The Role of Statistical and Taste Discrimination in Racial Disparities

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Shamena Yasmin Anwar

Dissertation Director: Professor Hanming Fang

May 2006

UMI Number: 3214167

ABSTRACT

The Role of Statistical and Taste Discrimination in Racial Disparities

Shamena Yasmin Anwar

2006

The theoretical literature on discrimination has proposed two main reasons why differential outcomes can arise for observationally equivalent individuals of different races. Taste-based theories postulate that differences in outcomes can develop if economic agents have a distaste or prejudice towards a particular racial group. On the other hand, statistical discrimination models attribute these outcomes to decision-making evaluators having incomplete information about the individuals under consideration. If evaluators believe that an individual's unobservable skill level is correlated with their racial background, they will have an economic incentive to take an individual's racial group into account when assessing their skill level.

Although there is an extensive empirical literature studying discrimination, most studies have focused on quantifying how much of the outcome differentials between racial groups can be attributed to discriminatory behavior. Very few studies have attempted to empirically determine whether these practices have arisen from taste or statistical discrimination. This is an important distinction because the effectiveness of policies to reduce discriminatory behavior depends upon the type of discrimination that is present. My dissertation provides empirical evidence on why discrimination arises by distinguishing between these two theories. The first chapter investigates whether people statistically discriminate when evaluating the skill level of others in the environment of the television game show *Street Smarts*. The second chapter develops an empirical test that determines whether racial differences in motor vehicle searches are due to racial prejudice on the part of police troopers. The third chapter evaluates whether taste discrimination on the part of customers might be responsible for the wage differential between black and white professional basketball players.

Contents

List of Tables

List of Figures

Acknowledgments

There are many people without whom this dissertation would not have been possible. I am deeply indebted to Hanming Fang for his exceptional guidance throughout this process. His constant support and encouragement were more than I could have ever asked for. I would also like to acknowledge the tremendous contributions Joseph Altonji and Patrick Bayer have made. Their detailed feedback and helpful suggestions have contributed significantly to my work. I am also extremely grateful to Panle Jia and Deran Ozmen for the advice and mentoring they provided me throughout the past six years.

The data utilized in Chapter 2 of the dissertation (co-authored with Hanming Fang) was obtained from the Florida Highway Patrol. We wish to thank Dick Zeller for providing us with the data. We are grateful to Kate Antonovics, Richard Freeman, and Nicola Persico for many helpful discussions regarding this project. We especially thank the Co-Editor of *American Economic Review* and four anonymous referees for many detailed and constructive comments on an earlier draft.

On a personal note, I would like to thank my parents, Mary and Zahur, and my brother, Mekhail, for both inspiring me to pursue a Ph.D. and instilling the belief in me that I could accomplish this goal. I could not have gotten through these past six years without the unconditional love and support they provided me. Finally, I would like to express my sincere gratitude to Jason Wong for his infinite patience and positive outlook on things. I cannot thank him enough for the emotional support he provided me on a day to day basis.

Introduction

There are two leading theories as to why disparities can arise between observationally equivalent individuals from different racial and gender groups. Taste-based theories attribute these disparities to economic agents being prejudiced against certain groups. Statistical discrimination theories ascribe these disparities to evaluators using an individual's group membership as a proxy for their unobservable skill level. My dissertation empirically investigates the role of taste and statistical discrimination in the disparities we observe. It is composed of three chapters, each of which examines a different area where an individual's group membership is likely to affect their outcome.

Chapter 1 tests whether statistical discrimination plays a role in skill assessment when evaluators have limited information. It is difficult to determine this in a standard labor market hiring setting, because the effects of statistical discrimination on hiring practices are empirically similar to the effects of taste discrimination. To circumvent this problem, I instead use data from the television game show *Street Smarts.* Due to the unique incentive structure present in this game I am able to develop a research design that can credibly distinguish between these two types of discrimination. This identification strategy is used to test for statistical discrimination both descriptively and within a formal structural model.

Both analyses indicate that contestants' statistically discriminate when evaluating the skill level of other individuals. In particular, I find that contestants perceive the unobservable skill level of blacks and females to be lower than that of non-blacks and males, respectively. These results imply that statistical discrimination may also play a significant role in hiring decisions.

Chapter 2, co-authored with Hanming Fang, investigates whether police troopers exhibit racial prejudice in their decision to search minority motorists at a higher rate than white motorists. One prominent approach to test for racial prejudice is the outcome test, which involves comparing the marginal search success rates of different racial groups of motorists. However because the marginal motorist is unobservable, researchers have not been able to accurately implement this test. We use a unique data set which contains demographic information about both the motorists searched on Florida highways and the troopers that conducted each search during a two-year period. The data also records the outcome of each search. Exploiting the information we have about trooper race allows us to develop a simple theoretical model of trooper search behavior. This model makes predictions we can use to design empirical tests for whether troopers of different races are monolithic in their search behavior, and whether they exhibit relative racial prejudice in motor vehicle searches. Our test of relative racial prejudice relies only on comparing the search outcomes of the average motorist searched, which is observable, and thus provides a partial solution to the well known inframarginality and omitted-variables problems that have plagued the empirical application of outcome tests in the past. When applied to the data, our tests soundly reject the hypothesis that troopers of different races are monolithic in their search behavior, but the tests fail to reject the hypothesis that troopers of different races do not exhibit relative racial prejudice.

Chapter 3 estimates the mean salary differential between equivalent black and white basketball players in the National Basketball Association (NBA) for the 2000-01 season, and tests whether fan preference for white basketball players (customer discrimination) might be a factor in this wage differential. The NBA provides a unique environment to estimate wage differentials because, unlike standard labor market situations, the productivity statistics of players are observable to the researcher. I find that, among players who are in at least their second contract, black players made 24% less than equivalent white players. Fan preference for white players is estimated by looking at the effect the racial composition of the road team has on the home team's attendance. Because the home team's preferences are unlikely to affect the racial composition of the road team, this identification strategy alleviates the endogeneity problem that is present in previous studies which measure fan preference as the effect the home team's racial composition has on the home team's attendance. Overall I find no significant evidence that fans prefer white players. However, when fan preference is estimated separately for each team, I find that a few teams' fans do have a strong preference for white players. This indicates that while customer discrimination might be a factor in the wage premium that is paid to white players, it is unlikely to be the main factor.

Chapter 1

Testing for Statistical Discrimination: Evidence from the Game Show *Street Smarts*

1.1 Introduction

Statistical discrimination can arise in screening situations where an individual must make a decision about someone else based on limited information. A prime example occurs with hiring in the labor market. In order to choose a potential employee from a group of applicants, employers will have to assess each applicant's skill level based on the information that is observable to them. If employers believe that the distribution of unobservable skills is different among racial and gender groups, defined as *cultural* groups, they will have an incentive to take these characteristics into account when they make their skill assessment.[1] This will cause employers to treat observationally equivalent applicants from

[1] Unobserved skill level should not be confused with inherent ability. Unobservable skill refers to factors like quality of education and past job experience, which will affect an individual's learned skill level but are

different cultural groups differently.

Despite the important implications statistical discrimination can have, it has proven to be very difficult to empirically determine whether this practice is a factor in hiring decisions. This is primarily because the effects of statistical discrimination on hiring practices are similar to the effects of another type of discrimination based on employer prejudice, termed taste-based discrimination. While employers will want to select the applicant they view as having the highest skill level, their selection decision will also depend on any personal prejudice they have towards the applicant's cultural group. From a researcher's standpoint, it is very difficult to tell if the negative effect a particular applicant's cultural background has on an employer's hiring decision occurs because the employer views them as having a relatively worse unobservable skill level or because employers have a personal prejudice towards them. It is important to distinguish between these two types of discrimination because the effectiveness of policies to reduce discriminatory behavior will depend on the type of discrimination present.[2],[3]

The difficulty in distinguishing between the effects of statistical and taste discrimination arises because both have the same effect on an employer's hiring decisions. To circumvent this problem, I instead attempt to determine whether statistical discrimination is a factor in skill assessment in the unusual environment of the television game show *Street Smarts.*

not typically observed by employers.

[2]By stating that the effect of group membership on an employer's hiring decisions reflects discrimination, I am implicitly assuming that researchers have all of the information about applicants' observable skill that employers have at the time they make their decisions. If researchers didn't have full information, then this effect would not necessarily reflect discrimination.

[3]If statistical discrimination is present, one possible policy remedy would be to require employers to gather more objective information about applicants. This policy would have no effect if it is instead taste discrimination that is driving hiring disparities.

During this game, contestants earn money by predicting how three random people, known as street savants, answered basic trivia questions. Like a hiring situation, any use of statistical or taste discrimination by contestants will affect the decisions they make in the game. The main contribution of this chapter is that, due to the unique incentive structure present, I am able to develop a new research design that can credibly distinguish between these two types of discrimination within this game.[4]

There are two unique features about this show which allow me to accomplish this. The first is that at the beginning of the third round, contestants will have to select the street savant who they perceive as having the most extreme skill level (i.e., the savant who has either the highest or lowest skill level). The second feature is that contestants' perception of a savant's skill level when they make this selection will depend on both their perception about the savant's unobservable skill and the savant's performance in answering five questions during the first two rounds of the game.

The combination of these two features will cause statistical and taste discrimination to have different effects on a contestant's choice of savant at the beginning of the third round. Specifically, if contestants statistically discriminate against a particular group of savants, the effect of a savant's cultural group on their likelihood of being selected will depend on their prior performance. In contrast, if there is taste discrimination against this same group, the effect of a savant's cultural group on their likelihood of being selected will be independent of their prior performance. This identification strategy is used to test for statistical discrimination both descriptively and within a formal structural discrete choice

[4]Another nice feature about using this game show is that we are privy to the same information about savants that contestants have when they make decisions. This will reduce the presence of omitted variable bias.

model.

Both the descriptive and structural analyses indicate that contestants strongly statistically discriminate against black and female savants. Furthermore, the structural results indicate that contestants perceive the differences in unobservable skill between these cultural groups to be quite large. At the beginning of the game contestants feel a college educated non-black male savant has an 89% chance of answering a question correctly. A college educated black male savant is perceived to have a 62% chance, a college educated non-black female has a 54% chance, and a college educated black female has a 27% chance. I also find evidence of taste-based discrimination towards female savants.

These perceptions about unobservable skill are admittedly identified from a case study and are thus not directly applicable to real world settings. However they strongly imply that statistical discrimination may play a significant role in hiring decisions also. Furthermore, the methodology proposed here provides insight as to what types of situations would be necessary in the labor market to be able to test for statistical discrimination.

The remainder of this chapter is organized as follows. Section 1.2 discusses the related literature. Section 1.3 describes the game *Street Smarts* in detail. Section 1.4 outlines the identification strategy. Section 1.5 presents the results from the descriptive analysis, while Section 1.6 presents the structural model and estimation results. Section 1.7 provides empirical support for the assumptions necessary for the identification strategy. Section 1.8 concludes. Tables and figures are included in an appendix in Section 1.9.

1.2 Related Literature

While it has been difficult to distinguish between statistical and taste discrimination in the labor market, there have been some studies that have accomplished this in other areas, although the methods developed cannot be directly applied to labor market settings. In Chapter 2 of this dissertation my co-author Hanming Fang and I try to determine why minority motorists are more likely to be searched by highway patrol officers. We fail to reject the null hypothesis that officers do not exhibit racial prejudice in their search decisions. List (2004) tries to determine why minorities are likely to receive inferior offers in situations which require market negotiations. Conducting a field experiment in the setting of sports card dealing, he finds this behavior arises due to statistical discrimination–dealers have different perceptions about the reservation wage distributions of minorities and whites.

The approach I take in this paper follows the general strategy proposed by Levitt (2004). He tries to distinguish between the effects of statistical and taste discrimination using contestant behavior on the game show *Weakest Link*.[5] Similar to *Street Smarts*, this game show provides a unique environment where statistical discrimination will have a different effect on contestant behavior than taste discrimination will. He finds no evidence of discrimination towards women or blacks, but does find some evidence that Hispanic contestants are treated adversely due to statistical discrimination.

This study improves on Levitt's study in two key ways. The first is that contestant behavior on *Weakest Link* is strategic, as there are many different factors that affect a con-

[5] Antonovics et. al (2003) also perform a similar study to Levitt's using data from *Weakest Link*.

testant's decisions. In contrast, there is virtually no strategic behavior between contestants on *Street Smarts*. This allows me to structurally model contestants' decisions and actually estimate the perceptions contestants have about the unobservable skill level of different cultural groups. This is helpful when trying to determine how large an impact statistical discrimination is likely to have.

The second issue is that the identification strategy used in Levitt's paper relies on the idea that in the beginning of the game contestants vote off the weakest players, but towards the end of the game they will vote off the strongest players. But a separate study by Antonovics et. al. (2003) has shown that there is no empirical support that this is how contestants actually vote, which sheds doubt on the identification argument. In contrast, I will show strong empirical support for the identification assumptions necessary in my model.

1.3 Background on *Street Smarts*

Street Smarts is a game show in which two contestants compete for a winner take all prize by predicting what other people know. Prior to the show, the host goes out to popular locations and talks separately to three random people.[6] Each of these three people, termed street savants, are asked the same set of trivia questions. These trivia questions are all designed to have the same (moderately easy) difficulty level but they span many different categories, including general knowledge, entertainment, sports and more. Table 1.1 shows the six different question categories, as well as some examples of each.[7] The

[6] While the show is based in southern California, the host travels all over the U.S. to interview people.

[7] The questions in the game do not have official categories. The categorization in this table reflects my subjective classification.

two contestants playing the game will then have to predict whether these street savants answered the questions right or wrong. The game consists of four rounds, with each round introducing a slightly different structure.

At the start of the show, the host introduces the two contestants and viewers can observe their race, gender and age.[8] The contestants are then introduced to the savants they will be making predictions about by being shown a short interview between the host and each of the three street savants. Contestants will typically learn the occupation of each of the savants during the course of the interview, and can observe their race, gender and age. Because there is such a wide variety of occupations, I assume that contestants use a savant's occupation level as an indicator of their education level (i.e., whether they went to college or not).

During Round 1 the host reads a question and contestants are told that one of the three street savants answered the question right and the other two got the question wrong.[9] The contestants must simultaneously predict which street savant got the question right.[10] Contestants then observe how each street savant answered the question, by viewing the videotaped clip of the host asking the savant the particular question.[11] There are three

---

In Round 3, 39.6% of the questions asked are "general knowledge", 20.4% are "entertainment", 4.9% are "sports", 9.8% are "slang", 6.5% are "childrens' interests", and 18.8% are "miscellaneous". The category "miscellaneous" groups together about eight different categories, each of which comprises less than 4% of the total amount of questions.

[8] Often, viewers also learn contestants' occupation.

[9] During the first season of the show Round 1 was slightly different than it was in later seasons, because it was not necessarily the case that only one savant got the question right. Contestants were just told to pick which savant they thought had answered the question correctly and it was sometimes the case that two of the three savants answered the question correctly.

[10] Contestants are not told the answer to the trivia question before they must make their prediction.

[11] Contestants will always observe how the savant they chose answered the question, but if a savant is not

questions in this round and each right prediction by a contestant is worth \$100. There are no penalties for wrong answers.

During Round 2 the host picks two of the three street savants and then reads a question. Contestants are told that one of the two savants got the question right and the other got the question wrong. Contestants must simultaneously predict which of the two savants got the question wrong, and are then shown how each of the two savants answered the question. There are three questions in this round, and for each question a different combination of two of the three street savants is chosen. Each question is worth \$200 with no penalties for wrong answers.

At the start of Round 3 contestants must each choose a savant that they will have to make predictions about during this round. The two contestants cannot choose the same savant, so the player trailing at the start of the round chooses first, and the other contestant chooses their savant from the remaining two.[12] The contestant that chose first is then told a question, and will have to predict whether the savant they picked answered the question right or wrong. They are then shown how their savant answered the question. The other contestant is then told a different question and must predict how their savant answered this question. This process repeats until both contestants have predicted how their savants answered three questions. Each correct prediction is worth \$300 in this round with no penalties for wrong answers.

---

picked, the host does not always show that savant's videotaped clip. But because the structure of the game forces two of the contestants to get the question wrong, and one to get it right, contestants will usually know how all three savants answered the question, even if they don't observe all of their answers.

During the first season, though, when up to two players could potentially have answered the question correctly contestants will not always know how each savant answered the question.

[12] If both players have the same score at the start of the third round, the player that won the backstage tie breaker before the start of the show gets to choose first.

In Round 4 the host reads a question to the contestants. Each contestant will then pick a street savant and predict if they answered the question right or wrong and wager an amount of money that cannot exceed what they have currently earned in the game. If the contestant predicts the savant's answer correctly they will add the amount of money they wagered to their total. If the contestant incorrectly predicts the savant's answer they will lose the amount of money they wagered from their total. This concludes the game–the player with the most money wins and keeps their win total and the losing player gets nothing.[13]

The data was collected by manually videotaping episodes and transcribing them. In total, data for 299 episodes were collected, resulting in a total of 598 contestants and 897 street savants.[14] Tables 1.2 and 1.3 provide information about the demographic characteristics of both the contestants and savants. Table 1.2 shows that 62.7% of the contestants are white, 22.4% are black, 6.5% are Hispanic, 6.4% are Asian, and 2.0% are unknown. For the savants, 59.4% are white, 24.1% are black, 7.9% are Hispanic, 5.7% are Asian, and 2.9% are unknown.

One can see that both genders are, for the most part, equally represented in the game: for contestants, 51.8% are female and 48.2% are male, while for savants 55.1% are female, and 44.9% are male. Information on both contestant and savants' occupation level was

[13]Producers have complete control over how this game is put together–they choose which savant answers to show and the order in which these answers appear. This opens up the possiblity that producers could manipulate the game to try and trick contestants. I will discuss the possible implications this could have in Section 1.7.3.

[14]*Street Smarts* went on for five seasons, beginning in Fall 2000 and finishing in Spring 2005. The episodes used include all of the episodes from the fifth season (which were recorded from a local television station), and random episodes from each of the previous four seasons (which were recorded from the cable television station *Game Show Network* which shows reruns of *Street Smarts*).

transformed into whether or not they might be perceived to be college educated or not.[15] The percentages of college educated people are very similar across both contestants and savants, with 24.1% of the contestants and 23.4% of the savants being college educated. In terms of age, people were classified into two groups: less than or equal to 35 years old, and greater than 35.[16] One can see that the age composition of both the contestants and savants is young, with 89.1% of the contestants and 75.3% of the savants less than or equal to 35 years old.

Table 1.3 provides a more in-depth look at the demographic characteristics shown in Table 1.2. The first panel of Table 1.3 shows how the male/female composition differs across the races for both contestants and savants. For example, we can see that about half of the white contestants (50.9%) are female, while 61.2% of the white savants are female. The second panel of Table 1.3 shows how the different racial and gender groups of contestants and savants differ across education levels. One can see that a higher percentage of blacks are college educated than whites are across both contestants and savants. Because in the general population blacks, on average, have a lower education level than whites, this indicates that the show might be using a sample of blacks that is more educated relative to whites than a random sample from the general population would have been. We can also see that while male contestants tend to be more educated than female contestants, female savants are more educated than male savants, although the differences are not too large.

[15] The Occupational Outlook Handbook was used to determine this in the following way: if an individual's occupation requires a college degree or if the individual says they are currently in college, they are classified as college educated. Individuals with occupations that do not require a college degree, or whose occupations are unknown, are classified as non-college educated.

[16] Contestants and savants do not state their age during the show, so I subjectively grouped these individuals into two age groups. Although this is a rather difficult visual classification to make, it is the best we can do without explicit age data.

1.4 Identification Strategy

The goal of the empirical estimation is to identify the perceptions cultural groups of contestants have about the average unobservable skill of one group of savants relative to another. It will also identify whether contestants have any taste-based preference for interacting with certain cultural groups of savants.[17] This will require us to distinguish between the effects of statistical and taste discrimination. Although contestants make many decisions regarding savants throughout the game, their choice of savant at the beginning of the third round is the only one that will allow us to distinguish between the two types of discrimination. There are two unique features about this particular decision which enable this.

The first feature is that contestants selection of savant is based on how extreme the savant's skill level is. Recall that during the third round contestants will have to make predictions about how this savant answers three questions. Thus, they will want to choose the savant whose future answers will be easiest to predict. This means they will choose the savant who they feel is either most likely to get the next question right or most likely to get the next question wrong.[18]

The second feature of this game is that a contestant's perception of a savant's skill level when they make this selection will depend on both their prior perception of the savant's

[17]Although we will be able to identify both taste and statistical discrimination, the focus of this chapter is more on testing for the latter. This is primarily because we are testing for this in a second-best situation and trying to apply the results to a more standard labor market situation. The nature of taste discrimination in this show is quite different than it would be in a hiring setting, as contestants only interact with savants for a short period of time. In contrast, one might think that the nature of statistical discrimination in this show is more representative of a hiring situation. Both situations involve an evaluator making a skill assessment of another individual based on limited information.

[18]There is no other decision in the game where contestants select a savant based on the extremeness of their skill level.

skill level, as well as the savant's performance in answering five questions during the first two rounds of the game. The contestant's prior perception about the savant's skill level is just their perception about the savant's unobservable skill.[19]

It is the combination of these two features which will cause statistical and taste discrimination to have different effects on a contestant's choice of savant in the third round. To see why, consider the following example: suppose we have two groups of savants, termed Group A and Group B, and that contestants statistically discriminate against Group A. In particular, they have a very low prior perception about savants from Group A, and a very high prior perception about savants from Group B. The following diagram shows how the perception of the two groups' skill level and their likelihood of being selected at the beginning of the third round depend on their performance in the first two rounds:

	Group A's skill level	Group B's skill level	Group contestants are most likely to select
high performance	average	high	Group B
low performance	low	average	Group A

Examining the second column, one can see that among Group A savants, those that perform extremely well in the first two rounds (answering basically all of their five questions correctly) will be considered to be average skilled at the beginning of the third round. This is because contestants combine their low prior perception about these savants with a high

[19]The prior perception a contestant has about a savant is their perception of their skill level at the beginning of the game before they have observed how they have answered any questions. The only reason these perceptions should differ between two otherwise equivalent savants from different cultural groups is if the level of unobservable skill is perceived to differ among the two cultural groups.

observed signal. Since at this stage contestants want to select the savant with the most extreme skill level, these savants will not be likely to be selected. The savants in Group A who perform extremely poorly in the first two rounds (answering most or all of their five questions wrong) will be considered to be very low skilled overall, and thus will be likely to be selected.

With savants in Group B we get the opposite pattern. Those that perform well in the first two rounds will be considered highly skilled overall and will be likely to be selected. Those that perform poorly will be considered to be average skilled and so will not be likely to be selected.

Thus, as the last column shows, among savants who perform extremely well in the game, Group B savants will be more likely to be selected. Among savants who perform poorly, Group A savants will be more likely to be selected.

Now suppose that contestants have the same prior about both groups (meaning there is no statistical discrimination), but instead have a personal prejudice against Group A savants. The following diagram shows how contestants' selection decision depends on the savant's performance during the first two rounds:

	Group contestants are most likely to select
high performance	Group B
low performance	Group B

Contestants having taste discrimination against Group A savants means they will not want to interact with them and so would prefer not to select them in the third round. This means that among savants with both high and low performance levels Group B savants will be more likely to be selected.

By comparing the selection pattern under statistical and taste discrimination the following idea should be clear: with statistical discrimination the effect a savant's cultural group has on their likelihood of being selected depends on their past performance, while with taste discrimination this effect is independent of past performance. This is the key distinction that will allow us to use this decision to distinguish between these two types of discrimination.[20]

Finally, it should be pointed out that several assumptions have been made in this section, both implicitly and explicitly, in order to ensure the validity of this identification strategy. In particular, we have assumed that contestants will choose the savant with the most extreme skill level, use a savant's past performance as a predictor of future performance, treat all questions as though they come from the same category, and use their true prior perception about each group of savants when making a decision. Empirical support will be shown for each of these assumptions in Section 1.7.

1.5 Descriptive Analysis

The diagram of statistical discrimination shown above indicates that the prior perception of a particular group will determine how likely they are to be chosen at high performance

[20] We can compare this decision contestants make to a hiring decision made in the labor market. Note that in the labor market we typically think employers will want to select the applicant that has the highest skill level. Examining the diagram for statistical discrimination, one can see that this would imply that among savants who perform poorly, Group B savants would be more likely to be selected (because average skill is greater than low skill). Thus the choice pattern will be the same under both statistical and taste discrimination, and so we cannot distinguish between them.

levels and at low performance levels. This implies that examining how likely a group is to be chosen at these different performance levels should allow us to infer something about the prior perception of that group. Comparing these rough estimates of prior perceptions about each group will then allow us to conduct a more descriptive test for statistical discrimination.

Table 1.4 shows how the probability of a person in a particular racial or gender group being selected in the third round depends on their performance in the first two rounds. The upper-left cell entry means that of the black savants that answered between zero to twenty percent of their questions correctly in the first two rounds, 45.8% of them were selected by contestants at the beginning of the third round.[21],[22] To get a more clear idea about how this selection probability depends on past performance we can graph these trends and compare them among groups. Figure 1.1 compares the trends for black versus non-black savants.[23] The downward sloping trend present for black savants means that as they perform better, their probability of selection decreases. This implies that contestants' prior perception about them is below .5.[24] In contrast, the trend for non-black savants is upward sloping–the probability that they are selected steadily increases as performance improves.

[21]If a savant was selected by the first contestant they were represented in the data set once and coded as being selected. If a savant was not selected by the first contestant but was selected by the second, they were represented in the data set twice and coded once as not being selected and once as being selected. Finally if a savant was not selected by either contestant, they were represented in the data set twice and coded both times as not being selected.

[22]Technically, there should be six previous performance categories instead of three (corresponding to whether savants got zero, one, two, three, four, or five questions correct). Due to the paucity of observations that were in these cells, these six categories were condensed into three.

[23]Due to the limited number of savants that fall into the racial category outside of blacks and whites, these savants were combined with whites into the "non-black" category.

[24]If the prior about blacks is less than .5, then when they perform poorly they will be considered to be more extreme than they will when they perform well. Thus they will be more likely to be selected at lower performance levels.

This corresponds to a prior perception that is greater than .5. These differences in prior perceptions imply that contestants are statistically discriminating against black savants.[25]

Figure 1.2 compares the trends for female versus male savants. For males, there is an upward sloping trend which is consistent with a prior perception about them that is greater than .5. For females, the pattern is slightly U-shaped so that the more extreme their performance is in either direction, the more likely they are to be selected. This coincides with a prior perception of around .5. Thus this descriptive analysis indicates it is also likely that contestants are statistically discriminating against female savants.

1.6 Structural Analysis

Although the descriptive analysis is informative, it is unlikely that the perceptions we are identifying are entirely accurate. We are measuring the probability that a given person is selected as simply whether or not they were chosen in the third round. This does not control for the characteristics of the other two savants that were available to be chosen, nor does it control for any of the other characteristics about the savant besides their race and gender. Since these factors will affect the likelihood that a given savant is chosen it will be necessary to control for them.

Another disadvantage is that we cannot explicitly estimate what the prior perception about each racial and gender group is. This is important because it will give us an idea about how large the effects of statistical discrimination are likely to be.

[25] The sign of the slope (positive or negative) is what indicates what the prior perception of the group of savants is. Note that the magnitude of the slope will determine how much weight is put on the signal. The steeper the slope, the more weight that is put on the signal (the selection probability is more responsive to the signal). Thus if the trend line for one group was steeper than for another it would indicate that contestants have more faith in their prior perception about the latter group.

In order to solve both of these problems it will be necessary to do a more formal analysis. Specifically, I will estimate a structural discrete choice model of a contestant's choice of savant at the beginning of the third round. This model will formalize the identification strategy that was presented in Section 1.4 and allow us to explicitly estimate what the prior perception about each racial and gender group is.

This section is organized as follows: Section 1.6.1 presents the structural model, Section 1.6.2 discusses the estimation procedure, and section 1.6.3 presents the results.

1.6.1 The Model

I assume that contestants view each question a savant answers as a Bernoulli trial, where $p_i \in (0,1)$ denotes the true probability that savant i will get each question correct. This requires that contestants do not view questions as being category-specific.[26] Because contestants do not know the questions they will have to make predictions about at the time they choose their savant, this is a reasonable assumption.[27],[28],[29] One can think of p_i as a

[26] If questions are considered to be category-specific, then a savant's p_i will depend on the type of question.

[27] Note that this assumption does not require contestants to make predictions based on p_i when they actually learn the question in the third round. It only requires contestants to behave this way when they are choosing their savant at the beginning of the round.

[28] Because contestants will learn the question before they make an actual prediction about the savant, one critique of this specification is that contestants might choose savants who are the most extreme in *each* category, rather than choosing savants whose skill level in answering the average question asked is the most extreme. In Section 1.7.2 I will provide empirical evidence supporting why the model I specify here is an accurate representation of contestant behavior.

[29] The primary reason contestants' choice of savant at the beginning of the third round is used to identify prior perceptions is because this is the only decision that will allow us to distinguish between statistical and taste discrimination. However, there is one other key advantage to using this decision: since it is reasonable to assume that contestants will be selecting the savant based on how extreme they are at answering the average question, contestants will be making the same decision in every game. This will allow us to aggregate the data from all shows together.

Note that all other decisions that contestants make in the game occur after they have already heard a particular question. This means their predictions will at least be partly based on how they believe that particular savant can answer that particular question. Since these questions will differ during a show and savant's true skill level in answering trivia questions asked in the game. Contestants do not know a savant's true p_i and thus have to use their perception of this.

I assume that contestants view the distribution of p_i among savants to be beta distributed with parameters α and β.[30] The beta distribution works well here because it lies on the support $(0, 1)$, and has a very flexible shape. Since it is unlikely that contestants will view savants with different observable characteristics as coming from the same skill distribution, I assume that contestants perceive the particular beta distribution savants come from, and thus the parameters α and β, as being a linear combination of the savant's race, gender, and education.

The beta distribution represented by the parameters α_i and β_i is the prior distribution that a contestant views savant i as coming from. Before the third round starts, contestants will also have observed how the savant has answered questions in rounds 1 and 2. They will use this information to update this prior perception about the distribution a savant comes from.

The number of questions a savant answers correctly in the first two rounds is distributed binomially. Since the beta distribution and the binomial distribution are a conjugate pair, contestants posterior distribution for each savant is also beta distributed, albeit with new

---

across shows, we will not be able to aggregate data from all shows together.

[30]The $beta(\alpha, \beta)$ *pdf* is:

$$f(p \mid \alpha, \beta) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1}, 0 < p < 1, \alpha > 0, \beta > 0$$

where $B(\alpha, \beta)$ denotes the beta function,

$$B(\alpha, \beta) = \int_0^1 p^{\alpha-1}(1-p)^{\beta-1} dp$$

The mean and variance of the beta distribution are given in (1.3) and (1.4).

parameters that incorporate the information provided by this signal. The new posterior distribution a contestant views savant i as coming from is:

$$Beta(y_i + \alpha_i, n_i - y_i + \beta_i)$$

where y_i is the number of questions savant i answered correctly and n_i is the number of questions savant i answered. I assume that at the start of the third round contestants perceive savant i's skill level, $\widehat{p_i}$, in playing the game to be the mean of this posterior distribution:

$$\widehat{p_i} = \frac{y_i + \alpha_i}{\alpha_i + \beta_i + n_i} \tag{1.1}$$

This can be rewritten in a more intuitive way:

$$\widehat{p_i} = \left(\frac{n_i}{\alpha_i + \beta_i + n_i}\right) \underbrace{\left(\frac{y_i}{n_i}\right)}_{\text{signal}} + \left(\frac{\alpha_i + \beta_i}{\alpha_i + \beta_i + n_i}\right) \underbrace{\left(\frac{\alpha_i}{\alpha_i + \beta_i}\right)}_{\text{prior mean}} \tag{1.2}$$

This shows that the posterior mean $\widehat{p_i}$ is just a weighted average of the mean of the prior distribution and the signal sent during the first two rounds of the game.

With this specification, the parameters we will be estimating will identify how the cultural background of the savants affect the shape of the prior distribution. These parameters will be difficult to interpret, though, because it is not at all intuitive what an increase in α or β really means. To facilitate understanding, we could instead directly parameterize the mean (μ) and variance (σ) of the prior distribution. In order to do this we need to rewrite α_i and $(\alpha_i + \beta_i)$ in terms of μ_i and σ_i.[31] The mean μ_i and variance σ_i of the prior beta

[31] Since only α_i and $(\alpha_i + \beta_i)$ enter into the expression for $\widehat{p_i}$, it is easier to parameterize α_i and $(\alpha_i + \beta_i)$

distribution are related to the parameters in the following way:

$$\mu_i = \frac{\alpha_i}{\alpha_i + \beta_i} \tag{1.3}$$

$$\sigma_i = \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)^2 (\alpha_i + \beta_i + 1)} \tag{1.4}$$

Using this relationship we can rewrite α_i and $(\alpha_i + \beta_i)$ as:

$$\alpha_i = \mu_i \left(\frac{\mu_i(1-\mu_i)}{\sigma_i} - 1 \right) \tag{1.5}$$

$$\alpha_i + \beta_i = \left(\frac{\mu_i(1-\mu_i)}{\sigma_i} - 1 \right) \tag{1.6}$$

We can then rewrite $\widehat{p_i}$ as:

$$\widehat{p_i} = \left(\frac{n_i}{\frac{\mu_i(1-\mu_i)}{\sigma_i} - 1 + n_i} \right) \left(\frac{y_i}{n_i} \right) + \left(\frac{1}{1 + \frac{n_i}{\left(\frac{\mu_i(1-\mu_i)}{\sigma_i} - 1\right)}} \right) (\mu_i) \tag{1.7}$$

The mean, μ_i, will be a linear combination of the characteristics of the savant. In particular:

$$\mu_i = \omega_1 + \omega_2 \cdot COLLEGE_i + \omega_3 \cdot BLACK_i + \omega_4 \cdot FEMALE_i \tag{1.8}$$

where $COLLEGE_i$ indicates savant i has attended or is currently attending college, and $BLACK_i$ and $FEMALE_i$ indicate whether the savant is black and/or female, respectively. The parameter ω_1 will measure a contestant's perception about the mean of the prior skill distribution of a non-black male savant who is not college educated. This corresponds to the contestant's view of how likely that savant is to answer a question correctly at the

instead of α_i and β_i.

beginning of the game.[32]

Technically the variance σ_i will also be a function of the same characteristics specified in (1.8), although it will have different parameters. Unfortunately, though, it will be difficult to identify this many parameters with the amount of data that is present. To simplify the estimation, I assume that $\sigma_i = \sigma$.[33]

In order to determine whether statistical discrimination is a factor in skill assessment we need to identify a contestant's perception of one cultural group of savant's average unobservable skill level relative to another. The main reason the cultural group of the savant affects a contestant's prior perception of their average skill level, once all observable characteristics are controlled for, is because average unobservable skill differs between these cultural groups. The parameters ω_3 and ω_4 identify what these relative differences are. Thus finding either $\omega_3 < 0$ or $\omega_4 < 0$ will indicate that contestants statistically discriminate against black and female savants, respectively.

At the beginning of the third round, the utility a contestant gets from selecting savant i will be a function of both the extremeness of the savant's skill level and any taste

[32]There are several other variables that are observed but not controlled for in (1.8) simply to keep down the number of parameters that need to be estimated. Ideally, we would want to include additional controls for savants who are neither white nor black, savants who have an unknown education level, and savants who are older than 35. The race and gender of the contestants are also likely to affect their skill perception of a particular savant and thus we would want to include controls for these also.

[33]Identifying the parameters composing the variance would essentially reveal how much faith contestants have in their prior perception about a given cultural group of savants. The results would be particularly interesting if we found that contestants had more faith in their prior about savants from their own cultural group.

discrimination they may have towards the savant:[34],[35]

$$U_i = \gamma \cdot \max(\widehat{p}_i, 1 - \widehat{p}_i) + \theta_1 \cdot BLACK_i + \theta_2 \cdot FEMALE_i + \epsilon_i \tag{1.9}$$

If contestants have a personal prejudice against black or female savants they will get less utility from picking them. However, this personal prejudice should not affect their perception of the savant's skill level, $\widehat{p}_i$, and so it will enter into the utility function separately.[36] The parameters θ_1 and θ_2 will measure this taste based prejudice against black and female savants.

Thus, the race and gender group of the savant will enter into a contestant's choice function twice. The impact a savant's group status has on the contestant's perception of their skill level (identified by the parameters ω_3 and ω_4) reflects statistical discrimination, while the direct impact their group status has on the contestant's choice of savant (identified by θ_1 and θ_2) picks up taste discrimination. To see how we can separately identify $\omega_3, \omega_4, \theta_1$ and θ_2, note that the utility specification given in (1.9) is just a formalization of the intuitive

[34]Note that the savant with the largest value of $\max(\widehat{p}_i, 1 - \widehat{p}_i)$ will have the most extreme skill level.

[35]There are two other factors that one might think will affect a contestant's choice of savant that are not represented in (1.9). The first is that contestants might select the savant whose answers they predicted best in the first two rounds. This is not a valid control, as contestants actually predict all savants' answers in the first two rounds equally well. Recall that at the same time contestants predict a savant will answer a question correctly, they are simultaneously predicting that another savant will answer the question incorrectly. Thus if their prediction is incorrect, they are wrong about both savants.

Another factor that might enter (1.9) is that contestants might select the savant who is most like them, because they think their answers will be easier to predict. Technically, this should not matter. Say a contestant and savant are alike and both have a 50% chance of getting a given question correct. Just because the contestant knows the answer to a specific question doesn't change the fact that the savant still only has a 50% chance of answering the question correctly. The idea is that they may both answer the same percentage of questions correctly, but they will not necessarily answer the same questions correctly. However, if contestants do believe choosing savants similar to them will help, this preference will show up in the taste parameters. It should not affect the identification of statistical discrimination.

[36]This assumes contestants get no disutility from perceiving other cultural groups' skill level accurately.

identification example presented in Section 1.4. When statistical discrimination is present the effect of a savant's group status on contestants' choice will depend upon the savant's prior performance y_i. In contrast, when taste discrimination is present, the effect of a savant's group status on contestants' choice will be independent of prior performance. This key distinction between statistical and taste discrimination allows the separate identification of each of their effects.

1.6.2 Estimation Procedure

In order to determine the parameters $\gamma, \omega_1, ..., \omega_4, \sigma, \theta_1$, and θ_2, I estimate a discrete choice model where the contestant's choice of savant at the beginning of the third round will be used to identify these parameters. I use maximum likelihood estimation, which is designed to find the parameters that make the choice of savant contestants make as likely as possible.

As laid out in Section 1.6.1 the utility contestants receive from selecting savant i is:

$$\begin{aligned} U_i &= \gamma \cdot \max(\widehat{p}_i, 1 - \widehat{p}_i) + \theta_1 \cdot RACE\_MATCH_i \\ &\quad + \theta_2 \cdot GENDER\_MATCH_i + \epsilon_i \end{aligned}$$

For simplicity we can define the observable portion of utility as V_i:

$$\begin{aligned} V_i &= \gamma \cdot \max(\widehat{p}_i, 1 - \widehat{p}_i) + \theta_1 \cdot RACE\_MATCH_i \\ &\quad + \theta_2 \cdot GENDER\_MATCH_i \end{aligned}$$

Contestants will choose savant i if and only if $U_i \geq U_k \; \forall k \neq i$, where k denotes the three savants who could potentially be chosen.[37] Thus, the probability that contestants choose savant i is:

$$\begin{aligned} P_i &= \Pr\left(U_i \geq U_k, \forall k \neq i\right) \\ &= \Pr\left(V_i + \epsilon_i \geq V_k + \epsilon_k, \forall k \neq i\right) \\ &= \Pr(\epsilon_k \leq \epsilon_i + V_i - V_k, \forall k \neq i) \end{aligned} \tag{1.10}$$

In order to determine the probability of the event in (1.10) happening, we need to specify the distribution of the error term ϵ_k. I assume that each ϵ_k is independently, identically distributed extreme value so that the choice probability specified above will be logit. This assumption means that the unobserved portion of utility for one alternative is unrelated to the unobserved portion of utility for another alternative (Train, 2003). The unobservable utility for choosing a given savant is likely to reflect a contestant's positive or negative taste for qualities about the savant that are unobservable to the researcher, such as their personality. Since it is very uncommon for people to have the same personality, we would not expect to know how a contestant felt about the second savant's personality just by knowing how they felt about the first. Thus specifying the choice probability as logit should not be too restrictive.

[37] Only the contestant that selects first will choose from among all three savants. The contestant that selects second only chooses among the remaining two. I will take this into account when I perform the actual estimation.

With a logit choice probability, (1.10) will have a simple closed form solution given by:[38]

$$P_i = \frac{e^{V_i}}{\sum_k e^{V_k}}$$

Now that we know the probability that a contestant will choose a given savant, we need to determine the probability that a contestant will choose the savant that they were actually observed to choose. This is given by:

$$\prod_i (P_i)^{x_i}$$

where $x_i = 1$ if a contestant chose savant i and zero otherwise. If we assume that each contestant's choice of savant is independent of that of other contestants, then the probability that each contestant in the sample will choose the savant that they were actually observed to choose can be given by the following likelihood function:

$$L(\gamma, \omega, \sigma, \theta) = \prod_{j=1}^{J} \prod_i (P_i)^{x_i}$$

where γ, ω, σ, and θ are vectors containing the parameters of the model, j denotes an individual contestant, and J is the total number of contestants in the data set. The log-likelihood function is then:

$$LL(\gamma, \omega, \sigma, \theta) = \sum_{j=1}^{J} \sum_i x_i \ln P_i \tag{1.11}$$

Our estimator will be the values of γ, ω, σ, and θ that maximize this function, subject

[38] See Train (2003) for a derivation of this logit choice probability.

to the necessary constraints. The fact that the underlying parameters, α and β, of the beta distribution must both be strictly positive implies two constraints on the transformed parameters, μ and σ. Namely,[39]

$$0 < \mu_i < 1, \forall i \tag{1.12}$$

$$0 < \sigma < \mu_i (1 - \mu_i), \forall i \tag{1.13}$$

Because the log-likelihood function specified is not continuously differentiable, the Nelder-Mead search algorithm was used to find the parameters.

To verify that the estimation method used uncovers the true parameters, a Monte Carlo simulation was performed. These results are shown in Table 1.5. The specified parameters are the underlying parameters that were used to create the simulated data. Although the actual data set only contains 598 observations, this simulated data set contains 15,548 observations. (The initial observations were replicated twenty-six times.) The estimated parameters were obtained by performing the above estimation method on the simulated data. The estimates are reasonably close to the true parameters, indicating that this estimation method does fairly well at finding the correct solution.

[39] From (1.3) we can see that, since both $\alpha_i > 0$ and $\beta_i > 0$, this will require $0 < \mu_i < 1$. From (1.5) and (1.6) we can see that in order to have both $\alpha_i > 0$ and $(\alpha_i + \beta_i) > 0$ we must have:

$$\frac{\mu_i (1 - \mu_i)}{\sigma_i} - 1 > 0$$

Solving for σ_i, we see that $\sigma_i < \mu_i (1 - \mu_i)$. From equation (1.4) it is evident that $\sigma_i > 0$.

1.6.3 Structural Results

Table 1.6 presents the results of the structural estimation. I have data on 299 shows, which results in a total of 598 observations (in every show, each of the two contestants will select a savant). The parameters $\omega_1, ..., \omega_4$ correspond to the demographic variables of the savant that compose the mean of a savant's prior skill distribution, μ_i. The parameters θ_1 and θ_2 estimate the impact of taste discrimination.

Before examining the parameters associated with the cultural group of the savant it is useful to check whether the other results seem realistic. Most importantly, we want to check the parameter γ, which will tell us how utility depends on the extremeness of a savant's skill level. The results show that γ is positive and significant, indicating that the more extreme a savant's skill level is, the more utility a contestant gets from selecting them. This is consistent with the model. The parameter σ tells us how much weight contestants put on the performance of the savant when forming their posterior perception of their skill level. A value of $\sigma = .02$ can be interpreted as follows: if a contestant's prior perception about a savant was .5, and the savant answered five out of five questions correctly in the first two rounds, their posterior perception of them would be .65. This is also consistent with what we would expect.

The parameter θ_1 is not statistically significant, indicating there is no taste-based discrimination against black savants. In contrast, the parameter θ_2 is negative and statistically significant, indicating contestants have taste discrimination against female contestants.

The parameter ω_1 is equal to .64, which corresponds to a contestant's perception of the prior mean of a non-college educated, non-black male savant's skill distribution. Recall that this corresponds to the perceived probability that that savant will get a given question correct at the beginning of the game. All the other ω's are relative to this. To find the prior perception about any other group one would add the parameter associated with that group to that of the reference group (ω_1). The parameter ω_2 is equal to .24, which indicates that contestants believe that savants who are college educated have a 24% higher chance of getting a question correct. This implies that contestants view savant performance in this game as depending on some form of learned skill, as learned skill level should be positively related to having a college education.

The parameters ω_3 and ω_4 will pick up the effects of statistical discrimination. I find that both parameters are negative and strongly significant, indicating that contestants have a lower prior perception of both black and female savants. This is consistent with the results from the descriptive analysis.

To get an idea of how large these differences in perceptions are, we can use these parameter estimates to determine what the prior perception is about each cultural group. Figure 1.3 shows how the prior perceptions about college educated savants differs among each racial and gender group. At the beginning of the game, contestants feel a non-black male savant has an 89% chance of answering a question correctly, a black male has a 62% chance, a non-black female has a 54% chance, and a black female has a 27% chance. These are relatively large differences, indicating that statistical discrimination is a very important factor in contestants' decisions. I will discuss what these results might imply about labor market hiring situations in the concluding remarks in Section 1.8.

1.7 Support for Model Assumptions

To ensure the validity of the identification strategy it is important to empirically verify the key assumptions that were made earlier. Section 1.7.1 examines whether contestants choose the most extreme savant. Section 1.7.2 examines whether contestants take into account the categories of the questions when making decisions. Finally, Section 1.7.3 tests whether contestants believe savants' past performance is predictive of their future performance, and if they are likely to be using their true beliefs about savants when making decisions.

1.7.1 Do Contestants Select the Most Extreme Savant?

The idea that contestants will select a savant based on the extremeness of their skill level is crucial to the identification strategy. To examine whether contestants actually behave in this way, we can estimate a simple discrete choice model of a contestant's selection of savant at the beginning of the third round.[40] The utility a contestant gets from selecting a savant is specified as:

$$\begin{aligned} U_i &= \lambda_1 \cdot \max(\widetilde{p_i}, 1-\widetilde{p_i}) + \lambda_2 \cdot \widetilde{p_i} + \epsilon_i \qquad (1.14) \\ \text{where } \widetilde{p_i} &= (\frac{y_i}{n_i}) \end{aligned}$$

Column 1 in Table 1.7 presents the results of this estimation when $\lambda_2 = 0$. The parameter λ_1 is positive and statistically significant, implying that a contestant's decision depends on how extreme the savant's performance is. To further make sure this is all that

[40] If contestants prefer to pick the most extreme savant in terms of skill, they will also show a preference to pick the savant with the most extreme performance. This is because a savant's perceived skill level will depend on their performance.

is important, we can allow utility to depend only on $\widetilde{p}_i$. Column 2 presents these results. We find that λ_2 is almost zero, indicating that it is the extremeness of performance that is the most important factor in contestants' decisions.

Finally, Column 3 presents the results with the full utility specification. We find that both λ_1 and λ_2 are positive and significant in this case. While we wouldn't expect λ_2 to be significantly positive, what is most important is that we find $\lambda_1 > \lambda_2$. These results indicate that if contestants were given the choice between two equally extreme savants, where one was a high performer and one was a low performer, they would prefer the high performer.

1.7.2 Category-Specific Questions

Earlier it was assumed that when contestants choose their savant at the beginning of third round they view each question that will come up as being the same. Although we do not expect them to make predictions like this when they actually hear the question, it was claimed they might behave this way because at this point in the game they do not know what the questions will be. The potential problem with this specification is that we are implicitly assuming that contestants behave as though they will not learn the question before they must make their predictions about savant answers, which is not the case.

The following example will show why the assumption made might be problematic: suppose a given savant is perceived as doing very well at one type of question, such as general knowledge, but doing relatively poorly at another type of question, such as sports. If a savant was asked questions of both types in the first two rounds, and answered all of the general knowledge questions correctly and all of the sports questions incorrectly, then their overall performance would be average. Under the model I specified, contestants would then be unlikely to choose this savant because they want the savant with the most extreme skill level. But because contestants know they will learn the question before having to make a prediction, this savant would be ideal to choose.

The above example indicates that it would be more ideal for contestants to take into account that questions come from different categories and choose the savant who is the most extreme in every category. I will now present an alternative category-specific model which coincides with this notion. This model requires that contestants have a perception of a savant's skill level at answering each of the six types of questions, as well as know the likelihood that the next question will be from a given category.[41] The utility that contestants would get from each savant would then depend on a weighted average of how extreme the savant's skill level was in each category, where the weights would represent how often each type of question came up. More formally, we could specify the utility contestants get from choosing savant i to be:

$$U_i = \sum_{c=1}^{C} (\%C) \cdot \max(\widehat{p_i^c}, 1 - \widehat{p_i^c}) + \epsilon_i$$

where c indicates the category of question, $\%C$ indicates the percent of total questions asked that are of a given category c, and $\widehat{p_i^c}$ is a contestant's perception of a savant's skill level in answering questions of category c.[42]

Although theoretically this category-specific model seems more plausible, it will be dif-

[41]See Table 1.1 for information about the six categories of questions, including their relative frequencies in the game.

[42]For simplicity, this specification ignores taste discrimination, although this could be easily added in.

ficult to estimate this many parameters. While before we had four parameters $(\omega_1, ..., \omega_4)$ that composed the mean of the prior skill distribution, we would now have four parameters for each of the six categories. With the amount of data that is present identification would be extremely difficult. Furthermore, just because we think contestants would do better to choose a savant in this way does not mean this is what they actually do. This would be a much more complex calculation on their part which would require a lot more information than the previous specification.

In order to determine which model contestants are more likely to use, we can take advantage of the following idea: if they are using the alternative category-specific model to select their savant when they don't know the question, then it is extremely likely that they will use information about category performance when they make predictions during the third round. Thus we can test whether contestant predictions in the third round depend more on the overall previous performance of the savants or the category-specific previous performance.

Table 1.8 shows the results of regressing a contestant's predictions about a savant's answer in the third round on the overall percent of questions they got correct in the first two rounds, and the category-specific percent of questions they got correct in the first two rounds. For example, if the question asked in Round 3 was a sports question, category-specific performance is the percent of sports questions the savant answered correctly in the first two rounds. If the savant had not previously faced a sports question, this observation was eliminated. The dependent variable is an indicator that takes on the value of one if contestants predict that a savant will get a given question in Round 3 correct. Results from logit estimation of this regression clearly show contestants base their predictions in the third round on the overall past performance of the savant and not the category-specific performance. The coefficient on overall percent correct is positive and strongly significant while the coefficient on category percent correct is actually negative, although it is statistically insignificant. Thus if contestants behave this way when they know the questions, it is rather likely they will also behave this way when choosing their savant without knowing the questions.

1.7.3 Manipulation of the Game

Up until now it has been implicitly assumed that the producers of *Street Smarts* do not manipulate the game. However, due to the way in which the game is put together, producers will have ample opportunity to do this. Prior to the actual in-studio game, savants are asked numerous questions. From these, producers will select which ten questions to use in the game, as well as the order in which they will be shown. This could lead to two potentially serious problems.

Producers could select questions in such a manner that savants who would be perceived as highly skilled perform relatively poorly during the game. If contestants realized this, they would have no incentive to base their decisions on their true perception of the savant's skill level. This would prevent us from identifying a contestant's true perceptions about a savant's unobservable skill. A second problem would arise if producers select the ordering of questions such that savants who do well in the first two rounds do poorly in the third. Contestants aware of this would not positively update their perception of a savant who performed well in the first two rounds. This is one of the key identifying features that allows us to distinguish between statistical and taste discrimination. If contestants do not positively update their beliefs in this manner it would invalidate the identification strategy used here.

To determine whether producers manipulate the game, I regress a savant's actual performance in the third round on their previous performance, their age group, education level, race, and gender. The results are shown in the first column of Table 1.9.[43] The dependent variable is the percent of the three questions in the third round that a savant answers correctly. If the game is not being manipulated we would expect two things. First, the coefficient on education should be positive, as this is the only characteristic about savants that is objectively related to their skill level in the game.[44] Second, the coefficient on performance in the first two rounds should also be positive. The results from an ordered logit estimation show that this is clearly not the case. A savant's education level is negatively related to their performance, although this is not statistically significant. Furthermore, past performance is strongly *negatively* related to future performance.

Producers' manipulation of the game will only result in the two problems mentioned earlier if contestants are aware the game is being manipulated. Recall that the only assumptions made were with respect to contestant behavior: contestants must believe past performance predicts future performance, and that players who are more skilled will perform better in the game. Thus these assumptions do not necessarily require that the game not

[43]The variable *young* refers to savants who are less than or equal to 35 years old. The variable *other* indicates savants who are neither white or black.

[44]We would expect college educated savants to perform better in this game because about 40% of the questions pertain to general knowledge, which they should be better at. There is also no reason to believe that they would do worse than non-college educated savants in answering the remaining 60% of the questions.

be manipulated, but only that contestants do not realize this.[45] The results shown in the second column of Table 1.9 indicate that contestants do not play the game as though it is being manipulated. The dependent variable is now a contestant's predictions of the percent of the three questions in the third round that a savant answers correctly. One can see that contestants believe that a savant's performance in the first two rounds positively predicts their performance in the third round. They also believe that more educated savants will perform better in the game. These results support both of the assumptions we have made about contestant behavior.

1.8 Conclusion

In the past it has proven quite difficult to empirically identify the role of both taste and statistical discrimination in labor market hiring decisions. This is primarily because both types of discrimination can lead to the same outcome, where the affected group is less likely to be hired. This chapter attempts to distinguish between taste and statistical discrimination by finding an alternative situation, contestant behavior on the game show *Street Smarts*, where this problem doesn't arise. This game has two unique features that distinguish it from other settings: contestants will make decisions based on how extreme a savant's skill level is, and contestants will update their prior perceptions about savants with information they receive during the game. As was shown, the combination of these two features will lead to taste and statistical discrimination having different effects on contestant behavior, enabling us to separately identify each of their roles.

[45] If the game was manipulated and contestants failed to realize this, it would only mean contestants do a poor job of predicting savants' answers. This is not important, because we do not care about how accurate their perceptions of savants are–we only care about identifying their perceptions.

This identification strategy was used to test for statistical discrimination both descriptively and in a formal structural model. Both analyses show that statistical discrimination plays a strong role in contestants' behavior. In particular, I find that contestants have a significantly lower prior perception of the skill level of both black and female savants. The structural estimation results also indicated that contestants have a distaste towards interacting with female savants.

The results concerning statistical discrimination will not directly reflect what is happening in labor market situations, as the skills that are required in each situation will be quite different. However, we can still garner importance from these results. In particular, they show that statistical discrimination plays a strong role in skill assessment in this game. Because hiring situations also rely on employers making skill assessments based on limited information, it is likely that statistical discrimination will be a factor there, too. This implies that policies that require employers to gather more objective information about applicants, which would reduce their reliance on cultural background, might be effective in reducing hiring disparities.

Finally, the methodology used in this chapter can provide insight about what types of situations in the labor market could potentially be exploited in the future to test for both statistical and taste discrimination. The unique feature about this game show is that contestants make decisions based on how extreme a savant's skill level is. While it is unlikely that we would find this feature in a labor market setting, there are situations where we might expect to find the opposite feature. Namely, there are certain occupations in which we might expect employers to hire medium-skilled applicants (as opposed to the most highly-skilled applicants), because they feel high-skilled workers will leave if a better job opportunity comes along. The same identification arguments that were presented here would be valid in that setting also. Situations like this could potentially be exploited in future research.

1.9 Appendix: Tables and Figures

Table 1.1: Examples of Questions Listed by Category

Category	Examples of Questions
general knowledge	What state was founded by William Penn?
	What letter is silent in the word autumn?
	In what U.S. city would you find the gateway arch?
entertainment	What is James Bond's numeric code name?
	The show *Nip/Tuck* is about people in what profession?
	Rapper Marshall Mathers is better known as whom?
sports	What sport do the LA Clippers play?
	In baseball, what is a grand slam?
	Tony Hawke is a legend in what sport?
slang/sayings	In slang, what happened if you blew a fuse?
	Complete the saying: looks can be deceiving?
	In slang, what does it mean if someone's got your back?
childrens' interests	In the nursery rhyme, what is the London Bridge doing?
	Grranimals is a popular line of kids what?
	On Sesame Street, what color is Grover's fur?
miscellaneous	What color is pesto sauce?
	What store's motto is "always low prices, always"?
	What tasty herb is believed to ward off vampires?
	What university is nicknamed the Fighting Irish?

Table 1.2: Demographic Characteristics of Contestants and Savants

		Contestants	Savants
race	white	62.7%	59.4%
	black	22.4%	24.1%
	Hispanic	6.5%	7.9%
	Asian	6.4%	5.7%
	other/unknown	2.0%	2.9%
gender	male	48.2%	44.9%
	female	51.8%	55.1%
education	non-college	75.9%	76.6%
	college	24.1%	23.4%
age	≤ 35	89.1%	75.3%
	> 35	10.9%	24.7%

NOTE: The sample consists of 598 contestants and 897 savants.

Table 1.3: Detailed Demographic Characteristics of Contestants and Savants

		Contestants		Savants	
		male	female	male	female
race	white	49.1%	50.9%	38.8%	61.2%
	black	41.0%	59.0%	53.7%	46.3%
	Hispanic	69.2%	30.8%	60.6%	39.4%
	Asian	47.4%	52.6%	43.1%	56.9%
	other/unknown	33.3%	66.7%	57.7%	42.3%
		non-college	college	non-college	college
race	white	77.6%	22.4%	76.9%	23.1%
	black	73.1%	26.9%	75.9%	24.1%
	Hispanic	76.9%	23.1%	77.5%	22.5%
	Asian	65.8%	34.2%	76.5%	23.5%
	other/unknown	89.3%	16.7%	73.1%	26.9%
gender	male	72.6%	27.4%	78.9%	21.1%
	female	79.0%	21.0%	74.7%	25.3%

Table 1.4: Relationship Between Previous Performance and Selection in the Third Round

	proportion of savants selected in the third round			
proportion correct	black	non-black	male	female
0-.2	.458 (107)	.368 (302)	.389 (203)	.393 (206)
.4-.6	.455 (220)	.385 (801)	.429 (436)	.378 (585)
.8-1	.368 (19)	.500 (46)	.482 (27)	.447 (38)

NOTE: Number of observations in parentheses.

Table 1.5: Monte Carlo Simulation

	specified parameters	estimated parameters
ω_1 (*constant*)	0.55	*0.5615 (.006)
ω_2 ($COLLEGE_i$)	0.2	*0.1959 (.013)
ω_3 ($BLACK_i$)	-0.2	*-0.199 (.013)
ω_4 ($FEMALE_i$)	-0.2	*-0.2104 (.013)
θ_1 ($BLACK_i$)	-0.5	*-0.4376 (.047)
θ_2 ($FEMALE_i$)	-0.5	*-0.5023 (.029)
γ	6	*5.9289 (.226)
σ	0.025	*0.0258 (.001)
log-likelihood	-13,245	
observations	15,548	

NOTE: Standard errors are in parentheses.

* denotes a parameter significant at the 10% level.

Table 1.6: Structural Estimation Results

	parameters
$\omega_1\,(constant)$	*0.6448 (.079)
$\omega_2\,(COLLEGE_i)$	*0.2447 (.114)
$\omega_3\,(BLACK_i)$	*-0.2702 (.092)
$\omega_4\,(FEMALE_i)$	*-0.3446 (.093)
$\theta_1\,(BLACK_i)$	0.1671 (.153)
$\theta_2\,(FEMALE_i)$	*-0.2208 (.134)
γ	*2.0143 (.854)
σ	*0.0204 (.009)
log-likelihood	-529.1
Observations	598

NOTE: Standard errors are in parentheses.

* denotes a parameter significant at the 10% level.

Table 1.7: The Effect of Previous Performance on a Contestant's Choice of Savant

	(1)	(2)	(3)
λ_1	*.5376 (.251)	—	*1.0244 (.258)
λ_2	—	.0239 (.0539)	*.3627 (.19)
observations	598	598	598

NOTE: Standard errors are in parentheses.

* denotes a parameter significant at the 10% level.

Table 1.8: The Effect of Savants' Average and Category-Specific Performance on Contestants' Predictions in Round 3

	questions predicted correct in third round
overall percent correct	*1.22 (.371)
category percent correct	-0.118 (.163)
constant	*-.468 (.146)
observations	1242
Pseudo-R^2	.0067

NOTE: Standard errors are in parentheses.

* denotes a parameter significant at the 10% level.

Table 1.9: The Relationship Between Savant Characteristics and their Actual and Predicted Performance in Round 3

	(1) percent correct in third round	(2) percent predicted correct in third round
percent correct in rounds 1 & 2	*-2.84 (.498)	*1.80 (.448)
$young_i$	-.290 (.195)	.153 (.184)
$college_i$	-.064 (.194)	*.290 (.183)
$other_i$	-.065 (.232)	-.076 (.216)
$black_i$	.012 (.195)	*-.312 (.184)
$female_i$	-.176 (.164)	-.224 (.155)
observations	598	598
Pseudo-R^2	.0348	.0186

NOTE: Standard errors are in parentheses.

* denotes a parameter significant at the 10% level.

Figure 1.1: Relation Between Selection and Performance for Black vs. non-Black Savants

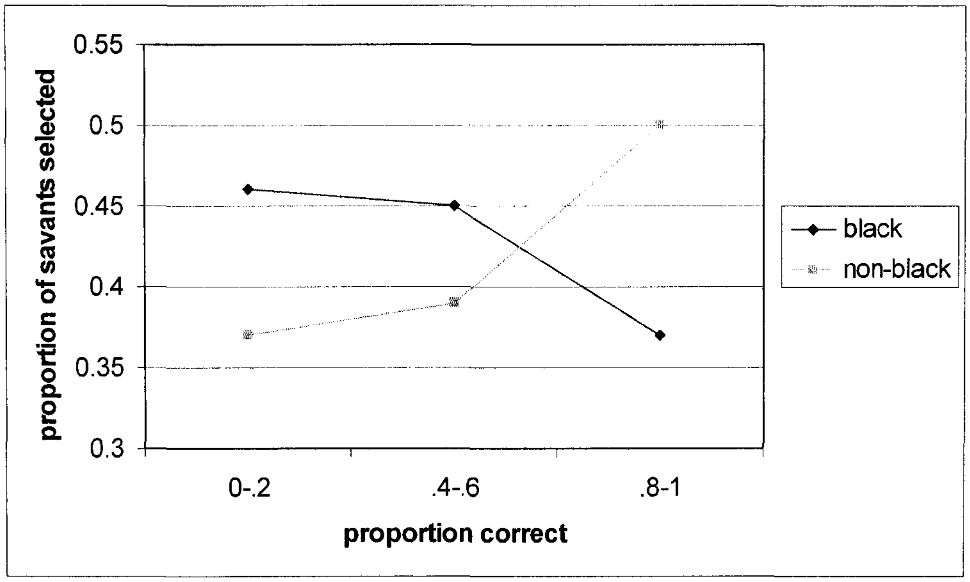


Figure 1.2: Relation Between Selection and Performance for Female vs. Male Savants

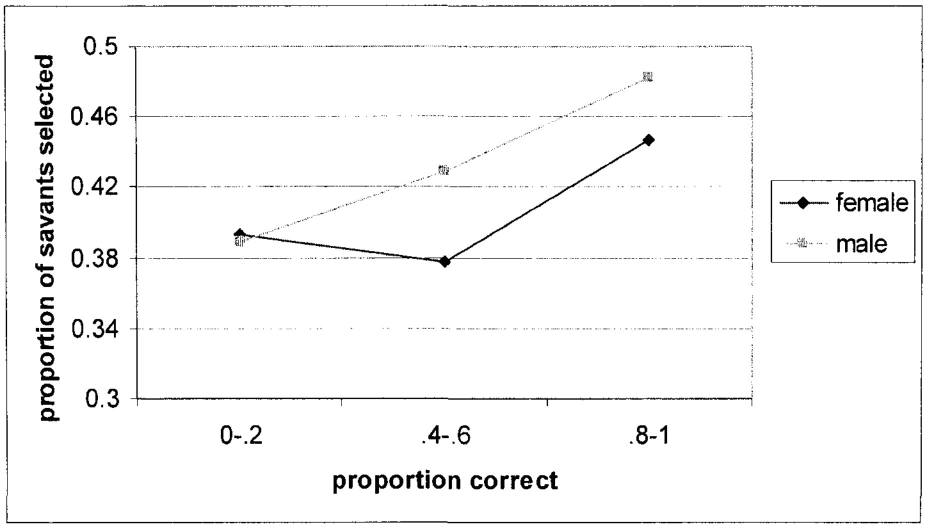
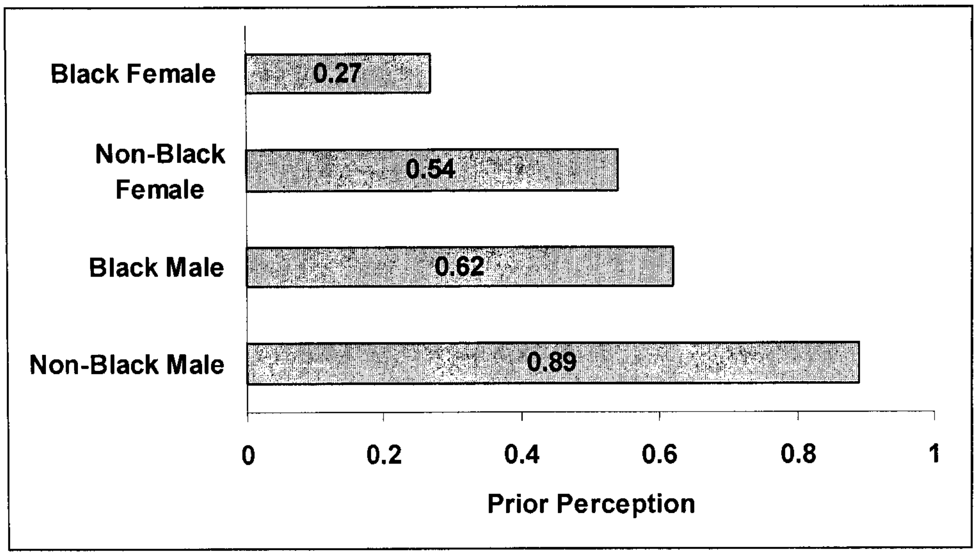


Figure 1.3: Prior Perception of College Educated Savants from Different Cultural Groups



Chapter 2

An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence (co-authored with Hanming Fang)

2.1 Introduction

Black motorists in the United States are much more likely than white motorists to be searched by highway troopers. Several recent lawsuits against state governments have used this racial disparity in treatment as evidence of "racial profiling," a term that refers to the police practice of using a motorist's race as one of the criteria in their motor vehicle search decisions. Racial profiling originated with the attempt to interdict the flow of drugs from Miami up Interstate 95 to the cities of the Northeast. For example, in 1985 the Florida Department of Highway Safety and Motor Vehicles issued guidelines for police on

"The Common Characteristics of Drug Couriers," in which race/ethnicity was explicitly mentioned as one characteristic (Engel et al., 2002). While the initial motivation for such guidelines may have been to increase the troopers' effectiveness in interdicting drugs, it also unfortunately opened up the possibility for troopers to engage in racist practices against minority motorists.

Following the public backlash generated by several cases in the 1990s such as *Wilkins v. Maryland State Police* [1996] and *Chavez v. Illinois State Police* [1999], almost all highway patrol departments have denounced using race as a criterion in stop and search decisions. But many citizens, especially minorities, are skeptical of this claim: motor vehicle search decisions, by their very nature, are made in the midst of face-to-face interactions, and thus it is simply hard to imagine that troopers can block the race and ethnicity information that a motorist presents. Moreover, data on trooper searches continue to show that they tend to search a higher proportion of minority motorists than white motorists. As is now well known, however, racial disparities in the aggregate rates of stops and searches do not necessarily imply *racial prejudice* (see, for example, Knowles et al., 2001, and Engel et al., 2002). If, for example, black drivers are more likely than white drivers to carry contraband, then the aggregate rate of stops and searches would be higher for black drivers even when race was not a factor in troopers' decision-making. Moreover, racial profiling may also arise if police attempt to maximize successful searches and race helps predict whether a driver carries contraband. This situation is called *statistical discrimination* in the terminology of Kenneth Arrow (1973).

How can we empirically distinguish racism from statistical discrimination? This question has garnered enormous public and academic interest (see, for example, National Research Council 2004), but it is also challenging, partly as a result of data limitations. For example, unless truly random searches are conducted, researchers typically will not observe the *true* proportion of drivers who carry contraband. Furthermore, ethnographic studies such as Sherman (1980) and Riksheim and Chermak (1993) have shown that many situational factors, including suspects' demeanor in the police-citizen encounter, influence police behavior. Such data are also typically unavailable. Because we have no way of controlling for all of the legitimate factors that might cause minority drivers to be searched with higher probability than white drivers, it becomes very difficult to determine the true motivation behind racial profiling with the available data.

One prominent approach that has been used to distinguish between racial prejudice and statistical discrimination is the "outcome test," whose idea originated in Becker (1957).[1] In the context of motor vehicle searches, the outcome test is based on the following intuitive notion: if troopers are profiling minority motorists due to racial prejudice, they will search minorities even when the returns from searching them, i.e., the probabilities of successful searches against minorities, are smaller than those from searching whites. More precisely, if racial prejudice is the reason for racial profiling, then the success rate against the *marginal* minority motorist (i.e., the last minority motorist deemed suspicious enough to be searched) will be lower than the success rate against the *marginal* white motorist. In contrast, if racial profiling results from statistical discrimination (i.e., if the troopers are profiling to maximize the number of successful searches), then the optimality condition would require

[1] Becker (1993a, 1993b) further elaborated on this idea and Ian Ayres (2001) presented several interesting applications.

that the search success rate against the marginal minority motorist be equal to that against the marginal white motorist. While this idea has been well understood, it is problematic in empirical applications because researchers will never be able to directly observe search success rates against marginal motorists. This is due to the fact that we cannot identify the marginal motorist, since accomplishing this would require having complete information on all of the variables that troopers use in determining the suspicion level of motorists. Because of this *omitted variables* problem, we can only observe the average success rate of searches against white and minority motorists, and not the marginal success rate. Since the equality of marginal search success rates does not imply and is not implied by the equality of the average search success rates, we cannot determine the relationship between the marginal search success rates of white and minority motorists by looking at average success rates. In past literature this has been referred to as the *infra-marginality* problem. These problems severely limit the rigorous application of the outcome test idea, especially in situations where the decision or the outcome is dichotomous.[2]

A seminal paper by Knowles et al. (2001, KPT hereafter) provides the first solution to the infra-marginality problem associated with the outcome test. They develop a simple but elegant theoretical model about motorist and police behavior and show that in equilibrium the infra-marginality problem may not arise. In their model, motorists differ in their characteristics, including race and possibly other factors that are observable to troopers but may or may not be available to researchers. Troopers decide whether or not to search

[2]See Stephen L. Ross and John Yinger (1999 and 2002, Chapter 8) and Ayres (2002) for detailed discussions of the infra-marginality problem in the context of mortgage lending and police practices respectively. In fact, in the case of mortgage lending, Ross (2003) and Ross and Yinger (2002, Chapter 8) argue that the complete elimination of the omitted variable bias results in a test with no power.

motorists while motorists decide whether or not to carry contraband. In this "matching pennies"-like model they show that if troopers are not racially prejudiced, all motorists, if they are searched at all, must in equilibrium carry contraband with equal probability regardless of their race and other characteristics. Thus in their model there is *no difference* between the marginal and the average search success rates. A nice feature of the KPT model is that it allows the motorists of different races to have different distributions of characteristics, as long as those characteristics are observable to the police (though they may not be observable to the researcher). Motorists with different characteristics may have different costs and benefits from carrying contraband, but these differences only imply that in equilibrium troopers will search motorists with different characteristics at different rates, which in fact provides the necessary deterrence to ensure that all motorists will carry contraband with equal probabilities. Because the infra-marginality problem does not arise at all in the equilibrium of KPT's simple model, they provide a solid theoretical basis for an empirical test based on the comparison of the *average* search success rates by the race of the *motorists*, a statistic typically available to researchers. A lower average search success rate implies racial prejudice against that group. Applying their test to a data set of 1,590 searches on a stretch of the I-95 in Maryland from January 1995 through January 1999, they find no evidence of racial prejudice against African-American motorists, but do find evidence of racial prejudice against Hispanic motorists.

While KPT's model provides a good starting point to distinguish between racial prejudice and statistical discrimination empirically, there are a couple of drawbacks to their theoretical model which cast doubt on the validity of their empirical test. First, KPT's model predicts that all motorists of a given race, if they are ever searched, will carry contraband with equal probability regardless of their other characteristics that may be observed by the police. This is the vital prediction that allows them to equate the average search success rate in a given racial group of motorists to the marginal search success rate, thus avoiding the infra-marginality problem. This prediction, however, also implies that a motorist's characteristics other than race should provide no information about the presence of contraband when a trooper decides whether to search. This implication of police behavior goes against trooper guidelines which require them to base their search decisions on the information the motorist presents to the trooper at the time of the stop, including the motorist's personal characteristics, their demeanor, and the contents of their vehicle that are in plain view, etc. (see, e.g., Sherman 1980 and Riksheim and Chermak 1993). KPT's basic model assumes that motorists' characteristics are exogenous, thus ruling out the plausible scenario that a motorist's actions when stopped are intimately related to whether or not he or she is carrying contraband. This is not just a minor quibble about details: once we allow the motorists' actions when stopped to enter into the officers' search decisions, the infra-marginality problem reappears into the empirical analysis. Our main contribution to the racial profiling literature in this chapter is that we develop a more realistic model of trooper behavior that allows officers to use information that they gather about motorists during traffic stops when they make their search decisions. We exploit the theoretical implications of our model to propose an alternative empirical test to detect racial prejudice *in the presence of* potential infra-marginality and omitted-variables problems associated with outcome tests.

The second issue we have concerning KPT (and this field of research in general) is that they implicitly assume that all troopers' behavior is *monolithic.* This assumption may not be valid.[3] Most existing data sets on police behavior do not contain detailed information about the trooper characteristics and thus it is assumed that all troopers, regardless of their race, have the same racial prejudice against minority motorists.[4] Donohue and Levitt (2001), in their study on arrest patterns and crime, find that the racial composition of a city's police force has an important impact on the racial patterns of arrests, suggesting that police behavior (or information they possess) is not monolithic. Within the framework of KPT, an invalid monolithic trooper behavior assumption can lead to wrong conclusion about whether officers are racially prejudiced. Imagine a world in which minority troopers are racially prejudiced against white motorists, while white troopers are prejudiced against minority motorists. It is possible that when examining the aggregate search outcomes of white and minority troopers, we would reach a conclusion that the police as a whole are not racially prejudiced. But this may seriously underestimate the harassment experienced by both white and minority motorists. This chapter deviates from the KPT model and embraces the possibility that police behavior may vary by their racial group, which is our second main contribution to the existing literature. As will be shown later, the variation of trooper behavior by their race will provide the key additional information that allows us to develop our empirical test.[5] We are able to relax the typical assumption that requires

[3]A formal definition of monolithic behavior is given in Section 2.3.

[4]The Maryland data set KPT used has only very limited information about troopers (see KPT 2001 and Samuel R. Gross and Katherine Y. Barnes 2002)

[5]We are grateful to an anonymous referee for clarifying this important point.

troopers to behave monolithically because we have a *unique* data set of highway stops and searches conducted by the Florida Highway Patrol that contains information on both the race of the trooper making the stop as well as the information about the motorist that is stopped.

The information we present in Table 2.1 further illustrates why it is unrealistic to impose the assumption that all troopers behave monolithically in our Florida data set.[6, 7] Panels A and B, respectively, show the search rate given stop and the average search success rate against motorists for the different combinations of motorist and trooper racial groups. The first row of Panel A shows that, of the white motorists stopped by white, black and Hispanic troopers, respectively 0.96, 0.27 and 0.76 percent of them were searched. Monolithic search behavior requires all troopers to search white motorists at the same rate, which clearly is not the case. Thus any empirical test that relies on this assumption will not be valid for our data set. The "All Troopers" column in Panel B of Table 2.1 also contains the information used in KPT's test.[8] Because the average search success rates against both black and Hispanic motorists fall below that against white motorists, KPT would conclude that troopers have racial prejudice towards black and Hispanic motorists. But, when we admit the possibility that the unobservable characteristics among motorists of different races may differ (in a possibly arbitrary way), we will argue that even such strong disparities in search rates and average search success rate may not prove racial prejudice.

[6]Table 2.1 is our main table and we will discuss it more in Section 2.5. The formal test of monolithic behavior is presented in Section 2.5.4.

[7]The numbers in the column labeled "All Troopers" are calculated directly from the raw data, but the numbers in the columns labeled "Trooper Race" are calculated from reweighted samples constructed from the raw data. See Section 2.4.3 for details about how we construct the reweighted samples.

[8]More discussion of the KPT test on this data set is provided in Section 2.5.7.

Our model of trooper search behavior follows the spirit of labor market statistical discrimination models (see, e.g., Coate and Loury 1993). Police officers observe noisy but informative signals about whether or not a driver carries contraband when they decide if a search is warranted. Guilty drivers, i.e., drivers who actually carry contraband, are more likely than innocent drivers to generate suspicious signals. A police officer incurs a cost of search $t(r_m; r_p)$ that depends on both his/her own race r_p and the race of the motorist r_m. Troopers of a particular race, say r_p, are said to be racially prejudiced if their cost of searching motorists depends on the race of the motorist.[9] The police force exhibits non-monolithic behavior if the cost of searching motorists of a given race r_m depend on the race of the trooper. Troopers are assumed to make their search decisions to maximize the number of successful searches (or arrests). The optimal decision of a race-r_p police officer in deciding whether a race-r_m motorist should be searched satisfies a threshold property: motorists should be searched if and only if their posterior probability of being guilty exceeds the search cost of race-r_p officers against race-r_m motorists, $t(r_m; r_p)$. We show that the police officers exhibit monolithic behavior if and only if both the search rate and average search success rate against any given race of motorists are independent of the race of the troopers conducting the search. Moreover, if none of the racial groups of troopers are racially prejudiced, then the ranking *over the race of troopers* of search rates and average search success rates against a given race of motorists should not depend on the race of the motorists. That is, if troopers of race r_p have a higher search rate (and lower average search success rate) against race-r_m motorists than troopers of race r_p', then race-r_p troop-

[9]We assume that race is the only characteristic of troopers that is likely to affect their search behavior. This is a plausible assumption because we are examining if troopers search white and minority motorists differently, so the race of the trooper is the most likely characteristic to affect their search patterns.

ers should also have a higher search rate (and lower average search success rate) against race-r'_m motorists than race-r'_p troopers. We use these theoretical predictions of the model to design empirical tests for both monolithic behavior and racial prejudice. The key idea of our empirical test is as follows: when there is no racial prejudice the race of motorists should not affect the *ranking* of search rates and search success rates *over officer races.*

An additional desirable feature of our model is that it has direct implications on the ranking of both the search rates and the average search success rates, and thus our model could potentially be refuted by the data we have available. It is also important to point out, though, that our test can only detect what we term to be *relative* racial prejudice and not absolute racial prejudice. This is because when the ranking of search rates and search success rates over officer races depends on the race of the motorists, we know that at least one of the racial groups of officers is using racial prejudice, but we cannot identify which group it is. Thus all we can conclude is that one group of troopers is more racially prejudiced relative to another group of troopers, instead of an absolute conclusion which would identify which groups of troopers were racially prejudiced.[10]

The implementation of our empirical tests relies on data sets that have race information on both troopers and motorists.[11] While such data has not been available for use in earlier empirical studies on racial profiling, we are able to obtain a data set from the Florida Highway Patrol which contains information on all vehicle stops and searches conducted on Florida highways between January 2000 and November 2001, together with the demograph-

[10] More discussions on this issue is provided in Section 2.4.2 when we discuss the power of our test.

[11] While our tests can in principle be implemented with only search data (by looking only at average search success rates), having data on all stops would be more desirable because we can then examine whether our model is refuted by the data by providing supporting evidence from the search rates.

ics of the trooper that conducted each stop and search. In implementing our empirical tests, we find strong evidence that the Florida Highway Patrol troopers do not exhibit monolithic behavior, but we fail to reject the hypothesis that troopers of different races do not exhibit relative racial prejudice.

The remainder of the chapter is structured as follows. Section 2.2 provides some additional discussion of the related literature. Section 2.3 presents and analyzes our model of trooper search behavior. Section 2.4 proposes empirical tests based on the theoretical predictions of the model. Section 2.5 describes the data set from the Florida Highway Patrol, presents our test results, and contrasts our results with those using KPT's test. Finally, Section 2.6 concludes. In Appendix A (located in Section 2.7) we present a simple equilibrium model of drug carrying behavior to show that our focus on trooper behavior in Section 2.3 is not problematic. Tables and figures are included in Appendix B in Section 2.8.

2.2 Related Literature

Dharmapala and Ross (2004) and Antonovics and Knight (2004) also discussed the possible shortcomings of the KPT model.[12] Dharmapala and Ross (2004) point out that KPT's test does not generalize if potential drug carriers may not be *observed* by the police or if there are different levels of drug offense severity.[13] Under those circumstances KPT's test fails because the infra-marginality and omitted variables problems re-emerge. More

[12] Hernandez-Murillo and Knowles (2004) use KPT's framework and semi-parametric bounds to reject the official explanation that lower hit rates on minorities are due to higher rates of non-discretionary search using Missouri's annual aggregate traffic-stop report for the year 2001. Dominitz and Knowles (2004) consider tests of racial prejudice when officers are assumed to minimize crime.

[13] KPT recognized this issue in their footnote 16.

specifically, the equilibrium of the KPT model under those circumstances may involve a group of motorists carrying drugs with probability one (being a "dealer") even when they are searched with probability one whenever the troopers observe them. If the probability of being a "dealer" is higher for minorities, then the average success rate against minorities should be greater than that for whites under statistical discrimination, and equal average success rates would actually indicate taste discrimination, contrary to KPT's conclusion.

Antonovics and Knight (2004) argued that KPT's test may not be robust when its model is generalized to allow for trooper heterogeneity.[14] They also proposed using data with both motorist and officer information. As we do in our paper, they show that if officers of different races have the same search cost against motorists of a given race, then the search rate against these motorists should be independent of the officers' race. They run a Probit regression using data from the Boston Police Department where the dependent variable is an indicator for whether a search took place for a given stop, and the explanatory variables include some observable characteristics of the driver and officer and a dummy variable indicating whether there is a racial mismatch between the officer and the driver. In their baseline regression, they find a positive coefficient on the "racial mismatch" variable, indicating that officers are more likely to conduct a search against motorists of races different from their own. They interpret this finding as evidence of racial prejudice. We argue in Section 2.4.2 that their interpretation of the evidence may be misleading. It is also useful to point out that their data is from the Boston Police Department and consists mainly of stops and searches in *local* neighborhoods. There are two potential problems with such data. First, as Hernandez-

[14]However, Persico and Todd (2004) show that, if officers' goal is to maximize search success rates rather than total number of successes, KPT's test can be generalized to allow for police heterogeneity.

Murillo and Knowles (2004) argued, many stops and searches conducted in local streets are in response to specific crime reports. In these situations, officers tend to have less discretion over who they search. Second, as argued by Donohue and Levitt (2001), for stops and searches conducted in local neighborhoods, it is much more likely that officers of different races may possess different amounts of information about motorists, as residents in the neighborhood may be more willing to share information with officers with the same race as them. In contrast, our data consists only of stops and searches conducted on highways, and as a result the above two issues are less concerning.

We would also like to point out that, besides the "outcome test" approach, a large field of literature has used a different statistical test, known as the "benchmarking test," to test whether troopers impose disparate treatment on motorists of different races.[15, 16] The benchmarking test typically compares the shares of racial or ethnic minorities in the population to their shares in the sample of motorists selected for discretionary stops and searches by police. The main drawback of the benchmark test is that it cannot determine if racial disparities arise out of racial prejudice or statistical discrimination. Furthermore, the benchmark test suffers from two main problems. The first problem is called the *denominator* problem, which refers to the question of what should be the right benchmark to compare the stop and search rates. It ideally should be the racial or ethnic composition of drivers on the road, but such information is typically unavailable. The second problem is the *omitted-*

[15] A refined version of this test uses regressions to estimate the probability of being searched as a function of race and other observable characteristics that may be related to propensity to commit crimes. Fridell (2004) provides a comprehensive review of different benchmarks in this approach.

[16] There are parallel and closely related approaches to test for disparate treatment in the literature on mortgage lending discrimination (see Ross and Yinger, 2002, and Ross, 2003, for comprehensive literature reviews). Paired-audit is a third frequently used method in the context of housing market, mortgage lending and car purchases (Ayres, 2001).

variables problem. If there exist certain characteristics whose distributions are correlated with motorists' race or ethnicity and if such characteristics may be observed by police but not available to researchers, benchmarking tests will not be completely informative about whether motorists' race affected the search decision.

2.3 The Model

We now present a simple model of trooper search behavior that underlines the empirical work in Section 2.5.[17] There is a continuum of troopers (interchangeably, police officers) and motorists (interchangeably, drivers). Let r_m and $r_p \in \{M, W\}$ denote the race of the motorists and the troopers respectively, where M stands for minorities and W for whites.[18] Suppose that among motorists of race $r_m \in \{M, W\}$, a fraction $\pi^{r_m} \in (0, 1)$ of them carry contraband.[19]

The information that is available to an officer when he or she makes the search decision consists of the motorist's race and many other characteristics pertaining to the motorist. Such characteristics may include, for example, the gender, age and residential address of the driver, the interior of the vehicle that is in the trooper's view, the smell from the driver or the vehicle, whether the driver is intoxicated, the demeanor of the driver in answering the trooper's questions, the make of the car, whether the car has an out-of-state plate, whether the car is rented or owned, location and time of the stop, as well as the seriousness of the

[17] Borooah (2001) and Bjerk (2004) develop somewhat related models of policing behavior.

[18] In the empirical part of the paper, we will examine three racial or ethnic groups: whites, blacks, and Hispanics. For now, though, we group blacks and Hispanics together as minorities for ease of exposition.

[19] For the purpose of deriving our empirical test, we will assume that π^{r_m} is exogenous. In Appendix A, we present an equilibrium model in which π^{r_m} is endogenously determined.

reason for the stop, etc.[20] Note that while the police officer observes all the characteristics in the decision to search, a researcher will typically have access to only a small subset of them. We assume, however, that the police officer will use a *single-dimensional index* $\theta \in [0,1]$ that summarizes all of the information that these characteristics indicate about the likelihood that a driver may be carrying contraband.[21] We assume that, if a driver of race $r_m \in \{M, W\}$ actually carries contraband, then the index θ is randomly drawn from a continuous probability density distribution $f_g^{r_m}(\cdot)$; if a race r_m driver does not carry contraband, θ would be randomly drawn from $f_n^{r_m}(\cdot)$. [The subscripts g and n stand for "guilty" and "not guilty," respectively.] Without loss of generality, we can assume that the two densities $f_g^{r_m}$ and $f_n^{r_m}$ satisfy the strict monotone likelihood ratio property (MLRP), i.e., for $r_m \in \{M, W\}$,

MLRP: $f_g^{r_m}(\theta) / f_n^{r_m}(\theta)$ is strictly increasing in θ.

The MLRP property on the signal distributions essentially means that a higher index θ is a signal that a driver is more likely to be guilty.[22] To the extent that there may be obviously guilty drivers (for example, if illicit drugs are in plain view), we assume that:

Unbounded Likelihood Ratio: $f_g^{r_m}(\theta) / f_n^{r_m}(\theta) \to +\infty$ as $\theta \to 1$.

[20] The questions the trooper will ask the motorist are typically focused on where the motorist is headed and the purpose of their visit. In listening to the response the trooper will try to discern how nervous or defensive the motorist is, and how logical the motorist's response is.

[21] It is useful to think that troopers aggregate their observed variables into the index θ by assigning them different weights. The weights troopers assign to a particular variable, however, can be different for motorists of different races. For example, Hispanic motorists in Florida tend to have more limited English skills than Whites. Thus the weights on English skills in the formation of θ should differ for Hispanic and White drivers.

[22] For any one dimensional index θ, we can always reorder them according to their likelihood ratio $f_g^{r_m}(\theta) / f_n^{r_m}(\theta)$ in an ascending order. Thus the MLRP assumption is with no loss of generality.

The MLRP also implies that the cumulative distribution function $F_g^{r_m}(\cdot)$ first order stochastically dominates $F_n^{r_m}(\cdot)$, which implies that drivers who actually carry contraband are more likely to generate higher and thus more suspicious signals. We think this single dimensional index formulation summarizes the information that is available to troopers when they make their search decisions on the highway in a simple but realistic manner.

Each police officer can choose to search a vehicle after observing the driver's vector (r_m, θ), where r_m is the driver's race and θ is the single-dimensional index that summarizes all other characteristics observed during the stop. We assume that a trooper wants to maximize the total number of convictions (or the number of drivers found carrying illicit contraband) minus a cost of searching cars. This is an important assumption because it requires that police officers always use any statistical information contained in the race of the motorist in their search decisions.[23]

Let $t(r_m; r_p)$ be the cost of a police officer with race r_p searching a motorist with race r_m, where $r_p, r_m \in \{M, W\}$. We normalize the benefit of each arrest (or successful drug find) to equal one, and scale the search cost to be a fraction of the benefit, so that $t(r_m; r_p) \in (0, 1)$ for all r_m, r_p. It is worth emphasizing that, different from KPT, we allow the troopers' cost of searching a vehicle to depend on the races of both the motorist and the officer, and thus we can directly confront the possibility that police officers may not be monolithic in their search behavior.

We now introduce some definitions. First, a police officer of race r_p is said to be *racially*

[23]This is also the police objective postulated in KPT. It is a plausible assumption because awards (such as Trooper of the Month honors) and/or promotion decisions are partly based on troopers' success in catching motorists with contraband. This assumption rules out the possibility that some officers ignore the race of a motorist even when it provides useful information. See Section 2.3.2 for more discussion on this key assumption.

prejudiced if he or she exhibits a preference for searching motorists of one race. Following KPT, we model this preference in the cost of searching motorists.[24, 25]

Definition 2.1. *A police officer of race r_p is racially prejudiced, or has a taste for discrimination, if* $t(M; r_p) \neq t(W; r_p)$.

Next, we say that police do not exhibit *monolithic behavior* if officers of different races do not use the same search criterion when dealing with motorists of some race.

Definition 2.2. *The police officers do not exhibit monolithic behavior if* $t(r_m; M) \neq t(r_m; W)$ *for some* $r_m \in \{M, W\}$.

Note that a monolithic police force does not mean that they are not racially prejudiced: it could be that police officers of both races are equally prejudiced against some race of motorists. Likewise, a non-monolithic police force does not necessarily imply that some racial group of troopers are racially prejudiced: it could be that each group of troopers has the same search cost against all groups of motorists, but that search costs depend on the race of the trooper.

2.3.1 Theoretical Implications

Let G denote the event that the motorist searched is found with illicit drugs in the vehicle. When a police officer observes a motorist of race r_m and signal θ, the posterior probability

[24] Strictly speaking, we should have a broad interpretation of the search cost $t(r_m; r_p)$. For example, the cost of decoding the demeanor may be smaller if $r_m = r_p$. We are not able to distinguish such cost differences from racial prejudice.

[25] We are interpreting racial prejudice as "consequential animus" in the terminology of Ayres (2001, Chapter 3). In other contexts such as mortgage lending, racial prejudice may be manifested as "association animus," i.e., a lender may be prejudiced against borrowers of a given race by not being willing to engage in transactions with them. We believe that "consequential animus" is an appropriate interpretation of racial prejudice in motor vehicle searches. We thank an anonymous referee for bringing this distinction to our attention.

that such a motorist may be guilty of carrying contraband, $\Pr(G|r_m,\theta)$, is obtained via Bayes' rule:

$$\Pr(G|r_m,\theta) = \frac{\pi^{r_m} f_g^{r_m}(\theta)}{\pi^{r_m} f_g^{r_m}(\theta) + (1-\pi^{r_m}) f_n^{r_m}(\theta)}.$$

It immediately follows from the MLRP that $\Pr(G|r_m,\theta)$ is monotonically increasing in θ. From the unbounded likelihood ratio assumption, we know that $\Pr(G|r_m,\theta) \to 1$ as $\theta \to 1$.

The decision problem faced by a police officer of race r_p when facing a motorist with race r_m and signal θ is thus as follows:

$$\max\{\Pr(G|r_m,\theta) - t(r_m;r_p);0\} \tag{2.1}$$

where the first term is the expected benefit from searching such a motorist and the second term is the benefit from not searching, which is normalized to zero. Thus the optimal decision for a trooper of race r_p is to search a race-r_m motorist with signal θ if and only if

$$\Pr(G|r_m,\theta) \geq t(r_m;r_p).$$

From the monotonicity of $\Pr(G|r_m,\theta)$ in θ, we thus conclude:

Proposition 2.1. *A race-r_p police officer will search a race-r_m motorist if and only if*

$$\theta \geq \theta^*(r_m;r_p)$$

where $\theta^*(r_m;r_p)$ is uniquely determined by

$$\Pr(G|r_m,\theta^*(r_m;r_p)) = t(r_m;r_p).$$

Moreover, the search threshold $\theta^*(r_m; r_p)$ *is monotonically increasing in* $t(r_m; r_p)$.

Proposition 2.1 says that the probability of a successful search for the *marginal* motorist is equal to the cost of search. Any infra-marginal motorist will have a higher search success probability. In what follows, we will refer to $\theta^*(r_m; r_p)$ as the *equilibrium search criterion* of race-r_p police officers against race-r_m motorists. We define the *equilibrium search rate* of race-r_p police officers against race-r_m motorists as $\gamma(r_m; r_p)$, which is given by

$$\gamma(r_m; r_p) = \pi^{r_m}\left[1 - F_g^{r_m}(\theta^*(r_m; r_p))\right] + (1 - \pi^{r_m})\left[1 - F_n^{r_m}(\theta^*(r_m; r_p))\right]. \tag{2.2}$$

The *equilibrium average search success rate* of race-r_p police officers against race-r_m motorists, denoted by $S(r_m; r_p)$, is given by

$$S(r_m; r_p) = \frac{\pi^{r_m}\left[1 - F_g^{r_m}(\theta^*(r_m; r_p))\right]}{\pi^{r_m}\left[1 - F_g^{r_m}(\theta^*(r_m; r_p))\right] + (1 - \pi^{r_m})\left[1 - F_n^{r_m}(\theta^*(r_m; r_p))\right]}. \tag{2.3}$$

We say that race-r_p police officers exhibit *statistical discrimination* if they have no taste for discrimination and yet they use different search criterion against motorists with different races.

Definition 2.3. *Assume* $t(M; r_p) = t(W; r_p)$. *Then race-*r_p *police officers exhibit statistical discrimination if* $\theta^*(M; r_p) \neq \theta^*(W; r_p)$.

Officers will choose to use statistical discrimination if the distribution of the signal θ among white and minority motorists is different. When these distributions differ and $t(M; r_p) = t(W; r_p)$ (as assumed), Proposition 2.1 implies that the race-r_p police will choose search criteria $\theta^*(M; r_p)$ and $\theta^*(W; r_p)$ so that the marginal search success rates against white and minority motorists are both equal to the search cost. This typically implies that $\theta^* (M; r_p) \neq \theta^* (W; r_p)$. One reason why the distribution of the signal θ might be different across motorists of different races is that one group might be more likely to carry contraband. For example, if minority drivers are more likely to carry contraband ($\pi^W < \pi^M$), then it will be optimal for a non-prejudiced officer to search relatively more minority drivers (assume everything else is the same for white and minority drivers), and thus they will set $\theta^* (M; r_p) < \theta^* (W; r_p)$. Another reason why the distribution of θ might be different for whites and minorities is that $f_g^{r_m} (\theta)$ and $f_n^{r_m} (\theta)$ can differ between motorist races.

Now we derive some simple implications of the model that will serve as the basis of our empirical test. First, note that if police officers are monolithic, then the cost of searching any given race of motorists is the same, regardless of the race of the officer. That is, $t (W; W) = t (W; M)$ and $t (M; W) = t (M; M)$. If we assume that white and minority troopers face the same population of white motorists and the same population of minority motorists, then Proposition 2.1 implies that both races of officers will use the same search criterion against a given race of motorists,[26] so that $\theta^* (W; W) = \theta^* (W; M)$ and $\theta^* (M; W) = \theta^* (M; M)$. Thus, following from the formula for the search rate (2.2) and average search success rate (2.3), we have:

Proposition 2.2. *If the police officers exhibit monolithic behavior, then* $\gamma (r_m; M) = \gamma (r_m; W)$ *and* $S (r_m; M) = S (r_m; W)$ *for all* $r_m \in \{M, W\}$.

Next, if *none* of the police officers are racially prejudiced, then it immediately follows

[26] In Section 2.4.3 we describe a resampling procedure to empirically deal with data sets in which this assumption may be invalid in the raw data.

from Definition 2.1 that the ranking of $t\left(r_m; M\right)$ and $t\left(r_m; W\right)$ does not depend on the motorist's race r_m, regardless of whether or not troopers are monolithic.[27] We can illustrate the implication of this using an example where white troopers find searching both minority and white motorists more costly than minority troopers do. More formally this can be written as $t\left(M; M\right) = t\left(W; M\right) < t\left(M; W\right) = t\left(W; W\right)$.[28] Because the search threshold given in Proposition 2.1 is monotonically increasing in $t\left(r_m; r_p\right)$ and both white and minority troopers face the same population of white and minority motorists, this implies that $\theta^*\left(M; M\right) < \theta^*\left(M; W\right)$ and $\theta^*\left(W; M\right) < \theta^*\left(W; W\right)$. Because the equilibrium search rate given in (2.2) is monotonically decreasing in $\theta^*\left(r_m; r_p\right)$, we immediately have that $\gamma\left(M; M\right) > \gamma\left(M; W\right)$ and $\gamma\left(W; M\right) > \gamma\left(W; W\right)$, so that race-$M$ officers' search rates will be higher for both races of motorists. Similarly, if $t(M; M) = t(W; M) > t(M; W) = t(W; W)$, then race-$M$ officers' search rates will be lower for both rates of motorists than race-W officers. Finally, if $t(M; M) = t(W; M) = t(M; W) = t(W; W)$, then race-$M$ officers' search rates will be equal to those of race-W officers for both races of motorists.

We can also show that if none of the police officers are racially prejudiced, then the rank order of average search success rates between white and minority troopers for any race of motorists should also be independent of the motorists' race. Recall the previous example where white troopers had a higher overall search cost than minority troopers. We showed

[27] Consider, for illustrative purposes, the case where $t(W; M) < t(W; W)$. Since race-M officers are assumed not to be racially prejudiced, we have $t\left(W; M\right) = t\left(M; M\right)$. Similarly since race-$W$ officers are not racially prejudiced, we have $t\left(W; W\right) = t\left(M; W\right)$. Thus it must be the case $t(M; M) < t(M; W)$. Thus $t\left(r_m; M\right) < t\left(r_m; W\right)$ for all r_m. Similar arguments show that if $t(W; M) > t(W; W)$, then we must have $t(M; M) > t(M; W)$; and if $t(W; M) = t(W; W)$ then we must have $t(M; M) = t(M; W)$. Thus the ranking of $t(r_m; M)$ and $t(r_m; W)$ does not depend on the motorist's race r_m.

[28] Note that the relationship $t(M; r_p) = t(W; r_p)$ does not imply that $\theta^*(M; r_p) = \theta^*(W; r_p)$, because troopers can be engaged in statistical discrimination.

this would imply that $\theta^*(M;M) < \theta^*(M;W)$ and $\theta^*(W;M) < \theta^*(W;W)$. The average search success rate with a search criterion θ^* against race-r_m motorist is simply

$$\frac{\pi^{r_m}\left[1 - F_g^{r_m}(\theta^*)\right]}{\pi^{r_m}\left[1 - F_g^{r_m}(\theta^*)\right] + (1 - \pi^{r_m})\left[1 - F_n^{r_m}(\theta^*)\right]},$$

and one can show that it is strictly increasing in θ^*.[29] Thus we have $S(W;M) < S(W;W)$ and $S(M;M) < S(M;W)$. That is, the ranking of $S(r_m;M)$ and $S(r_m;W)$ does not depend on r_m.

The above discussion is summarized in the following proposition:

Proposition 2.3. *If neither race-M nor race-W of police officers exhibit racial prejudice, then neither the ranking of $\gamma(r_m;M)$ and $\gamma(r_m;W)$ nor the ranking of average search success rates $S(r_m;M)$ and $S(r_m;W)$ depends on $r_m \in \{M,W\}$. Moreover, for any r_m, the ranking of $\gamma(r_m;M)$ and $\gamma(r_m;W)$ should be the exact opposite of the ranking of $S(r_m;M)$ and $S(r_m;W)$.*[30]

In our model if race-r_p troopers are not racially prejudiced, we know that race-r_p troopers' *marginal* search success rate against white motorists will be equal to their marginal search success rate against minority motorists. But because in our model the marginal

[29]To see this, note that it will be strictly increasing in θ^* if and only if

$$H(\theta^*) = \frac{1 - F_g^{r_m}(\theta^*)}{1 - F_n^{r_m}(\theta^*)}$$

is strictly increasing in θ^*. Note that, after some simplification,

$$H'(\theta^*) = \frac{\int_{\theta^*}^{1}\left[f_n^{r_m}(\theta^*) f_g^{r_m}(\theta) - f_g^{r_m}(\theta^*) f_n^{r_m}(\theta)\right] d\theta}{\left[1 - F_n^{r_m}(\theta^*)\right]^2}.$$

From MLRP, we know that, for all $\theta > \theta^*$, $f_g^{r_m}(\theta)/f_n^{r_m}(\theta) > f_g^{r_m}(\theta^*)/f_n^{r_m}(\theta^*)$, thus the integrand in the numerator is always positive. Hence $H'(\theta^*) > 0$.

[30]The last statement in Proposition 2.3 holds regardless of whether or not troopers are racially prejudiced.

motorist's guilt probability is smaller than that of the infra-marginal motorists, we cannot conclude that race-r_p troopers' *average* search success rate against white motorists will be equal to their average search success rate against minority motorists. This is in stark contrast to KPT's model where there is no distinction between marginal and average motorists. Nonetheless, Proposition 2.3 provides *robust* testable implications of our model based on rank orders of observable statistics – the search rates and the average search success rates.[31]

The contrapositive of Proposition 2.3 is simply that, if the ranking of $\gamma(r_m; M)$ and $\gamma(r_m; W)$, or the ranking of $S(r_m; M)$ and $S(r_m; W)$, depend on r_m, then *at least one racial group* of the troopers exhibit racial prejudice. Without further assumptions, it is not possible to determine which group of troopers are racially prejudiced.

2.3.2 Discussion of the Model

Assumption on the Signal Distributions

Our model allows the signal distributions $f_g^{r_m}$ and $f_n^{r_m}$ to be specific to the racial group of the drivers. This flexibility is important if we intend to use our model as a basis for an empirical test. As we explained in the introduction, black and white drivers may exhibit different characteristics in their encounters with highway troopers, and thus imposing f_g^M and f_n^M to be equal to f_g^W and f_n^W, respectively, would be a very strong assumption and may be empirically implausible. Indeed, it is possible for example that minority drivers not carrying contraband might tend to be more nervous during a stop than whites. Also

[31] Proposition 2.3 provides testable implications on the rank orders of both search rate and average search success rates. In this regard, our test is in agreement to Ross and Yinger (2002, Chapter 8) in the context of mortgage lending discrimination, where they emphasize the inextricable link between loan approval decisions and loan performance (see also Ross, 1997).

note that, since θ is most likely not observable by researchers, we do not want to impose parametric distributional assumptions. While sharper tests may be designed if we were to impose more parametric distributional restrictions on $f_g^{r_m}$ and $f_n^{r_m}$, the desirable feature of our test is its *robustness.*

Despite this flexibility, our formulation does assume that the signals of race-r_m motorists are drawn from the same distributions independent of police officers' race. For example, we do not allow for the possibility that minority drivers will present a signal that is drawn from one distribution when they are stopped by a minority trooper and another signal that is drawn from a different distribution when they are stopped by a white trooper. This would be a suspicious assumption, for example, if the stops and searches occur on *local* streets. As argued in Donohue and Levitt (2001), a black community may be more willing to cooperate with a black officer, and thus black officers may obtain more information about a black motorist on the streets. However, we maintain that this is a realistic assumption in highway searches. When stopping a black driver on highways, a trooper typically does not have any other citizens to rely on for additional information. Thus any informational advantage that black officers have about black motorists on local streets may not extend to the highways. Thus as long as white and black troopers observe the same list of characteristics and summarize them in the same way, our assumption will be valid.

One may also argue that minority drivers might be more nervous with white officers than they are with minority officers, regardless of whether or not they are carrying contraband. But as long as white officers properly take this fact into account, they should put a lower weight on the observed nervousness from a black motorist when they formulate the signal index θ. Thus this argument does not necessarily invalidate our assumption that $f_g^{r_m}$ and $f_n^{r_m}$ do not depend on the race of the police officers r_p.

Assumptions on the Officers' Optimization Problem

We assume in officers' optimization problem (2.1) that they maximize the total number of convictions minus a cost of searching cars. We also assume that officers exploit all statistically valid racial inferences in making their search decisions. Our assumption that troopers will always use the race of motorists as a factor in deciding whether or not to search is not necessarily at odds with the official policies on racial profiling. Most highway patrol departments prohibited using race as the primary cause for police-citizen contact, but did not rule out using it as one of many factors. For example, California Highway Patrol prohibits racial profiling which it defined as occurring "when a police officer initiates a traffic or investigative contact based primarily on the race/ethnicity of the individual."[32] Federal courts have ruled that race cannot be the only basis for search and seizure, but it can be one among other factors (see, for example, *Whren v. United States* [1996] and *United States v. Waldon* [2000]).

Finally, in officers' optimization problem (2.1), we assume that they do not have search capacity constraints and thus they judge each stopped vehicle individually to determine whether it is worth a search. But if officers did have a search capacity constraint they would choose to search only the most suspicious motorists. In reality, however, capacity constraints are not likely to be important: in our data, an officer has on average less than 7 searches in a span of almost two years. One might also think that an officer may also

[32] See California Highway Patrol Public Contact Demographic Data Summary (pp. 1).

care about the quality of the contraband found and that this should be reflected in their objective function, but unfortunately we do not have such information in our data set and thus cannot include this in our model.[33]

Assumption on the Pool of Motorists Faced by Troopers of Different Races

In the model, we assume that the fraction of race-r_m motorists carrying contraband $\pi^{r_m} \in (0,1)$ does not depend on the race of the troopers searching them. That is, we assumed that the pools of motorists faced by troopers of different races are the same. This assumption may not be empirically valid if white and minority troopers are systematically assigned to patrol in different locations and different times of the day (indeed, our raw data indicate that this is the case–see Tables 2.6 and 2.7). In Section 2.4.3 we describe a resampling procedure to deal with this problem empirically.

2.4 Empirical Tests

2.4.1 Test for Monolithic Trooper Behavior

Proposition 2.2 suggests a test for whether troopers of different races exhibit monolithic search behavior that is implementable even when researchers have no access to the signals θ observed by troopers in making their search decisions. Under the null hypothesis that police officers exhibit monolithic behavior, then, for any race of drivers, the search rates and average search success rates against drivers of that race should be independent of the race of the troopers that conduct the searches. That is, under the null hypothesis of monolithic

[33]The Maryland data set used by KPT does contain the quantity of drugs found in the searches (see Knowles et al., 2001).

trooper behavior, we must have, for all $r_m \in \{M, W\}$,

$$\gamma(r_m; M) = \gamma(r_m; W), \tag{2.4}$$

$$S(r_m; M) = S(r_m; W). \tag{2.5}$$

Any evidence in violation of any of these equalities would reject the null hypothesis.

It is worth pointing out that both equalities (2.4) and (2.5) hold if and only if the null hypothesis is true. To illustrate why this is true we need to show that when the null hypothesis is not true we will never satisfy equality (2.4) and (2.5). Without loss of generality, suppose that troopers are not monolithic in their search behavior against white motorists ($r_m = W$). That is, $t(W;W) \neq t(W;M)$. If $t(W;W) > t(W;M)$, then, because both white and minority troopers face the same population of white motorists, we know from Proposition 2.1 that $\theta^*(W;W) > \theta^*(W;M)$, i.e. white troopers will use a more strict search criterion than minority troopers when searching white motorists. This then simultaneously implies that $\gamma(W;W) < \gamma(W;M)$ and that $S(W;W) > S(W;M)$, following from the proof in footnote 29. Thus the test using either (2.4) or (2.5) has an asymptotic power of one.

Moreover, the relationship between search rates and average search success rates suggests that, in principle, our model can be refuted. According to our model, whenever $\gamma(W;W) < \gamma(W;M)$, this must be because $\theta^*(W;W) > \theta^*(W;M)$ which directly implies that $S(W;W) > S(W;M)$. Thus if the rank order between the search rates between racial groups of troopers for a given race of motorists is not exactly the opposite of the rank order between the average search success rates, then we know that at least some of the conditions of our model are not satisfied.[34]

2.4.2 Test for Racial Prejudice

Proposition 2.3 suggests a test for whether some racial groups of troopers exhibit racial prejudice in their search behavior. Under the null hypothesis that none of the racial groups of troopers have racial prejudice, it must be true that both the ranking of search rates for a given race of motorists r_m across the races of troopers $[\gamma(r_m; M)$ and $\gamma(r_m; W)]$ and the ranking of average search success rates $[S(r_m; M)$ and $S(r_m; W)]$ do not depend on $r_m \in \{M, W\}$. The null hypothesis will be rejected if the ranking of $\gamma(r_m; M)$ and $\gamma(r_m; W)$, or the ranking of $S(r_m; M)$ and $S(r_m; W)$, depends on the race of the motorists r_m.

Power of the Test

Two features of our empirical test for racial prejudice are worth discussing in further detail. First, our test will only indicate whether or not there is a "relative bias" among troopers. This is because when we do find evidence of racial prejudice, we only know that at least one racial group of officers are racially prejudiced, but cannot determine which one. Second, the power of our test is not one, even when the sample size goes to infinity. To illustrate, suppose that the truth is $t(M; M) = t(W; M) < t(M; W) < t(W; W)$. That is, race-$M$ officers are not racially prejudiced, but race-W officers are prejudiced against minorities (race-W officers' cost of searching minority motorists is smaller). In this case, race-W

[34] Of course, if the search rates between racial groups of troopers for a given race of motorists are equal, then the average search success rates between racial groups of troopers for a given race of motorists must also be equal.

officers will apply higher search criteria toward both races of motorists, and thus the race-W officers' search rates will be lower regardless of the race of the motorists. Therefore the null hypothesis is not rejected even it is false and we as a result commit a type-II error. In general, our test has an asymptotic power of zero if $[t(W;W) - t(W;M)][t(M;W) - t(M;M)] > 0$;[35] it has an asymptotic power of one if $[t(W;W) - t(W;M)][t(M;W) - t(M;M)] < 0$.

The low power of our test may be considered a weakness of our test. On the other hand, if we do find evidence against the null hypothesis, we can be quite confident that at least one racial group of troopers is racially prejudiced. If we were willing to assume that the signal distributions $f_g^{r_m}$ and $f_n^{r_m}$ do not depend on r_m, then one can derive more powerful tests for racial prejudice. Our test can be considered to be the robust implication from a plausible behavioral model that does not impose strong and unverifiable distributional assumptions.[36]

Finally it is worth pointing out that even though our test has low power, it is able to detect racial prejudice when we apply it to the Boston data analyzed in Antonovics and Knight (2004). Their Table 1 indicates that in their data, black officers' search rate is higher than white officers against white motorists, but white officers' search rate is higher than black officers against black motorists. This is a violation of the rank order independence for search rates, which indicates that at least one racial group of the officers are racially

[35]This will be true either when $t(W;W) > t(W;M)$ and $t(M;W) > t(M;M)$, or when $t(W;W) < t(W;M)$ and $t(M;W) < t(M;M)$. That is, our test will fail to detect relative racial prejudice if the troopers of some race have smaller (or larger) costs of searching drivers of any race than troopers of other races.

[36]In this regard, our position is similar to Manski (1995) who preached the tolerance of ambiguity in empirical research in social sciences.

prejudiced.[37]

Difference from Test of Antonovics and Knight (2004)

Now we relate our test of racial prejudice to the test proposed in Antonovics and Knight (2004). As we described in the introduction, they use evidence that police officers are more likely to conduct a search if the race of the officer differs from the race of the driver as evidence of racial prejudice. First, it is useful to point out that their test is *different* from our rank order test proposed above. Consider the following simple example. Suppose that $r_m, r_p \in \{W, M\}$ and let the search rates be as follows: $\gamma(M; M) = .05, \gamma(W; M) = .10, \gamma(M; W) = .20$ and $\gamma(W; W) = .15$. That is, minority officers are more likely to search white motorists than minority motorists, and white officers are more likely to search minority motorists than white motorists. Thus officers in this example are more likely to conduct a search if the race of the motorist is different from their own, causing Antonovics and Knight's test to conclude that racial prejudice is occurring. However, such patterns of search rates satisfy our rank independence condition, that is, $\gamma(r_m; W) > \gamma(r_m; M)$ for $r_m \in \{W, M\}$, and thus our test would *not* consider this as evidence of racial prejudice.[38] If we allow for arbitrary differences, including higher moments, in the signal distributions between white and minority motorists [determined by (π^W, f_g^W, f_n^W) and (π^B, f_g^B, f_n^B) respectively], a positive coefficient on "racial mismatch" can be consistent with the hypothesis

[37] Their paper did not present information about average search success rates which, as we remarked earlier, could have been used to potentially refute our model.

[38] When we ran Probit regressions as specified in Table 6 of Antonovics and Knight (2004) on our Florida data set, the coefficient on the "racial mismatch" variable is positive and significant (the point estimate is about 0.1 with a robust standard error of 0.013). Thus their test, in contrast to ours, would have concluded racial prejudice.

that both racial groups of officers are not prejudiced, even though they must behave non-monolithically. We would like to emphasize, however, that we do *not* mean to say that our test proves no racial prejudice: our conclusion is simply that no racial prejudice could not be ruled out by the data without making stronger and non-verifiable distributional assumptions on the signal distribution.

A second difference between Antonovics and Knight's (2004) test and ours is that we use both search rate and average search success rates in our test, while their test uses only search rates. Using both pieces of information permits us to potentially refute the behavioral model on which our test is based. We think this is an additional strength of our test (see Ross and Yinger, 2002, Chapter 8 for related discussion in the context of mortgage lending).

2.4.3 Resampling Procedure

As we mentioned in Section 2.3.2, our model assumes that the fraction of race-r_m motorists carrying contraband $\pi^{r_m} \in (0,1)$ does not depend on the race of the troopers searching them. Our raw data, summarized in Tables 2.6 and 2.7, indicates that white and minority troopers are systematically assigned to patrol in different locations and at different times of the day, and thus might face different populations of motorists. We will now explain an empirical method that can possibly resolve this problem so that we can use our empirical test even when the raw data does not satisfy this condition. For illustration purposes, suppose that there are two trooper stations, denoted 1 and 2, each with 100 officers. Suppose that in station 1, 80 officers are white and 20 are minorities; in station 2, 60 officers are white and 40 are minorities. Thus, on average 70 percent of the troopers are white and 30 percent are minorities. If the motorists that drive through the patrol areas of stations 1 and 2 differ in their characteristics, then the assumption that on average white and minority troopers face the same pool of motorists may be invalid. To deal with this issue we create reweighted samples in the following way. We keep all the minority officers (20 of them) in station 1, but randomly select 47 out of the 80 white officers. Similarly, we keep all the white officers (60 of them) in station 2, but randomly select 26 out of the 40 minority officers. Thus we create a reweighted sample of 107 white officers and 46 minority officers. Among the 153 officers in the artificial sample, (roughly) 70 percent of them are whites and 30 percent are minorities, and they are equally likely to be assigned to stations 1 and 2. We can calculate the various search rates and average search success rates in this reweighted sample. To alleviate the sampling error, we use independent resampling to create a list of such reweighted data sets.[39]

This resampling method can effectively ensure that, when we calculate the search rates and average search success rates, the white and minority officers in the sample are assigned to different trooper stations with equal probability. Thus on average, white and minority officers are facing the same pool of motorists.

[39] As pointed out by a referee, our resampling procedure is reminiscent of the "reweighting" method proposed in DiNardo et al. (1996) in the context of decomposing the effects of institutional and labor market factors on changes in the U.S. wage distributions. The main difference from our procedure is that we use independent multiple resampling to alleviate sampling error.

2.5 Empirical Results

2.5.1 Data Description

We now apply the tests described above to data from the Florida State Highway Patrol. The Florida data is composed of two parts. The first is the *traffic* data set that consists of all the stops and searches conducted on all Florida highways from January 2000 to November 2001. For each of the stops in the data set, it includes (among other things) the date, exact time, county, driver's race, gender, ethnicity, age, reason for stop, whether a search was conducted, rationale for search, type of contraband seized, and the ID number of the trooper who conducted the stop and/or search. This part of the data is similar to those used in earlier studies of racial profiling (e.g., KPT, 2001, and Gross and Barnes, 2002).[40] The *unique* feature of our data set is the second part, which is the *personnel* data that contains information on each of the troopers that conducted the stops and searches in the traffic data set, including their ID number, date of birth, date of hiring, race, gender, rank, and base troop station. We merge the traffic data and the personnel data by the unique trooper ID number that appears in both data sets. The merged data set thus provides information about the demographics of the trooper that made each stop and search. After eliminating cases in which there was missing information on the demographics of the trooper that conducted the stop, we end up with 906,339 stops and 8,976 searches conducted by a total of 1,469 troopers.[41] Florida State Highway Patrol troopers are assigned to one of ten

[40]Even though KPT have data on stops, they did not use them in their analysis. Gross and Barnes (2002) provided some basic statistics about the stop data.

[41]We also eliminated cases where the race of the motorist and trooper was neither white, black, or Hispanic, since there are not enough observations of the other racial groups to consider them.

trooper stations. Except for trooper station K, which is in charge of the Florida Turnpike, all other stations cover fixed counties. Figure 2.1 shows the coverage area of different troop stations.

2.5.2 Descriptive Statistics

Tables 2.2 and 2.3 summarize the means of the variables related to the motorists in our sample. Of the 906,339 stops we observe, 66.5 percent were carried out against white motorists, 17.3 percent against Hispanic motorists, and 16.2 percent against black motorists. In all race categories of the motorists, male motorists account for at least 67 percent of the stopped motorists. Among all the motorists that were stopped, 48 percent were in the 16-30 age group, 33.6 percent were in the 31-45 age group and 18.3 percent were 46 and older. Close to 90 percent of stopped motorists have in-state license plates, and close to 70 percent of the stops were conducted in the daytime (defined to be between 6am and 6pm).

Of the 8,976 searches we observe, 54.6 percent were performed on white motorists, 23.4 percent on Hispanic motorists, and 22.1 percent on black motorists. In all race categories, more than 80 percent of searches were performed on male motorists, and overall, 84.8 percent of searches were against male drivers. Among the motorists that were searched, 58.4 percent were in the 16-30 age group, 31.7 percent were in the 31-45 age group and only 9.9 percent were in the 46 and older age group. Vehicles with in-state plates account for 85.7 percent of the searches, and 52.5 percent of the searches were conducted at night (recall 30.3 percent of the stops were at night). 79.2 percent of searches were not successful (they yielded nothing). Drugs were the most common contraband seized in successful searches (15.1 percent of total searches), followed by alcohol/tobacco (2.1 percent) and drug paraphernalia (1.5 percent).

Tables 2.4 and 2.5 summarize the means of variables related to the troopers in our sample. The first column of Table 2.4 shows that in our data, blacks, Hispanics and whites account for 13.7, 10, and 76.3 percent of the troopers respectively. 89 percent of the troopers are male. From Tables 2.4 and 2.5 we see that white troopers conducted 73 percent of all stops and 86 percent of all searches. The corresponding numbers for black troopers are 16 and 4.6 percent; for Hispanic troopers they are 11.4 and 9.5 percent. Female troopers conducted 9.3 percent of all stops and 6.9 percent of all searches.

2.5.3 Examining the Assumption that Troopers Face the Same Population of Motorists

Before we conduct our tests of monolithic behavior and racial prejudice we first examine whether a crucial assumption of our test, that all troopers face the same population of motorists, is satisfied in the raw data (before resampling). This assumption, of course, is not directly testable, because $\pi^{r_m}, f_g^{r_m}(\theta), f_n^{r_m}(\theta)$ and θ are all unobservable. The best we can do is to examine the distribution of observable motorist characteristics faced by troopers of different races. Table 2.6 shows the proportions of stopped motorists with given characteristics faced by troopers of different races. The characteristics of motorists reported in the table include race, gender, age, and time of stop. For each row, we also report in the last column the p-values for Pearson χ^2 tests of the null hypothesis that the proportions of stopped motorists with the characteristics specific to that row are the same for all three racial groups of troopers. As one can see, the hypothesis that troopers of different races face the same population of motorists can be statistically rejected in the raw data, even though the differences are numerically quite small. One may suspect that the reason that troopers of different races are stopping motorists with different characteristics is that black, Hispanic and white troopers are assigned to different troops. For example, Hispanic troopers are likely to have an over-representation in Troop E (covering Miami in Dade County) relative to Troop A and H (covering counties in the Florida Panhandle). Indeed, Table 2.7 shows that the allocations of troopers of different races to different troops, and time of assignment, do not seem random in the raw data. For this reason, we think it is important to conduct the resampling methods we described in Section 2.3.2.[42] By construction, in the reweighted data we created with the resampling method, troopers of a given race are assigned to different troops with the same probabilities. The Pearson χ^2 test also reveals that in the reweighted sample troopers of different races are assigned to night shifts with the same probability. Thus we can maintain our hypothesis that the distribution of the observable characteristics of the stopped motorists faced by troopers are the same in the reweighted sample. We report our test results below using data from the reweighted samples.

2.5.4 Empirical Results for the Test of Monolithic Trooper Behavior

Our main empirical results are presented in Table 2.1 in the introduction. Panel A shows two facts: first, regardless of motorists' race, white officers search the highest percentage of the motorists they stop, and black officers search the lowest percentage; second, for all

[42]One may argue that all of the stops occurred on Florida highways, and the drug flow in Florida tends to go from Miami (a city in the southern tip of Florida) to cities in the northeastern United States; that is, drug couriers are moving throughout Florida (except for possibly the panhandle). Thus troopers stationed in different areas are likely to face similar population of drivers, and the differences in the stopped motorists' characteristics reflect the differences in stop behavior of the troopers of different races, rather than the differences in the driver population. It is plausible, but in this paper we take the population of stopped motorists as given.

officers' races, the percentage of black motorists searched is higher than Hispanic motorists, which in turn is higher than white motorists.

We now implement our test for determining whether troopers of different races exhibit monolithic behavior in their search decisions. Recall that we said that if troopers are monolithic they will all search a given race of motorists at the same rate. Thus we need to test whether or not the search rates that are given in Panel A of Table 2.1 differ among trooper racial groups for a given group of motorists. To accomplish this, we compute the p-values from the Pearson χ^2 test under the null hypothesis that troopers of all races search race-r_m motorists with equal probability. These p-values are shown in Table 2.1. Specifically, the Pearson's χ^2 test statistic under the null hypothesis all troopers with race in $\mathcal{R}$ search race-r_m motorists with equal probability is given by

$$\sum_{r_p \in \mathcal{R}} \frac{\left(\gamma \widehat{(r_m; r_p)} - \gamma \widehat{(r_m)}\right)^2}{\gamma \widehat{(r_m)}} \sim \chi^2 (R-1),$$

where $\gamma \widehat{(r_m; r_p)}$ is the estimated search probability of race-r_p officers against race-r_m motorists, $\gamma \widehat{(r_m)}$ is the estimated search probability against race-r_m motorists unconditional on the race of the officer, and R is the cardinality of the set of troopers' race categories $\mathcal{R}$. The p-value for a given motorist race gives the significance level above which we can reject the null hypothesis that the three search rates corresponding to that row are equal, which is the prediction under the null hypothesis of monolithic behavior. Because all the p-values are less than 0.001, this provides strong evidence against monolithic trooper behavior.

Panel B presents the average search success rate for given motorist/trooper race pairs. The first finding from Panel B is exactly converse to the first finding from Panel A: for any given motorist race, black officers' average search success rate is higher than that of Hispanic officers, which in turn is higher than that of white officers. To test for monolithic behavior, we need to see whether all racial groups of troopers have the same average search success rate against a given racial group of motorists. The p-value in each row is from the Pearson χ^2 test under the null hypothesis that troopers of all races have the same average search success rate against motorists of the race in that specific row. Again the Pearson χ^2 test statistic under the null hypothesis that all troopers with race in $\mathcal{R}$ have the same average search success rate against race-r_m motorists is given by

$$\sum_{r_p \in \mathcal{R}} \frac{\left(S\widehat{(r_m;r_p)} - S\widehat{(r_m)}\right)^2}{S\widehat{(r_m)}} \sim \chi^2(R-1),$$

where $S\widehat{(r_m;r_p)}$ is the estimated average search success rate of race-r_p officers against race-r_m motorists, and $S\widehat{(r_m)}$ is the estimated average search success rate against race-r_m motorists unconditional on the race of the officers. All p-values are less than .001, which again provides strong evidence against monolithic trooper behavior.

The second finding from Panel B is that, for all officers, the average search success rate is highest against white motorists, followed in order by black and Hispanic motorists. Though this finding is not directly related to our test for monolithic behavior, it provides strong support for our modelling assumption that the distributions of unobservable characteristics for motorists of different races may be very different, not only in means but also in higher moments. For example, Panel A shows that black officers search about the same percentage of white and Hispanic motorists (0.27 vs. 0.28), but their average search success rate against white motorists are much higher than that for Hispanic motorists (39.4 vs. 21.0).

2.5.5 Empirical Results for the Test of Racial Prejudice

We have so far provided strong evidence that troopers do not exhibit monolithic search criteria when deciding whether to search motorists of a given race. Now we describe the results from our test for racial prejudice as described in Section 2.4.2. Under the null hypothesis that none of the racial groups of troopers are racially prejudiced, we argued that the rank order over the search rates $\gamma(r_m; W), \gamma(r_m; B)$ and $\gamma(r_m; H)$, and the rank order over the average search success rates $S(r_m; W), S(r_m; B)$ and $S(r_m; H)$, should both be independent of r_m. From the estimated mean search rates and average search success rates in Table 2.1, we know that, for all $r_m \in \{W, B, H\}, \gamma\widehat{(r_m; W)} > \gamma\widehat{(r_m; H)} > \gamma\widehat{(r_m; B)}$, and $S\widehat{(r_m; W)} < S\widehat{(r_m; H)} < S\widehat{(r_m; B)}$. We can use a simple Z-statistic to formally test whether

$$\gamma(r_m; W) \quad > \quad \gamma(r_m; H) > \gamma(r_m; B), \tag{2.6}$$

$$S(r_m; W) \quad < \quad S(r_m; H) < S(r_m; B). \tag{2.7}$$

For example, let the null hypothesis be $\gamma(r_m; W) = \gamma(r_m; H)$. We can test it against the one-sided alternative hypothesis $\gamma(r_m; W) > \gamma(r_m; H)$ by using

$$Z = \frac{\gamma\widehat{(r_m; W)} - \gamma\widehat{(r_m; H)}}{\sqrt{\frac{\text{SVar}_W}{n_W} + \frac{\text{SVar}_H}{n_H}}}$$

where n_W and n_H are the number of stops conducted by white and Hispanic officers, respectively, against race-r_m motorists, and SVar_W and SVar_H are respectively the sample variances of the search dummy variables in the samples of stops against race-r_m motorists conducted by white and Hispanic officers. By the Central Limit Theorem (due to our large sample size), Z has a standard normal distribution under the null hypothesis. The null will be rejected in favor of the alternative at significance level α if $Z \geq z_\alpha$ where $\Phi(z_\alpha) = 1 - \alpha$. When $r_m = W$, the value of the Z-statistic is 27.4 under the null, thus we can reject it in favor of the alternative $\gamma(W; W) > \gamma(W; H)$ at a significance level close to 0. Similarly, for the test of the null hypothesis $\gamma(W; H) = \gamma(W; B)$ against $\gamma(W; H) > \gamma(W; B)$, we obtain a Z-statistic of 65, thus again rejecting the null in favor of the alternative. Implementing this test to other races of motorists, we find that the evidence supports inequality (2.6).

Analogously we can formally test inequality (2.7) by using a Z-test

$$Z' = \frac{\widehat{S(r_m; W)} - \widehat{S(r_m; H)}}{\sqrt{\frac{\text{SVar}'_W}{n'_W} + \frac{\text{SVar}'_H}{n'_H}}} \sim N(0,1), \tag{2.8}$$

where n'_W and n'_H are the number of searches against race-r_m motorists conducted by white and Hispanic officers respectively, and SVar'_W and SVar'_H are respectively the sample variances of the search success dummy variables in the sample of searches against race-r_m motorists conducted by white and Hispanic officers. The null will be rejected in favor of the alternative at significance level α if $Z' \leq -z_\alpha$ where $\Phi(z_\alpha) = 1 - \alpha$. For example, when we consider white motorists, we obtain a Z-statistic of -324.1 for white and Hispanic officers, and thus we are able to reject the null in favor of the alternative $S(W; W) < S(W; H)$ at a significance level essentially equal to 0. Likewise, we can reject the null $S(W; H) = S(W; B)$ in favor of the alternative $S(W; H) < S(W; B)$ at a significance level close to 0 (with a Z-statistic of -254). Implementing this test to other races of motorists, we find that the evidence supports inequality (2.7).

To summarize, we cannot reject the null hypothesis that troopers do not exhibit relative racial prejudice. Of course, we would like to emphasize caution in interpreting our finding: while we do not find definitive evidence of racial prejudice, it is still possible that some or all groups of troopers are racially prejudiced. If the latter is true, then we have committed a type-II error as a result of the weak test.

2.5.6 Other Empirical Implications

It is interesting to note some additional implications from the tests we conducted above. First of all, inequality (2.6) implies that the search criterion used by troopers against race-r_m motorists have the ranking

$$\theta^* (r_m; W) < \theta^* (r_m; H) < \theta^* (r_m; B) .$$

In light of Proposition 2.1, this implies a ranking over the search costs: for any r_m,

$$t (r_m; W) < t (r_m; H) < t (r_m; B) .$$

That is, white troopers seem to have smaller costs of searching motorists of any race, followed by Hispanic troopers. Black troopers have the highest search costs.

Second, as we mentioned at the end of Section 2.4.1, our model is refuted if, for each r_m, the rank order of the search rates against race-r_m motorists $\gamma (r_m; W), \gamma (r_m; B)$ and $\gamma (r_m; H)$ is not exactly the opposite of the rank order of the corresponding average search success rates $S (r_m; W), S (r_m; B)$ and $S (r_m; H)$. As we showed above, the statistical evidence in our data does not refute our model.

2.5.7 Replicating KPT's Test

It is useful to contrast our findings with those from KPT's test. Recall that KPT's test relies on the prediction from their model that, under the null hypothesis of no racial prejudice, the average search success rates should be independent of the motorists' race. The last column in Panel B of Table 2.1 shows the average search success rate for different races of motorists in the *raw* data, and Table 2.8 shows the p-values from a Pearson χ^2 test on the hypothesis that the average search success rates are equal across various race groupings. Their test immediately implies that the troopers show racial prejudice against black and Hispanic motorists, especially Hispanics. However, as we argued, this conclusion is only valid if their model of motorist and trooper behavior is true. We would like to emphasize, though, that our test does not necessarily refute the presence of racial prejudice. Our results are simply that, without strong (and possibly untenable) assumptions, we cannot confidently prove the presence of relative, let alone absolute, racial prejudice.

2.6 Conclusion

Black and Hispanic motorists in the United States are much more likely than white motorists to be searched by highway troopers. Is this apparent racial disparity driven by racist preferences by the troopers, or by motives of effectiveness in interdicting drugs? This chapter presents a simple but plausible model of police search behavior, and we define racial prejudice, statistical discrimination and monolithic trooper behavior within the confines of our model. We then exploit the theoretical predictions from this model to design empirical tests that address the following two questions. Are police monolithic in their search behavior? Is racial profiling in motor vehicle searches motivated by troopers' desire for effective policing (statistical discrimination) or by their racial prejudice (racism)? Relative to the seminal research in Knowles et al. (2001), our model allows troopers of different races to behave differently, thus allowing us to examine non-monolithic trooper behavior; moreover, our model does not yield, and the subsequent empirical test does not rely on, the convenient, but in our view unrealistic, implication that all drivers of the same race carry contraband with the same probability regardless of characteristics other than race, which is the vital prediction underlying their tests. We also propose a resampling method to deal with raw data sets where one of the major assumptions underlying our model and empirical tests is violated. Our tests require data sets with race information about both the motorists and troopers. When applied to vehicle stop and search data from Florida, our tests soundly reject the hypothesis that troopers of different races are monolithic in their search behavior, but fail to reject the hypothesis that troopers of different races do not exhibit relative racial prejudice. Finally we would like to emphasize that our test for racial prejudice is relatively conservative in that we may not always conclude there is racial prejudice when it is actually present. Although our test is a low-power one, which implies a high probability that a type-II error will occur, the positive side of this is that when we do find evidence of racial prejudice it is rather conclusive.

This chapter only focuses on the officers' search decisions. But the trooper must first stop the motorist prior to a search. In our analysis, we took the sample of cars that are stopped as our population and focus solely on determining racial prejudice in troopers' search decisions. Given data limitations, examining the possibility of racial prejudice in highway stops is beyond the scope of this paper. However, it is possible that the racial prejudice of police officers are reflected in their stop decisions as well as (or instead of) their search decisions. Because our model allows for general differences in the unobservable distributions among motorists of different races, the presence (or lack thereof) of racial prejudice at the stop level should not affect our conclusions about *additional* racial prejudice in the search decisions. Investigating racial bias in stops is clearly an important topic for future research, when suitable data sets that include random samples of drivers on the road become available.

Finally, our proposed test of racial prejudice in motor vehicle searches in the presence of infra-marginality and omitted variable problems may also be applicable to detect racial prejudice in mortgage lending. The analog of troopers of different races in the context of mortgage lending is different banks operating in the same metropolitan area. For example, the analog of search rates by driver races will be mortgage denial rates by applicant races, and the analog of search success rates is mortgage default rates. The tests developed in this paper suggest that the comparison of mortgage denial rates and default rates for banks operating in the same metropolitan area can potentially reveal (relative) racial prejudice on the part of banks.

2.7 Appendix A: A Model with Endogenous Drug Carrying Decisions

In Section 3 we assumed that the proportion of motorists in race group r_m is exogenously given as $\pi^{r_m} \in (0,1)$. For the purpose of testing for monolithic behavior and racial prejudice, this partial equilibrium approach suffices. However, for other purposes such as public policy considerations like reducing crimes and the "war on drugs," one may want to know how any changes in trooper behavior may affect the motorists' drug carrying decisions.[43] One needs an equilibrium model to address such questions. In this appendix, we propose a simple model. We show that closing our partial equilibrium model in Section 2.3 is easy; moreover, such an equilibrium model has nice equilibrium uniqueness properties under reasonable conditions. This is in contrast to the labor market statistical discrimination models where multiple equilibria naturally arise and are the driving force for statistical discrimination (see, among others, Coate and Loury 1993).

Consider a single motorist race group r_m, and two trooper racial groups, r_p and r_p'.[44] Suppose that in the trooper population a fraction α is of race r_p and the remainder fraction $1-\alpha$ is of race r_p'. Suppose that Nature draws for each driver a utility cost of carrying contraband $v \in \mathbb{R}_+$ from CDF G with a continuous density. The utility cost v represents feelings of fear experienced by a driver from the act of carrying contraband. If a driver carries contraband and is not caught, he/she derives a benefit of $b > 0$. If a guilty driver

[43] See Persico (2002) for an analysis on how racially blind search policies may affect the total crimes committed by motorists.

[44] Because we are only considering one racial group of motorists, we will omit r_m from the subsequent notation. Having more than one racial groups of motorists will not change any of the results below.

is searched and thus arrested, he/she experiences an additional cost (over and above v) of c_g. If a driver does not carry contraband, he/she does not incur the utility cost of v. But the inconvenience experienced by an innocent driver when he/she is searched is denoted by c_n. Naturally we assume that $c_g > c_n$. We assume that a driver's realization of v is his or her private information; b, c_g and c_n are constants known to all drivers and police officers. Each driver decides whether to carry contraband.

As before, we normalize the benefit of each arrest to the police officer to be one, and for notational simplicity, the cost of search for a race-r_p trooper is written as $t_p \in (0,1)$ and that for a race-r_p' trooper is $t_p' \in (0,1)$. As in Section 2.3, troopers observe noisy but informative signals regarding whether or not a driver is carrying contraband: if a driver is guilty, the signal $\theta \in [0,1]$ is drawn from PDF $f_g(\cdot)$; if the driver is not guilty, then θ is drawn from PDF $f_n(\cdot)$. As before f_g/f_n is strictly increasing in θ. Let F_g and F_n denote the corresponding CDFs of f_g and f_n. We assume that a trooper wants to maximize the total number of convictions minus the cost of searching cars.

We first suppose that a proportion π of drivers choose to carry contraband and analyze the optimal search behavior of the troopers. Let $\Pr(G|\theta)$ denote the posterior probability that a driver with signal θ is guilty of carrying illicit drugs, which is given by

$$\Pr(G|\theta,\pi) = \frac{\pi f_g(\theta)}{\pi f_g(\theta) + (1-\pi) f_n(\theta)}.$$

A race-r_p trooper will decide to search a driver with signal θ if and only if

$$\Pr(G|\theta,\pi) - t_p \geq 0;$$

which, from the MLRP, is equivalent to $\theta \geq \theta_p^*(\pi)$ where $\theta_p^*(\pi) \in [0,1]$ is the unique solution to

$$\Pr(G|\theta,\pi) = t_p.$$

Obviously $\theta_p^*(\pi)$ is strictly decreasing in π. Similarly, race-r_p' troopers will search a motorist if and only if the motorist's signal θ exceeds $\theta_{p'}^*(\pi)$ where $\theta_{p'}^*(\pi)$ solves

$$\Pr(G|\theta) = t_p'.$$

Now suppose that race-r_p and race-r_p' troopers use search criteria of θ_p^* and $\theta_{p'}^*$ respectively. The expected payoff of a driver with utility cost v from carrying contraband is given by

$$\overbrace{\left[\alpha F_g\left(\theta_p^*\right)+(1-\alpha)F_g\left(\theta_{p'}^*\right)\right]b}^{\text{Term 1}}-\overbrace{\left\{\alpha\left[1-F_g\left(\theta_p^*\right)\right]+(1-\alpha)\left[1-F_g\left(\theta_{p'}^*\right)\right]\right\}c_g}^{\text{Term 2}}-v$$

where Term 1 is the probability of not being caught multiplied by the benefit from drugs if the motorist is not caught. Note that a fraction α of the troopers are of race-r_p and use a search criterion of θ_p^*, and $1-\alpha$ of the troopers use $\theta_{p'}^*$. Thus the expected probability of not being caught is $\alpha F_g\left(\theta_p^*\right)+(1-\alpha)F_g\left(\theta_{p'}^*\right)$. Term 2 is the expected probability of being caught multiplied by the cost of being caught with illicit drugs. Of course, the driver suffers a disutility v whenever he or she carries drugs.

The expected payoff of a driver, whose utility cost is v, from not carrying contraband is simply the inconvenience cost of being searched by mistaken troopers:

$$-\left\{\alpha\left[1-F_n\left(\theta_p^*\right)\right]+(1-\alpha)\left[1-F_n\left(\theta_{p'}^*\right)\right]\right\}c_n.$$

Thus a driver with utility cost realization v will decide to carry illicit drugs if and only if $v \leq v^*\left(\theta_p^*, \theta_{p'}^*\right)$ where

$$\begin{aligned} v^*\left(\theta_p^*, \theta_{p'}^*\right) &= \left[\alpha F_g\left(\theta_p^*\right) + (1-\alpha) F_g\left(\theta_{p'}^*\right)\right] b - \left\{\alpha\left[1 - F_g\left(\theta_p^*\right)\right] + (1-\alpha)\left[1 - F_g\left(\theta_{p'}^*\right)\right]\right\} c_g \\ &\quad + \left\{\alpha\left[1 - F_n\left(\theta_p^*\right)\right] + (1-\alpha)\left[1 - F_n\left(\theta_{p'}^*\right)\right]\right\} c_n. \end{aligned} \tag{2.9}$$

Thus if the troopers follow search criteria θ_p^* and $\theta_{p'}^*$ respectively, the proportion of drivers who will choose to carry contraband is given by $G\left(v^*\left(\theta_p^*, \theta_{p'}^*\right)\right)$.

An equilibrium of the model is a triple $\left(\pi, \theta_p^*, \theta_{p'}^*\right)$ such that:

$$\Pr\left(G|\theta_p^*, \pi\right) = t_p \tag{2.10}$$

$$\Pr\left(G|\theta_{p'}^*, \pi\right) = t_{p'} \tag{2.11}$$

$$G\left(v^*\left(\theta_p^*, \theta_{p'}^*\right)\right) = \pi \tag{2.12}$$

The existence of equilibrium follows directly from Brouwer's Fixed Point Theorem. Now we show that in fact for any CDF G with non-negative support (i.e., $v \in \mathbb{R}_+$), the equilibrium is *unique*. Suppose that there are two equilibria in which the proportion of guilty motorists are π and $\tilde{\pi}$ with $\pi > \tilde{\pi}$. Observe from (2.9) that $v^*(0,0) = c_n - c_g < 0$ and

$$\begin{aligned} \frac{\partial v^*\left(\theta_p^*, \theta_{p'}^*\right)}{\partial \theta_p^*} &= \alpha c_n f_n\left(\theta_p^*\right)\left[\frac{f_g\left(\theta_p^*\right)}{f_n\left(\theta_p^*\right)} \frac{b + c_g}{c_n} - 1\right], \\ \frac{\partial v^*\left(\theta_p^*, \theta_{p'}^*\right)}{\partial \theta_{p'}^*} &= \alpha c_n f_n\left(\theta_{p'}^*\right)\left[\frac{f_g\left(\theta_{p'}^*\right)}{f_n\left(\theta_{p'}^*\right)} \frac{b + c_g}{c_n} - 1\right]. \end{aligned}$$

By the MLRP, we know that there exists $\left(\widehat{\theta_p^*}, \widehat{\theta_{p'}^*}\right) \in [0,1)^2$ such that $v^*\left(\theta_p^*, \theta_{p'}^*\right)$ is strictly increasing in both θ_p^* and $\theta_{p'}^*$ when $\left(\theta_p^*, \theta_{p'}^*\right) > \left(\widehat{\theta_p^*}, \widehat{\theta_{p'}^*}\right)$. Since $v^*(0,0) < 0$ and the support of G is non-negative, we have $G(v^*(0,0)) = 0$. Moreover, $G\left(v^*\left(\theta_p^*, \theta_{p'}^*\right)\right)$ will be zero for all $\left(\theta_p^*, \theta_{p'}^*\right) \leq \left(\widehat{\theta_p^*}, \widehat{\theta_{p'}^*}\right)$. Thus any $\left(\theta_p^*, \theta_{p'}^*\right) \leq \left(\widehat{\theta_p^*}, \widehat{\theta_{p'}^*}\right)$ cannot be part of the equilibrium (because if $\pi = 0$, the optimal thresholds should be 1 from the troopers' best response). Thus in both equilibria of the model, we must have $\left(\theta_p^*, \theta_{p'}^*\right) > \left(\widehat{\theta_p^*}, \widehat{\theta_{p'}^*}\right)$ and $\left(\tilde{\theta}_p^*, \tilde{\theta}_{p'}^*\right) > \left(\widehat{\theta_p^*}, \widehat{\theta_{p'}^*}\right)$. That is, both equilibria lie in the region where $v^*(\cdot, \cdot)$ is strictly increasing in both arguments. If $\pi > \tilde{\pi}$, equilibrium conditions (2.10) and (2.11) imply that $\theta_p^* < \tilde{\theta}_p^*$ and $\theta_{p'}^* < \tilde{\theta}_{p'}^*$, therefore $0 < v^*\left(\theta_p^*, \theta_{p'}^*\right) < v^*\left(\tilde{\theta}_p^*, \tilde{\theta}_{p'}^*\right)$. But then it implies that $\tilde{\pi} > \pi$, a contradiction.

2.8 Appendix B: Tables and Figures

Table 2.1: Search Rates and Average Search Success Rates against Motorists of Different Races

Motorist's Race	Trooper Race: White	Trooper Race: Black	Trooper Race: Hispanic	p-value	All Troopers
Panel A: Search Rate Given Stop (percent)					
White	0.96 (6.68E-4)	0.27 (7.73E-4)	0.76 (9.26E-4)	<0.001	0.81 (.090)
Black	1.74 (1.30E-3)	0.35 (1.42E-3)	1.21 (2.28E-3)	<0.001	1.35 (.115)
Hispanic	1.61 (1.46E-3)	0.28 (0.76E-3)	0.99 (3.03E-3)	<0.001	1.34 (.115)
Panel B: Average Search Success Rate (percent)					
White	24.3 (9.43E-3)	39.4 (5.57E-2)	26.0 (2.28E-2)	<0.001	25.1 (.434)
Black	19.9 (1.26E-2)	26.0 (5.32E-2)	20.8 (2.67E-2)	<0.001	20.9 (.407)
Hispanic	8.5 (9.78E-3)	21.0 (4.55E-2)	14.3 (6.63E-2)	<0.001	11.5 (.319)

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

Table 2.2: Means of Variables Related to Motorist Stops

Motorists' Characteristics	Stops: All Stops	Stops: By Motorist Sex: Female	Stops: By Motorist Sex: Male
Black	.162 (.368)	.327 (.470)	.673 (.470)
Hispanic	.173 (.378)	.225 (.417)	.775 (.471)
White	.665 (.472)	.319 (.466)	.681 (.466)
Female	.304 (.460)	1.00 (.00)	0.00 (.00)
Male	.696 (.460)	0.00 (.00)	1.00 (.00)
Age:			
16-30	0.481 (0.500)	0.325 (0.468)	0.675 (0.468)
31-45	0.336 (0.472)	0.295 (0.456)	0.705 (0.456)
46+	0.183 (0.386)	0.269 (0.444)	0.731 (0.444)
License Plate:			
In-state	0.899 (0.302)	0.310 (0.462)	0.690 (0.462)
Out-of-state	0.101 (0.302)	0.252 (0.434)	0.748 (0.434)
Time:			
Day (6am-6pm)	0.697 (0.459)	0.316 (0.465)	0.684 (0.465)
Night	0.303 (0.459)	0.275 (0.447)	0.725 (0.447)
Number of Observations:	906,339	275,527	630,812

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

Table 2.3: Means of Variables Related to Motorist Searches

Motorists' Characteristics	All Searches	Searches By Motorist Sex: Female	Searches By Motorist Sex: Male
Black	.221 (.415)	.146 (.354)	.851 (.354)
Hispanic	.234 (.423)	.098 (.296)	.902 (.296)
White	.546 (.498)	.178 (.382)	.822 (.382)
Female	.152 (.359)	1.00 (.00)	0.00 (.00)
Male	.848 (.359)	0.00 (.00)	1.00 (.00)
Age:			
16-30	0.584 (0.493)	0.149 (0.356)	0.851 (0.356)
31-45	0.317 (0.465)	0.162 (0.368)	0.838 (0.368)
46+	0.099 (0.299)	0.136 (0.343)	0.864 (0.343)
License Plate:			
In-state	0.857 (0.350)	0.155 (0.362)	0.845 (0.362)
Out-of-state	0.143 (0.350)	0.132 (0.338)	0.868 (0.338)
Time:			
Day (6am-6pm)	0.475 (0.499)	0.161 (0.367)	0.839 (0.367)
Night	0.525 (0.499)	0.144 (0.351)	0.856 (0.351)
Contraband Seized:			
None	0.792 (0.406)	0.155 (0.362)	0.845 (0.362)
Drugs	0.151 (0.358)	0.137 (0.344)	0.863 (0.344)
Paraphernalia	0.015 (0.122)	0.156 (0.364)	0.844 (0.364)
Currency	0.003 (0.051)	0.174 (0.388)	0.826 (0.388)
Vehicles	0.010 (0.100)	0.154 (0.363)	0.846 (0.363)
Alcohol/Tobacco	0.021 (0.142)	0.151 (0.359)	0.849 (0.359)
Weapons	0.006 (0.078)	0.055 (0.229)	0.945 (0.229)
Other	0.003 (0.049)	0.318 (0.477)	0.682 (0.477)
Number of Observations:	8,976	1,364	7,612

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

Table 2.4: Means of Variables Related to Trooper Stops

	Troopers	Stops		
Troopers'	All	All	By Trooper Sex	
Characteristics	Troopers	Stops	Female	Male
Black	0.137	0.160	0.115	0.885
	(0.344)	(0.366)	(0.319)	(0.319)
Hispanic	0.100	0.114	0.070	0.930
	(0.300)	(0.318)	(0.256)	(0.256)
White	0.763	0.726	0.092	0.908
	(0.425)	(0.446)	(0.289)	(0.289)
Female	0.106	0.093	1.00	00.00
	(0.307)	(0.291)	(0.00)	(0.00)
Male	0.894	0.907	00.00	1.00
	(0.307)	(0.291)	(0.00)	(0.00)
Ranks:				
Captain	0.022	0.002	0.239	0.761
	(0.148)	(0.041)	(0.426)	(0.426)
Lieutenant	0.070	0.013	0.023	0.977
	(0.255)	(0.112)	(0.151)	(0.151)
Sergeant	0.145	0.062	0.054	0.946
	(0.352)	(0.241)	(0.226)	(0.26)
Corporal	0.147	0.112	0.068	0.932
	(0.354)	(0.316)	(0.252)	(0.252)
LEO	0.602	0.810	0.101	0.899
	(0.490)	(0.392)	(0.301)	(0.301)

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

Table 2.5: Means of Variables Related to Trooper Searches

Troopers' Characteristics	Searches: All Searches	By Trooper Sex: Female	By Trooper Sex: Male
Black	0.046	0.044	0.956
	(0.208)	(0.206)	(0.206)
Hispanic	0.095	0.025	0.975
	(0.293)	(0.155)	(0.155)
White	0.859	0.076	0.924
	(0.348)	(0.265)	(0.265)
Female	0.069	1.00	00.00
	(0.254)	(0.00)	(0.00)
Male	0.931	00.00	1.00
	(0.254)	(0.00)	(0.00)
Ranks:			
Captain	0.002	0.474	0.526
	(0.046)	(0.513)	(0.513)
Lieutenant	0.007	0.000	10.000
	(0.081)	(0.000)	(0.000)
Sergeant	0.053	0.052	0.948
	(0.224)	(0.223)	(0.223)
Corporal	0.071	0.030	0.970
	(0.257)	(0.170)	(0.170)
LEO	0.866	0.073	0.927
	(0.341)	(0.261)	(0.261)

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

Table 2.6: Distribution of Characteristics of Stopped Motorists by Trooper Race in the Raw Data

Motorist's Race	Motorist's Characteristics	White Troopers	Black Troopers	Hispanic Troopers	p-value
White	Male	0.679	0.684	0.701	<0.001
	Night stops	0.288	0.272	0.318	<0.001
	Age: 16-30	0.471	0.460	0.445	<0.001
	Age: 31-45	0.325	0.341	0.349	00.02
Black	Male	0.671	0.667	0.686	<0.001
	Night stops	0.332	0.308	0.354	<0.001
	Age: 16-30	0.514	0.514	0.507	0.001
	Age: 31-45	0.340	0.344	0.356	00.03
Hispanic	Male	0.783	0.774	0.761	<0.001
	Night stops	0.322	0.288	0.393	<0.001
	Age: 16-30	0.516	0.497	0.494	<0.001
	Age: 31-45	0.350	0.363	0.355	0.01

Table 2.7: Proportion of Troopers with Different Races by Troop and Time Assignment in the Raw Data

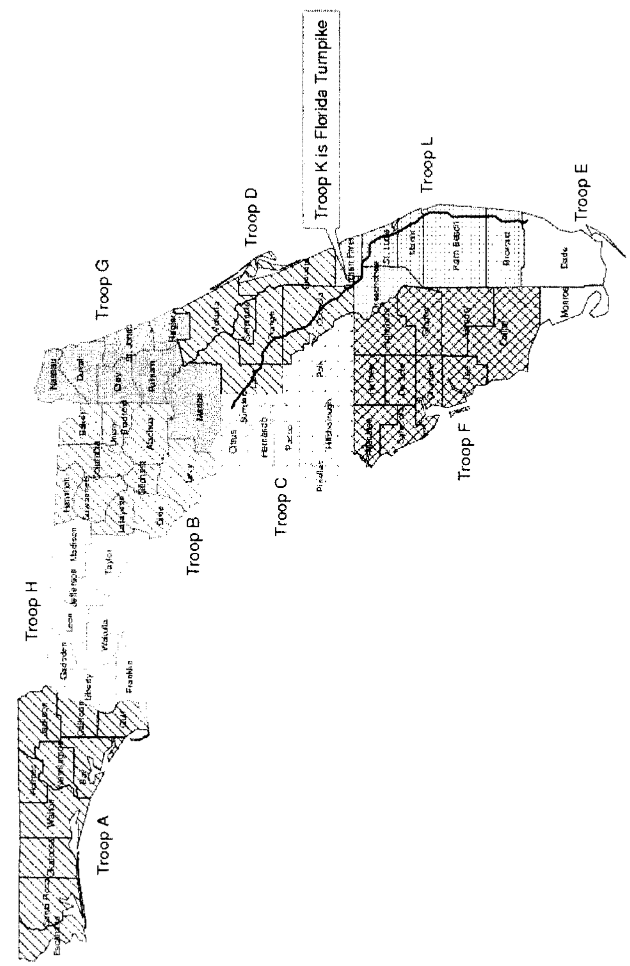
	Troopers' Race		
	White	Black	Hispanic
Troop			
A	0.930 (0.256)	0.054 (0.227)	0.016 (0.124)
B	0.889 (0.316)	0.081 (0.274)	0.030 (0.172)
C	0.816 (0.389)	0.116 (0.321)	0.068 (0.253)
D	0.793 (0.406)	0.117 (0.322)	0.090 (0.287)
E	0.412 (0.494)	0.236 (0.426)	0.352 (0.479)
F	0.880 (0.326)	0.056 (0.231)	0.063 (0.245)
G	0.833 (0.374)	0.135 (0.343)	0.032 (0.176)
H	0.886 (0.320)	0.114 (0.320)	00.00 (0.00)
K	0.698 (0.461)	0.147 (0.355)	0.155 (0.364)
L	0.603 (0.491)	0.298 (0.459)	0.099 (0.300)
% Night Stops	0.283 (0.172)	0.284 (0.192)	0.349 (0.179)

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

Table 2.8: p-values from Pearson's Chi-Squared Tests on the Hypothesis that Average Search Success Rates are Equal Across Various Groupings: KPT Test

Groupings	Average Search Success Rate
White, Black, Hispanic	< 0.001
White, Black	< 0.001
White, Hispanic	< 0.001
Black, Hispanic	< 0.001

Figure 2.1: Troop Station Coverage Map



Chapter 3

Wage and Customer Discrimination in the Professional Basketball Labor Market

3.1 Introduction

Wage discrimination against minorities in the labor market has always been a hotbed of interest because of its large social implications, but it is a question that is often very hard to answer. Most studies of discrimination model a worker's wages to be a function of their ability, productivity, and a variety of demographic variables including race. The effect that a worker's race has on their wages is then taken to be the measure of discrimination. But the reliability of this estimate of wage discrimination depends on how well we can control for worker productivity. If we cannot control for worker productivity accurately then it is very likely that the unexplained wage differential between blacks and whites that we are attributing solely to discrimination could also be partly due to a productivity differential between them that we cannot observe. In most labor markets there is no productivity data available on workers and so the best proxy for this measure becomes a worker's education level and work experience. To the extent that these variables do not pick up all of the productivity differences between blacks and whites, these estimates of wage discrimination may not be very reliable.

In order to resolve this problem many economists have turned to the sports labor market to study wage discrimination. The sports labor market differs from most other labor markets because it provides an extensive set of statistics on productivity. Specifically, in the National Basketball Association (NBA), detailed statistics are kept on the player's performance in every game they play in, which should provide an accurate, comparable, and objective measure of productivity.[1] This reduces the impact of the omitted [productivity] variable bias and increases the reliability of the estimates of racial wage discrimination.

At first glance, the NBA seems like the last place where wage discrimination against black players might be occurring, because most of the highest paid players are black. But past studies have shown that this is not necessarily the case. Kahn and Sherer found that, after controlling for a variety of productivity and market characteristics, white players earned 20% more than black players on average in the 1985-86 NBA season and that this difference was significant (1988). Other studies have repeated Kahn and Sherer's experiment for more recent years but they have not found any significant racial wage premium for whites or blacks after controlling for player productivity: during the 1990-91 season whites earned 9.1% more than blacks on average (Bodvarsson and Partridge 2001); in the 1994-95 season

[1] If the productivity statistics were not objective then they might themselves reflect discrimination.

whites actually earned 1.3% less than blacks on average (Hamilton 1997); and in the 1996-97 season whites earned 6% more than blacks on average (Johnson 1999), although all of these percentages are statistically insignificant. So it seems that in the mid-1980s whites enjoyed a substantial wage premium over black players but that this premium has disappeared in the 1990s.

What can account for the wage premium that whites received in the 80s? Becker proposed that blacks would receive lower pay than whites for the same work if they are a victim of any or all of the following: co-worker discrimination, employer discrimination, and customer discrimination. He predicted that under competitive forces only customer discrimination could survive.[2]

Customer discrimination can arise in the NBA if fans prefer to see white players over black players with equal ability, and thus will be willing to pay a premium to see them. Profit maximizing owners will then be willing to pay more for white players because they get a greater return on them. Consequently fan preferences will be reflected in a wage premium for white players. If customers in most NBA markets prefer to see white players over black players with equal ability, then black players will not be able to segregate into markets in which they are not discriminated against, and so this wage premium will persist under

[2]Statistical discrimination is another form of discrimination that will lead to wage differences between equivalent blacks and whites that will persist in the long run. Statistical discrimination arises when uncertainty about a worker's productivity encourages the employer to use statistics about the average performance of the [racial] group to predict a particular worker's productivity. If employers were uncertain about the productivity of two players, one black and one white, and if white players were more productive than black players on average, then the white player would earn more than the black player on the basis of their group membership. Statistical discrimination against blacks is not a plausible phenomenon in the NBA–the readily available productivity statistics on each player make it unlikely that employers would be uncertain about the productivity of particular players. Furthermore, since on average black players are actually more productive than white players in the NBA, statistical discrimination would lead to black players earning more than white players, after controlling for productivity, which is not what we observe.

competition. Thus customer discrimination is a conceivable cause of the wage differential observed between black and white players.

In their pioneering study of customer discrimination in the NBA, Kahn and Sherer use a panel data set containing the attendance and other characteristics of all NBA teams for the years from 1980-1986 to try to estimate the degree to which fans prefer to see white players. Specifically, they regress a particular team's home attendance in a given year on the percent of the team that is white and on other relevant characteristics of the team for that year that would be expected to affect home attendance (such as winning percentage, ticket price, etc.). The coefficient on the race of the team is what they use as their estimate of customer discrimination. In their study they find that replacing one black player with an equivalent white player would raise each team's total home attendance by 8,000 to 13,000 fans per season from 1980-1986. One potential problem with their specification, however, is that team owners might determine the racial makeup of the team based on their fans' preferences, which would cause the race of the team to be endogenous and bias their estimates of customer discrimination.

In this chapter I investigate whether wage and customer discrimination are a factor in the NBA during more recent times. The key contribution is the development of a new method to test for customer discrimination that is not subject to the endogeneity issues of Kahn and Sherer's method. Using the intuition that for a particular NBA team game by game attendance should predominantly be affected by the road team, I regress per game attendance on the racial composition of the road team, along with a number of other relevant controls. Because the racial composition of the road team should not be affected by the preferences of the home team fans, this should eliminate the endogeneity problem present in the previous specification.

The findings suggest that during the 2000-01 NBA season, among players that were in at least their second contract, white players made 24% more than equivalent black players, a result which is strongly significant. In terms of customer discrimination, I find no evidence of fan preference for white players on the visiting team during the 2000-01 and 2001-02 seasons when data from all teams are combined together. However, when fan preference is estimated separately for each team, I find that certain teams' fans (namely, Houston, Milwaukee, and Phoenix) do have a strong preference for white players. This indicates that while customer discrimination might be a factor in the wage premium that is paid to white players, it is unlikely to be the main factor.

This chapter is organized as follows. Section 3.2 presents the strategy that will be used to estimate wage and customer discrimination. Section 3.3 presents the results of this estimation, and Section 3.4 concludes. Tables and figures are included in an appendix in Section 3.5.

3.2 Methodology

3.2.1 Measuring Wage Discrimination

To determine the wage premium given to white basketball players in the 2000-01 NBA season, I will estimate the following model using the ordinary least squares (OLS) estimator:

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot C_i + \beta_3 \cdot R_i + \varepsilon_i \tag{3.1}$$

where i refers to the player, Y is a measure of the player's salary, X is a vector of player characteristics, C is a vector of characteristics pertaining to the player's contract, R is a binary variable which takes the value of 1 if the player is white and 0 if the player is black, and ε_i is an error term.

The dependent variable Y is the log of the average annual salary that a player earned over the course of the contract they were currently in during the 2000-01 season. These contracts are usually anywhere from three to five years long; players' salaries tend to be structured so that they either get more money later in the deal or more money earlier in the deal. Using the player's average annual salary over the course of their contract, as opposed to their salary for a particular year, should be a more accurate measure of what the team feels the player is worth.

The vector of characteristics X includes the following variables: the seasons played and a quadratic term of seasons played ($SEASONS$ and $SEASONS^2$, respectively), an indicator for players who are centers ($CENTER$), field goal percentage ($FGPCT$), per game rebounds ($REBOUNDS$), per game assist to turnover ratio (ASS/TO), per game steals ($STEALS$), per game blocks ($BLOCKS$), and per game points scored ($POINTS$).

With the exception of $CENTER$ and seasons played, each of these statistics are averaged over the four years prior to the player signing their current contract.[3] This differs from previous studies (Kahn and Sherer, 1988) which used the player's career averages up until the current season. The specification used here is likely to be more realistic for several

[3] For example, if a player signed their current contract at the beginning of the 1998-99 season, then their average salary during this contract would be regressed on the average of their productivity statistics from the 1994-95 season through the 1997-98 season. If players had only played between one and three years in the league at the time they signed their contract, their career averages before they signed their current contract were used.

reasons. Most players in the NBA sign long term deals (usually at least five years), so if, for example, they are in the third year of that contract then their current income should have no direct connection with their statistics for the past two years, because their pay was set before they played those two years. It is also unlikely that a ten year veteran's current income is based on his statistics from his first year, because it is often the case that a players' skill set improves or declines over their career so that their first year statistics are no longer a good predictor of their ability. Furthermore, when players sign their first contract they have yet to play a single NBA game so their salary during this period cannot be a reflection of their NBA statistics. While choosing to average a player's statistics over the previous four years before they signed their contract may seem like an arbitrary number, it is long enough to account for the consistency of a player's performance, while short enough to not hold a player accountable for his performance earlier in his career.

The variables $SEASONS$ and $SEASONS^2$ are included because while a player's salary does depend on the number of seasons he has played, the relationship is not linear—in the beginning, more seasons played usually contributes to higher earnings, but after a point the player becomes worth less as he gets older due to the physical wear and tear on his body. Whether or not a player is a center is included to allow for the possible wage premium that might be paid to centers due to their scarcity. The other statistics included in X are all obvious indicators of on-court productivity.[4],[5]

The vector of contract related variables C includes indicator variables for whether the player signed their contract after the NBA lockout in 1998 ($LOCKOUT$), is earning the minimum wage ($MINWAGE$), or is a free agent who resigned with their previous team ($RESIGN$).[6]

The salary structure rules in the NBA are actually quite complex, but the gist of them are as follows: each team has a soft salary cap, meaning the sum of players' salaries are not allowed to go over this cap unless they are resigning their own player.[7] Thus I controlled for whether or not a player resigned with their own team in the variable $RESIGN$. The NBA has a minimum wage specified for players depending on the number of years they have been in the league. This might cause players to get paid more than what teams feel they are worth and so whether or not a player was making minimum wage is controlled for in the variable $MINWAGE$. The NBA changed its collective bargaining agreement after the

[4]Controls for the number of games played, minutes played per game, and number of playoff games played were left out due to the possible endogeneity of these variables. These variables all share the common trait that someone else besides the player is determining their value.
For example, the team's coach decides how many minutes that a player will play. If a player had some unmeasured characteristic that caused him to get a higher salary which would also cause him to play more minutes (a quality like floor leadership) then this variable will be correlated with the error term and we will have an endogeneity problem.

[5]One possible problem with these productivity statistics is that they cannot measure the actual contribution of a player to the team's performance and hence we cannot assess their overall importance to the team. Although statistics are important to team performance, some players have characteristics like leadership which cannot be measured in statistics. There are also complementarities that exist between players (which is the case when a statement such as "Player X makes every one around him better" is said), causing a player to be worth a lot more on a team in which he fits in with other players better. But since the only measure of team worth we have is through a player's statistics we cannot take these factors into account. If blacks tended to excel in these unmeasured characteristics, then this would bias our estimate of wage discrimination against blacks downward.

[6]This includes "sign-and-trades", whereby the player resigns with the team they had been with previously and is then immediately traded to another team.

[7]There are a few other exceptions, but this is the most important one.

lockout in 1998. The main impact this had on player's wages was that the new agreement restricted the amount that players who were resigning with their previous team could sign for, and it significantly raised the minimum wage.[8] This caused wages for players who signed their contracts after the lockout to be a lot lower, and so this factor is controlled for in the variable $LOCKOUT$.[9]

The players in the sample play on 27 teams located in 26 cities.[10] One might expect salaries to differ between equivalent players that signed contracts with teams located in different cities simply because of market factors affecting the revenues of the team.[11] For example, we would expect teams with a higher home attendance, a higher winning percentage and located in a city with a higher population (which indicates that the team is in a larger market) to earn higher revenues, and thus possibly be able to pay their players more for a given level of productivity. We might also expect teams' markets to differ in many

[8]Previous to this, players could be resigned for any amount, because the team was allowed to go over its salary cap to resign the player. This caused players to demand extremely high salaries, and the team was forced to pay this if they wanted to keep the player. These restrictions still allow the player's former team to sign them for more than they could get elsewhere, but it is no longer an unlimited amount.

[9]The reason why signing after the lockout causes players' wages to fall in the sample is because we do not observe players making the minimum wage who signed their contracts before the lockout (these players do not typically sign long term deals), and so we are only picking up the effect of the players that signed long term deals (which are the better players). Thus in our sample $LOCKOUT$ will affect players' wages negatively.

[10]Players that either signed their contract with or were currently playing on the two Canadian teams (Toronto Raptors and Vancouver Grizzlies) were excluded from the sample. These teams must be excluded from the customer discrimination analysis since it is hard to get data on their market that is at a comparable level to the market data available for cities in the United States. They are excluded here also so that we can compare the results of the wage and customer discrimination analyses.

[11]About 25% of the sample currently plays for a team that is different than the team they signed with (i.e. they were traded). When a player is traded the terms of their contract remain the same, although another team is now responsible for paying them. I chose to control for the market stats pertaining to the team the player signed their contract with because I felt this was where the exact salary determination was made. When I controlled for the characteristics of the team the player currently plays on instead, the results did not change much.

other ways than these that are hard to observe and control for: some teams might be closer to their salary cap limit which would restrict the amount they can pay players, some cities (like Los Angeles) have much higher tax rates than other cities so they might have to pay players more in gross terms in order to compete with other teams in attracting players, and some team owners may just have preferences towards paying players less money than other teams would. Since these are all valid ways in which teams differ (and since there are likely to be many more), I will estimate this model controlling for team fixed effects. This should control for all of the observed and unobserved differences between teams.

In this model, the coefficient β_3 will be the estimate of wage discrimination. It will give the wage premium that is paid to a white player over a black player at the mean of the wage distribution, after holding productivity, contract and market characteristics constant.[12] Finally, it should be pointed out that because of the way the productivity variables are defined, all players in the sample must have played in the NBA before signing their current contract; this effectively restricts the sample to those players who signed their contract in the free agent market.[13] Robustness checks will be done to determine if the results found are specific only to the free agent market or if they apply to all NBA players.

[12] About 4% of the players in the dataset were European players. These players were excluded from the sample to make it easier to directly compare blacks with whites.

[13] In this paper, any player who is in at least their second contract is defined as a player who signed their contract in the free agent market. Typically after a player completes their first contract they become a restricted free agent. In this situation teams are free to bid on them, but their current team has the opportunity to match this salary. After their second contract, players become completely free to sign with whichever team they want. In both situations, though, players have much more inflence over the salary they can get than they do in their first contract.

3.2.2 Measuring Customer Discrimination

As a first attempt to estimate customer discrimination against black NBA players, I will estimate the following model with OLS using panel data on fan attendance for all NBA teams from the 1990-91 season through the 2000-01 season:

$$\begin{aligned}
Log(ATTEND_{it}) &= \alpha_0 + \alpha_1 \cdot YEAR_{it} + \alpha_2 \cdot WINPCT_{it} \\
&+\alpha_3 \cdot PREVWIN_{it} + \alpha_4 \cdot STARS_{it} \\
&+\alpha_5 \cdot ARENA_{it} + \alpha_6 \cdot PRICE_{it} \\
&+\alpha_7 \cdot PCTWHITE_{it} + \alpha_8 \cdot INCMSA_{it} \qquad (3.2) \\
&+\alpha_9 \cdot RACEMSA_{it} + \alpha_{10} \cdot POPMSA_{it} \\
&+\alpha_{11} \cdot LOCKOUTYR_{it} + \varepsilon_{it}
\end{aligned}$$

where i refers to the particular team and t refers to the particular year.[14]

This model is similar to the one used in most previous studies that have tested for customer discrimination (Kahn and Sherer, 1988; Hamilton, 1997; Johnson, 1999; and Bodvarsson and Partridge, 2001). *ATTEND* (the total home attendance of a team in a particular year) is a good reflection of fan (customer) preferences. In addition, home attendance has a direct impact on team revenue, so it is likely that fan preferences here will be reflected in player's salaries, which is exactly the link we need to propose that customer discrimination against black players will cause them to be paid lower on average than white

[14] The Toronto Raptors and Vancouver Grizzlies are excluded from the sample.

players.[15]

$WINPCT$ (the number of wins the team had in the current year), $STARS$ (the number of players on the team that were all-stars) and $PREVWIN$ (the number of wins the team had last year) are all indicators of how popular the team is—the higher these indicators are, the greater fan attendance should be. $ARENA$ (the total capacity of a team's home arena), $POPMSA$, $INCMSA$, $RACEMSA$, (population, real per capita income, and percentage of the population that is black, respectively, in the Standard Metropolitan Statistical Area where one's team was located) and $PRICE$ (the average ticket price to a team's home game) are all indicators of the market the team is located in and thus should have an impact on fan attendance. To control for any other unobserved heterogeneity between teams, this model will be estimated with team fixed effects.[16] $LOCKOUTYR$ is a dummy variable, which takes on a value of 1 if the year is 1999 and 0 otherwise. This controls for the NBA lockout in 1999 which reduced the number of home games from 41 to 25.

With this specification, the coefficient on $PCTWHITE$ (the percentage of the players on the team that are white) will be the estimate of customer discrimination. It will tell us how total annual home attendance is affected when the racial composition of the home

[15]In addition to home attendance, team revenue also comes from selling team merchandise and television deals, both of which are affected by fan preferences, but it would be extremely hard to get data on these variables. While home attendance is not always the best measure of total fan appeal (since presumably only relatively wealthy people will go to NBA games), most of team revenue comes from home attendance so it is likely that the preferences of the fans who attend the games will be the most dominant in determining player's salaries.

[16]As was indicated in the wage regression section, cities may differ in other ways than that which we can control for and these differences may affect the home attendance of these teams. For example, teams that have direct public transportation into the arena, are in a more densely populated area (so that the distance of the arena is not as far), and which do not have many other substitutes (other professional sports teams in the area) are more likely to have higher home attendance, holding all else constant. Also, while the different racial makeup of the populations in different cities was controlled for, there might be other ways in which the distribution of these populations differ (such as in the age distribution) which would also affect home attendance.

team is altered, holding all else constant.

Although this approach has been used in previous customer discrimination studies, there are a couple problems with it. The main problem is that the variable $PCTWHITE$ is likely to be endogenous: owners might know the preferences of their fans and thus will choose the racial composition of their team accordingly. This will bias the estimate of customer discrimination, α_7, towards zero.[17] A second problem is that the price variable is also likely to be an endogenous regressor. Teams that are likely to have some unmeasurable quality about them that will cause team attendance to be high (an effect that is included in the error term) will also tend to have high ticket prices, because this "quality" will allow the team to command higher prices.

One way to solve both of these problems and get a more accurate estimate of customer discrimination is to take a completely different approach. In particular, we could look at fan attendance for a particular team at a game by game level. For that particular team, fan attendance should depend primarily on the characteristics of the visiting team, once we have controlled for how the winning percentage of the home team varies from game to game. We can then test how home team attendance varies in response to the racial composition of the visiting team. Using data from the 2000-01 and 2001-02 NBA seasons I estimated the following model, controlling for team fixed effects:

$$\ln A_i = \lambda_0 + \lambda_1 \cdot G_i + \lambda_2 \cdot H_i + \lambda_3 \cdot R_i + \lambda_4 \cdot PCTWHITE_i + \varepsilon_i \tag{3.3}$$

[17]For example, let's say that over the ten year time period included in the panel data set fan preference in Utah for white players has increased. If team owners know this, a rational response for them would be to have more white players on the team. It is plausible then that attendance would not change over this time period (holding all other variables constant). This would make it seem as if the coefficient on the $PCTWHITE$ variable should be zero.

where i refers to a particular game, A is the attendance of a particular game, G refers to game controls (the year the game was played in [$YEAR$] and whether or not the game was played on a weekend [$WEEKEND$]), H refers to home team controls (the winning percentage of the home team going into the game [$HOMEWIN$]), R refers to road team controls (the winning percent of the road team going into the game [$ROADWIN$], the road team's winning percentage from the previous season [$PREVWIN$], the number of past and present all-stars on the road team [$STARS$], and how deep the road team went in the playoffs the year before [$PREVPLOF$], and $PCTWHITE$ refers to the percent of the road team that is white. This is measured as the percent of the total minutes in the season that were played by white players on that team.

Each of the 27 teams in the sample plays 82 home games (41 for each of the two seasons). The main reason we would expect attendance to vary game by game for a given home team is because of the characteristics of the road team, which are controlled for in (3.3). But since this specification combines all 27 teams together, there might be many other reasons why attendance could differ between two different games (from different home teams). Thus it will be necessary to use a fixed effects approach, as this will only look at the game-by-game variation for a given home team, and average the results over all the teams in the sample. With this specification, λ_4 will be the estimate of fans' preference for white players.

Note that the endogeneity of the price variable is not a problem anymore. It is no longer necessary to use price as a regressor since game by game attendance for a particular team will not depend on the price of the ticket (this is constant across games). More importantly, the specification in (3.3) also takes care of the endogeneity of the team racial composition variable, because we would not expect the road team's owners to choose the racial composition of their team based on the home team's fan preferences.

Additionally, we might want to determine whether or not fan preference for white players is the same in all cities. In particular, we might want to see if cities that have a larger percentage of whites have a stronger preference for white players.[18] Since the racial composition of a particular city hardly changes from year to year, we cannot estimate an interaction term between $PCTWHITE$ and that variable and still used a fixed effects approach. To get around this problem I will instead run a separate regression for each of the 27 teams.[19] We can then determine if the fan preference for white players in each city is correlated with the percent of whites in the city.

3.3 Estimation

3.3.1 The Wage Discrimination Model

Data was collected on all non-European players who were under contract for the 2000-01 NBA season. Player salary data was obtained from the newspaper *USA Today* (12/08/00), and player performance data was taken from NBA.com, the NBA's official website.

Table 3.1 presents the mean values for both black players and white players for most of the variables that were used in the wage regression. The mean values for the productivity statistics for black players are, with a couple exceptions, higher than those for white players, although not all of these differences are significant. Black players also earn higher salaries,

[18] We might expect this if we thought it was only white fans that preferred white players, and that black fans actually preferred black players.

[19] Because I have two years of data, there will be 82 observations for each team regression.

are less likely to earn the minimum wage, and are more likely to have resigned with their own team.

Table 3.2 presents the OLS estimates for the wage regression. In all three specifications team fixed effects are taken into account and the dependent variable is the log of a player's average salary during their current contract. Column 1 presents the results of specification (3.1) from the methodology section: a player's average annual salary is regressed on the average of their statistics from the four years prior to the signing of their contract. As expected, a player's productivity statistics are all positively related to their salary, although only one of these effects is statistically significant (averaging an extra point results in a 7% pay increase). The main coefficient we are interested in is $RACE$, which estimates the unexplained wage differential between white and black players. It indicates that, on average, white players make 24% more than black players, an effect which is statistically significant.

These results strongly imply that white players are paid a premium over equivalent black players. As pointed out earlier, however, due to the way the productivity statistics are constructed the sample used here consists solely of players that signed their contracts in the free agent market. To determine if this effect is present among all players, we could instead regress a player's salary on the career average of their productivity statistics measured up through the 1999-2000 season. This would allow us to expand the sample to include all players except those that were rookies in the 2000-01 season. Column 2 presents the results of this specification. One can see that the wage premium whites receive falls quite a bit, and the effect is no longer significant.[20]

[20]The specification in Column 2 also includes two additional controls not listed in the table. The first is a control for players making the maximum wage allowed. This happens quite often among players in their first contract, because the NBA sets a very low limit on the salaries that can be paid to these players. The

But there could be two reasons the results in Column 2 differ from those in Column 1. The first is that the sample in Column 2 includes more players, and the second is that the construction of the productivity variables differs. To determine which one drives the results, Column 3 runs the same specification as in Column 2, except now the sample is restricted to free agents. As one can see, the coefficient on $RACE$ is similar to that from Column 1, which indicates that the difference in the construction of the productivity variables are not having a significant impact on the results. Rather, it seems that the wage premium tends to be present only among free agents, and thus declines when we expand the sample to include players who are in their first contract.[21]

The main result found here of wage discrimination among players who signed their contracts in the free agent market, but not among all players, is somewhat counterintuitive. Free agents are allowed to sign with any team they want, so one would think that black players would sign with the team that discriminates against them the least, causing the estimate of wage discrimination to fall among this sample. One possible reason we might get these results is that teams do not know what the NBA productivity of players will be when they sign them to their first contract, but they have a much better idea about this when they sign them to succeeding contracts. Thus statistical discrimination might be present when players are signed to their first contract. Since white players are on average less productive

second is a control for players who signed their contract in the free agent market but did not resign with their own team.

[21]Due to the presence of large outliers among NBA player's salaries one might make the case that a median regression might actually be more appropriate. Median regressions were run, but they do not change the results significantly. One possible reason for this is because while the actual salary distribution is skewed due to the presence of large outliers, the log of this salary distribution (which is what is used in this paper) is actually very close to a normal distribution, thus causing the results of the mean and median regression to be very similar.

than black players, white players might receive lower salaries than they otherwise would have been. When they sign later contracts, though, statistical discrimination is no longer present, and so their wages will be higher for a given productivity level as a result. Thus among all players the wage premium in favor of whites could be reflecting positive fan preference for white players tempered by statistical discrimination which adversely affects white players, while among free agents the wage gap in favor of whites only reflects the positive fan preference for white players, thus causing the wage gap to be higher among free agents.

One other possible explanation for these results has to do with the NBA salary structure for players in their first contracts. These players are subject to maximum salaries that are specified by the NBA and that are based on the draft position of the player. In fact, most players in their first contract are making the maximum they are allowed to. This will reduce the amount of variability that is present among these players, which could depress the wage premium white players in their first contract would have received.

3.3.2 The Customer Discrimination Model

To estimate customer discrimination it was necessary to collect data from several sources. Data on the racial composition and population of each team's SMSA were taken from the Census Bureau's Government Information Sharing Project, while data on real per capita income was obtained from the Bureau of Economic Analysis. Data on the price of tickets was taken from Team Marketing Research. All monetary variables were put in terms of 1990 dollars. Game attendance was obtained from ESPN.com. Finally, the racial composition of each team was obtained from *The Complete Handbook of Pro Basketball*, a book published each year which contains pictures of players on every team roster for that year. The race of each player was determined by looking at their picture. All other data was obtained from NBA.com.

Table 3.3 presents the OLS estimates of the standard home attendance regression specified in (3.2), controlling for team fixed effects. This model was estimated using data from the years 1991-2001, and determines how a particular team's attendance depends on its characteristics. The results indicate that a team's winning percentage, previous winning percentage, and arena capacity are all positively and significantly related to their home attendance. If a team increases its winning percentage by one percent, home attendance should increase by .24%.[22] The coefficient on $PCTWHITE$ will indicate whether or not customer discrimination is taking place. The results imply that increasing the percentage of white players on the home team by one percent will increase that team's fan attendance by .09%. This effect is statistically significant, implying customers do have a preference for white players.

Although this model is the standard one run in past studies of customer discrimination, it can be problematic because both $PRICE$ and $PCTWHITE$ are endogenous.[23] This might give us an inaccurate estimate of customer discrimination. As discussed in the methodology section, one way to eliminate these endogeneity issues would be to estimate an alternative model, which was specified in (3.3). Game by game home attendance is regressed on the home team's winning percentage and a variety of characteristics of the

[22] All of the percentage variables mentioned in this paper were recorded in proportions in the dataset, so the coefficients should be interpreted accordingly.

[23] The results clearly indicate that the $PRICE$ variable is endogenous; it is postive and significant, while a simple demand story would predict the price effect to be negative.

road team, including their racial composition, taking team fixed effects into account. Table 3.4 presents the results of this estimation that was run using OLS with data from the 2000-01 and 2001-02 seasons.[24]

The results indicate that, for a particular team, their winning percentage as well as the road team's winning percentage are both positively related to the attendance for a particular game. In addition, road teams that had been in the playoffs the previous year, and games that took place on weekends were a bigger crowd draw. All of these results are as expected. To determine if customer discrimination is present, we can look at the coefficient on $PCTWHITE$, which is close to zero and statistically insignificant. This indicates that a one percent increase in the percent of the visiting team's minutes that are played by white players will not change fan attendance significantly, implying that customers do not have a preference for white players.[25]

Thus, when the endogeneity of the $PCTWHITE$ variable is taken into account, fan preference for white players falls to almost zero. This is somewhat surprising, because endogeneity of the team's racial composition should bias the estimate of fan preference towards zero, so we would expect the coefficient on $PCTWHITE$ to increase in magnitude once this problem is corrected for. One possible explanation for these results lies in the fact that these two models are measuring different things. It could be the case that fans

---

[24] Although there are 29 teams in the NBA, I eliminated the two Canadian teams (the Toronto Raptors and the Vancouver Grizzlies). The Charlotte Hornets were also eliminated from the analysis because they were in the process of moving that year and fans were angry, causing attendance to be unusually low. Also, the New York Knicks, Sacramento Kings, and Washington Wizards were eliminated from the analysis because they sold out all of their home games during this time period, resulting in no variation in the dependent variable.

[25] While this is the most accurate measure of $PCTWHITE$, many other specifications were tried, such as the percent of the players that are white, percent of the starters that are white, and percent of the top ten players (in terms of minutes played) that are white. The results did not change appreciably.

care a lot about the racial composition of the home team (which is measured in (3.2)), but do not care much about the racial composition of the road team (which is measured in (3.3)). Thus if we could correct for the endogeneity of the home team's racial composition in such a way that we could still measure how home team attendance depends on the racial composition of the home team, we might indeed find that fan preference for white players increases. So even though the results indicate that there is no preference for whites on the road team, we cannot definitively conclude that this means there is no preference for whites on the home team.

To determine whether fan preference for whites on the road team varies as the percent of blacks in the home team's SMSA changes, specification (3.3) was estimated separately for each team.[26] Figure 3.1 shows how the coefficient on the $PCTWHITE$ variable for each of these team regressions (represented on the y-axis) is related to the percent of blacks in the home team's SMSA. Almost two thirds of the teams have positive coefficients, which indicates a positive preference for white players. A few of these coefficients are statistically significant (Houston, Milwaukee, and Phoenix), indicating that some teams have a strong preference for white players on the road team, although among all teams there is basically no preference. We might expect that as the percentage of blacks in a city increases, the fan preference for white players would decrease, which would imply a negative relationship between these two variables. As the figure shows, though, there is basically no relationship between these two factors, implying that fan preference for whites is independent of the racial composition of the city.[27]

[26] There are 82 observations for each team regression.

[27] The correlation between these two variables is -.054, but is not statistically significant.

3.4 Conclusion

This chapter tests whether white professional basketball players earned a wage premium during the 2000-01 season, and, if so, whether customer discrimination was one cause for this. I accomplish this by developing a new method to test for customer discrimination in which preference for white players is measured by how home team fan attendance varies in response to the racial composition of the road team. This method solves the endogeneity issue that was present in past studies which measured fan preference by seeing how home team attendance depended on the racial composition of the home team.

The findings presented here suggest that white players who are in at least their second contract earn a significant wage premium over similar black players. In terms of customer discrimination I find that, overall, fans do not prefer to see white players on the opposing team. However, when estimating customer discrimination separately for each team I find that several teams do have a preference for white players. I also find that customers may have a preference for white players on their own team. Overall, these results are consistent with weak evidence of customer preference for white players.

These results are consistent with Becker's theory that the presence of customer discrimination will cause wage differentials to arise.[28] One way to find stronger evidence of a direct causal link between these two types of discrimination would be to gather more extensive salary data and run wage regressions separately for each team. This could then be compared with the estimates of customer discrimination for each team. Finding that the specific teams that preferred white players were also the ones that paid the highest

[28] Any situation where the amount of wage discrimination is at least as great as the amount of customer discrimination will be consistent with Becker's theory.

premiums to white players would provide much stronger empirical support for Becker's theory.

These results, however, are not consistent with the idea that customer discrimination is the only cause of wage discrimination: while I find only marginal evidence of customer discrimination, there is strong evidence of wage discrimination. This indicates it is highly likely that other factors are driving this wage premium. Future work in this area could attempt to determine what these factors are, and whether they can explain why the wage premium disappears when the sample is expanded to include all players.

3.5 Appendix: Tables and Figures

Table 3.1: Mean Values of the Wage Regression Sample

	Blacks	Whites
ln *salary**	14.85	14.54
SEASONS	5.77	5.53
*CENTER**	0.13	0.34
FGPCT	0.449	0.45
*REBOUNDS**	4.33	3.53
ASS/TO	1.34	1.37
*STEALS**	0.87	0.62
BLOCKS	0.56	0.49
*POINTS**	10.3	7.12
*LOCKOUT**	0.79	0.89
*MINWAGE**	0.09	0.19
*RESIGN**	0.5	0.36
observations	245	53

NOTE: * indicates mean difference between black and white players is significant at 10% level.

Table 3.2: OLS Results for the Wage Discrimination Model

	free agents (4yr stats avg) (1)	all players (career stats) (2)	free agents (career stats) (3)
constant	*13.3 (.392)	*13.9 (.338)	*12.8 (.445)
$SEASONS$	.037 (.038)	*.166 (.032)	*.100 (.040)
$SEASONS^2$	*-.005 (.003)	*-.010 (.002)	*-.008 (.002)
$CENTER$	.160 (.123)	*.171 (.097)	*.206 (.126)
$FGPCT$	.674 (.766)	.017 (.627)	1.16 (.867)
$REBOUNDS$	.031 (.026)	*.068 (.021)	*.047 (.027)
ASS/TO	.095 (.067)	*.163 (.051)	*.172 (.069)
$STEALS$	.159 (.118)	.011 (.094)	.086 (.123)
$BLOCKS$	.138 (.094)	*.160 (.078)	.155 (.098)
$POINTS$	*.070 (.010)	*.056 (.008)	*.073 (.010)
$RACE$	*.240 (.097)	.103 (.073)	*.221 (.097)
$LOCKOUT$	.093 (.093)	*-.305 (.081)	.051 (.091)
$MINWAGE$	*-1.36 (.123)	*-.849 (.112)	*-1.31 (.125)
$RESIGN$	*.310 (.074)	.252 (.161)	*.356 (.074)
R^2	.789	.833	.791
observations	231	297	233

NOTE: Standard errors are in parentheses.

* denotes a parameter significant at the 10% level.

Table 3.3: OLS Results for the Standard Customer Discrimination Model

	annual attendance
constant	*12.4 (.250)
YEAR	*.008 (.004)
WINPCT	*.240 (.053)
PREVWIN	*.099 (.048)
STARS	-.005 (.008)
ARENA	*.000 (.000)
PRICE	*.006 (.001)
PCTWHITE	*.094 (.056)
INCMSA	.000 (.000)
RACEMSA	-.614 (.795)
POPMSA	.000 (.000)
LOCKOUTYR	*-.532 (.020)
R^2	.879
observations	295

NOTE: Standard errors are in parentheses.

* denotes a parameter significant at the 10% level.

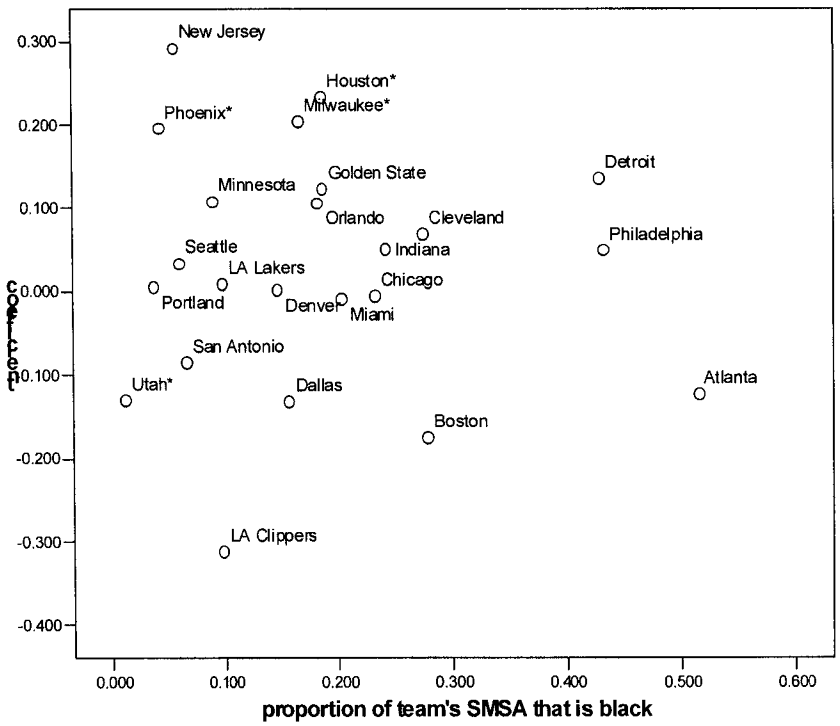
Table 3.4: Effect of the Percent White of the Road Team on the Home Teams' Attendance

	game attendance
constant	*9.40 (.028)
YEAR	-.004 (.007)
HOMEWIN	*.375 (.041)
ROADWIN	*.113 (.033)
WEEKEND	*.076 (.007)
STARS	.004 (.003)
PREVWIN	-.037 (.047)
PREVPLOF	*.032 (.004)
PCTWHITE	.018 (.041)
R^2	.597
observations	1850

NOTE: Standard errors are in parentheses.

* denotes a parameter significant at the 10% level.

Figure 3.1: Preference for White Players versus Racial Composition of SMSA



NOTE: * indicates that the coefficient (on PCTWHITE) is significant at 10% level

Bibliography

Altonji, Joseph G., and Rebecca M. Blank (1999): "Race and Gender in the Labor Market," in *Handbook of Labor Economics,* edited by David Card and Orley Ashenfelter, Amsterdam: Elsevier Science, 3143-3259.

Altonji, Joseph G., and Charles R. Pierret (2001): "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, Vol. 116, 293-312.

Antonovics, Kate, Peter Arcidiacono, and Randall Walsh (2004): "Race and Gender Discrimination in the Weakest Link," Unpublished Manuscript, Durham, N.C.: Duke University: Department of Economics.

Antonovics, Kate L. and Brian G. Knight (2004), "A New Look at Racial Profiling: Evidence from the Boston Police Department." NBER Working Paper No. 10634.

Arrow, Kenneth (1973), "The Theory of Discrimination." in O. Ashenfelter and A. Rees eds., *Discrimination in Labor Markets*, Princeton, New Jersey: Princeton University Press, pp. 3-33.

Ayres, Ian (2001), *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination.* Chicago: University of Chicago Press.

Ayres, Ian (2002), "Outcome Tests of Racial Disparities in Police Practices." *Justice Research and Policy*, Vol. 4, Special Issue, pp. 131-142.

Becker, Gary S. (1957), *The Economics of Discrimination.* Chicago: University of Chicago Press.

Becker, Gary S. (1993), "The Evidence against Blacks Doesn't Prove Bias." in *Business Week*, April 19, available on line at www.businessweek.com.

Becker, Gary S. (1993), "Nobel Lecture: The Economic Way of Looking at Behavior." *Journal of Political Economy*, 101(3), pp. 385-409.

Bjerk, David (2004), "Racial Profiling, Statistical Discrimination, and the Effect of a Colorblind Policy on the Crime Rate." Unpublished paper, McMaster University.

Bodvarsson, Orn B. and Mark D. Partridge, (2001), "A Supply and Demand Model of Co-Worker, Employer and Customer Discrimination," *Labour Economics*, Volume 8 (3), 389-416.

Borooah, Vani K. (2001), "Racial Bias in Police Stops and Searches: An Economic Analysis." *European Journal of Political Economy*, 17(1), pp. 17-37.

Brown, Robert W. and R. Todd Jewell, (1994), "Is There Customer Discrimination in College Basketball? The Premium Fans Pay for White Players," *Social Science Quarterly*, Volume 75 (2), 401-413.

Casella, George and Roger L. Berger (1990). *Statistical Inference,* Duxbury Press.

Coate, Stephen and Glenn C. Loury (1993), "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review,* 83(6), pp. 1220-1240.

Cornell, Bradford, and Ivo Welch (1996): "Culture, Information, and Screening Discrimination," *Journal of Political Economy*, Vol. 104, 542-571.

Desser, Arna, James Monks, and Michael Robinson, (1999), "Baseball Hall of Fame Voting: A Test of the Customer Discrimination Hypothesis," *Social Science Quarterly*, Volume 80 (3), 591-603.

Dharmapala, Dhammika and Stephen L. Ross (2004), "Racial Bias in Motor Vehicle Searches: Additional Theory and Evidence." *Contributions to Economic Analysis & Policy*, 3(1), Article 12.

Dillingham, Alan E., Marianne A. Ferber, and Daniel S. Hammermesh (1994): "Gender Discrimination by Gender: Voting in a Professional Society," *Industrial and Labor Relations Review*, Vol. 47, 622-633.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, 64(5), pp. 1001-1044.

Dominitz, Jeff and John Knowles (2004), "Crime Minimization and Racial Bias: What Can We Learn from Police Search Data?" Unpublished paper, Carnegie Mellon University and University of Pennsylvania.

Donohue, John J. and Steven D. Levitt (2001), "The Impact of Race on Police and Arrests." *Journal of Law and Economics*, 44(2), pp. 367-394.

Engel, Robin S., Jennifer M. Calnon, and Thomas J. Bernard (2002), "Theory and Racial Profiling: Shortcomings and Future Directions in Research." *Justice Quarterly*, 19(2), pp. 249-273.

Farrell, Amy, Jack McDevitt, and Deborah Ramirez (2000), "A Resource Guide on Racial Profiling Data Collection Systems: Promising Practices and Lessons Learned." U.S. Department of Justice.

Fridell, Lorie A. (2004), *By the Numbers: A Guide for Analyzing Race Data From Vehicle Stops.* Police Executive Research Forum, Washington, D.C.

Gross, Samuel R. and Katherine Y. Barnes (2002), "Road Work: Racial Profiling and Drug Interdiction on the Highway." *Michigan Law Review*, Vol. 101, pp. 651-754.

Hamilton, Barton Hughes, (1997), "Racial Discrimination and Professional Basketball Salaries in the 1990s," *Applied Economics*, Volume 29, 287-296.

Hernandez-Murillo, Ruben and John Knowles (2004), "Racial Profiling or Racist Policing?: Distinguishing between Competing Explanations for Discrimination." *International Economic Review*, 45(3), pp. 959-990.

Hollander, Zander, *The Complete Handbook of Pro Basketball*, Signet, years 1988, 1989, 1992, 1994, and 1997.

Irani, Daraius, (1996), "Estimating Customer Discrimination in Baseball Using Panel Data: 1972-1991," in John Fizel, Elizabeth Gustafson and Lawrence Hadley (eds) *Baseball Economics: Current Research*, Greenwood, 47-61.

Johnson, Sean D., (1999), "Wage Discrimination in the National Basketball Association: Is There Discrimination Based on Race?" *Villanova Sports and Entertainment Law Journal*, Volume 6 (1).

Kahn, Lawrence M., (1991), "Discrimination in Professional Sports: A Survey of the Literature," *Industrial and Labor Relations Review*, Volume 44 (3), 395-418.

Kahn, Lawrence M. and Peter D. Sherer, (1988), "Racial Differences in Professional Basketball Players' Compensation," *Journal of Labor Economics*, Volume 6, 40-61.

Knowles, John, Nicola Persico and Petra Todd (2001), "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy,* 109(1), pp. 203-228.

Levitt, Steven D., and Stephen J. Dubner (2005): *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, New York, Harper Collins Publishers.

Levitt, Steven D. (2004): "Testing Theories of Discrimination: Evidence from *Weakest Link*," *Journal of Law and Economics*, 431-452.

List, John A. (2004), "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *Quarterly Journal of Economics*, 119(1), 49-89.

Manski, Charles (1995), *Identification Problems in the Social Sciences.* Cambridge: Harvard University Press.

Nardinelli, Clark and Curtis Simon, (1990), "Customer Racial Discrimination in the Market for Memorabilia: The Case of Baseball," *Quarterly Journal of Economics*, Volume 105 (3), 575-595.

National Research Council (2004), *Measuring Racial Discrimination.* Panel on Methods for Assessing Discrimination, Rebecca M. Blank, Marilyn Dabady and Constance F. Citro, Eds.

Neumark, David, (1999), "Wage Differentials by Race and Sex: The Roles of Taste Discrimination and Labor Market Information", *Industrial Relations*, July.

Persico, Nicola (2002), "Racial Profiling, Fairness, and Effectiveness of Policing." *American Economic Review*, 92(5), pp. 1472-97.

Persico, Nicola and Petra Todd (2004), "Using Hit Rates to Test for Racial Bias in Law Enforcement: Vehicle Searches in Wichita." Unpublished paper, University of Pennsylvania.

Raphael, Steven, Michael A. Stoll, and Harry J. Holzer (2000): "Are Suburban Firms More Likely to Discriminate against African-Americans?," *Journal of Urban Economics*, Vol. 48, 485-508.

Riksheim, Eric C. and Steven M. Chermak (1993), "Causes of Police Behavior Revisited." *Journal of Criminal Justice*, 21(2), pp. 353-382.

Ross, Stephen L. (1997), "Mortgage Lending Discrimination and Racial Differences in Loan Default: A Simulation Approach." *Journal of Housing Research*, 8(2), pp. 277-297.

Ross, Stephen L. (2003), "What is Known about Testing for Discrimination: Lessons Learned by Comparing across Different Markets." Working Paper 2003-21R, University of Connecticut.

Ross, Stephen L. and John Yinger (1999),"The Default Approach to Studying Mortgage Discrimination: A Rebuttal," in Margery A. Turner and Felicity Skidmore, eds., *Mortgage Lending Discrimination: A Review of Existing Evidence.* Urban Institute: Washington D.C.

Ross, Stephen L. and John Yinger (2002), *The Color of Credit: Mortgage Discrimination, Research Methodology and Fair-lending Enforcement.* Boston: MIT Press.

Ruud, Paul A. (2000): *An Introduction to Classical Econometric Theory*, New York, Oxford University Press.

Sherman, Lawrence W. (1980), "Causes of Police Behavior: The Current State of Quantitative Research." *Journal of Research in Crime and Delinquency*, 17(1), pp. 69-100.

Stoll, Michael A., Steven Raphael, and Harry J. Holzer (2004): "Black Job Applicants and the Hiring Officer's Race," *Industrial and Labor Relations Review*, Vol. 57, 267-287.

Stone, Eric W. and Ronald S. Warren Jr., (1999), "Customer Discrimination in Professional Basketball: Evidence From the Trading-Card Market," *Applied Economics*, Volume 31, 679-685.

Train, Kenneth (2003): *Discrete Choice Methods With Simulation*, Cambridge University Press.