On the observational implications of taste-based discrimination in racial profiling

William A. Brock, Jane Cooley, Steven N. Durlauf *, Salvador Navarro
Department of Economics, University of Wisconsin at Madison, United States

ARTICLE INFO

Article history:
Available online 25 June 2011

JEL classification:
C5
K4

Keywords:
Racial profiling
Discrimination
Identification

ABSTRACT

This paper contributes to a growing literature that attempts to determine whether disparities in police stops and searches of potential criminals of different races stem from taste-based discrimination. The key challenge in making this evaluation is that police officers have more information than the econometrician and thus racial disparities in police behavior may result from these unobservable factors rather than discrimination. We develop a general equilibrium model of police and potential criminal behavior that encompasses key models in the literature. We highlight the assumptions needed for existing methods of detecting racial discrimination to hold. In particular, we show that when there are increasing costs to search, existing tests for discrimination can give incorrect results. Given the potential importance of these costs, we then propose some alternate methods for detecting racial bias in police behavior.

© 2011 Elsevier B.V. All rights reserved.

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% of stops of the New York police department in 2006 were nonwhite, and 51% of these were black. These differentials are often referred to as racial profiling. Racial profiling continues to be an active area of research 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 et al. (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.

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

¹ This latter differential may be attributed to statistical discrimination, which is explicitly not our focus.
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 equals 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 also illustrate a set of assumptions through which police officer heterogeneity can be used to identify relative bias in our model.2

The profiling 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 group 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 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 are 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 which we use 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 Section 3 develops a generalization of Persico’s (2009) version of the KPT model, which we use to study how taste-based discrimination affects equilibrium search success rates. We also discuss AF’s (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 lead to both point and partial identification results. Section 6 concludes.

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.

3 We do not claim a general observational equivalence result nor do we argue that field experiments would not be able to reveal taste-based discrimination in racial profiling.
contraband) and a population of police officers each of whom chooses a strategy for stopping individuals who will be searched.

Each member \( j = 1, \ldots, N \) of the civilian population is a member of some group. The characteristics of a group are described as a triple, \( (r, c_o, c_r) \), 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 \( c_r \) 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 \( n_{r,c_o,c_r} \) 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 \( \epsilon_j \), 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 a group attribute. A primary identification challenge is the presence of the characteristics that the police officer observes that the researcher does not (\( c_o \) as opposed to \( \epsilon_j \)).

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

\[
EU_c(r, c_o, c_r, \tau_{c_o,c_r}, \epsilon_j) = \pi_{r,c_o,c_r} U_{c,S}(r, c_o, c_r, \epsilon_j) + (1 - \pi_{r,c_o,c_r}) U_{c,NS}(r, c_o, c_r, \epsilon_j),
\]

where \( \pi_{r,c_o,c_r} \) is the probability of being searched if one is a member of the group. As noted above, \( \epsilon_j \) 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.

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

\[
\pi_{r,c_o,c_r} = \Pr(EU_c(r, c_o, c_r, \tau_{c_o,c_r}, \epsilon_j) > 0),
\]

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 \( i = 1, \ldots, M \) chooses group-specific search rates \( s_{r,t,c_o,c_r} \). We are not explicit here about the constraints that search rates must satisfy. In most of our analysis, search rates will 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 \( s_{t,r,c_o,c_r} \) to a group characterized by \( (r, c_o, c_r) \) is

\[
\pi_{r,c_o,c_r} V_{C}(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{t,r,c_o,c_r}) + (1 - \pi_{r,c_o,c_r}) V_{NC}(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{t,r,c_o,c_r}).
\]

Here \( V_{C}(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{t,r,c_o,c_r}) \) denotes the utility of searching a guilty member of the group and \( V_{NC}(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{t,r,c_o,c_r}) \) denotes the utility of searching an innocent member. We include the individual search rate \( s_{t,r,c_o,c_r} \) 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 \( a_{t,r,c_o,c_r} \) denote the number of members of a group searched by officer \( i \), i.e. \( a_{t,r,c_o,c_r} = s_{t,r,c_o,c_r} n_{r,c_o,c_r} \). Aggregating across police determines the probability of being searched via

\[
s_{t,r,c_o,c_r} = \int s_{r,t,c_o,c_r} d_i = \frac{\int a_{t,r,c_o,c_r} d_i}{n_{r,c_o,c_r}}.
\]

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

\[
\pi_{t,r,c_o,c_r} = \pi \left( r, c_o, c_r, s_{t,r,c_o,c_r} \right).
\]

Once we are explicit about the restrictions on search rates that an officer must fulfill, the description of the crime choices of civilians and the search choices of police officers is complete.

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.

2.2. 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

\[
V_{C}(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{t',r,c_o,c_r}) = V_{C}(p_{o,i}, p_{u,i}, r', c_o, c_r, s_{t',r,c_o,c_r}) \quad \text{if} \quad s_{t,r,c_o,c_r} = s_{t',r,c_o,c_r}
\]

and

\[
V_{NC}(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{t',r,c_o,c_r}) = V_{NC}(p_{o,i}, p_{u,i}, r', c_o, c_r, s_{t',r,c_o,c_r}) \quad \text{if} \quad s_{t,r,c_o,c_r} = s_{t',r,c_o,c_r}.
\]

Notice that we do not require that \( \pi_{t,r,c_o,c_r} = \pi_{t',r,c_o,c_r} \). Officers may prefer to search one race or another because of differences in guilt probabilities; these appear in the expected utility, Eq. (3), for 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 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 requires the analyst to know both the distribution of characteristics of the individuals who are actually searched and 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 \((c_o)\), 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 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.

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 problem general, it is still based on parametric assumptions on the officer objective functions. 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 and searches and their outcomes. A heuristic justification for this view is the following. Suppose that \(V_u(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{r,c_o,c_r}) = 0\) and that \(V_c(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{r,c_o,c_r})\) 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 \(p_{o,i}\):

\[
\bar{s}_{p_{o,i},r,c_o} = \int \bar{s}_{p_{o,i},p_{u,i},r,c_o,c_r} dF_{p_{o,i},c_o,p_{u,i},r,c_r}. 
\]

(8)

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

\[
g_{r,c_o} = \pi \left( r, c_o, c_r, s_{r,c_o,c_r} \right) \int \bar{s}_{p_{o,i},p_{u,i},r,c_o,c_r} dF_{p_{o,i},c_o,p_{u,i},r,c_r}.
\]

(9)

Since one is free to choose any distribution functions \(F_{p_{o,i},c_o,p_{u,i},r,c_r}\) and \(F_{u,c_r}\) and any utility function that does not depend on race, \(V_c(p_{o,i}, p_{u,i}, c_o, c_r, s_{r,c_o,c_r})\), 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 \(V_c(p_{o,i}, p_{u,i}, c_o, c_r, s_{r,c_o,c_r})\) and \(V_c(p_{o,i'}, p_{u,i'}, c_o, c_r, s_{r,c_o,c_r})\) together with associated distribution functions \(F_{p_{o,i},c_o,p_{u,i},r,c_r}\) and \(F_{p_{o,i'},c_o,p_{u,i'},r,c_r}\) to produce any ranking across the two officers. The problem is that solely restricting \(V_c(p_{o,i}, p_{u,i}, c_o, c_r, s_{r,c_o,c_r})\) 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.

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 for observational implications of taste-based discrimination, we extend KPT using a device of Persico (2009). 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 implicit informational assumptions in their analyses. Relative to our abstract formulation, we assume that

\[
V_u(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{r,c_o,c_r}) = \left( \beta_{r,c_o,c_r} + b_{r,c_o,c_r} s_{r,c_o,c_r} \right) s_{r,c_o,c_r}^{-1}
\]

(10)

and

\[
V_c(p_{o,i}, p_{u,i}, r, c_o, c_r, s_{r,c_o,c_r}) = b_{r,c_o,c_r} s_{r,c_o,c_r}^{-1}.
\]

(11)
From this formulation, one can interpret $\beta_{i\tau,c_i,c_o}$ as the (extra) utility of searching a guilty person in distinction to $b_{i\tau,c_i,c_o}$, the utility of the act of searching itself. Since the total number of searches of members of a group by officer $i$ is $n_{i,c_i,c_o}\beta_{i\tau,c_i,c_o}$, the overall expected utility from a given allocation of searches across groups by officer $i$ is

$$\int (\beta_{i\tau,c_i,c_o} n_{i,c_i,c_o} + b_{i\tau,c_i,c_o}) dc_o dc_r dr.$$ (12)

For our specification, the term $\beta_{i\tau,c_i,c_o}$ 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. Additionally, we introduce the parameter $b_{i\tau,c_i,c_o}$ in order to allow for payoffs to depend on the act of stopping a civilian regardless of his guilt. This factor is arguably the most 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 $\alpha = 1$ with $\beta_{i\tau,c_i,c_o} = \beta_i$ and $b_{i\tau,c_i,c_o} = 0$; it also embeds the KPT preferences, with $\alpha = 1$, $\beta_{i\tau,c_i,c_o} = \beta$ and $b_{i\tau,c_i,c_o} = b_i$. We are also interested in cases where $\alpha < 1$. 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 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 determinant. 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.

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. 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. Sanga (2009) provides some evidence that the litigation affects police behavior. He replicates the KPT analysis, extending it to 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 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 is 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. A recent example of psychological research of this type is (Warm et al., 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 $s_{i\tau,c_i,c_o}$ is that it implies that a police officer would experience lower stop costs if he were to focus on cruder group designations than $(r, c_o, c_i)$. This is also consistent with existing psychological evidence and in fact provides a basis for understanding why individuals stereotype, see Macrae et al. (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 $\alpha$ is more defensible than another, but rather to argue that there do not exist strong a priori reasons to privilege $\alpha = 1$ over $\alpha < 1$. As argued earlier, 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 $\alpha$, 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, Eqs. (6) and (7), to the preferences embedded in Eq. (12), the absence of taste-based discrimination on the part of officer $i$ requires that

$$\beta_{i\tau,c_i,c_o} = \beta_{i\tau',c_i,c_o} = \beta_{c_i,c_o},$$

and

$$b_{i\tau,c_i,c_o} = b_{i\tau',c_i,c_o} = b_{c_i,c_o}.$$ (13)

We thus distinguish between discrimination with respect to the utility of searching a guilty driver versus an innocent driver.  

---

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.

9 To be precise, KPT does not constrain the number of officer stops to meet some fixed time, so $\alpha$ 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.
We now describe optimal police behavior. Officers are assumed to have a fixed number of searches $T$ which they allocate across groups,\(^{12}\)

\[ \int n_{r,c_o,c_u} \pi_{r,c_o,c_u} \delta c_o dc_u dr \leq T. \]  

(14)

The Lagrangian for the officer's problem is therefore

\[ \int \left( (\beta_{r,c_o,c_u} \pi_{r,c_o,c_u} + b_{r,c_o,c_u}) n_{r,c_o,c_u} s_{r,c_o,c_u} \right. \]

\[ \left. - \mu_B r n_{r,c_o,c_u} \delta c_o dc_u dr + \mu_B r T \right) \]  

(15)

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

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

\[ s_{r,c_o,c_u} = \frac{\left( \beta_{r,c_o,c_u} \pi_{r,c_o,c_u} + b_{r,c_o,c_u} \right) \int_{T} n_{r,c_o,c_u} \delta c_o dc_u dr + \mu_B r T}{\int_{T} (\beta_{r,c_o,c_u} \pi_{r,c_o,c_u} + b_{r,c_o,c_u}) n_{r,c_o,c_u} \delta c_o dc_u dr}. \]

(16)

Aggregating across police gives group-specific search rates,\(^{13}\) see Eq. (17) given in Box I on the next page, where $dF_{\beta_{r,c_o,c_u}}$ denotes the distribution of taste parameters across police 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 Eq. (18) given in Box II on the next page. Given the nonlinearity in the individual equilibrium stop choices, it is evident from Eq. (18) that even if the median or mean officer is unbiased, this does not imply that the presence of bias fails to affect aggregate group-specific search rates.

Let $g = 1, \ldots, G$ index the groups that can be formed from all the combinations of $(r, c_o, c_u)$ in the population. Assume the total number of groups, $G$, is finite. Eq. (18) defines a mapping of the space

\[ \Omega = \{ s_1 \in [0, MT/n_1] \times (s_2 \in [0, MT/n_2]) \times \cdots (s_G \in [0, MT/n_G]) \}. \]

(19)

into itself. A fixed point set of search rates in (18) constitutes a Nash equilibrium. The following proposition establishes existence under mild conditions. We therefore state.

**Proposition 1 (Existence of Nash equilibrium).** Assume $\pi (r, c_o, c_u, s_{r,c_o,c_u})$ is continuous in $s_{r,c_o,c_u} \in [0, MT/n_{c_o,c_u}]$. Then a Nash equilibrium exists in aggregate officer stop choices across groups.

**Proof.** The claim is immediate from Brower’s fixed point theorem since $G$ has been assumed to be finite. \(\Box\)

3.1. 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 Eq. (16) that, as one takes the limit $\alpha \rightarrow 1$ from below, officer's available searches $T$ become entirely concentrated on a single group, determined by the maximum value of $\beta_{r,c_o,c_u} \pi_{r,c_o,c_u} + b_{r,c_o,c_u}$ across groups. 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. 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. The first order condition from the officers’ problem in (15) implies

\[ (\beta_{r,c_o,c_u} \pi_{r,c_o,c_u} + b_{r,c_o,c_u}) s_{r,c_o,c_u}^{-1} = (\beta_{r',c_o,c_u} \pi_{r',c_o,c_u} + b_{r',c_o,c_u}) s_{r',c_o,c_u}^{-1}. \]

(20)

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

\[ \beta_{r,c_o,c_u} = \beta_{r',c_o,c_u} \text{ and } b_{r,c_o,c_u} = b_{r',c_o,c_u}. \]

(21)

We introduce this level of parameter homogeneity as it implies that in equilibrium, each officer chooses an identical cross group distribution of search rates.\(^{14}\) The average and individual group search rates thus coincide for each group. We assume homogeneous preferences for most of this section. We return to officer heterogeneity in Section 3.4.

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

\[ (\beta_{c_o,c_u} \pi (r, c_o, c_u, s_{r,c_o,c_u}) + b_{c_o,c_u}) s_{r,c_o,c_u}^{-1} = (\beta_{c_o,c_u} \pi (r', c_o, c_u, s_{r',c_o,c_u}) + b_{c_o,c_u}) s_{r',c_o,c_u}^{-1}. \]

(22)

Eq. (22) 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

\[ \pi (r, c_o, c_u, s) = \pi (r', c_o, c_u, s) . \]

(23)

This condition means that given the group characteristics $(c_o, c_u)$ 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.

Combining (22) and (23), 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

\[ \pi_{r,c_o,c_u} = \pi_{r',c_o,c_u} \text{ and } s_{r,c_o,c_u} = s_{r',c_o,c_u}. \]

(24)

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 \((r, c_o, c_u)\) 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 \(r\), \(n_r \int \pi (r, c_o, c_u, s, r, c_o)\) \(S_{r, c_o, u, d|c_o|/r}\) to the total number of searched people of race \(r\), \(n_r \int S_{r, c_o, u, d|c_o|/r}\). That is, the hit rate for race \(r\), \(h_r\), is defined as

\[
h_r = \frac{\int \pi (r, c_o, c_u, s, r, c_o) S_{r, c_o, u, d|c_o|/r}}{\int S_{r, c_o, u, d|c_o|/r}}. \tag{25}
\]

Note that \(dF_{c_o, u, d|c_o|/r}\), 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.

Given (18), our functional forms restrict the race level hit rate to fulfill Eq. (26) given in Box III on the next page. It is evident that, unless \(dF_{c_o, u, d|c_o|/r} = dF_{c_o, u, d|c_o|/r'}\), it is possible that

\[
h_r \neq h_{r'}. \tag{27}
\]

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

\[
h_r \neq h_{r'}. \tag{28}
\]

unless \(dF_{c_o, u, d|c_o|/r} = dF_{c_o, u, d|c_o|/r'}\).

This illustrates an essential restriction in the KPT framework relative to our framework. 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. In equilibrium, \(\pi (r, c_o, c_u, s, r, c_o)\) is constant and factors out of (25). Our specification does not have this feature. Indeed it is clear from (25) that equal hit rates across race or across observable characteristics generally do not hold without this feature, so the link between (26) and (28) 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).

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), Bohnholz and Heckman (2005) have pioneered this recognition by showing how differences in higher moments in \(dF_{c_o, u, d|c_o|/r}\) and \(dF_{c_o, u, d|c_o|/r'}\) 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 (24) 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 taste is informative about guilt probabilities at equal policing levels, i.e.

\[
\pi (r, c_o, c_u, s) > \pi (r', c_o, c_u, s). \tag{29}
\]

In our setting, unless \(\alpha = 1\), equal guilt probabilities imply taste-based discrimination (i.e. (21) does not hold). This is immediate from the equilibrium condition (20). Furthermore in the absence of taste-based discrimination, the equilibrium search rates implied by (20) require

\[
\pi_{r, c_o, c_u} > \pi_{r', c_o, c_u} \quad \text{and} \quad S_{r, c_o, u} > S_{r', c_o, u}. \tag{30}
\]

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. Equal guilt probabilities would necessarily imply that \(h_r > h_{r'}\) 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 from allowing officer utility to directly depend on the distribution of searches across groups. This is a plausible difference and would arise, for example, from returns to scale in the search process. It follows that hit rates for coarser population subdivisions are not necessarily equal even if taste-based discrimination is absent.

One response to our argumentation is that KPT’s 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

---

15 The same argument applies if one conditions on equal stop rates as well as \((r, c_o)\).
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.\(^\text{16}\)

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.

**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 suggests that KPT’s or Persico’s (2009) methodological analyses are incorrect, but rather we illustrate 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 functional form for group-specific guilt probabilities, one where race plays no direct intrinsic role in guilt

\[
\pi(r, c_o, c_u, s_{r,c_o,c_u}) = \pi(c_o, c_u) s_{r,c_o,c_u}^{-\gamma}.
\]

Second, we assume that utility only derives from stops of the guilty,

\[
b_{i(r,c_o,c_u)} = 0, \quad \forall i, r, c_o, c_u.
\]

These assumptions allow for a closed form solution to group level search rates,

\[
s_{r,c_o,c_u} = \frac{\left(\beta_o c_o \pi(c_o, c_u)\right)^{1-\gamma} MT}{\int \left(\beta_c c_o \pi(c_o, c_u)\right)^{1-\gamma} n_{c_o} c_o d c_o d c_o}.
\]

We may now draw several implications. Take two groups \((r, c_o, c_u)\) and \((r', c_o, c_u)\). From (33), 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

\[
h_{r,c_o} = \frac{\int \pi(r, c_o, c_u, s_{r,c_o,c_u}) s_{r,c_o,c_u} dF_{c_u|c_o} c_o}{\int s_{r,c_o,c_u} dF_{c_u|c_o} c_o} (34)
\]

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 \(h_{r,c_o}\), in Eq. (34), has the closed form solution

\[
\int \int \pi(c_o, c_u) s_{r,c_o,c_u}^{-\gamma} dF_{c_u|c_o} d c_u = \int \pi(c_o, c_u) \times \left(\frac{\left(\beta_o c_o \pi(c_o, c_u)\right)^{1-\gamma} MT}{\int \left(\beta_c c_o \pi(c_o, c_u)\right)^{1-\gamma} n_{c_o} c_o d c_o d c_o}\right)^{1-\gamma} dF_{c_u|c_o} c_o. (35)
\]

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 \(dF_{c_u|c_o}\) renders (35) race dependent.

Similar arguments apply to \(\int s_{r,c_o,c_u} dF_{c_u|c_o} c_u\), the denominator of (34), as it obeys

\[
\int s_{r,c_o,c_u} dF_{c_u|c_o} c_u = \int \frac{\left(\beta_o c_o \pi(c_o, c_u)\right)^{1-\gamma} MT}{\int \left(\beta_c c_o \pi(c_o, c_u)\right)^{1-\gamma} n_{c_o} c_o d c_o d c_o} dF_{c_u|c_o} c_o. (36)
\]

The presence of \(\pi(c_o, c_u)\) and integration against \(dF_{c_u|c_o}\) in the RHS of this expression also can render (36) race dependent.

3.2. 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. Dominitz and Knowles (2006) 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

$$\min_{s_{r,c_0},c_0} \int \left( (\pi (r, c_0, c_u, s_{r,c_0,c_0}) - \mu s_{r,c_0,c_0} n_{c_0,c_0} dc_0 dc_u dr) + \mu s_{r,c_0,c_0} \right) \text{MT},$$

where the subscript on $\mu$, 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. (31) is modified to

$$\pi (r, c_0, c_u, s_{r,c_0,c_0}) = \pi (r, c_0, c_u) s_{r,c_0,c_0}^{\gamma}$$

for race to have intrinsic predictive value in group-specific guilt. In this case the guilt rates across groups are related to search rates by

$$\frac{\pi (r, c_0, c_u)}{\pi (r', c_0', c_u')} s_{r,c_0,c_0}^{\gamma} = \frac{s_{r',c_0',c_0'}}{s_{r',c_0',c_0'}}.$$  \hspace{1cm} (38)

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 $\beta_{r,c_0,c_0} \equiv \beta, \beta_{r',c_0,c_0} \equiv 0$. The ratio of equilibrium guilt rates in the noncooperative model is

$$\frac{\pi (r, c_0, c_u)}{\pi (r', c_0', c_u')} s_{r,c_0,c_0}^{\gamma} = \frac{s_{r',c_0',c_0'}}{s_{r',c_0',c_0'}}.$$  \hspace{1cm} (39)

As $\alpha \to 0$, the ratios of guilt probabilities of the social planner and noncooperative equilibrium coincide as the search rates coincide. Recall that $\alpha$ 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 $\alpha$ 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 (38). But this also suggests that distinguishing between the types of equilibria is difficult when (38) is a good approximation of the guilt rate process and when marginal increases in search rates are extremely costly to an officer.

3.3. 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 (32) and that police officers are homogeneous so that $\beta_{r,c_0,c_0} = \beta_{r',c_0,c_0}$. We also assume that

$$\pi (r, c_0, c_u, s_{r,c_0,c_0}) = \pi (r, c_0, c_u) s_{r,c_0,c_0}^{\gamma}.$$  \hspace{1cm} (41)

holds. In analogy to (33), the unique group-specific search rate is

$$s_{r,c_0,c_0} = \frac{\left( \beta_{r,c_0,c_0} \pi (r, c_0, c_u) \right)^{\frac{1}{1+\gamma r}} \text{MT}}{\int \left( \beta_{r,c_0,c_0} \pi (r', c_0', c_u') \right)^{\frac{1}{1+\gamma r'}} n_{r',c_0',c_0'} dc_0' dc_u' dr'}. \hspace{1cm} (42)$$

From (42), the equilibrium ratio of policing on group $(r, c_0)$ to group $(r', c_0')$ is

$$s_{r,c_0} = \frac{\int \left( \beta_{r,c_0,c_0} \pi (r, c_0, c_u) \right)^{\frac{1}{1+\gamma r}} dF_{c_0|r,c_0}}{\int \left( \beta_{r,c_0,c_0} \pi (r', c_0', c_u) \right)^{\frac{1}{1+\gamma r'}} dF_{c_0'|r',c_0'}} \hspace{1cm} (43)$$

when taste-based discrimination is absent, which leads to

**Proposition 5** (Lack of Information about Taste-Based Discrimination from Differences in Search Rates Applied to Different Races). Let $\gamma_{r,c_0,c_0}$ be any observed ratio of policing on race $r$ and race $r'$ where the observed characteristics $c_0$ are the same. Then, there exist values of $\beta_{r,c_0,c_0}, dF_{c_0|r,c_0}, dF_{c_0'|r',c_0'}, \pi (r, c_0, c_u)$, and $\pi (r', c_0, c_u')$ such that (43) holds.

As an extreme example suppose $\beta_{r,c_0,c_0} \pi (r, c_0, c_u)$ and $\beta_{r,c_0,c_0} \pi (r', c_0, c_u')$ 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 (42) that is germane to our assumptions only place positive mass on these two values. For this case, the ratio of search rates obeys $s_{r,c_0} = \frac{Pr_{r,c_0}}{Pr_{r',c_0}}$ where the $Pr_{r,\cdot}$ 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 (43) when we are free to choose $\beta_{r,c_0,c_0}, dF_{c_0|r,c_0}, dF_{c_0'|r',c_0'}, \pi (r, c_0, c_u)$, and $\pi (r', c_0, c_u')$ such that (42) 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.

3.4. 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. 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.

Parallel to our Proposition 2, AF concludes that, given these unobservables, equal hit rates across observed groups is no longer an implication of the absence of taste-based discrimination. Our setting, however, does not restrict the role of the unobservables

\[ \int \left( \beta_{r,c_0,c_0} \pi (r, c_0, c_u) \right)^{\frac{1}{1+\gamma r}} \text{MT} \]

\[ \int \left( \beta_{r,c_0,c_0} \pi (r', c_0', c_u') \right)^{\frac{1}{1+\gamma r'}} n_{r',c_0',c_0'} dc_0' dc_u' dr' \]

\[ \frac{\int \left( \beta_{r,c_0,c_0} \pi (r, c_0, c_u) \right)^{\frac{1}{1+\gamma r}} dF_{c_0|r,c_0}}{\int \left( \beta_{r,c_0,c_0} \pi (r', c_0', c_u) \right)^{\frac{1}{1+\gamma r'}} dF_{c_0'|r',c_0'}} \]
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 KFT’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, 145 and 146, Eqs. (6) and (7)) in our notation is

\[ s_{i,r,c_1} > s_{i,r,c_2} \quad \text{and} \quad h_{i,r,c_1} < h_{i,r,c_2} \Rightarrow s_{i,r',c_1} > s_{i,r',c_2} \]

and

\[ h_{i,r',c_1} < h_{i,r',c_2}. \]  (44)

In words, if officer \( i \) searches a potential criminal of race \( r \) at a lower rate with a higher 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 \((r, \pi_c, \pi_d)\). Similar to AF, we restrict the role of the unobservables as follows. First, we assume that \( \pi_c \) is a scalar random variable. We further restrict its role so that the main implication of AF’s monotone likelihood assumption holds, namely that \( \pi_{r,c_1,c_2} \) is monotonically increasing in \( \pi_c \). Second, we mimic AF’s officers’ preferences by imposing \( \beta_{i,r,c_1,c_2} = 1 \) and \( b_{i,r,c_1,c_2} = b_i \), 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

\[ \frac{s_{i,r,c_1}}{s_{i,r,c_2}} = \int \left( \frac{\pi_{r,c_1}}{\pi_{r,c_2}} + b_i \right) \frac{1}{\pi_{r,c_2}} \, dc_{i,r,c} > 1, \]  (45)

then it does not necessarily follow that

\[ \frac{s_{i',r,c_1}}{s_{i',r,c_2}} = \int \left( \frac{\pi_{r',c_1}}{\pi_{r',c_2}} + b_{i'} \right) \frac{1}{\pi_{r',c_2}} \, dc_{i',r,c} > 1, \]  (46)

even if \( \frac{\pi_{r,c_1}}{\pi_{r,c_2}} < 1 \) and \( \pi_{r,c_1,c_2} = \pi_{0,c_2} \). Since the distribution of the unobservables depends on race, condition (45) does not determine whether (46) holds.

Thus, AF’s rank condition (44) 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 in our model than in AF. Starting with a situation where (45) and (46) hold, one can change the distribution of unobservables, while maintaining \( \pi_{r,c_1,c_2} \) is monotonically increasing in \( \pi_c \) and violate the rank condition.

AF’s condition only holds when the distributions do not depend on race \( \left( \frac{dc_{i,r,c}}{dc_{i',r,c}} = \frac{dc_{i,r,c}}{dc_{i',r,c}} \right) \) and the guilt probabilities are independent of race \( \left( \pi_{r,c_1,c_2} = \pi_{0,c_2} \right) \), since nothing depends on race anymore.

4. Uncovering biased officers

In this section we consider the problem of a police chief who follows two commonly suggested policies in order to eliminate biased officers.\(^{18}\) In the first case, he makes 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. We want to understand whether biased officers survive if either policy is implemented. We assume that the police chief has access to the same information as an officer regarding the criminals, i.e. to \((r, \pi_c, \pi_d)\), so that the distinction between \( c_0 \) and \( c_i \) is not relevant and let \( c = (c_0, c_i) \).

4.1. Equal search rates

Suppose that a police chief implements a policy under which 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 \( b_i = 0 \). From (16) it follows that the ratio of search rates for two groups differing only by race is

\[ \frac{s_{i,r,c}}{s_{i',r,c}} = \frac{\beta_{i,r,c} \pi_{r,c}}{\beta_{i',r,c} \pi_{r,c}}. \]  (47)

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.

However, 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 such that they are equal across groups that differ only by race, i.e.,

\[ s_{i,r,c} = s_{i',r,c} = s_{i,c}. \]  (48)

The officer chooses search rates \( s_{i,r,c} \) to maximize his utility \( \int \beta_{i,r,c} \pi_{r,c} n_{r,c} \, dc_r \) subject to both (48) and the time constraint \( \int n_{r,c} s_{i,r,c} \, dc_r \leq T \).

Replacing (48) into the time constraint we can rewrite it as

\[ \int s_{i,c} \int n_{r,c} \, dc_r \leq T \quad \text{and} \quad \int n_{r,c} s_{i,c} \, dc_r \leq T. \]  (49)

If we also replace (48) into the utility function, we have the modified objective function

\[ \int \left( \frac{\alpha}{\pi_c} \int \beta_{i,r,c} \pi_{r,c} n_{r,c} \, dc_r \right) \, dc. \]  (50)

Eqs. (49) and (50) produce the Lagrangian for the officer’s problem

\[ \int \left( \frac{\alpha}{\pi_c} \int \beta_{i,r,c} \pi_{r,c} n_{r,c} \, dc_r - \mu \beta_{i,r,c} \pi_{r,c} \pi_{r,d} \right) \, dc + \mu \pi_{r,d} T, \]  (51)

where the subscript on \( \mu \) is the multiplier for the officer’s time constraint under (48), refers to this being the equal search rates problem. The solution to (51) produces an optimal search rate given by

\[ s_{i,c} \equiv s_{i,c} = \left( \frac{\int \beta_{i,r,c} \pi_{r,c} n_{r,c} \, dc_r}{\int n_{r,c} s_{i,c} \, dc_r} \right)^{\frac{1}{\alpha}}. \]  (52)

By construction any officer, whether he is racially biased or not, can survive this policy.

\(^{18}\) We do not need to take a stance on the source of unobservables.

\(^{18}\) In this section, we do not explicitly prove existence of an equilibrium for the constrained problems we present. Instead, we simply ask whether, if an equilibrium exists, biased officers can adjust their behavior and still act on their bias.
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 $\beta_{r,c} = \beta_{r}$ 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 Eq. (52) 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.

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 refer to this issue in Section 5.

4.2. 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

$$\pi_{r,c} = \pi_c \quad \forall r.$$  \hspace{1cm} (53)

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

Suppose now that officers respond to the introduction of the policy. In this case, the officer chooses search rates $s_{i,r,c}$ to maximize his utility $\int \beta_{i,r,c} \pi_{r,c} n_{i,c} s_{i,r,c}^2 dc dr$ subject to both (53) and the time constraint $\int n_{i,c} s_{i,r,c} dc dr \leq T$. Replacing (53) into the utility function, we obtain $\int (\pi_c \int \beta_{i,r,c} n_{i,c} s_{i,r,c}^2 dc) dr$. The officer’s problem is then

$$\int (\pi_c \beta_{i,r,c} n_{i,c} s_{i,r,c}^2 - \mu_{EH} n_{i,c} s_{i,r,c}) dr dc + \mu_{EH} T,$$

where $\mu_{EH}$ is the Lagrange multiplier for the equal hit rates problem. The solution to this problem is given by the optimal search rate,

$$s_{i,r,c} = \left(\frac{(\pi_c \beta_{i,r,c})}{\int (\pi_c \beta_{i,r',c})^{1/\gamma} n_{i,c}^{\gamma} dc^{1-\gamma}}\right)^{1/\gamma} T.$$  \hspace{1cm} (55)

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

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

$$\frac{s_{i,r,c}}{s_{i,r',c}} = \left(\frac{\beta_{i,r,c}}{\beta_{i,r',c}}\right)^{1/\gamma}.$$  \hspace{1cm} (56)

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. Search rates at the race level must obey

$$s_{i,r} = \frac{\int (\pi_c \beta_{i,r,c})^{1/\gamma} dF_{i|r}}{\int (\pi_c \beta_{i,r',c})^{1/\gamma} dF_{i|r'}},$$

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

4.3. 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 $\int \beta_{i,r,c} \pi_{r,c} n_{i,c} s_{i,r,c}^2 dc dr$ subject to (48) and (53) and the time constraint $\int n_{i,c} s_{i,r,c} dc dr \leq T$. In parallel to earlier reasoning substitute (48) into the time constraint to get (49). Substituting (48) and (53) into the utility function produces

$$\int \left(\frac{s_{i,r,c}^2}{\pi_c} \int \beta_{i,r,c} n_{i,c} dr dc \right) dc.$$

The officer chooses search rates to maximize the Lagrangian

$$\int \left(\frac{s_{i,r,c}^2}{\pi_c} \int \beta_{i,r,c} n_{i,c} dr - \mu_{EH} n_{i,c} s_{i,r,c} - \mu_{EH} T\right) dc + \mu_{EH} T,$$

where $\mu_{EH}$ is the Lagrange multiplier forcing equal hit and search rates. The solution is

$$s_{i,r,c} \equiv s_{i,c} = \left(\frac{(\pi_c \int \beta_{i,r,c} n_{i,c} dr)^{1/\gamma}}{\int (\pi_c \int \beta_{i,r',c} n_{i,c} dr)^{1/\gamma} dc^{1-\gamma}}\right)^{1/\gamma} T.$$  \hspace{1cm} (60)

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.

$$\frac{s_{i,r,c}}{s_{i',r,c}} = \left(\frac{\int \beta_{i,r,c} n_{i,c} dr}{\int \beta_{i',r,c} n_{i,c} dr}\right)^{1/\gamma}.$$  \hspace{1cm} (61)

As before, bias is reflected in higher hit rates for groups containing more members of the particular race against which the officer is biased. Unlike the case where officers only equalize search rates, disparities in guilt probabilities across races do not play a role.

The 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 Sections 3 and 4, these results do not depend on the existence of unobservables since the police chief has access to the same information as officers. 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. Officers use variables that are correlated with race in their search allocations in order to indirectly sample races differentially.

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 (32), i.e. that officers only care about arresting guilty criminals.

5.1. Information from observable group search rates with homogeneous officers

Our first example focuses on restrictions placed on search rates for observed groups $(r, c)$. Assume that officer preferences are homogeneous so that $\beta_{r,c,q} = \beta_{q}$. Suppose further that race has no intrinsic value in predicting guilt probabilities, so that (31) holds. Finally, assume that unobservables are such that $\psi_{o,r} = \pi_{c} (c_{o} = \beta_{o})$ and $\beta_{q} = \beta_{q, c_{o}}$ are both monotonically increasing in $c_{o}$.

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 \( df_{\alpha | r, c_o} \) first order stochastically dominates \( df_{\alpha | r', c_o} \) 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

\[
\frac{\int s_{t, r, c_o} df_{\alpha | r, c_o}}{\int s_{t, r', c_o} df_{\alpha | r', c_o}} > 1. \tag{63}
\]

(b) If \( df_{\alpha | r, c_o} = df_{\alpha | r', c_o} \), then the ratio in (63) is unity.

Proof. From (36) calculate the ratio \( \int s_{t, r, c_o} df_{\alpha | r, c_o} / \int s_{t, r', c_o} df_{\alpha | r', c_o} \); note that the term \( \int (\beta_{i, r, c_o} \pi (c_o')) \frac{1}{\pi (c_o')} dc_o \) 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. For high \( c_o \) individuals like Gates, policy makers might assume roughly the same conditional distributions \( df_{\alpha | r, c_o} \) and \( df_{\alpha | r', c_o} \). Note that Proposition 6b 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, suppose previous studies have shown that the observed average 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.

5.2. 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. Assume that the officer’s preferences do not vary over unobservable characteristics, i.e. \( \beta_{i, r, c_o} = \beta_{i, r', c_o} \). The relative search rates for two officers \( i, i' \) satisfy

\[
\frac{s_{i, r, c_o}}{s_{i', r', c_o}} = \frac{\int \left( \frac{\beta_{i, r, c_o} \pi (c_o)}{\pi (c_o')} \right) \frac{1}{\pi (c_o')} df_{\alpha | r, c_o}}{\int \left( \frac{\beta_{i', r', c_o} \pi (c_o)}{\pi (c_o')} \right) \frac{1}{\pi (c_o')} df_{\alpha | r', c_o}} = \left( \frac{\beta_{i, r, c_o}}{\beta_{i', r', c_o}} \right) \frac{1}{\pi (c_o')} \frac{1}{\pi (c_o')} \frac{\int \pi (c_o) df_{\alpha | r, c_o}}{\int \pi (c_o) df_{\alpha | r', c_o}}. \tag{64}
\]

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

\[
\frac{n_{r, c_o}}{n_{r', c_o}} = \left( \frac{\beta_{i, r, c_o}}{\beta_{i', r', c_o}} \right) \frac{1}{\pi (c_o')} \frac{1}{\pi (c_o')} \frac{\int \pi (c_o) df_{\alpha | r, c_o}}{\int \pi (c_o) df_{\alpha | r', c_o}}. \tag{65}
\]

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 since

\[
h_{i, r, c_o} = \frac{\int \pi (r, c_o) \pi (r, c_o) \frac{1}{\pi (r, c_o)} df_{\alpha | r, c_o}}{\int \left( \beta_{i, r, c_o} \pi (r, c_o) \right) \frac{1}{\pi (r, c_o)} df_{\alpha | r, c_o}} = \frac{\int \pi (r, c_o) \pi (r, c_o) \frac{1}{\pi (r, c_o)} df_{\alpha | r, c_o}}{\int \left( \beta_{i', r', c_o} \pi (r, c_o) \right) \frac{1}{\pi (r, c_o)} df_{\alpha | r, c_o}}, \tag{66}
\]

and so do not depend on officer bias.

5.3. 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, which 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 \( \beta_r = 1 \) (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

\[
s_r = \frac{(\beta_r \pi_r) \frac{1}{\pi_r} T}{\pi_r \left( n_r + \pi_r \frac{1}{\pi_r} n_r' \right)}, \quad \text{and} \quad \frac{\beta_r \pi_r}{\pi_r}, \tag{67}
\]

\[
s_r = \frac{(\beta_r \pi_r) \frac{1}{\pi_r} T}{\pi_r \left( n_r + \pi_r \frac{1}{\pi_r} n_r' \right)}, \tag{68}
\]

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

\[
s_r = \frac{\pi_r \frac{1}{\pi_r} T}{\pi_r \left( n_r + \pi_r \frac{1}{\pi_r} n_r' \right)}, \tag{69}
\]

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

\[
s_r = \frac{(\beta_r \pi_r) \frac{1}{\pi_r} MT}{\pi_r \left( n_r + \pi_r \frac{1}{\pi_r} n_r' \right)} p + \frac{\pi_r \frac{1}{\pi_r} MT}{\pi_r \left( n_r + \pi_r \frac{1}{\pi_r} n_r' \right)} (1 - p), \tag{69}
\]

and

\[
s_r = \frac{\pi_r \frac{1}{\pi_r} MT}{\pi_r \left( n_r + \pi_r \frac{1}{\pi_r} n_r' \right)} p + \frac{\pi_r \frac{1}{\pi_r} MT}{\pi_r \left( n_r + \pi_r \frac{1}{\pi_r} n_r' \right)} (1 - p). \tag{70}
\]
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 Eqs. (69) and (70) give us a system of 2 equations with 3 unknowns: \( p \), \( \alpha \), and \( \beta_r \). From this system one can find identification regions for the 3 parameters. For example, for each point in a grid of values \( \alpha < 1 \), Eqs. (69) and (70) imply values for \( p \) and \( \beta_r \). If either \( p \not\in [0,1] \) 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 situations where 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 \( \beta_r \) is so large that type \( r \) officers search their race exclusively on race \( r \), so that their search rates are \( \frac{1}{\pi_r} \). In this case the observed search rates become

\[
s_r = \frac{MT}{n_r} p + \frac{\frac{1}{\pi_r} MT}{\frac{1}{\pi_r} n_r + \frac{1}{\pi_r} n_r} (1 - p)
\]

(71)

and

\[
s_r' = \frac{\frac{1}{\pi_r} MT}{\frac{1}{\pi_r} n_r + \frac{1}{\pi_r} n_r} (1 - p).
\]

(72)

Adding (71) and (72) and rearranging terms

\[
p = \frac{1 - s_r - s_r'}{MT - 1}.
\]

(73)

Eq. (73) 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 \( \frac{1}{\pi_r} \) may not be consistent with the data. One way to sharpen the bound is to replace \( \frac{1}{\pi_r} \) 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.

### 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 claims about the absence of taste-based discrimination and relationships between searches and guilt is sensitive to functional form assumptions and to assumptions about the nature of unobserved group level characteristics. 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 can lead to vastly different claims about the presence or absence of taste-based discrimination. In particular, we highlight how the presence of unobservables affects the robustness of their tests. Heckman (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.

How can the racial profiling literature constructively proceed? One way 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 this strategy. A systematic development of models that span the plausible assumptions concerning the 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. This approach leads to model averaging methods; see Brock et al. (2003) for defense of this approach.

A different route is to reformulate tests of taste-based discrimination in a partial identification context; we give an example in Section 5.3 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. Theoretical work on partial identification in abstract profiling environments has been done by Brock (2006) and Manski (2006), and is an appropriate conclusion to this paper.

### References


