiling, Fairness, and the Effectiveness of Policing,” *American Economic Review*, 92, 5, 1472-1497.

Persico, N., (2009), “Racial Profiling? Detecting Bias using Statistical Evidence,” *Annual Reviews in Economics*, 1, 229-254.

Ridgeway, G., (2007), “Analysis of Racial Disparities in the New York Police Department’s Stop, Question, and Frisk Practices,” RAND Corporation.

Risse, M. and R. Zeckhauser, (2004), “Racial Profiling,”*Philosophy and Public Affairs*, 32, 2, 131-170.

Sanga, S., (2009), “Reconsidering Racial Bias in Motor Vehicle Searches: Theory and Evidence,” *Journal of Political Economy*, 117, 6, 1155-1159.

Sklansky, D., (2008), *Democracy and the Police*, Stanford: Stanford University Press.

Warm, J., R. Parasuraman, and G. Matthews, (2008), “Vigiliance Requires Hard Work and is Stressful,” *Human Factors*, 50, 3, 433-441.