THE ECONOMICS AND MEASUREMENT OF RACIAL BIAS IN LAW ENFORCEMENT

By

BRADY PATRICK HORN

A dissertation submitted in partial fulfillment
of the requirements of the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Economic Sciences

MAY 2009

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of BRADY PATRICK HORN find it satisfactory and recommend that it be accepted.

Jill J. McCluskey, Ph.D., Chair

Ron C. Mittelhammer, Ph.D.

Nicholas P. Lovrich, Ph.D.

ACKNOWLEDGEMENTS

My sincere thanks go to my advisor Dr. Jill McCluskey. I have been fortunate to have such a remarkable mentor, who I am indebted to for her immense help in all aspects of my research and professional growth. I would like to thank Dr. Ron Mittelhammer for help with methodology and statistics. I am grateful to Dr. Rodney Fort and Dr. Jonathan Yoder for insightful comments, mentorship and tutelage. I would also like to thank Dr. Nicholas Lovrich and Michael Gaffney for guidance on issues relating to the operation of the criminal justice system. Finally, I would like to acknowledge the Washington State Patrol for providing access to their breathalyzer data.

THE ECONOMICS AND MEASUREMENT OF RACIAL BIAS IN LAW ENFORCEMENT

Abstract

By Brady Patrick Horn, Ph.D.
Washington State University
May 2009

Chairs: Jill J. McCluskey

This dissertation consists of three independent but related papers. The first paper develops a new empirical approach to measure racial bias which features fixed effects generated at an agency level to account for differences in the regional cost of assessments, as well as mitigate omitted variable bias. This empirical approach is applied to a unique dataset, which contains the results of all breathalyzer tests administered in Washington State from 2003 through 2006. Applying both the fixed effect test and conventional tests for racial bias to the Washington State Patrol dataset, I find evidence of racial inequality exhibited in assessments of both black and Asian motorists. However, the traditional test for racial bias suggests that police officers are racially biased in favor of Hispanic motorists. In contrast, using the fixed effects model, this outcome is no longer significant. This finding provides evidence that it is important to incorporate heterogeneous regional costs of assessment into models of racial bias.

The second paper explores a significant problem of studying racial bias in law enforcement called the infra-marginality problem. That is, if the underlying distributions of the probabilities of guilt differ by race, the link between assessment outcomes and racial bias will be imperfect. In this paper I model the potential for the distribution of characteristics to differ by race, and then use simulations to compare how these differences in characteristics impact

existing tests for racial bias. I find that individually the two most common tests for racial bias will be unreliable tests for racial bias; however, a test can be constructed which combines the two, and that test will be robust to the criticism of infra-marginality.

The third paper of my dissertation employs a representative voter model to study the referendum that led to the construction of Qwest Field in Seattle. Voters "believed" the all-or-nothing threat that characterizes all sports facility votes. In addition, there are proximity values since the odds of voting yes decrease with distance from the proposed facility site and increase with population. All-in-all, the nine closest counties, also more highly populated and richer counties, had their way in this election.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Motivations

The Civil Rights Act of 1964 created legal cause of action in Federal Courts in claims of discrimination on the basis of age, race, or gender under the Equal Protection Clause of the 14th Amendment to the U.S. Constitution. Although it is a widely held belief that racial bias exists in law enforcement, litigants using this cause of action have found it difficult to prove that an officer singled out a suspect because of race. In this dissertation I study racial bias exhibited in drinking and driving enforcement. I take advantage of a tremendous dataset provided by the Washington State Patrol which contains information on every DUI assessment conducted in Washington state from January 2003 to October 2006. A challenge to measuring racial bias with these data, as with most criminal justice datasets, is that we only have access to information about individuals who are assessed for a crime rather than every individual who could have been assessed for the same criminal offense. If the propensity to commit crime, which is unobserved, differs by race, the link between assessment outcomes and the marginal assessment decision will be imperfect. This is known as the infra-marginality problem. A potential solution to the infra-marginality problem was proposed by Knowles, Persico, and Todd (2001). In studying drug trafficking enforcement they show that if motorists and officers act strategically with one another, racial bias can be tested by comparing the average success rates officers exhibit with different racial groups. If officers are significantly less successful with motorists of a certain racial group it would suggest racial bias.

In the first paper (chapter two) I extend this model to allow for differences in regional costs of assessment. I show that if different regions have different assessment costs, and are at the same time correlated with race, testing for racial bias using average success rates will be unreliable. Instead, I suggest a parametric test for racial bias which uses fixed effects, calculated at the agency level to account for differences in regional cost of assessments. An advantage of this test is that it allows more information about the motorist to be included, thus reducing the potential for omitted variable bias. I then apply both Knowles, Persico, and Todd's test for racial bias and the fixed effect logit test to the Washington State data, and compare the relative effectiveness of both tests.

In the second paper (chapter three) I further investigate the infra-marginality problem. While, this criticism is well established, besides the obvious potential for statistical bias, little is known about the effect that differences in the propensity to commit crime, by race, have on the existing tests for racial bias. In this paper I create a model of racial bias which allows the distribution of the propensity to commit crime to differ by race. Unfortunately, little is known about these distributions, so to make inferences about the effect changes in these distributions have on the existing tests for racial bias I use simulations. From these simulations a simple test is derived which is robust to the infra-marginality problem. I then apply this robust test for racial bias to the Washington state data.

The third paper (fourth chapter) is a relatively independent paper, which studies the outcome of the statewide referendum that led to the construction of Qwest Field, a football stadium in Washington state. An interesting element of this referendum is the all-or-nothing threat of losing an NFL franchise. Paul Allen purchased a $10 million option to buy the Seahawks and made it clear he would only exercise this option if a new, publicly funded stadium was built for the team. If the option expired, the team would likely be relocated. I implement a representative voter model to determine what voter characteristics, including geographic characteristics, impacted voting outcomes.

Summary of Findings

In the first paper, applying Knowles, Persico, and Todd's test for racial bias I find that officers are, on average, less successful with black and Asian motorists than white motorists. According to the hypothesis of the test this suggests racial bias against both of these racial groups. However, curiously, the test suggests that officers are significantly more accurate in their assessments of Hispanic drivers, which suggests racial bias against white motorists in favor of Hispanic motorists. In comparison, implementing the fixed effect logit model, I no longer find statistically significant evidence of racial bias in favor of Hispanic motorists. This finding suggests that the low-income, highly Hispanic middle part of Washington state has a higher cost of assessment than other parts of the state, a fact which can bias the conditional means test, but is appropriately accounted for in the fixed effects of the parametric model. This analysis provides evidence that it is important to incorporate heterogeneous costs of assessments into models of racial bias. Additionally, I find that officers are most successful with middle-aged motorists, slightly less successful with female motorists, and more successful at night.

In the second study I find that, when the distribution of characteristics differs by race, both search rate and find rate tests are unreliable tests for racial bias. When there are more, high probability outcomes, search rate tests will overestimate racial bias, and find rate tests will underestimate racial bias. However, if both tests suggest racial bias it cannot be due to differences in the distribution of characteristics by race. Applying this test to Washington state breathalyzer data I find that officers appear to be racially biased against black motorists; however, the opposite results are found with Hispanic motorists. This could suggest inverse racial bias against white motorists in favor of Hispanic motorists; however, it could also be due to a larger noise component associated with assessments of motorists of this racial group.

In the third paper I find that in Referendum 48 voters "believed" the all-or-nothing threat that characterizes all sports facility votes. The odds of a yes vote decreased with distance from the proposed facility site and increased with population (at a decreasing rate); and the odds of a yes vote were higher in counties with more college-educated voters. I also found that the odds of a yes vote increased in counties with higher unemployment. All-in-all, the nine counties closest to the proposed construction (these were also more highly populated and richer counties) had their way in this election.

CHAPTER TWO

MEASURING RACIAL AND ETHNIC BIAS IN DRINKING UNDER THE INFLUENCE ENFORCEMENT

Summary

Although it is a widely held belief that racial bias exists in law enforcement, litigants have found it difficult to prove that an officer singled out a suspect because of race. This paper models the interaction between motorists and police and develops an alternative empirical approach to measure racial bias. In previous work, perfect observability by police is implicitly assumed, and it has been further assumed both that motorists have control over discretionary characteristics and there are uniform regional costs of police assessment. We construct an empirical test that compares officer find rates while accounting for additional signals used by officers. We use fixed effects generated at the agency level that account for deviations in the regional cost of assessment as well as mitigate the impact of omitted variable bias. The empirical application of the model is based on a unique dataset containing the results of all breathalyzer tests administered from 2003 through 2006 by the Washington State Patrol. We test for the existence of racial bias exhibited by officers conducting assessments for driving under the influence of alcohol (DUI). Both the conditional means tests and our new parametric test indicate that officers are less accurate with black and Asian motorists. However, the conditional means test suggests police officers are more accurate with Hispanic motorists than white motorists, which is consistent with bias in favor of Hispanic motorists. In contrast, using our parametric test that

incorporates fixed effects, the effect of Hispanic motorists is not significant, which we interpret as likely due to the low-income, highly Hispanic middle part of the state having a higher cost of assessment than other parts of the state. We suggest that the higher assessment cost leads to a statistical bias in the conditional means test, but is accounted for in our fixed effects model, leading to a more accurate assessment of racial bias.

I. Introduction

The Civil Rights Act of 1964 created legal cause of action in Federal Courts in claims of discrimination on the basis of age, race, or gender under the Equal Protection Clause of the 14th Amendment to the U.S. Constitution. Although it is a widely held belief that racial bias exists in law enforcement, litigants using this cause of action have found it difficult to prove that an officer singled out a suspect because of race. Over 40 years later, the question remains: How does one enforce a rule for equitable treatment that is extremely difficult to operationalize? Certainly, some degree of progress has been made in understanding racial discrimination. Field studies conducted in labor, housing, and automobile markets have provided clear evidence of both intentional and unintentional racial bias (Bertrand and Mullainathan 2002; Neumark 1996; Yinger 1995; Zhao 2005; Zhao, Ondrich and Yinger 2006). Laboratory experiments conducted by psychologists have provided evidence of subtle, axiomatic forms of racial bias (Greenwald, Oakes and Hoffman 2002). Unfortunately, proving racial bias outside the lab has proven considerably more challenging. Market-based studies are usually limited by omitted variables bias. The correlation in the United States between race and other demographic variables such as income and education compounds this problem (Kennedy et al., 1998). Yet, the need for a quantitative test for racial bias is greater than ever. In 1992 the Supreme Court issued a series of decisions on case law which resulted in the need for more stringent evidence than in the past to support claims of racial bias. How to measure racial bias remains a research question of considerable currency.

The question of racial bias is particularly critical for the field of law enforcement, where over the last 20 years police agencies frequently have been accused of racial bias in carrying out their public safety duties. With the help of organizations such as the American Civil Liberties Union (ACLU), National Association for the Advancement of Colored People (NAACP), and the Urban League, a number of plaintiffs have been successful in cases brought against state and local police authorities. However, claims of racial bias that lead to verdicts of guilt on the part of public agencies are in the minority. For example, in 2007 the State of California dismissed all 322 racial profiling cases it faced because they were found to be “unfounded” or lacked sufficient evidence (*LA Times* 2007). The rationale given for the blanket dismissals is that it is almost impossible to prove that an officer singled someone out because of race alone. It is generally agreed that more detailed data are required to provide proof of the existence of racial bias.

Fortunately, the quality of data available for use in investigating issues relating to racial bias in traffic stops is increasing. As of 2004, 29 states instituted data collection processes to record motorists’ race or ethnicity during traffic stops. In addition, technological innovations in information gathering, storage, and retrieval allow officers to capture more details associated with traffic stop assessments. These data allow for more accurate quantitative tests of racial bias to be carried out than was possible in the past.

Attempts to analyze quantitatively claims made of racial bias in law enforcement have mainly focused on two statistics: the “search rate” and the “find rate.” The search rate is calculated by dividing the number of individuals in a racial group who are assessed (or searched) for a crime by the total number of individuals within that racial group in the general population of interest. All else equal, significantly different search rates across racial groups are suggestive of police officers targeting some racial group more than others. In the economics literature, this is referred to as “statistical discrimination” (Arrow 1973). The use of this statistic has been criticized for two major reasons. First, the population of a racial group at risk of an assessment is difficult to obtain. This is referred to as the “denominator problem.” Second, racial disparities in assessment rates do not necessarily imply racial bias. Crime rates and race have been historically correlated (Blumstein 1982, 1993; Knowles 2001; Engel 2002).

The find rate statistic (also called “outcomes test”) is calculated by dividing the number of individuals in a racial group found to be guilty by the total number of individuals searched in that racial group. If there is no racial bias, then the expected probability of guilt of the marginal suspect should be equal across racial groups. Using this statistic is appealing because it avoids many of the search rate’s inherent pitfalls such as unequal crime rates across racial groups. Becker (1957) examined discrimination in labor markets and focused on biased outcomes rather than biased intentions. Becker’s outcomes tests used wage rates to test for racial bias. Becker also noted that if banks discriminate against minorities, one should expect loans made to minorities to have lower default rates (for a discussion, see Ayers 2002). A major problem with taking this approach, however, is that researchers are typically unable to observe the marginal suspect. Moreover, owing to an omitted variable dilemma, average find rates are typically used.

However, inequality of the average find rates does not necessarily imply inequality of the marginal find rates. This problem has been referred to as the "infra-marginality" problem (Anwar and Fang 2006).

In a seminal paper, Knowles, Persico, and Todd (2001, hereinafter KPT) provide an answer to the infra-marginality impasse with their economic analysis of racial bias in drug-trafficking enforcement in Maryland. In a "matching-pennies" type game, KPT assume that police officers choose which cars to search based on weighted motorists' characteristics in order to maximize the number of convictions relative to the cost of search. In turn, motorists consider the expected net benefit from trafficking drugs. Since motorists know the officers' search criteria, they adjust their discretionary characteristics in order to minimize the probability of being searched. Racial prejudice or discrimination is suggested when the officer's cost of assessing one racial group is lower than the cost of assessing another racial group. A mixed-strategy equilibrium is derived for each racial group in which police officers randomize over whether to search, and motorists randomize over whether to traffic drugs. With the random searches, racial bias can then be tested by simply comparing the average find rate across racial groups. Using drug enforcement data from Maryland, KPT do not find evidence of racial bias toward black motorists, but they do find some evidence of bias toward Hispanics.

KPT made a significant impact on how racial bias is quantified. Their test can be implemented even with incomplete datasets and it effectively addresses problems with earlier work. However, some researchers have expressed criticisms of their assumptions. Dharmapala and Ross (2004), for example, argue that officers do not observe motorists with probability one. This disrupts the single mixed-strategy equilibrium found in KPT because some motorists will traffic drugs regardless of the probability of assessment. Consequently, subsequent find rate tests will have potentially significant bias.

Anwar and Fang (2006) question the assumptions that motorist characteristics are discretionary and that trooper behavior is monolithic. KPT's model implicitly assumes discretionary characteristics, which may be defensible for drug trafficking but may not be for many other law enforcement settings. If motorists do not have the ability to manipulate their characteristics, then the KPT equilibrium breaks down. If motorists' characteristics other than race provide no information on the probability of guilt, officers could pull over automobiles at random with equal success. This is not only unlikely, but goes strongly against established law enforcement guidelines. This criticism is also expressed by Dominitz (2003). For KPT's model to be valid, motorists "…must not only know all of the information that officers use but also know how officers use this information to determine search rates. Otherwise, the find rate tests are unreliable," (Dominitz, page 429). In other words, unless motorists have complete information about the characteristics used by police officers and the corresponding weights applied to those characteristics, a uniform, mixed-strategy equilibrium will not exist. Motorist characteristics will again become informative, and find rate tests will be biased.

Examining a different law enforcement setting requires alternative assumptions and leads us to develop a new approach to the measurement of racial bias. We construct an empirical test that compares officer find rates while accounting for additional signals used by officers. We use fixed effects generated at the agency level that account for deviations in the regional cost of assessment as well as mitigate the impact of omitted variable bias. Our first assumption is that in driving under the influence of alcohol (DUI) enforcement, it is unlikely that all motorists' characteristics other than race are discretionary to the motorist. For example, the level of intoxication, *ex post*, but before the decision to drive, cannot be manipulated except by time. Second, we allow for regional differences in the cost of assessment, which is appropriate when the data set is obtained from a large and heterogeneous geographic area. Thirdly, we follow Dharmapala and Ross (2004) in allowing for the possibility that officers do not observe all motorists, so that it is quite possible that a drunk driver may not cross the path of any police officers.

We extend the racial bias assessment literature by developing a test for racial bias that assumes that some characteristics are not discretionary to the motorist and is robust to differences in regional costs of assessment. For comparison, we apply both conditional means tests and our fixed-effects tests to a unique dataset provided by the Washington State Patrol. Both tests generate results suggesting unequal treatment of both Asian and black motorists relative to white motorists. However, using the conditional means test, police officers are indicated to be more accurate with Hispanic motorists than white motorists, which would suggest bias *in favor of* Hispanic motorists. Our fixed effects model does not conclude such reverse discrimination. An explanation for the divergence in results is that the fixed effects model correctly incorporates the regional variation in the costs of traffic stop assessment into the analysis of bias.

II. Theoretical Framework

The general set-up of our model has some similarities to KPT's model, but is adapted to the problem of DUI enforcement and modified through a number of alternative assumptions. A continuum of police officers and motorists is assumed. The race of the motorists is represented by $r \in \{W, B, H\}$, where W = White, B = Black, and H = Hispanic.

A. Exogenous Characteristic and Police Observation

For the sake of exposition, the motorists' characteristics are denoted, as in KPT, by a one-dimensional variable $c \in [0,1]$ which is determined by the driver's level of intoxication (considered to be exogenous at this point in time), and cannot be readily manipulated by the driver. Note this differs from KPT's characteristic *c*, which they assumed could be manipulated by the motorist. The officer's cost of a traffic stop assessment is represented by t_r, which may vary by race. Following KPT, the benefit of a successful assessment (apprehending a drunk driver) is normalized to one, and this normalization also scales the cost function to be a fraction of the benefit. An officer is racially prejudiced when his or her normalized cost of assessment differs by race, e.g. $t_B \neq t_W$. Motorists consider the benefit of driving after drinking, $v(c,r)$, the cost of a DUI[1] represented by $-j(c,r)$, and the probability of being assessed $\gamma(c,r)$. Following Dharmapala and Ross (2004), we assume that the motorist is observed by police with probability $m \in (0,1)$. The net expected payoff of drinking and driving is then calculated as follows:

[1] For purposes of this analysis, we assume that conviction will occur with certainty if the driver is above the .08% legal limit for alcohol. This assumption could be relaxed by adding uncertainty over conviction, but this is beyond the scope of the paper.

$$m\left\{\gamma(c,r)\left[-j(c,r)\right]+\left[1-\gamma(c,r)\right]v(c,r)\right\}+(1-m)v(c,r). \qquad (1)$$

The motorist will drive after drinking if this expression is positive, is willing to randomize if the expression is zero, and not drive after drinking if the expression is negative.

In deciding whether to assess each motorist the officers consider the cost of assessment t_r and the probability of guilt $P\left(G|c,r\right)$, where the derivative of this probability with respect to *c* is positive, i.e. $P_c\left(G|c,r\right)>0$. Thus the higher the level of c exhibited by a motorist, the greater the probability of guilt. The police officer's decision whether to assess motorists of type *c* and *r* is determined $\forall(c,r)$ by the solution to

$$\max_{\gamma(c,W),\gamma(c,B)\gamma(c,H)}\sum_r\int\left[P\left(G|c,r\right)-t_r\right]\gamma(c,r)\,f\left(c|r\right)dc. \qquad (2)$$

where $f\left(c|r\right)$ represents the density of the distribution of the characteristic c in (given) the r-type population.

It is evident from the maximization objective in (2) that if $P\left(G|c,r\right)-t_r>0$, then $\gamma(c,r)=1$ and the officer will choose to assess a motorist of type (r, c) for a DUI. If $P\left(G|c,r\right)-t_r<0$, $\gamma(c,r)=0$ and the officer will choose not to assess the motorist. If $P\left(G|c,r\right)-t_r=0$, the officer is indifferent and is willing to randomize the choice of assessment. Let c_r^* represent the marginal motorist for racial group *r* such that $P\left(G|c,r\right)-t_r=0$. All observed motorists of group *r* exhibiting characteristics greater than c_r^* are assessed, and all motorists of group *r* with a characteristic level under c_r^* are not assessed. This is depicted in Figure 1. If the cost of assessment is greater for a given race, it will increase the level of characteristics exhibited by the marginal motorist, and assessments will be restricted to motorists with higher probabilities of guilt. This is depicted in Figure 2.

In equilibrium, the motorist will drive if $c < c_r^*$, because the motorist's c is below the characteristic level of the marginal motorist, so that he/she will not be assessed if observed. Also, the motorist will drive if $\gamma < \frac{v(c,r)}{m(j(c,r)+v(c,r))}$ since in this case the probability of being observed by the police is low enough for the expected pay off from drinking and driving (1) to be positive. Since the equilibrium depends on c, motorist's characteristics are informative in assessing racial bias in DUI assessments.

In contrast, KPT base their test for racial bias on average find rates:

$$\sum_{r \in R} \frac{\left(\hat{P}_r - \hat{P}\right)^2}{\hat{P}} \sim \chi^2\left(R-1\right), \tag{3}$$

where *R* is the cardinality of the set of race categories, and $\hat{P}_r$ and $\hat{P}$ are the conditional and unconditional estimated find rates, respectively. If officers are significantly less accurate with black motorists, then racial bias is suspected. The statistic in (3) tests for racial bias without the use of any motorist characteristics other than race. This test is derived from a model in which motorists are able to manipulate their characteristics to minimize their probability of being assessed. The intuition in the case of drug trafficking is straight forward. For instance, one could speculate, at some given time, that drug traffickers choose to have tinted windows to avoid detection of wrongdoing. Officers recognizing the increased probability of success associated with searching motorists with tinted windows will increase their assessments of motorists exhibiting this characteristic. Once the find rate increases for motorists exhibiting tinted windows, drug traffickers will stop using vehicles with tinted windows to traffic drugs. This reaction renders tinted windows to be a meaningless enforcement signal. Extending this logic to every characteristic that can be manipulated by the potential drug trafficker eventually causes each characteristic to become uninformative.

This logic does not likely extend to DUI enforcement, however. First, drinking and driving is a more impulsive crime, whereas drug trafficking is a more premeditated/planned event. It is unlikely that all motorists who drink and drive have the foresight to manipulate their characteristics in order to be inconspicuous. Second, the expected benefit as well as the expected cost of drinking and driving is significantly lower than is the case with trafficking drugs. It may not be realistic for motorists to exhaust the necessary resources to uncover the characteristics and weights used by officers to predict guilt. The inability or inaction by even a subset of the population of motorists to alter their characteristics in order to minimize the probability of assessment will cause some motorists' characteristics to become informative, albeit noisy signals for officers.

B. Heterogeneous Regional Costs of Assessment

The marginal cost of DUI assessment may differ by location, and we will show in this section how this can influence the results of tests for racial bias. It may be the case that a more affluent region can afford more police officers per capita. Having additional officers available to deal with other costly crimes, an officer will have a lower cost of assessment for possible DUI infractions. Less predictably but with the same outcome, some police forces simply place a larger emphasis on DUI prevention than others. A region's sheriff and police chiefs can subjectively choose the level of emphasis placed on DUI prevention and allocate a higher percentage of the police force to DUI prevention. This encourages officers to assess more motorists out of the pool of potential drunk drivers. However, because officers are incentivized to assess motorists with positive, but lower, probabilities of guilt, the success rate exhibited by the regional police force will decrease. The officer's marginal cost of assessment, t_{rl}, can be allowed to differ by both race (r) and the location of assessment *(l)*, which will extend the preceding theoretical model to incorporate differences in the regional cost of assessment. In this case, the officer's counterpart to the optimization problem (2) can be expressed as

$$\max_{\gamma(c,W),\gamma(c,B)\gamma(c,H)} \sum_r \int \left[P\left(G \middle| c,r\right) - t_{rl} \right] \gamma\left(c,r\right) f\left(c \middle| r\right) dc \qquad (4)$$

In this extended model, if the motorist exhibits characteristics such that the expected benefit of a successful assessment is greater than the *regional cost* of assessment, then the officer chooses to assess the motorist for a DUI. Otherwise, the motorist is not assessed. All of the conditions that relate to the choice of assessment probabilities and the associated decisions with regard to assessment that were presented following (2) apply identically, with t_r replaced by t_{ri}. If the cost of assessment is greater in a given region, it will increase the level of characteristics exhibited by the marginal motorist. In KPT's model, racial prejudice is defined (see Definition 1, page 210) as unequal costs in searching motorists across races. For KPT's model, in regions with higher costs, assessments will only be conducted on motorists having higher probabilities of carrying contraband.

It is likely that the regional costs of assessment vary within a statewide dataset, especially across rural and urban regions. If the costs of assessment vary by region, and regions are unequal in their racial composition, means tests may be deceiving. Figure 3 offers a hypothetical example of a state containing two distinct regions (we will refer to them as Region 1 and Region 2) and two distinct racial groups, black and white. Assume the police in both regions are racially unbiased (using KPT's definition) and that Region 1 is predominantly white (80%), while Region 2 has equal proportions of black and white residents. A region in which there is no racial bias will have the same costs of assessment conditioned on race, thus $t_{W1} = t_{B1}$ and $t_{W2} = t_{B2}$. However, if the costs of assessment vary by region, the success rates can differ across race when the regions are aggregated. When we aggregate across regions in this hypothetical example, the statewide average find rate for white motorists is 81.3%, whereas the statewide average find rate for black motorists of 55.37%. Using a means test for prejudice, the researcher would infer racial bias when, by construction, we know that the police in both regions are not racially biased. Thus, if assessment costs differ by region, and if regions differ in their racial composition, then conditional aggregate find-rate tests will differ from those controlling for region.

III. Empirical Analysis

A. Data

In Washington state, the legal blood alcohol content (BAC) limit is 0.08%. Any assessed motorist with a BAC over this threshold is given a citation for DUI. Enforcement of this law is conducted in two stages. In the first stage, an officer observes the characteristics of a vehicle suspected of being operated by a driver that is under the influence. If sufficient characteristics are exhibited by the vehicle, the officer proceeds to the second stage and pulls over the motorist to conduct a field sobriety test. At this point, more informative signals such as the motorist's breath, speech, and motor skills are evaluated. If the officer deems the probability of guilt sufficiently high, the motorist is transported to a regional testing location. At the regional testing location, a calibrated and carefully maintained machine is used to conduct a breathalyzer test, and the results of two passes of the machine are documented, including the time of the test and demographic details about the motorist.

The multi-step enforcement process is represented in Figure 4. It is important to note that data are only collected in the final stage of the enforcement process. If the officer determines the probability of guilt is insufficient to administer a breathalyzer test, the motorist is released and this driver contact is not included in the dataset. Also note that while conducting the field sobriety test, officers are able to reconsider the probability of guilt. This renders many superficial signals observed in the first stage, such as speeding, uninformative in the decision to administer a breathalyzer.

The dataset used here contains all DUI assessments conducted in Washington state from January 2003 to October 2006. The variables include information on the race, gender, and age of the observed motorist, as well as time of assessment and assessing agency. We also know whether the assessing officer received advanced training for drug-induced impairment. Table 1 summarizes the variables used in our empirical tests. Of the 68,692 breathalyzer tests administered, almost 22 percent were performed on females. Just over 5 percent of the breathalyzer tests were administered on blacks, 2.71 percent were administered on Asians, and 3.26 percent were administered on Hispanics.[2] The variable *night* is an indicator variable equaling 1 if the assessment took place between 10:00 PM and 2:00 AM. Just over half of breathalyzers were administered during this time period. The indicator variable *late night* equals 1 if an assessment occurred in the time period between 2:00 AM and 4:00 AM, and just over 31 percent of the breathalyzers were administered during this time period. The variable *Drug Recognition Training* is an indicator variable equaling one if the officer is specially trained in drug recognition. About 11 percent of the breathalyzers were conducted by an officer having advanced drug recognition training.

Table 2 compares Washington's ethnic makeup with the composition of the breathalyzer data. Inferences can be made about the probability that a motorist of a particular race will be administered a breathalyzer. A difference between the percentage of breathalyzers administered to individuals in a specific racial group and the percentage of individuals within that racial group in the general population does not necessarily indicate racial discrimination. It is an indication of statistical discrimination in the sense of Arrow (1973). It should be noted that population percentages are an incomplete proxy for actual drivers on the road. With those caveats, we note

[2] Washington state started racial coding for Hispanic in 2005. In 2003 and 2004 Hispanic individuals were coded as white. The percentage for 2005 and 2006 is 9.2%.

that African Americans represent just over 5 percent of the breathalyzers administered, but constitute only about three-and-half percent of the population in Washington state. In terms of statistical discrimination, this finding means that black motorists are administered a breathalyzer 48 percent more often than the average motorist. Asians, on the other hand, are administered a breathalyzer 63 percent less often than the average motorist.

B. Empirical Model

In this section, we implement an alternative method for testing racial bias which accounts for unequal regional costs of assessment. Note that it is also possible to conduct aggregate means tests adjusting for unequal regional cost of assessment. One could condition the data on both region and race, and then calculate average find rates per race, per region. If, in each region, the conditional find rate was lower for one race, it would suggest racial bias. Unfortunately, regional sample size limitations may limit the implementation of this method, especially when datasets contain little racial diversity in distinct regions. Any future attempts to implement aggregate means tests adjusting for unequal regional cost of assessment would need to create a criterion for measuring racial bias using multiple aggregate means. Instead of carrying out within region comparisons a regression framework with fixed effects for each region is used. This method allows conditioning on many characteristics at once, and also allows the cost of assessment to vary by location.

The "find rate" test we implement closely follows that of KPT. Racial inequality is quantified by measuring the differences in success rates across racial groups. Racial inequality is defined as one racial group having a lower probability of guilt than another:

$$P\left(BAC \geq 0.08 \middle| c,B\right) \neq P\left(BAC \geq 0.08 \middle| c,W\right) \tag{5}$$

In contrast to KPT, we specify the probability of a successful assessment parametrically.[3] This is advantageous because it allows the probability of the BAC being above the legal limit to be a function of not only race but also of other characteristics that cannot be readily manipulated, such as age and gender.[4] Also, the probability of the BAC being above the legal limit can be specified as a function of the time of day, which is a beneficial feature of the model since the probability of drinking and driving is expected to be correlated with the time of day. In basic general form, the probability of guilt is specified as the following linear index model:

$$P\left(BAC \geq 0.08 \middle| c,R\right) = F\left(X_m\beta_m + X_0\beta_0 + X_t\beta_t\right) \tag{6}$$

where X_m indicates characteristics exhibited by the motorist, X_0 indicates characteristics exhibited by the officer, and X_t represents an indicator for the time of assessment. To estimate this binary choice model, a cumulative distribution function $F(\cdot)$ needs to be specified and then

[3] This is discussed as a statistical option by KPT.

[4] Using conditioned success rates, KPT was able to make inferences on characteristics other than race. This was done by conditioning the probabilities of guilt on characteristics other than race and other characteristics at the same time. For instance, the probability of successfully assessing a white woman compared to that of black woman.

optimized using maximum likelihood.

If differences in regional assessment cost have an effect upon success rates, and race is correlated with location, then omitting location will cause the parameters for race to be biased, rendering any test of racial bias unreliable. To capture the effect that location plays on success rates, fixed affects are incorporated to adjust for deviations in success rates by location. This is represented by:

$$P_l\left(BAC \geq 0.08 \middle| c,R,l\right) = F\left(\alpha_l + X_{lm}\beta_m + X_{l0}\beta_0 + X_{lt}\beta_t\right) \tag{7}$$

where α_l is a location-specific indicator variable. Estimation using location-specific indicator variables can be problematic for at least two reasons. First, depending on the level of aggregation for location, this could significantly increase the numbers of parameters to be estimated. However, statewide assessment datasets often have a large number of observations, which can accommodate reasonably high dimensional model parameterizations. A second potential drawback is the treatment of locations that have few assessments, especially those locations having mostly, or all positive or negative outcomes. Traditionally, locations with all positive or all negative outcomes are dropped.[5]

However, this technique is somewhat limited in its ability to control for correlations between multiple characteristics.

[5] An alternative method, unexplored in this paper, would be to aggregate locations with few observations into larger regional groups.

C. *Empirical Results*

For purpose of comparison, we first present results of tests based on the conditional average find rates. Following KPT, the null hypothesis of no racial bias is represented as:

$$P\left(G=1\middle|c,r\right)=P\left(G=1\right) \quad \forall r,c \tag{8}$$

Conditional find rates are calculated by race, and Pearson χ^2 tests are used to test for significance differences. Results are reported in Table 3 and Table 4. The results of this test indicate significant differences in success rates across races. Officers are 3.96% less accurate in their assessments of black motorists than white motorists. The differences in success rates are more pronounced with Asian motorists, where they are 5.72% less accurate in their assessments than for whites. However, curiously, the test suggests that officers are significantly more accurate in their assessments of Hispanic drivers. It is also found that gender has a subtle impact on accuracy, which was an effect also found by KPT in their analysis of Maryland drug trafficking. They hypothesized that the mostly male officers might potentially derive extra utility from assessing women. Finally, the results suggest that officers are more successful later at night.

Now consider, in comparison, our parametric method for measuring racial bias. To measure officer find rates, we construct an indicator variable taking the value of 1 if the assessed motorist's blood alcohol content is over 0.08, and 0 otherwise. We include the following motorist characteristics: gender, race, and age. Gender and race are represented by indicator variables. Age and age squared are included to capture potential nonlinear effects of age. Finally, an indicator variable for drug recognition training is incorporated into the regression, the hypothesis being that officers with drug recognition training will have higher assessment success rates because they are better able to recognize that a motorist is under the influence of drugs rather than alcohol.

Accounting for differences in regional costs of assessment is accomplished by incorporating the aforementioned fixed effects into the model. The potential location effects are specified at the agency level. Pragmatically, this is the smallest level of aggregation available in the dataset, allowing the most flexible parametric specification. However, there are a few reasons that officers in the same agency will incur the same cost of assessment. Monetarily, officers in one agency are tied to the same budget. Also, because officers are patrolling the same region, the number of officers per crime committed is equal at the agency level. Measuring the cost of assessment as time taken away from enforcing the next most costly crime, officers in the same agency would share a common cost of assessment. Finally, and least predictably, officers in the same agency report to the same sheriff or police chief who determines the level of emphasis on DUI enforcement for that municipality or county. All else held constant, an increase in the level of emphasis will motivate officers to decrease the probability of guilt of the marginal motorist, increasing the number of motorists assessed, and decreasing their agency's respective overall success rate. Using a fixed effects approach allows for differences in assessment costs to be captured by the agency-specific indicator rather than indirectly biasing the other parameter estimates. Marginal effects from the fixed-effect model can be interpreted as the difference in success rates for that particular group. If racial bias exists in DUI assessment, the parameter for the specific racial group will be negative and significant.

The results of the fixed-effects model, based on a logit distribution, are presented in Table 5. Many of the results from this model corroborate the results of the conditional means test, but there are some important differences, and there is also additional information relating to assessment accuracy provided by the parametric fixed-effects approach. We again find that officers are less successful with black and Asian motorists than with white motorists, officers are more successful in the late night, and officers are slightly less accurate with women than men. However, in the fixed-effects logit model, we are also able to make inferences about the effect of the motorist's age. The parameter values for *age* and *age-squared* are jointly significant, with their values being positive and negative, respectively. The curve peaks at an age of 41.67, which indicates that officers are most accurate with middle-aged motorists. As expected, officers are less accurate with young motorists, but surprisingly have decreased accuracy rates for elderly motorists. Also, using the parametric approach, the coefficient for Hispanic motorists is positive, but it is no longer statistically significant. This suggests that the low-income, highly Hispanic middle part of Washington state has a higher cost of assessment than other parts of the state, which can bias the conditional means test but was accounted for in the fixed-effects of the parametric model.

We qualify our results by noting that find-rate models with binary dependant variables can be limited in their ability to measure racial bias by the infra-marginality problem (Ayers 2002). If the distributions of guilt are systematically different by race, inference about the marginal decision using population averages may be inaccurate. However, Ayers (2002) provides two justifications for the use of find rates in police searches. First, it is difficult to articulate a reason why average success rates would not be a credible proxy for marginal success rates.[6] Secondly, irrespective of infra-marginality, unequal find rates imply disparate impact. Thus, even if officers are treating motorists equally, the current assessment process subjects minorities to a disproportionate number of unsuccessful assessments.

IV. Summary and Conclusions

Analyzing how race is used as a signal in law enforcement is both challenging and controversial. However, there is a growing demand for reliable statistical tests of racial bias. To meet this demand, many states are implementing more rigorous data collection programs. Significant progress has been made in analyzing racial bias in drug-trafficking enforcement. We extend this literature to study potential racial bias exhibited by officers enforcing DUI laws. However, specific differences between law enforcement settings, such as between the drug-trafficking context and the DUI context must be addressed. In the context of DUI enforcement, we argue that already intoxicated motorists are unable to fully adjust their characteristics to minimize the probability of assessment, so characteristics can provide meaningful signals.

We also extend the current literature to allow for differences in regional costs of assessment. A main finding of this analysis is that differences in regional cost of assessments

[6] It could be argued that potential genetic predispositions to alcoholism, by race, could cause some systematic differences in probability of guilt and, hence, bias assessment rate tests. For instance, our results could be biased by a group of white motorists who were obviously drunk. Assessments of this group would be extremely successful, causing the accuracy of white motorists overall to he higher. To test for this possibility, we estimated our model with the dataset restricted to observations with blood alcohol contents under twice the legal limit (.16). With this data subset, the estimated parameters for Asian and black remain negative and highly significant.

can confound the established conditional means tests. We implement a parametric test for racial bias which uses fixed-effects calculated at the agency level to account for differences in regional cost of assessments. An advantage of this test is that it allows more information about the motorist to be included, thus reducing the potential for omitted variable bias. We implement both KPT's conditional means test and our parametric test to analyze whether racial inequality has been exhibited against African American and Asian motorists. Any evidence of racial inequality could be derived from police bias; however, one must be cautious when accusing a police department of racist practices. Many confounding details have been suggested in terms of race and equality of search rates. Language barriers, cultural differences, and drug usage have all been suggested as factors that may contribute to decreased officer accuracy.

The results from the conditional means test and our fixed-effect parametric model share a number of similarities. However, the conditional means tests conclude reverse racial discrimination favoring Hispanic motorists. In contrast, using the fixed-effects model, the coefficient for Hispanic motorists is not significant, suggesting that the low-income, highly Hispanic middle part of the state has a higher cost of assessment than other parts of the state, which can bias the conditional means test but is accounted for in the fixed-effects of the parametric model.

Bibliography

Anwar, Shemena, and Hanming Fang. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96 (March 2006): 127-151.

Ayres, Ian. "Outcome Tests of Racial Disparities in Police Practices." *Justice Research and Policy* 4 (Fall 2002): 132-142.

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

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

Blau, Francine D., and Andrea H. Beller. "Black-White Earnings Over the 1970's and 1980's: Gender Differences in Trends." *Rev. Econ. and Statis.* 74 (May 1992): 276-286.

Cornwell, Christopher, and William N. Trumbull. "Estimation The Economic Model of Crime with Panel Data." *Rev. Econ. and Statis.* 76 (May 1994): 360-366.

Dharmapala, Dhammika, and Stephen L. Ross. "Racial Bias in Motor Vehicle Searches: Additional Theory and Evidence." *Contributions to Economic Analysis & Policy* 3 (2004): 1-21.

Dominitz, Jeff. "How Do the Laws of Probability Constrain Legislative and Judicial Efforts to Stop Racial Profiling." *American Law and Economics Review* 5 (Fall 2003): 412-432.

Greenwald, Anthony G., Mark A. Oakes, and Hunter G. Hoffman. "Targets of Discrimination: Effects of Race on Responses to Weapons Holders." *Journal of Experimental Social Psychology* 39 (2002): 399-405.

Hernandez-Murillo, Ruben, and John Knowles. "Racial Profiling or Racist Policing." *International Economic Review* 45 (August 2004): 959-989.

Kennedy, BP, I Kawachi, R Glass, D Prothrow-Stith. “Income Distribution, Socioeconomic Status, and Self Rated Health in the United States: Multilevel Analysis.” *British Medical Journal* 317 (1998): 917-921.

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

Knowles, John, Nicola Persico, and Petra Todd. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *J.P.E.* 109 (February 2001): 203-229.

National Research Council. *Measuring Racial Discrimination.* Rebecca M. Blank, Marilyn Dabady and Constance F. Citro, eds. Washington, DC: The National Academies Press, 2004.

Persico, Nicola. "Racial Profiling, Fairness, and Effectiveness of Policing." *American Economic Review* 92 (December 2002): 1472-1497.

Yinger, John. *Closed Doors, Opportunities Lost: The Continuing Cost of Housing Discrimination.* New York: Russell Sage Foundation, 1995.

Yinger, John. "Evidence of Discrimination in Consumer Markets." *J. of Econ. Perspectives* 12 (1998): 23-40.

Zhao, Bo, Jan Ondrich, and John Yinger. "Why do Real Estate Brokers Continue to Discriminate? Evidence from the 2000 Housing Discrimination Study." *Journal of Urban Economics* 59 (May 2006): 394-419.

Figure 2.1: Probability of Guilt is Equal to the Cost of Assessment for the Marginal Motorist

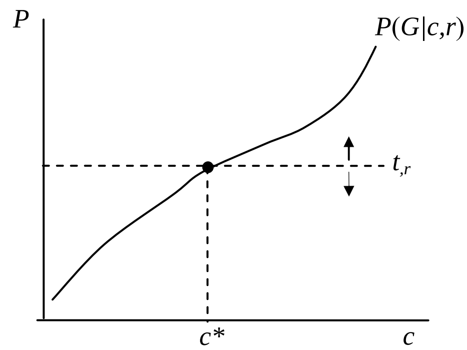


Figure 2.2: Different Assessment Costs across Racial Groups

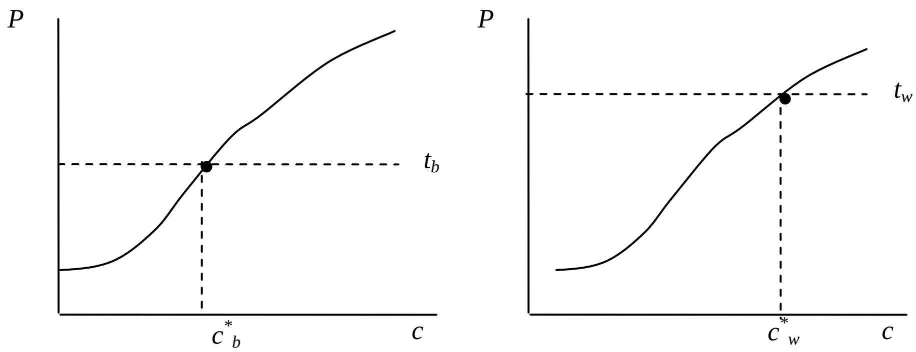
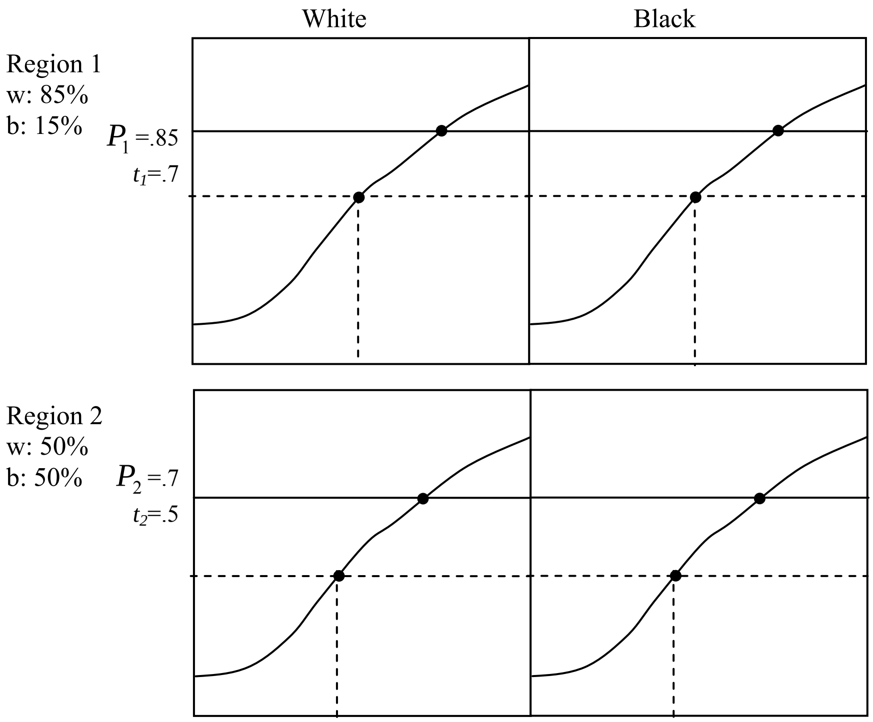


Figure 2.3: Hypothetical Example with Heterogeneous Costs across Regions



$$\overline{P}_W = \frac{.85(.85)+.5(.7)}{.85+.5} = .7944$$

$$\overline{P}_B = \frac{.15(.85)+.5(.7)}{.15+.5} = .7346$$

Figure 2.4: Outcomes from Assessment Decision

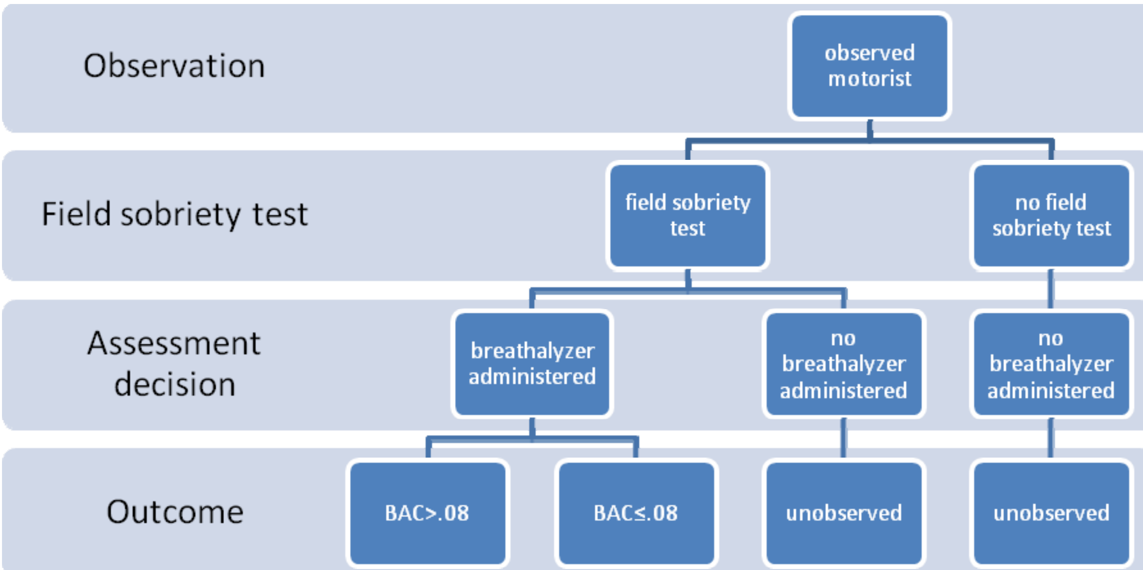


Table 2.1: Means of Variables Used in Analysis

	All observations	Female	White	Black	Asian	Hispanic
Age	34.78258	34.56477	35.12371	33.86344	32.88698	30.72289
Age Squared	1343.972	1315.736	1372.761	1253.544	1187.943	1016.794
Female	0.2188173	1	0.227225	0.160719	0.218532	0.066042
Black	0.051811	0.038055	0	1	0	0
Asian	0.0271793	0.027144	0	0	1	0
Hispanic	0.0326239	0.009846	0	0	0	1
Night (10PM-2AM)	0.5166832	0.502894	0.530461	0.457994	0.45849	0.400714
Late Night (2AM-4AM)	0.3104583	0.348813	0.297674	0.388873	0.422603	0.402499
Drug Enforcement Training	0.1044663	0.115162	0.103613	0.119135	0.101232	0.094601

Table 2.2: Comparison of Observational and Demographic Data

Ethnicity	Census% of Washington State's Population	Observations	% of Test Data
White	77.1%	58255	84.81%
Black	3.5%	3559	5.18%
Asian	8.8%	2241	3.26%
Hispanic	6.4%	1867	2.72%
Undetermined	4.2%	2770	4.03%
All races	100.0%	68692	100.00%

Note--census data is from the CDC and was last updated 2005.

Table 2.3: Conditional Breathalyzer Success Rates

	Observations	Percentage of Motorists Having a BAC>.08
All	68692	.8791417
Gender:		
Male	53661	.8816086
Female	15031	.8703346
Race:		
White	58255	.8815209
Black	3559	.8420905
Asian	1867	.8243171
Hispanic	2241	.9036145
Other	2770	.8938628
Race and gender:		
White male	45018	.8842907
White females	13237	.8721009
Black males	2987	.8433211
Black females	572	.8356643
Asian males	1459	.8245374
Asian females	408	.8235294
Hispanic males	2093	.9063545
Hispanic females	148	.8648649
Time:		
Day	11874	.8343439
Night	35492	.8794094
Late night	21326	.9036388

Table 2.4: Chi-squared Results

Groups	Pearson's chi-squared	p-value
Race:		
Black, Hispanic, Asian, and white	113.79	0.0000
Black and white	49.12	0.0000
Hispanic and white	10.15	0.0014
Asian and white	55.95	0.0000
Black and white, males only	45.09	0.0000
Black and white, females only	6.46	0.0110
Hispanic and white, males only	9.59	0.0020
Hispanic and white, females only	0.07	0.7933
Asian and white, males only	48.63	0.0000
Asian and white, females only	8.29	0.0040
Sex:		
Males and Females	14.05	0.0002
Time:		
Day, night, and late night	344.74	0.0000
Day and night	157.82	0.0000
Day and late night	343.99	0.0000

Table 2.5: Estimation Results of Fixed-Effect Logit Model

Variable	dy/dx	Std. Err.	P>|z|
Age	0.0094424	0.00026	0
Age Squared	-0.0001133	0.00001	0
Female	-0.022123	0.003	0
Black	-0.0534052	0.00948	0
Asian	-0.0832755	0.01345	0
Hispanic	0.0179427	0.01043	0.085
Night (10PM-2AM)	0.0549722	0.00521	0
Late Night (2AM-4AM)	0.0901645	0.00655	0
Drug Recognition Training	-0.0234658	0.00685	0.001

CHAPTER THREE

THE INFRA-MARGINALITY PROBLEM AND QUANTIFYING RACIAL BIAS WITH OUTCOME DATA

Summary

Studies of racial bias in law enforcement are almost always conducted with outcome data. A limitation of "outcome tests" is that if the underlying distributions of the probabilities of guilt differ by race the link between assessment outcomes and racial bias will be imperfect, and the existing tests for racial bias will no longer be reliable. In this article, we analyze a model that allows for the possibility for the distribution of characteristics to differ by race. We then use simulations to compare how these differences in characteristics impact existing tests for racial bias. If the distributions of guilt differ, we find that individually the two most common tests for racial bias, the search rate and find rate, will be unreliable tests. However, since the statistical biases go in opposite directions, we ascertain that using find rate tests in conjunction with search rate tests will result in a robust test for racial bias. A limitation of this test is that it may conclude no racial bias, when bias does exists; however, if evidence of racial bias is found, this finding is not susceptible to criticisms of the infra-marginality problem.

I. Introduction

Quantifying racial bias in law enforcement is challenging. The biggest obstacle that researchers face is that data sets are almost always limited to observations on individuals who are assessed for a crime, rather than all individuals who could have been assessed for a particular criminal offense. Researchers are resigned to make inferences about officers' assessment decisions based only on information about the individuals assessed. The inevitable criticism of any study which attempts to measure racial bias with outcome data is that the link between assessment outcomes and the marginal assessment decision is imperfect. The propensity of individuals to commit crime, which is unobserved, may deviate by race. If this is the case, then outcome tests will be unreliable. This criticism is known as the infra-marginality problem.

While, this criticism is well established, little is known about the effect that differences in the propensity to commit crime, by race, have on the existing tests for racial bias. In this article a model is developed which allows the distribution of the propensity to commit crime to differ by race. Simulation is used to determine the effect that differences in these distributions have on existing tests for racial bias. A unique solution to the infra-marginality problem is found, which is robust to differences in the propensity to commit crime, by race.

Attempts to analyze quantitatively claims made of racial bias in law enforcement have mainly focused on two statistics. The first statistic, which is called the search rate or "statistical discrimination," is predominantly used in the criminal justice literature. It is calculated by dividing the number of individuals in a racial group who are assessed (or searched) for a crime by the total number of individuals within that racial group in the general population of interest.

$$Search\ \ rate = \frac{\#\,searched_r}{population_r} \tag{1}$$

All else equal, significantly different search rates across racial groups are suggestive of police officers targeting some racial group more than others. Between 1996 and 2001, search rate statistics were used in at least thirteen studies of racial bias.[1] Most of these studies measured differences in search-rates by race in traffic stops, and all of these studies found at least a minor degree of disparity in search rates. Six studies concluded that these differences were due to racial bias, whereas the other seven acknowledged that there could be legitimate, race-neutral explanations for this observed disparity. The problem with linking differences in search rates directly with racial bias is that there may be differences in driving behaviors by race. This, of course, is a highly controversial proposition. Any evidence given about incarceration rates for example, could be interpreted by many as a "self-fulfilling" prophecy (Banks 2002). To allow for potential differences, if any, in driving characteristics by race, the literature has focused on trying to find the proper "base-rate" or denominators with which to compare with the number of assessments. The idea is that if the proper comparison group can be found, racial bias could be tested with outcome data. Unfortunately, this is a daunting task because little is known about the difference in the propensity to commit crime by race. For the search-rate test to be a reliable test for racial bias, more needs to be known about the differences in the propensity to commit crime by race, and how these differences affect search-rate tests. This issue is summarized by Engel,

[1] A review of these studies can be found in Engel 2002.

Calnon, and Bernard (2002) who write, "Ultimately, the problem with interpreting these results is that these traffic and field-interrogation data have been collected with-out the guidance of any theoretical frameworks. Researchers have simply counted things."[2]

The second test for racial bias, which is called the find rate or the outcomes test and was first proposed by Becker in (1957). The find rate compares the average success rate officers exhibit across racial groups.

$$Find\ \ rate = \frac{\#\,guilty_r}{\#\,searched_r} \tag{2}$$

If, all else equal, officers are less successful with a particular racial group, it suggests that officers are racially biased against that racial group. The idea is that if officers are racially biased, they are willing to assess minority motorists even when the returns from searching them are lower than that of searching whites.

Knowles, Persico, and Todd (2001, hereinafter KPT) contributes significant theoretical motivation for the use of the find rates to test for racial bias. In KPT's model, police officers choose which individuals to search based on weighted motorists' characteristics in order to maximize the number of convictions relative to the cost of search. In turn, motorists, who know the officer's search criteria, minimize the probability of being searched by adjusting their discretionary characteristics. Under these assumption and as a result of this strategic interaction, a single, mixed-strategy equilibrium can be derived for each racial group in which police officers

[2] Engel, Calnon, and Bernard (2002), p. 259.

randomize over whether to search, and motorists randomize over whether to traffic drugs. KPT show that under these conditions, motorists' characteristics have no predictive power on a motorist's probability of guilt, and differences in success rates by race will be directly related to racial bias.

The major criticism of KPT's model is that the end result is unrealistic. To use the find rate to test for racial bias, a researcher must believe that every possible motorist characteristic is completely unrelated to the probability of committing a crime. For this to occur, motorists must know all of the information that officers use and know how officer use this information to determine search rates (Dominitz 2003). This may be plausible in drug-trafficking enforcement, as KPT suggest. However, there are many law enforcement situations where this is unrealistic. For instance, in drinking and driving enforcement evidence has been provided that there are motorist characteristics correlated with drinking and driving. This suggests, at least in drinking and driving enforcement, that motorist's ability to manipulate their characteristics in order to minimize their probability of being searched is, at best, imperfect. Under these conditions, researchers studying racial bias will have to address potential differences in the distributions of the probability of guilt, by race, and the distributions of characteristics, by race. Unfortunately, little is known about these distributions or about how they may differ by race. Researchers only observe the average outcomes of assessment, and they do not observe the marginal assessment decision (e.g. the last motorist deemed suspicious enough to be searched).[3] Both the search rate and the find rate will be unreliable tests for racial bias. This dilemma is called the infra-

[3] This is especially problematic in situations where the outcome of the assessment decision is dichotomous (Ross and Yinger 2002; Ayers 2002).

marginality problem (Ayers 2002; Anwar and Fang 2006; and Ross and Yinger 2002).

A. Drinking and driving enforcement

The cost inflicted on society by drinking and driving is tremendous by any standard. In 2006, 13,470 people died in alcohol-impaired driving crashes. This accounted for nearly one-third (32%) of all traffic-related deaths in the United States.[4] According to National Highway Traffic Safety Administration (NHTSA), the annual cost of alcohol-related crashes totals more than $51 billion.[5] While the driving-impaired death figure seems large, it has been consistently decreasing over the last 20 years. Researchers argue that this decline is due, as least in part, to increased drinking and driving enforcement (Hause et al. 1977; Voas and Haus 1987; Blomberg 1992). Over the past 20 years, the nation's police departments have increased expenditures and manpower to combat drinking and driving substantially. A significant challenge of enforcing drinking and driving is the inability of police officers to directly observe the blood alcohol content (BAC) of a given vehicle operator. Officers must predict each motorist's BAC using observable characteristics or cues. If a motorist exhibits enough characteristics or cues by which

---

[4] Dept of Transportation (US), National Highway Traffic Safety Administration (NHTSA). Traffic Safety Facts 2006: Alcohol-Impaired Driving. Washington (DC): NHTSA; 2008 [cited 2008 Oct 22]. Available at URL: http://www-nrd.nhtsa.dot.gov/Pubs/810801.pdf

[5] Blincoe L, Seay A, Zaloshnja E, Miller T, Romano E, Luchter S, et al. The Economic Impact of Motor Vehicle Crashes, 2000. Washington (DC): Dept of Transportation (US), National Highway Traffic Safety Administration (NHTSA); 2002. Available at URL: http://www.nhtsa.dot.gov/staticfiles/DOT/NHTSA/Communication & Consumer Information/Articles/Associated Files/EconomicImpact2000.pdf

the officer feels that there is "probable cause" then the motorist is transported to a regional testing facility where an official breathalyzer test is administered.

Research has been conducted on identifying driving actions that are indicative of driving while impaired (Harris 1980). In order to help officers accurately identify impaired motorists, NHTSA conducted a study to identify driving and behavioral cues that are correlated with drinking and driving. This research involved two field studies spanning five law enforcement agencies. In these field studies, motorists' characteristics as well as assessment outcomes were recorded for every enforcement stop. In this way, the correlation between driving characteristics and the motorists' BACs were studied. Based on this research, the NHTSA developed 24 driving cues that are correlated with drinking and driving. Examples of driving characteristics correlated with drinking and driving (pre-stop cues) are: problems maintaining proper lane position, slow speed, and jerky braking. Once the vehicle has been pulled over, officers gain access to face-to-face characteristics, known as post-stop cues. Examples of post-stop cues are: swaying or unsteady balance, repeating questions or comments, and slurred speech. While these cues are valuable in predicting the motorists' BACs, the process is nonetheless fallible. Inevitably to enforce drinking and driving laws, society must rely on officer's best judgment on whether to assess a motorist for drinking and driving. Unfortunately, this subjective nature of drinking and driving enforcement opens the door for claims of racial bias.

II. Theoretical Framework

KPT generate a simple and elegant model of racial bias and derive a test for racial bias that uses outcome data. Paramount to their model is the strategic nature of the motorist and officer interaction. While it is likely that motorists and officers act strategically to a degree, the process is unlikely to be perfectly strategic. The criticisms offered by Anwar and Fang; Dominitz; and Ayers, along with the evidence provided by NHTSA, cast doubt on KPT's conclusion that motorists' characteristics provide no reliable information about the probability of guilt. In a situation where motorists are impaired, they are unable to completely respond to the probability of being searched. As a result, find-rate tests will be unreliable.

In this section a model of racial bias is developed. Next, an attempt is made to formalize the challenges of measuring racial bias with outcome data, specifically, dealing with the infra-marginality problem. In our model of officer assessment, we allow for the possibility of differences in the distribution of characteristics of guilt by race. The idea is that these additional features can be simulated to predict the outcome of search and find rate tests.

A. The officer's assessment decision and racial bias

Consider a society whose population consists of only two racial groups, black and white. We denote the races as $r \in \{B,W\}$ where W indicates white motorists and B indicates black motorists. For the sake of exposition, the motorists' characteristics are denoted, as in KPT, by a one-dimensional variable, $c \in [0,1]$. As c increases, the motorist exhibits a higher level of characteristics that are correlated with guilt. The officer's cost of a traffic stop assessment is represented by t_r, which may vary by race if the officer is prejudice against one race. If the officer assesses a motorist who has committed a crime, the officer derives a benefit, B_r, otherwise the benefit is zero. The officer's assessment decision can be represented by the expected payoff to assessing the motorist:

$$P(G(c,r)) * B_r - t_r \tag{3}$$

If this expression is positive, the officer will assess the motorist; if it is negative, the officer will not assess the motorist; and if this expression is equal to zero, the officer will be willing to randomize over the assessment decision. The minimal probability required by an officer to conduct an assessment can be calculated for each race.

$$P\left(G\left(c_w^*\right)\right) = \frac{t_w}{B_w} \tag{4}$$

$$P\left(G\left(c_b^*\right)\right) = \frac{t_b}{B_b} \tag{5}$$

Figure 1 represents the officer's assessment decision. An officer will assess every motorist of race r that has a characteristic level greater than c_r^* and will not assess motorists with a characteristic level less than c_r^*. If a motorist has characteristics identically equal to c_r^*, the officer will be indifferent between assessing motorist and not assessing the motorist. We refer to these motorists as the "marginal motorists."

Racial bias in drinking and driving assessment is defined as $P\left(G\left(c_b^*\right)\right) \neq P\left(G\left(c_w^*\right)\right)$. This could be a function of officers deriving a higher benefit from assessing one racial group over another, a lower cost of assessment of assessing one racial group over another or both. If officers are racially biased against motorists of one race, they will be willing to assess motorists from that racial group who have a lower probability of being guilty of drinking and driving. This situation is represented in Figure 2. The utility derived by the officer from an assessment is

$$\sum_{r=W,B} \int [B_r P(G(c,r) - t_r] \gamma(c,r) f(c,r) dc \tag{6}$$

Where $\gamma(c,r)$ is the probability of assessment, which is chosen by the officer, and $f(c,r)$ indicates the distribution of characteristics by race. To maximize utility, the officer chooses the probability of assessment, $\gamma(c,r)$. If $B_r P(G(c,r) > t_r$, the officer sets $\gamma(c,r) = 1$. If $B_r P(G(c,r) < t_r$, the officer set $\gamma(c,r) = 0$, and the officer is willing to randomize if $B_r P(G(c,r) = t_r$.

B. *Strategic interaction and potential differences in the distribution of characteristics by race*

In KPT's model, motorists are able to strategically manipulate all of their characteristics and their behaviors, in order to minimize the probabilities of being searched. Since all motorists that have characteristics greater than c_r^* will be assessed with probability equal to 1, it would take a "crazy" criminal to drive while exhibiting this level of characteristics. The motorist should never choose to drive with a characteristic level greater than c_r^*. Instead, the motorist will strategically update all of his/her characteristics in order to minimize his/her probability of being searched.

The end result of this strategic updating is that motorists' characteristics have no predictive power on probability of guilt. Under these conditions, a unique, mixed-strategy equilibrium can be derived in which motorists randomize over whether to drive after drinking, and officers randomize over whether to assess a given motorist. Since characteristics do not impact the motorist's probability of guilt, researchers do not have to worry about differences in the distribution of characteristics by race, and racial bias can be measured by simply comparing average success rates conditioned by race.

For KPT's model to be reliable, motorist characteristics must be completely unrelated to the probability of guilt. However, in drinking and driving enforcement, the NHTSA has provided evidence that some motorist characteristics are correlated with drinking and driving. This evidence suggests, at least in the drinking and driving enforcement area, that motorist's ability to manipulate their characteristics in order to minimize their probability of being searched is, at best, imperfect. Researchers studying racial bias in drinking and driving enforcement will have to address potential differences in the distribution of the probabilities of guilt by race and the distribution of other salient characteristics by race. Taking equation 4 and solving for $\gamma(c,r)$, the total expected utility for the officers is given by

$$\sum_{r=W,B} \int_{c_r^*}^{\infty} \left[B_r P(G(c,r) - t_r \right] f(c,r) dc \tag{7}$$

Figure 3 represents the assessment outcomes when the distributions of the probability of guilt and the distribution of characteristics differ by race. Panel A indicates the officer's assessment decision when the probability of guilt can differ by race. By construction, the officers are racially unbiased because $P\left(c_b^*\right) = P\left(c_w^*\right)$. Notice that this does not necessarily indicate that $c_b^* = c_w^*$. Officers then assess every motorist with characteristics greater than c_r^* for each race. Panel B represents the distribution of characteristics by race. The characteristics required for assessment, taken from panel A, are used in panel B to determine how many motorists of each race are assessed.

C. Search-rate and find-rate tests, and the infra-marginality problem

The infra-marginality problem arises when the distribution of characteristics differs by race or $f(c,b) \neq f(c,w)$. If this is the case, the relationship between the average motorist and the marginal motorist will be imperfect. Compounding this problem is the fact that the distribution of characteristics, especially the difference by race, is largely unknown. This is a troubling, persistent criticism of both search-rate and find-rate tests. In this section, we consider both the search-rate test and the find-rate test and investigate what conditions are necessary for them to be reliable tests for racial bias. The search rate can be more formally expressed as

$$\textit{Search rate} = \frac{\int_{c_r^*}^{\infty} f(c,r)dc}{\int_{o}^{\infty} f(c,r)dc} \tag{8}$$

Where $\int_{c_r^*}^{\infty} f(c,r)dc$ is the number of motorists who are searched for a crime, and $\int_{0}^{\infty} f(c,r)dc$ is the total number of motorists who are observed driving. Racial bias can be defined as

$$\frac{\int_{c_b^*}^{\infty} f(c,b)dc}{\int_{o}^{\infty} f(c,b)dc} \neq \frac{\int_{c_w^*}^{\infty} f(c,w)dc}{\int_{o}^{\infty} f(c,w)dc} \tag{9}$$

If the distributions of the probability of guilt and the distribution of characteristics are equal across races, then racial bias can be measured by simply comparing how often motorists of different racial groups are assessed. Notice that it is not required for the distribution of characteristics for each race to come from an identical data generating processes. The search rate will still be an accurate test for racial bias if the probabilities of guilt are linear combinations of one another. For instance if $f(c,b) = \delta * f(c,w)$, search rates will still be an accurate test for racial bias. Notice, in Figure 3, that officers are unbiased, $t_b = t_w$. However, comparing search rates, there is a disproportionate number of black motorists assessed per capita. Under the conditions of the test, this suggests racial bias against black motorists when there is no racial bias by construction. This occurs because, in this hypothetical example, a larger fraction of black motorists have a characteristic level that is greater than c_b^*. In fact, any difference in distribution that is not a linear combination can bias search rate tests.

Find rate tests can be expressed more formally as

$$Find\ rate = \frac{\int_{c_r^*}^{\infty} P\left(G(c|r)f(c|r)dc\right)}{\int_{c_r^*}^{\infty} f(c|r)dc} \tag{10}$$

Where $\int_{c_r^*}^{\infty} P\big(G(c|r) f(c|r)dc$ is the expected number of successful assessments for a particular racial group, and $\int_{c_r^*}^{\infty} f(c|r)dc$ is the total number of assessments conducted on a particular racial group. Racial bias is defined as

$$\frac{\int_{c_b^*}^{\infty} P(G(c|b)f(c|b)dc}{\int_{c_b^*}^{\infty} f(c|b)dc} \neq \frac{\int_{c_w^*}^{\infty} P(G(c|w)f(c|w)dc}{\int_{c_w^*}^{\infty} f(c|w)dc} \tag{11}$$

Again, the idea behind the find rate test is that, if characteristics are uncorrelated with committing a crime, officers will randomize when searching motorists. In equilibrium, officers will assess motorists at random with probability, $p^*(G|c,r)$. In this case, the probability of guilt has no correlation with characteristics, and can be specified as $P(G|r)$. In which case, find rates simplify to

$$P(G|r)\frac{\int_{c_r^*}^{\infty} f(c|r)dc}{\int_{c_r^*}^{\infty} f(c|r)dc} = P(G|r) = \frac{t_r}{B_r} \tag{12}$$

Equation 12 shows the relationship between find rate tests and racial bias when motorists' characteristics are uncorrelated with probabilities of guilt. Even if the distribution of characteristics differs by race, as long as the probability of guilt does not, there will be a direct relationship between find rate tests and racial bias. This depends on motorists' ability to manipulate their characteristics to minimize their probability of being assessed. Situations in which this occurs, potentially drug trafficking enforcement, are in fact rather rare. There are many law enforcement situations in which researchers will not be able to rely on this strong assumption. To be clear, if even a subset of motorists is unable to update all of their characteristics, the probability of guilt will depend on the characteristics exhibited by the motorist, and find rate tests will be unreliable.

D. Modeling racial bias in drinking and driving enforcement

In this section, we develop a model for racial bias that allows for differences in the distribution of probability of guilt by race and for differences in the distribution characteristics by race. In our model, the distribution of post-stop BAC is represented by $BAC_r(c)$, which can vary by race. Officers know the distribution of $BAC_r(c)$ for each race. However, they only observes a noisy, predicted BAC denoted $BAC_r^* = BAC_r(c) + \varepsilon_r$. Since officers know the true underlying distributions, the noise factor, $\varepsilon_r \sim f\left(0,\sigma_r\right)$, will be centered around zero. However, the variance can differ by race. The officer's goal is to assess motorists who are legally drunk. Following Equation 1, the officer's assessment decision is given by

$$P\left(BAC_r(c) > .08\right) * B_r - t_r \tag{13}$$

Following Equations 2 and 3, the conditions for the marginal motorists for each race, will be

$$P\left(BAC_w(c_w^*) > .08\right) = P_w^* = \frac{t_w}{B_w} \tag{14}$$

$$P\left(BAC_b(c_b^*) > .08\right) = P_b^* = \frac{t_b}{B_b} \tag{15}$$

Where P_r^* represents the exact probability for each race for which an officer is indifferent between assessing the motorist and not. Incorporating the minimum probabilities required for assessment of a given race, we can re-write the assessment decision as

$$A = \binom{1}{0} \text{ iff } P\left[\left(BAC_r(c) > .08\right)\right]\binom{>}{<} P_r^* \tag{16}$$

If the motorist exhibits characteristics such that the officer determines that the probability of guilt is greater than the minimal probability required to conduct an assessment, the officer will assess the motorist, and A=1. Otherwise the motorist will not be assessed. Using the predicted BAC and rearranging we derive a new equation which represents that marginal motorist

$$\Pr\left(\varepsilon_r < \left(BAC_r^*\left(c_r^*\right) - .08\right)\right) = P_r^* \tag{17}$$

The officer's assessment decision can be re-written as

$$A = \binom{1}{0} \text{ iff } Pr\left(\varepsilon_r < \left(BAC_r^*(c) - .08\right)\right)\binom{>}{<} P_r^* \tag{18}$$

In equation 18, the officer predicts the BAC of the motorist using his or her observable characteristics, c. If the probability that the motorist is drinking and driving is greater than that of the marginal motorist, the officer will assess the motorist. The officer accepts that under these conditions, sometimes he or she will assess a motorist who is not guilty of drinking and driving. This will happen if A=1 and $\varepsilon_r > BAC_r^*(c) - .08$. The officer's objective is given by

$$\sum_{r=W,B} \int_{c_r^*}^{\infty} \left[B_r P\left(\left(\varepsilon_r < \left(BAC_r^*(\mathrm{c}) - .08\right)\right) - t_r \right] f(c,r)dc \tag{19}$$

Under these conditions, differences in find-rate and search-rate tests by race may be caused by differences in five different variables, by race. Differences caused by t_r or B_r would indicate racial bias, however differences in $\varepsilon_r, BAC_r^*(\mathrm{c})$, or f (c,r) would not. Unfortunately, little is known about the distributions of any of these variables.

This model specification addresses an additional potential pitfall of current tests for racial bias, which have not been addressed in the literature. Officers may be able to predict the BAC of one racial group more accurately than another. This situation is modeled as differences in the variances in the predicted BAC by race or $\sigma_W \neq \sigma_B$. It has been shown that individuals are more able to interpret cues from individuals belonging to their own social groups (Blair 2002). Unequal assessment by race could be caused by a predominantly white police force that fares better at interpreting the characteristics of white motorists than interpreting the characteristics of minority motorists. This is not racial bias but this differential ability could affect search and find rate tests. Finally, note that, in the specific case in which $f(c,b) = f(c,w)$, $BAC_b(c) = BAC_w(c)$, and $\sigma_b = \sigma_w$, there is a direct relationship between the marginal motorist and the average success rates. Simple search-rate or find-rate tests could be used to measure racial bias.

III. Simulation

In this section, we study the impact that differences in the distributions of BAC and differences in error component structures have on search-rate and find-rate tests. To accomplish this we offer a series of simulations in which we can control P_r^* as well as the distributions of ε_r, $BAC_r^*(c)$, and $f(c,r)$. The goal of this analysis is to better understand this system and to derive a test for racial bias that is robust to differences in the distributions of ε_r, $BAC_r^*(c)$, and $f(c,r)$ by race. One limitation of this method is that assumptions must be made about the distribution of both the BAC and the error component. As with all statistical models, if one chooses a distribution incorrectly, the results can be biased.

The distribution of the number of motorists by race at each BAC level is given by BAC_r. In this set of simulations, the distribution of BAC_r is assumed to be a gamma distribution. The first appealing property of the gamma distribution is that it is non-negative. This is required since motorists cannot have a negative BAC. Second, it is easy to construct a gamma distribution that is skewed to the right. This is appropriate because in DUI assessments, there are a significant number of motorists with high blood alcohol contents. The distribution of the noise component, ε_r , is chosen as normal. In the simulations, we start with a base case (simulation 1) and then change parameters to find the effect that these changes have on the eventual distribution of assessed motorists and search and find rate tests. In doing this, we are better able to contrast the effect of racial bias with the effect of differences in distributions by race. Table 1 summarizes the simulations. In the first simulation, we set β=4 for the distribution of observed motorists. The variance of the noise component is chosen as 40, and officers only assess motorists they are at least 70% confident have a BAC greater than 0.08%. The distribution of the observed motorists and the corresponding assessed motorists for the base group is presented in Figure 4. From this simulation, it is apparent that the distribution of the assessed motorists is shifted to the right compared with the distribution of observed motorists. As expected, the motorists who are not assessed are largely on the left side of the distribution. Almost all of the motorists on the right tail of the distribution are assessed. In this simulation, officers assess 57.37% of the observed motorists and are 92.5% successful in their assessments.

In simulation 2 the BAC distribution and the distribution of the error component are unchanged, but the probability required for assessment is reduced from 70% to 50%. The decrease in the probability of guilt required for assessment for this group represents racial bias. Figure 5 gives the distribution of assessed motorists of simulations 1 and 2 and displays the corresponding statistical outcomes for each group. The noticeable difference between the two distributions is a leftward shift in simulation 2. The difference in the statistical outcomes between simulations 1 and 2 provides evidence for both search rate and finds rate tests. The hypothesis of search rate tests is that if officers are racially biased and everything else is equal, officers will assess the biased group with a higher frequency. Based on simulation results, this is, in fact, the case. The assessment rate for the base group is 57.37%, whereas the assessment rate for the biased group is 69.38%. Additionally, under these conditions, our results support the use of find rate tests. The hypothesis of find rate tests, everything else being equal, is that officers will be less accurate in the searches they conduct on the biased racial group. The accuracy exhibited by officers in the unbiased simulation is 92.5%, and it is 87.84% in the biased simulation. Comparing these two simulations adds evidence that under the conditions of identical distributions and identical noise components, racial bias can be tested using either search rate or find rate tests.

The reliability of find rate and search rate tests is disputed precisely because the distribution of the probabilities of guilt by race may differ. To determine the impact that different BAC distributions by race have on the distribution of assessed BAC levels and on observed statistical outcomes, we present simulation 3 and simulation 4. In both of these simulations, we maintain the parameter values identical to the base group, except for the β parameter. In simulation 3 we increase the β parameter to represent a lower BAC distribution. In simulation 4, we increase the β parameter to represent a higher BAC distribution. The distribution of observed and assessed motorists and the corresponding statistical outcomes for each group are presented in Figure 6. By inspection of Figure 6, if the BAC distributions differ by race, search rate and find rate tests will be inappropriate tests for racial bias.

In simulation 3, the assessment rate is greater than the base group, a finding which would suggest racial bias against group 3. However, the find rate is lower which suggests, conversely, racial bias against group one in favor of group three. By construction, there is actually no racial bias as $P_b^* = P_w^*$. The search rate is larger for simulation 3 because there is a larger percentage of the population of group 3 with a high BAC. The find rate is greater for group 3 than simulation 1 because the distribution of the right tail is fatter. With more highly likely outcomes, the success rate for group 3 will be greater. For instance, in this simulation, a motorist with a predicted BAC of twice the legal limit (0.16%) has a predicted probability of guilt equal to 97.72%. The opposite occurs in simulation 4. Since there is a smaller percentage of the population with high BACs and there are fewer highly likely outcomes, the find rate is lower and the assessment rate is higher. These two cases (simulations 2 and 3) illustrate the potential pitfalls of the infra-marginality problem.

More complicated are the cases when racial bias occurs *and* BAC distributions differ by race. In simulation 5, officers are racially biased, $P^* = .5$, and there is a lower BAC distribution, $\beta = 3$. In simulation 6, officers are racially biased , $P^* = .5$, but there is a higher BAC distribution, $\beta = 5$. The results are given in Figure 7. Simulation 5 results in both a lower search rate and a lower find rate. The search rate would incorrectly suggest racial bias against the base group, and the find rate would correctly imply racial bias against group 5. However, comparing simulation 6 with the base group, the exact opposite results are found. Both the search rate and the find rate are greater than the base group. Researchers using the search rate would correctly infer racial bias against group 6, and researchers using the find rate would incorrectly conclude racial bias against the base group. It is apparent from simulations 3 through 6 that in any situation where the distribution of BAC differs by race, search rate and find rate tests will be unreliable. In simulations 3 and 4, researchers may conclude there is racial bias when none exists, and in simulations 5 and 6 racial bias can be masked by unequal BAC distributions.

There is another issue that has largely not been addressed in the literature. Psychologists and sociologists have pointed out that individuals are more apt to properly assess signals and characteristics exhibited by their own group (Blair 2002). This is relevant to studies of racial bias, in that it is possible for officers to be more accurate in predicting the BAC of motorists that belong to their own group. A police force comprised predominantly of one race could be better at predicting the BAC of motorists of that same race. This may be a rationale used to explain lower find rates in assessments of minority groups. To model a potential decrease in prediction accuracy, in simulation 7 we increase the variance of the noise component. Figure 8 compares the high variance group with the base group. In this comparison officers are more successful with the group in which they are worse at predicting the actual BAC. This somewhat counterintuitive result happens because, when officers have more trouble predicting the BAC of a particular racial group, the officer still will only assess motorists that have a probability of guilt greater than P_r^*. Because of this, officers will assess fewer motorists of that racial group, but they will still assess the high probability motorists. Thus, when officers have more trouble predicting the BAC of a particular racial group, they will assess fewer motorists of that racial group, but they will be more accurate, on average, with the motorists from that group which they do assess.

IV. Empirical Test for Racial Bias When the Distributions of Characteristics Differ by Race

The simulation results are presented in Table 2. Holding the BAC distributions and the variance of the error terms constant, comparing simulation 1 and simulation 2, racial bias can be accurately tested using search-rate tests or find-rate tests. However, if the distribution of characteristics differs across races, these tests are no longer reliable. Comparing simulation 1 with simulation 3, the search-rate test suggests racial bias against group 3, and find-rate tests suggest racial bias against the base group when, by construction, there is no racial bias. Comparing simulation 1 with simulation 4, we find the opposite results, in which the search-rate tests suggest racial bias against the base group and find-rate tests suggest racial bias against group 4. It is apparent that, under these conditions, the use of search-rate or find-rate tests alone to test for racial bias is unreliable.

As such, a simple method to test for racial bias is proposed here, which is robust to differences in the distribution of characteristics by race. The intuition behind this test is that both search-rate tests and find-rate tests are biased from the differences in the distribution of characteristics by race, but in opposite directions. When there is a higher (lower) BAC distribution the search rate will be inflated (deflated) and the find rate will be deflated (inflated). If these two tests are used in conjunction with one another a test can be created which is robust to differences in the distribution of characteristics by race. The intuition is simple. If either the search rate is higher, or the find rate is lower, it could be due to differences in the distribution of characteristics by race. However, if both the find rate is lower and the search rate is higher for a particular racial group, it could not be due to differences in distributions of characteristics by race. Racial bias is then measured by

$$\sum_{r \in R} \left(\frac{\#\, searched_r}{population_r} > \frac{\#\, searched_B}{population_B} \,\&\, \frac{\#\, guilty_r}{\#\, searched_r} < \frac{\#\, guilty_B}{\#\, searched_B} \right) \tag{20}$$

where R is the set of race categories, and B represents a base group. This test is advantageous because in situations in which there is no racial bias, but there is a difference in distribution of BAC by race, this test will not suggest racial bias. However, in situations in which racial bias occurs in conjunction with different BAC distributions by race (as in simulations 5 and 6) racial bias may still be masked by the differences in distributions by race. Thus, in using this test, if evidence is found of racial bias it will not be the result of differences in BAC distributions by race. However, if no evidence of racial bias is found, it cannot be concluded that there is no racial bias.

V. Empirical Results

We now apply the test described above to breathalyzer data collected in Washington state. The dataset contains all DUI assessments conducted in Washington state from January 2005 to October 2006. As part of the official process when administering a breathalyzer, the officer transports the motorist to a regional testing facility where demographic details including race are recorded. The motorist blows into the breathalyzer in two separate trials. The results are sent electronically to the state lab where they are documented and recorded. The legal BAC limit in Washington state is 0.08%. Any assessed motorist with a BAC above this threshold is given a citation for DUI. Table 3 calculates both the search rate and the find rate for motorists, of different ethnicities and races, in Washington state. For the search rate we use census population figures, provided by the CDC, as the base rate. The find rate is calculated by dividing the number of motorists, of a given race, with a $BAC > .08\%$ by the total number of motorists or that race, who were administered a breathalyzer.

Since Washington state is composed predominantly of white motorists, we use white motorists as the comparison group. To check for racial bias, we compare the search rate and the find rate for each racial group with the search rate and find rate of white motorists. First, we find that Asian motorists have a lower find rate. Looking at that finding alone would suggest racial bias. However the search rate is also lower for Asian motorists. As a result, no such inference can be made. Asian motorists could have a lower BAC distribution compared to white motorists, but racial bias cannot be ruled out. The same finding, with opposite search rate and find rate results is observed with regard to Native American motorists. The search rate is higher for this group, which individually could be attributed to racial bias. However, the find rate is also higher. Native American motorists could have a higher BAC distribution, but racial bias cannot be ruled out. Comparing the outcomes of assessments conducted on black motorists and white motorists, we find that officers search black motorists more often and are less successful with these motorists. This is evidence of racial bias against this group which cannot be attributed to differences in distributions, by race. Finally, comparing the assessment outcomes of Hispanic motorists and white motorists, we find the officers search Hispanic motorists less often than they search white motorists. Additionally, officers are more successful with Hispanic motorists than they are white motorists. According to the hypothesis of the test, this would suggest racial bias against white motorists in favor of Hispanic motorists.

VI. Conclusions

It is difficult to quantify racial bias with outcome data. Since the propensity to commit crime is unobserved across racial categories, the existing tests for racial bias may be incorrect. In this article the possibility that the distribution of the probability of guilt to differ by race is modeled. We use simulations to compare how hypothetical differences in characteristics would affect existing tests for racial bias. We find that, under these varying conditions, the existing tests for racial bias are unreliable. However the two tests (search rates and find rates) are biased in the opposite directions. Combining these tests will result in a test that is robust to differences in BAC distributions by race. Applying this test to Washington state data, we find that officers appear to be racially biased against black motorists. Using the same data, we find evidence of inverse racial bias against white motorists in favor of Hispanic motorists.

Bibliography

Anwar, Shemena, and Hanming Fang. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96 (March 2006): 127-151.

Ayres, Ian. "Outcome Tests of Racial Disparities in Police Practices." *Justice Research and Policy* 4 (Fall 2002): 132-142.

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

Banks, Richard. "Beyond Profiling: Race, Policing, and the Drug War. *Stanford Law Review*, Forthcoming.

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

Blair, Irene. "The Malleability of Autonomic Stereotypes and Prejudice." *Personality and Social Psychology Review* 6: 242-261.

Blomberg, R.D. Lower BAC Limites for Youth: Evaluation of the Maryland 02 Law. Washington DC: National Highway Safety Administration, 1992.

Dharmapala, Dhammika, and Stephen L. Ross. "Racial Bias in Motor Vehicle Searches: Additional Theory and Evidence." *Contributions to Economic Analysis & Policy* 3 (2004): 1-21.

Dominitz, Jeff. "How Do the Laws of Probability Constrain Legislative and Judicial Efforts to Stop Racial Profiling." *American Law and Economics Review* 5 (Fall 2003): 412-432.

Dominitz, Jeff, and John Knowles. "Crime Minimization and Racial Bias: What Can We Learn from Police Search Data?" *The Economics Journal* 116 (November 2006): F368-F384.

Engel, R. S., Calnon, J. M., & Bernard, T. J. 2002. Theory and racial profiling: Shortcomings and future directions in research. *Justice Quarterly* 19(2):249-273.

Greenwald, Anthony G., Mark A. Oakes, and Hunter G. Hoffman. "Targets of Discrimination: Effects of Race on Responses to Weapons Holders." *Journal of Experimental Social Psychology* 39 (2002): 399-405.

Harris, Doulas. “Visual Detection of Driving While Intoxicated.” *Human Factors* 22 (1980): 725-732.

Hernandez-Murillo, Ruben, and John Knowles. "Racial Profiling or Racist Policing." *International Economic Review* 45 (August 2004): 959-989.

Kennedy, BP, I Kawachi, R Glass, D Prothrow-Stith. “Income Distribution, Socioeconomic Status, and Self Rated Health in the United States: Multilevel Analysis.” *British Medical Journal* 317 (1998): 917-921.

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

Knowles, John, Nicola Persico, and Petra Todd. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *J.P.E.* 109 (February 2001): 203-229.

National Research Council. *Measuring Racial Discrimination.* Rebecca M. Blank, Marilyn Dabady and Constance F. Citro, eds. Washington, DC: The National Academies Press, 2004.

Persico, Nicola. "Racial Profiling, Fairness, and Effectiveness of Policing." *American Economic Review* 92 (December 2002): 1472-1497.

Yinger, John. *Closed Doors, Opportunities Lost: The Continuing Cost of Housing Discrimination.* New York: Russell Sage Foundation, 1995.

Yinger, John. "Evidence of Discrimination in Consumer Markets." *J. of Econ. Perspectives* 12 (1998): 23-40.

Zhao, Bo, Jan Ondrich, and John Yinger. "Why do Real Estate Brokers Continue to Discriminate? Evidence from the 2000 Housing Discrimination Study." *Journal of Urban Economics* 59 (May 2006): 394-419.

Figure 3.1: Probability of Guilt is Equal to the Cost of Assessment for the Marginal Motorist

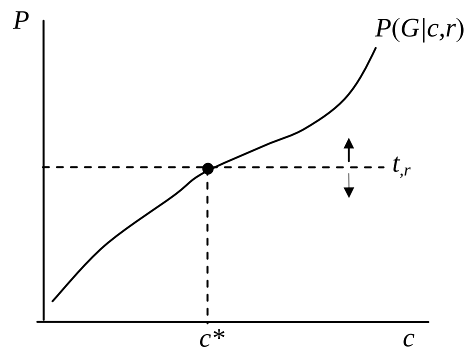


Figure 3.2: Different Assessment Costs across Racial Groups

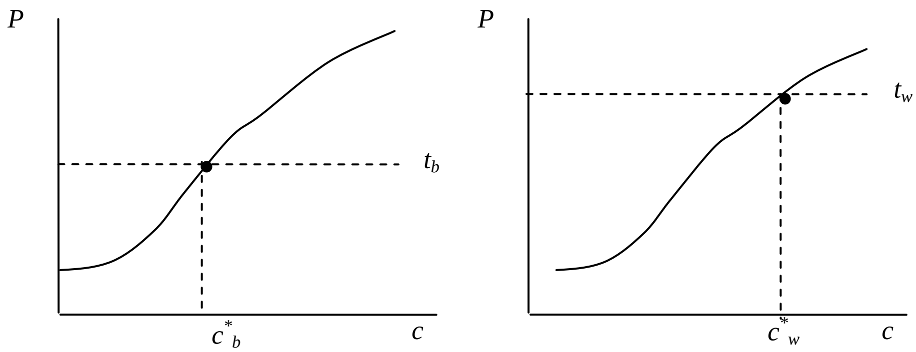


Figure 3.3: Assessment Outcomes When the Probability of Guilt and the Distribution of Characteristics Differ by Race

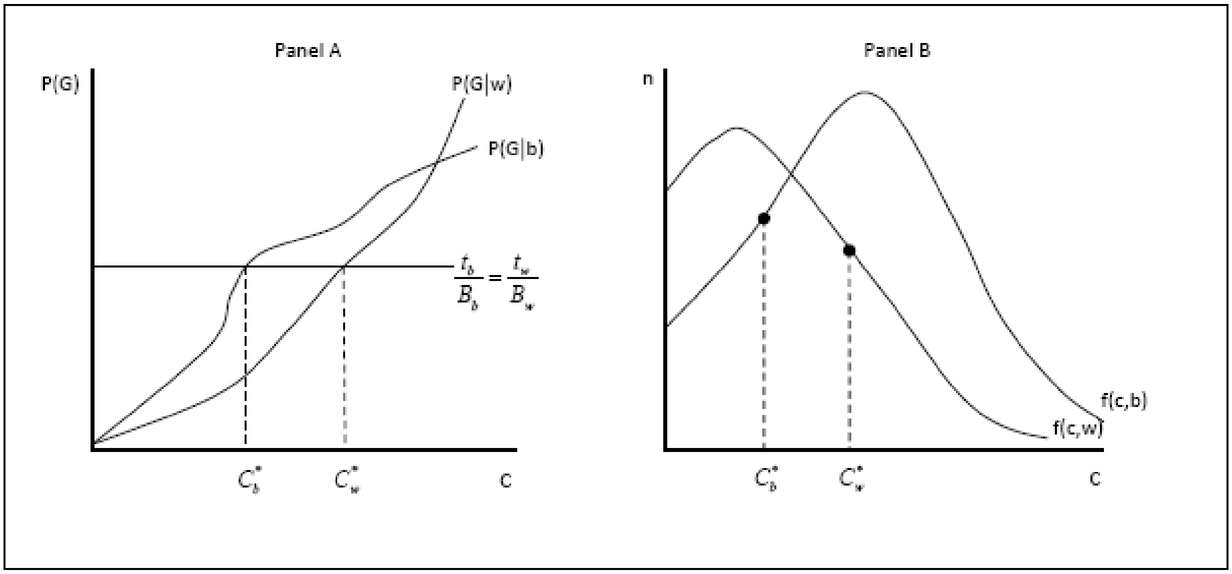


Figure 3.4: Distributions of BACs for Observed and Assessed Motorists in Simulation 1

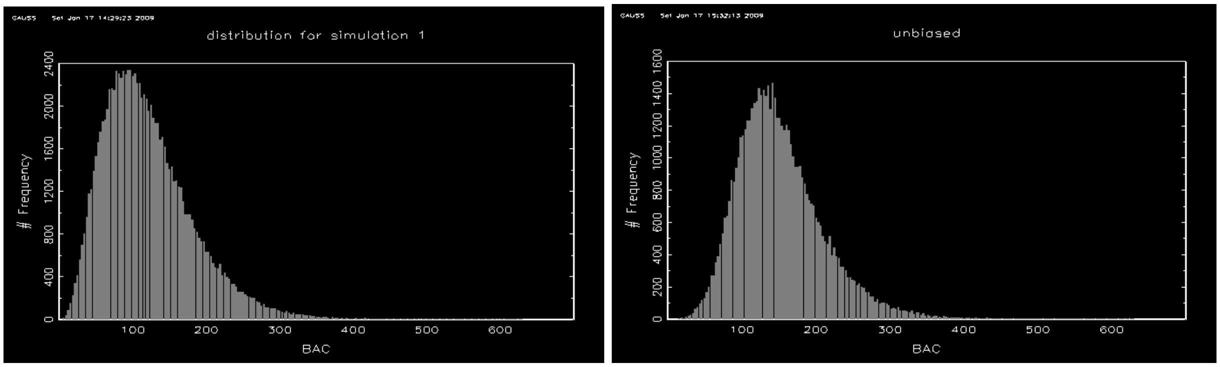
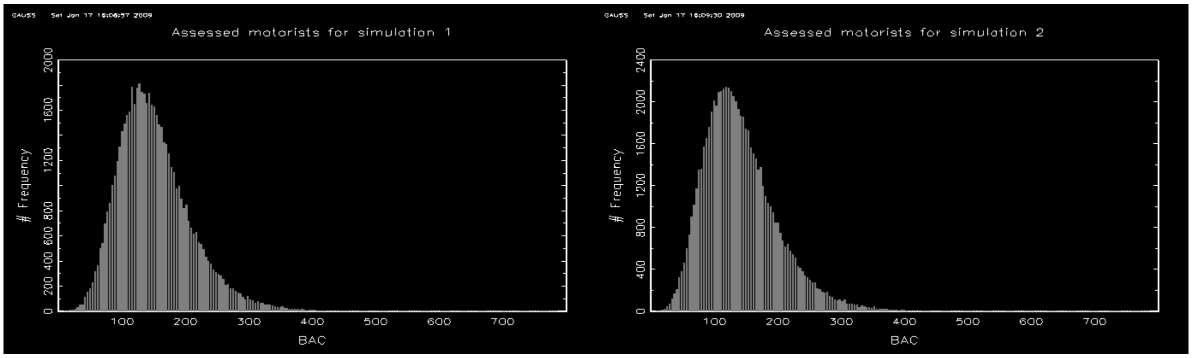
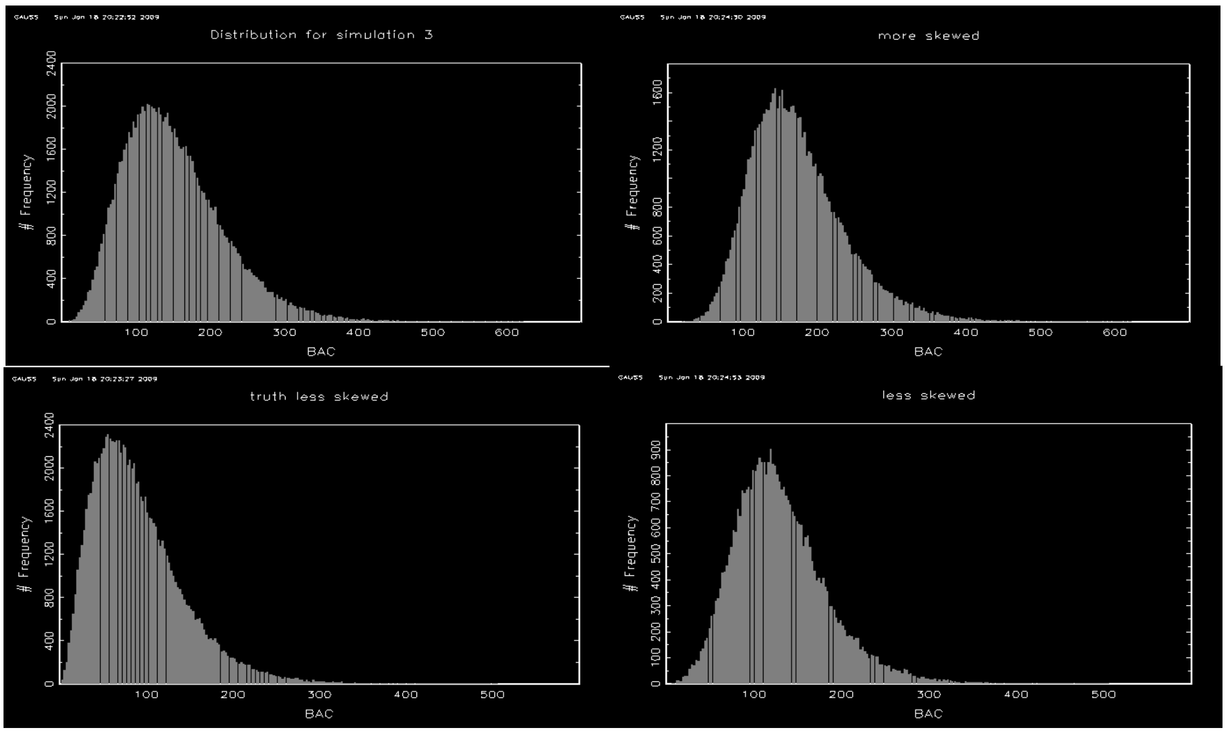


Figure 3.5: Distributions of BACs for Assessed Motorists in Simulation 1 and Simulation 2



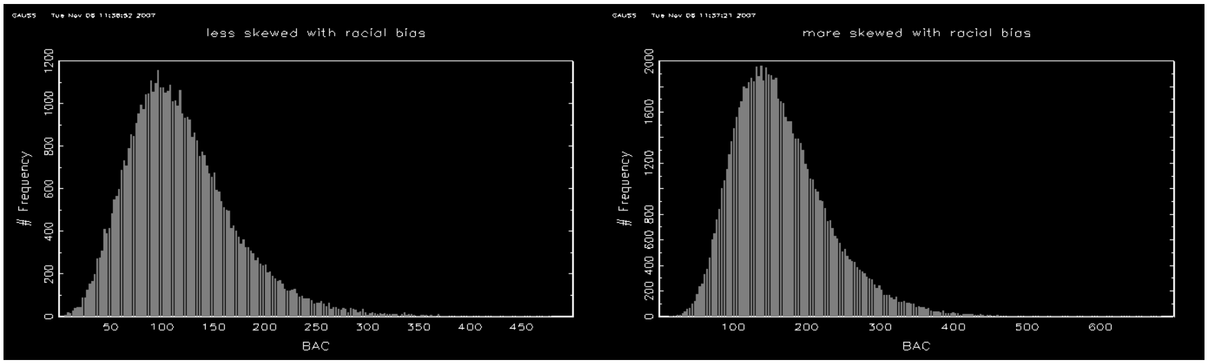
Simulation		Mean	Standard Deviation	Assessment Rate	Success Rate
1	Base group	151.05	57.69	57.37%	92.50%
2	Lower P_r^* (racial bias)	141.23	58.13	69.38%	87.84%

Figure 3.6: Distributions of BACs for Observed and Assessed Motorists in Simulation 3 and Simulation 4



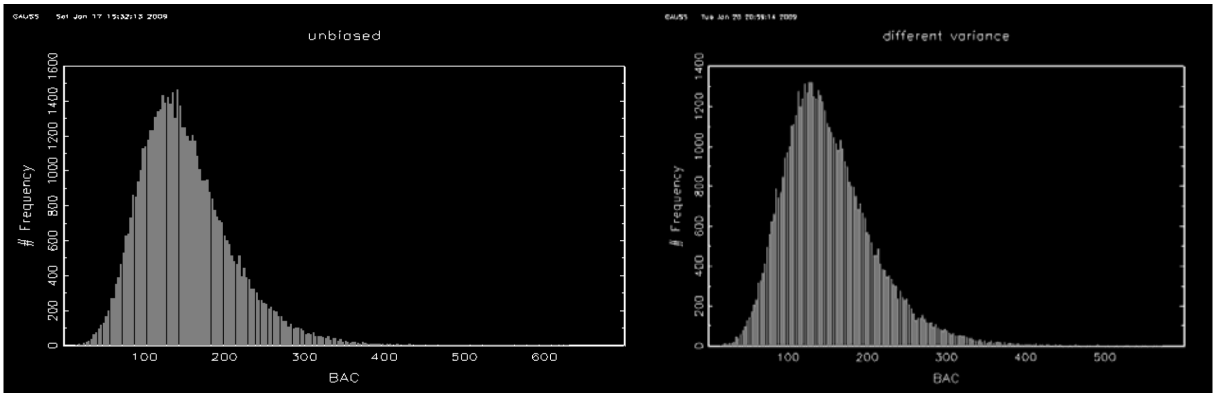
Simulation		Mean	Standard Deviation	Assessment Rate	Success Rate
1	Base group	151.05	57.69	57.37%	92.50%
3	Higher BAC distribution	172.81	64.03	72.28%	96.87%
4	Lower BAC distribution	130.22	52.54	39.74%	84.41%

Figure 3.7: Distributions of BACs for Assessed Motorists in Simulation 5 and Simulation 6



Simulation		Mean	Standard Deviation	Assessment Rate	Success Rate
1	Base group	151.05	57.69	57.37%	92.50%
5	Lower P_r^* and lower BAC	118.33	51.91	52.80%	76.35%
6	Lower P_r^* and higher BAC	164.54	64.12	81.80%	94.48%

Figure 3.8: Distributions of BACs for Assessed Motorists in Simulation 1 and Simulation 7



Simulation		Mean	Standard Deviation	Assessment Rate	Success Rate
1	Base group	151.05	57.69	57.37%	92.50%
7	Greater variance	153.94	57.19	54.11%	93.62%

Table 3.1: Parameter Values for each Simulation

Model	Representative of	$f_r(c)$	β	ε_r	P_r^*
1	Base	Gamma	4	(0,40)	70%
2	Racial Bias	Gamma	4	(0,40)	50%
3	More Skewed	Gamma	5	(0,40)	70%
4	Less Skewed	Gamma	3	(0,40)	70%
5	Less Skewed and Racial bias	Gamma	5	(0,40)	50%
6	Greater Variance	Gamma	4	(0,50)	70%

Table 3.2: Statistical results of for Simulations

Simulation		Mean	Standard Deviation	Assessment Rate	Success Rate
1	Base group	151.05	57.69	57.37%	92.50%
2	Lower P_r^* (racial bias)	141.23	58.13	69.38%	87.84%
3	More Skewed	172.81	64.03	72.28%	96.87%
4	Less Skewed	130.22	52.54	39.74%	84.41%
5	Lower P_r^* and Less Skewed	118.33	51.91	52.80%	76.35%
6	Lower P_r^* and More Skewed	164.54	64.12	81.80%	94.48%
7	Greater Variance	153.94	57.19	54.11%	93.62%

Table 3.3: Search and Find rate Test Applied to WSP Data

	Observations	Search Rate	Find Rate
White	18961	0.3852%	88.01%
Black	1388	0.5961%	83.86%
Asian	696	0.1606%	81.32%
Hispanic	2238	0.3681%	90.35%
Native American	502	0.4850%	92.03%

*search rate is conducted using census population figures provided by the CDC as the benchmark.

CHAPTER FOUR

WEALTH AND AGENDA CONTROL IN STADIUM FUNDING: THE 1997 QWEST FIELD REFERENDUM OUTCOME

Summary

Direct inspection of data on the Qwest Field referendum in the state of Washington, plus a county-level representative voter model of yes votes, support the following conclusion. The closest counties to the stadium, also the most highly populated and richest counties in the state, ruled this election while the costs were state-wide. In addition, there is some evidence that voters believed there was a relocation threat if the referendum failed. The odds of voting yes were higher in counties with more college-educated voters but, paradoxically, also in counties with higher unemployment.

I. Introduction

Dear fellow Washingtonians:

I've said from the start I wouldn't go forward with purchasing the Seahawks and building a new stadium and exhibition center without your approval. Knowing a "Yes" vote will be an act of trust, I'd like to share my commitments to this public/private partnership… Should we move forward, the new stadium and exhibition center will be a valuable asset – bringing our communities together and benefiting the state for decades to come.

—Paul Allen (Secretary of the State of Washington, 1997, p. 4).

Referendum voting outcomes have proven informative about economic behavior in many areas of government spending. Primarily, analysis has been in education, health care, and nuclear power. Here, we examine another large-scale public endeavor, provision of a stadium for a pro sports team owner. In particular, through inspection of the geographic distribution of yes votes, related population and income data, and estimation of a county-level representative voter model, we examine a referendum vote in the state of Washington to subsidize the building of what is now known as Qwest Field in Seattle. It is difficult to escape the conclusion that the closest counties to the stadium, also the most highly populated and richest counties, ruled this election even though the costs were borne state-wide.

Direct inspection of the geographic distribution of yes votes in the state suggests that support increased systematically with proximity to the stadium. Indeed, the highest level of yes votes occurred in the counties immediately proximate to the stadium. It also ends up that these counties are the most densely populated and richest counties in the state.

Results of estimating a county-level, "representative voter" model are entirely consistent with the inspection results—proximity and population really do rule. Adding to only one, less formal finding in the literature on sports facility referendum voting, there is some evidence that voters believed the relocation threat posed by failure of the referendum. Further, as found in previous works on city-level measures, the odds of a yes vote were higher in counties with more college-educated voters. But there was also a paradoxical result—the odds of a yes vote increased in counties with higher unemployment.

The paper proceeds as follows. In the next Section II, we give the background on the election. Section III contains our inspection of the geographic distribution of yes votes across the state and its relationship to population and income. The county-level, representative voter model is in Section IV, along with a data description and a summary of the results. Conclusions round out the paper in Section V.

II. Background

Referendum 48 was decided in a state-wide special election on June 17, 1997. The specific details of the stadium funding can be found in the 28 page *Official Washington Voter's Pamphlet* (Secretary of the State of Washington, 1997). Overall, the ballot stated that the stadium would cost approximately $425 million with a 76-24 public-private split ($323 million public money). The referendum passed by a slight 51.1 percent of the popular vote—820,364 yes; 783,584 no (Secretary of the State of Washington, 2009).

Some revenue elements to cover the public portion were added diversions from private spending; $95 million would be covered by a ticket tax, a parking tax, and a King County

(Seattle) room tax extension. The remaining elements in the public portion, although touted otherwise, were direct diversions of funds spent elsewhere across the state on public services. The most obvious of these was sales tax forgiveness amounting to $101 million. Less obvious was $127 million from a new lottery game. To the extent that new lottery games simply redistribute a given propensity in the population to gamble, this new game would divert funds from their previously allocated purpose. In 1997, lottery funds were dedicated to education construction projects for K-12 and higher education, economic development, problem gambling prevention and treatment, and the state's General Fund. Thus, there are impacts not just on private spending through new revenue devices, but also on the previous distribution of public spending.

The final element in the public portion is especially interesting; $27 million came in tax breaks to the builders of the stadium. Economically, it is difficult to determine the true cost of this $27 million "contribution." If the next best opportunity for these builders was a purely private endeavor, was $\frac{\$27\text{million}}{\$425\text{million}} \cong 6.4$ percent the "going rate" tax break on privately financed development? If the next best opportunity for these builders was a purely public endeavor, then this $27 million appears to be a phantom contribution; the public never would have borne this cost in the first place since, presumably, the same tax break would have applied. If the latter was the case, then the true cost of the stadium was actually $398 million and the public-private split was 74-26 ($296 million public).

Proponents did all they could to portray Referendum 48 as essential to keeping the Seahawks in Seattle, building on threats and actions by the previous owner, Ken Behring, to move the team to California (this and following details are in Fort 1999). Los Angeles was without an NFL team and various owner interests in the L.A. area were actively pursuing NFL teams. Behring tried to move the Seahawks to Hollywood Park just prior to the referendum episode. He was turned back by the NFL under a league-enforced cooling off period.

During the cooling off period, Paul Allen paid $10 million for an option to buy the Seahawks and made it clear he would only exercise this option if a new, publicly funded stadium would be built. If the option expired, the team would still belong to Behring whose past behavior predicted that the team would move. Thus, Allen's option to purchase left the underlying threat that voters would lose their team if they did not come through with the stadium subsidy by passing Referendum 48.

III. The Geographic Distribution of Yes Votes, Population, and Income

Figure 1 (reference) shows the geographic distribution of yes votes on Referendum 48 by color-coding the yes-vote percentage by county. We note that the Cascade Mountains split the state into what residents refer to as the West-side and the East-side. Figure 1 makes clear that the West-side carried the vote; proximity to the eventual site of Qwest Field in Seattle, King County, coincides with the strength of the vote in favor of Referendum 48. With the exception of Benton County, all are symmetrically distributed around Seattle, taking into account the presence of Puget Sound.

Adding the population and income data in Table 1, it is safe to characterize the counties that passed Referendum 48, separately portrayed in Table 2, as follows. They are 1) in the Northwest corner of the state, proximate to the eventual location of Qwest Field in Seattle, 2) the most heavily populated counties in the state, and 3) at the top of the state income distribution. There are nearly no exceptions to this characterization. Among counties that passed the referendum, Benton County is not in the Northwest corner, but it ranks 10^{th} in population and 5^{th} in county income, and Grays Harbor County stands out only for its low income ranking.

Among counties that did *not* pass the referendum, San Juan County appears to be in close proximity but actually is quite isolated as a group of very small islands reachable only by air or ferry (Island County is aptly named but in contrast to San Juan County has a bridge to Snohomish County). Clark, Spokane, and Yakima Counties are top-10 population areas but all are quite distant. In addition, Spokane and Yakima Counties are far down the income distribution. Clark and San Juan Counties have high income but Clark is quite distant from Seattle and, as just mentioned, San Juan is quite isolated.

For whatever reason, Benton County voters followed the West-side vote. It may be that Benton's large population (10^{th}, primarily in the "Tri-Cities" of Kennewick, Pasco, and Richland) and income (5^{th}) cause it to support the referendum despite its marginally manageable distance from Seattle. For example, Spokane County also is highly populated but quite far from King County and voted heavily against the referendum.

To make the story complete, and quite symmetrical, the counties with the lowest proportion of yes votes are in the farthest corners of the state away from King County. Not surprisingly, these regions also happen to have low population compared to the rest of the state (except for aforementioned Spokane County). The distribution of yes votes suggests that the more populated, more affluent West-side took full political advantage over the rest of the state. Nine counties symmetrically around the proposed location of the stadium, plus Benton County, had their way over the remaining 29 counties in the state since the costs were borne state-wide.

IV. "Representative Voter" Analysis

The calculus of the individual referendum voting decision was originally explored in detail by Borcherding and Deacon (1972), and Deacon and Shapiro (1975). As they point out, the main problem is that individual voting is not observed. Lacking individual voting data, it is typical to appeal to a pivotal voter like the "median voter" (Downs 1957). Under this choice, the individual calculus informs empirical analysis, but one only need analyze one representative individual, namely, the voter holding the median expenditure preference.

However, two other issues arise. What statistics represent the median voter? For example, does the pivotal median preference follow from being at the median in income? In addition, what is the appropriate level of aggregation for the statistical characterization of the pivotal voter?[1] Finally, there may be situations where actual spending outcomes are not the median voter's most preferred outcome. This is the well-known Romer and Rosenthal (1978) "reversion threat" under agenda control. Filimon, Romer and Rosenthal (1982), Fort (1988), and Fort (1997) list the many works that followed that original work (including more by Romer and Rosenthal). Chang and Turnbull (2002) summarize the numerous works where agenda control has mattered in empirical analyses of voting.

The literature on referendum voting in the sports context is not quite so extensive. Agostini, Quigley and Smolensky (1997) estimated a voting model at the precinct level for a ballpark in the San Francisco area, finding that socio-demographic data shape voting preferences. Fort (1997) studies stadium funding referenda for their general outcome characteristics and the importance of agenda control but does no formal estimation of voting outcomes. Brown and Paul (1999) argue in support of the classic concentrated benefits/dispersed costs public choice outcome for a city referendum in Cincinnati. Fort (1999) categorizes referenda in terms of their impact on spending levels compared to spending levels determined by direct democracy. Depken (2000) estimates fan loyalty and shows that it helps determine voting outcomes on nine city stadium elections. Coates and Humphries (2006) analyze city-level NFL stadium votes in Houston and Green Bay finding what they refer to as "proximity value;" being close to the facility increases yes votes. Dehring, Depken and Ward (2008) analyzed the new NFL Cowboy stadium in Arlington, Texas, finding that homeowners voted in favor of increased property values. But none of these works explicitly accounts for agenda control, and our work adds to the literature by examining a state-level election.

Following from the analysis of the county-level distribution of yes votes in the last section, we are curious about the determinants of these county-level results. We humor this curiosity by adopting a county-level aggregation. But we also try to account for the fact that there may be agenda control since the election was quite close (51.1% in favor) and closeness of the actual election outcome is indicative of agenda control (Fort 1988; Fort 1997). To capture the flavor of the story in Figure 1 more precisely, we employ the following county-level voting model:

$$\log\left(\frac{Y_i + \frac{1}{2}}{N_i + \frac{1}{2}}\right) = \alpha + \beta X_i + \varepsilon_i \text{, i = 1, …, 39 counties,} \qquad (1)$$

where Y_i = yes votes in county j, N_i = no votes in county i, X_i is the vector of explanatory variables for county i, α and β are parameters to be estimated, and ε is the error term. This "traditional" estimation technique is discussed fully in Pindyck and Rubinfeld (1997) and we use

[1 1] A comprehensive review is presented in Hoxby (2000).

the precision improvement of adding ½ in both the numerator and denominator as suggested by Cox (1970). The traditional technique estimates the model in expression (1) by ordinary least squares corrected for heteroskedasticity (for example, using "White's Correction").

In addition to the traditional technique, Theil's "group logit" approach is recently touted for estimating vote shares for representative voter models in the political science literature (Mikhailov, Niemi and Weimer2002). Each county differs in the level of votes cast and, thus, the variance of the vote in each county is different. To increase efficiency, a group logit model gives greater weight to the outcomes with smaller variances. For our problem, let P_i = population in county i and at issue is the improvement using a weight proportional to P_i. Under the group logit, feasible generalized least squares is used to estimate:

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha' + \beta' X_i + \varepsilon'_i \text{, } i = 1,\ldots,n \text{ counties, where } p_i = \frac{Y_i}{P_i} \qquad (2)$$

Turning to specification of independent variables, we follow the general line suggested in the voting literature: people vote in their own self-interest, subject to the price they will pay for the outcome. Our primary measure of net benefit is proximity (data descriptions and descriptive statistics are in Table 3): we expect the net benefits of a new stadium to be higher for residents of counties closer to the stadium, measured by the distance from a given county seat to Seattle.

The rest of the economic calculus involves price, income, and population. Price is problematic because of the multi-source revenue specification in the referendum (a variety of taxes, lottery and sales tax diversion, and tax forgiveness). So, we take the approach popular in the voting literature and identify groups whose welfare would be expected to change in predictable ways. Relevant to sales tax forgiveness, individuals dependent on existing unemployment programs, and other related state support programs, would not want to subsidize a professional football stadium, measured by unemployed per capita. We hypothesize that the elderly, measured by retired people per capita, should be opposed to paying for long-term capital projects since they are more likely to enjoy only a relatively shorter period of benefits. In addition, and especially for lottery diversions, families with children in school should feel schools threatened by fund diversion, measured by CHILDREN in school per capita.

For the remaining two variables, income is problematic. Despite the old econometric safe that "everything is correlated in one way or another," we find multicollinearity especially troublesome for income as shown in Table 4. An acceptable approach to this type of problem is to omit a variable and we do so (lamentably) for any explicit income measure in favor of the added explanation possible from the broader array of other independent variables. Population is a control variable and we allow for non-linear impacts with its squared value.

As stated in the introduction, it was possible that the election was subject to agenda control; vote yes or the team will leave. This threat was more than speculation since voters had just witnessed the near-move of the team to California by owner Ken Behring and it was reinforced in the purchase option choice by Paul Allen. We reiterate that the closeness of the election, 51.1 percent in favor, is indicative evidence of agenda control.

Of course, voters are free to assess the chance that this will really occur. To incorporate this possibility, we devise a proxy that measures the highest-level of football voters will be able to enjoy if the election failed. Voters can reach one of two Division 1A football alternatives, located nearly completely at the diagonal extreme across the state from each other—the University of Washington Huskies in Seattle (King County) and the Washington State University

Cougars in Pullman (Whitman County). Since the proposed NFL stadium and the University of Washington alternative are both in Seattle, we constructed the following variable. If the distance from a given county seat to the King County seat was smaller than the distance to the Whitman County seat, then CLOSEST = 0; else CLOSEST = distance between the county seat and the Whitman County seat (just a few miles from Pullman). If the coefficient estimate on this variable ends up positive, then the odds of a yes vote are larger the farther away is the college alternative. This would be consistent with voters taking to the booth the belief that the pro alternative would be lost in the event the election failed. If the coefficient estimate is zero, voters discounted that possibility.

Two other controls seem reasonable. Past work at the precinct-level on city measures found that education increased the probability of yes votes. Thus, BS (bachelor degrees per 1000 population) is employed as an independent variable for high demanders. We also hypothesize that general political leanings should effect any predisposition toward spending, measured by the ratio of the number of votes cast for the democratic candidate Clinton to the number of votes cast for Republican candidate Dole in the 1996 general Presidential election.

Our empirical results are in Table 5 for the traditional approach (expression (1)) and the group logit approach (expression (2)). Precisely the same specification of independent variables, and precisely the same data, makes R^2 a useful tool for comparing estimation techniques. Goodness-of-fit improves with the Theil group logit compared to the traditional approach (adjusted R^2 increases by 0.273, about 59%). In addition, more variables are significant and enter with hypothesized signs. For these data, the group logit approach to voting outcomes offers precisely the distinct improvement in estimation argued by Mikhailov, Niemi and Weimer (2002). The rest of our discussion proceeds relative to the group logit results in Table 5.

Most estimated coefficients are of expected sign. Odds of voting yes fall with DIST so we also find support for proximity value for voters. For our attempts to capture price effects, we find only that the odds of voting yes *increase* with UNEMP, counter to our expectations, and none of the other "price" variables matter. This is an unexpected outcome. Perhaps this is just multicollinearity since the simple correlation coefficient between UNEMP and BS is –0.68 (Table 4) and BS is statistically significant. Or perhaps the cost of voting for unemployed people is simply very low and they are bigger fans than employed people! The odds of a yes-vote increase at a decreasing rate with POP. All-in-all, these particular results are just what one would expect from our inspection of the distribution of yes votes, population, and income in the last section.

To go along with the possibility of agenda control indicated by the closeness of the election outcome, we find evidence that voters incorporated the loss of the team into their voting rather than discounting this threat. However, the evidence is not overwhelming; the coefficient on CLOSEST is positive but only marginally significant (90 percent level of confidence). That our education control BS is significant, increasing the odds of voting yes, is consistent with other research on voting at the precinct-level on city measures.

We have no explanation for the statistical insignificance of RETIRED, CHILDREN, and DEMO beyond the obvious—either our variables are not capturing what we intended, or this type of logic useful in past voting studies does not hold for this particular special election (perhaps sports really are different, after all).

V. Conclusions

On June 17, 1997 voters in the state of Washington passed Referendum 48 with 51.1

percent of the popular vote. Qwest Field was eventually constructed and professional football remained in the state of Washington. Eventual owner Paul Allen purchased a time sensitive option to buy the Seahawks, spent millions on advertising, and covered the cost of the special election. The election was clearly characterized by the threat that the team would be lost in the event of referendum failure by then-owner Ken Behring and by owner-in-waiting Paul Allen.

Direct inspection of the geographic distribution of yes votes indicates that proximity carried the election (with one county as an exception to this rule). Adding basic income and population data to this mix then suggests that more highly populated, richer counties, all proximate to the eventual stadium location in Seattle, ruled the election outcome.

In a county-level representative voter model, we discover a number of interesting things. First, the results of direct inspection are supported. As with proximity value findings by others analyzing city-level precinct level data, the odds of voting yes decrease with distance from the proposed facility. Second, odds of voting yes increase at a decreasing rate with population. Third, we find (weak) evidence that voters believed the threat that they would lose their team if the election failed—the election barely passed and the odds of voting yes increased the farther voters were from the next best, high-level, college football alternative. In addition, although we tried more than one variable to capture price and income impacts, only unemployment mattered and, paradoxically, increased the odds of a yes vote. As in other studies, education increased the odds of voting yes. Variables intended to capture voting by the elderly and parents of school-age children, as well as relative Democratic voting sentiments, all were insignificant.

Thus, the evidence is quite strong that ten of the thirty nine counties in the state passed the referendum. Nine of these ten counties were located symmetrically around the proposed location of the stadium. They all are richer and more densely populated than the rest of the counties in the state. This suggests that the more populated, more affluent, West-side of the state took full advantage of its political power over the rest of the state in building itself a new football stadium even though the costs were borne state-wide.

Bibliography

Agostini, S. J., Quigley, J.M., and Smolensky, E. 1997. "Stickball in San Francisco." In R. G. Noll and A. Zimbalist (Eds.) *Sports, Jobs and Taxes: The Economic Impact of Sports Teams and Stadiums*. Washington, D. C.: Brookings Institution, pp. 385-426.

Borcherding, T.E., and Deacon, R.T. 1972. "The Demand for the Services of Non-Federal Governments." *American Economic Review* 62:891-901.

Brown, C., and Paul, D.M. 1999. "Local Organized Interests and the 1996 Cincinnati Sports Stadia Tax Referendum." *Journal of Sport and Social Issues* 23:218-237.

Chang, C., and Turnbull, G.K. 2002. "Bureaucratic Behavior in the Local Public Sector: A Revealed Preference Approach." *Public Choice* 113:191-210.

Coates, D., and Humphreys, B.R. 2006. "Proximity Benefits and Voting on Stadium and Arena Subsidies." *Journal of Urban Economics* 59:285-299.

Cox, D.R. 1970. *The Analysis of Binary Data*. London: Metheun and Co.

Deacon, R., and Shapiro, P. 1975. "Private Preference for Collective Goods Revealed through Voting on Referenda." *American Economic Review* 65:943-955.

Dehring, C.A., Depken, C.A., II, and Ward, M.R. 2008. "A Direct Test of the Homevoter Hypothesis." *Journal of Urban Economics* 64:155-170.

Depken, C.A., II. 2000. "Fan Loyalty and Stadium Funding in Professional Baseball." *Journal of Sports Economics* 1:124-148.

Downs, A. 1957. *An Economic Theory of Democracy*. New York, NY: Harper and Row.

Filimon, R., Romer, T., and Rosenthal, H. 1982. "Asymmetric Information and Agenda Control: The Bases of Monopoly Power in Public Spending." *Journal of Public*

Economics 17:51-70.

Fort, R. 1988. "The Median Voter, Setters, and Non-repeated Construction Bond Issues." *Public Choice* 56:213-231.

Fort, R. 1997. "Direct Democracy and the Stadium Mess." In R. G. Noll and A. Zimbalist (eds.) *Sports, Jobs and Taxes: The Economic Impact of Sports Teams and Stadiums*. Washington, D. C.: Brookings Institution, pp. 146-177.

Fort, R.D. 1999. "Stadium Votes, Market Power and Politics." *University of Toledo Law Review* 30:419-441.

Hoxby, C.M. 2000. "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review* 90:1209-1238.

Mikhailov, N., Niemi, R.G., and Weimer, D.L. 2002. "Application of Theil Group Logit Methods to District-level Vote Shares: Tests of Prospective and Retrospective Voting in 1991, 1993, and 1997 Polish Elections." *Electoral Studies* 21:631-648.

Pindyck, R.S., and Rubinfeld, D.L. 1997. *Econometric Models and Econometric Forecasts, 4th Ed.* New York, NY: Irwin Professional.

Romer, T., and Rosenthal, H. 1978. "Political Resource Allocation, Controlled Agendas, and the Status Quo." *Public Choice* 33:27-43.

Secretary of the State of Washington. 1997. *Official Washington's Voter Pamphlet*.

Secretary of the State of Washington. 2009. "Elections & Voting." URL last accessed March 11, 2009: http://www.secstate.wa.gov/elections/results_report.aspx?e=11&c=&c2=&t=&t2=&p=&p2=&y= .

State of Washington. 2008a. "Historical/Current Data Set: Total Resident Population by Year by County, Washington, 1960 to 2008." Office of Financial Management, July, 2008. URL last accessed March 11, 2009: http://www.ofm.wa.gov/pop/coseries/default.asp

State of Washington. 2008b. "Median Household Income Estimates by County: 1989 to 2007 and Projection for 2008." Office of Financial Management, October, 2008. URL last accessed March 11, 2009: http://www.ofm.wa.gov/economy/hhinc/default.asp

U.S. Census Bureau. 2009. "County and City Data Book: 2000." URL last accessed March 11, 2009: http://www.census.gov/prod/www/ccdb.html .

Figure 4.1: Referendum Bill 48: Geographic Voting Outcomes

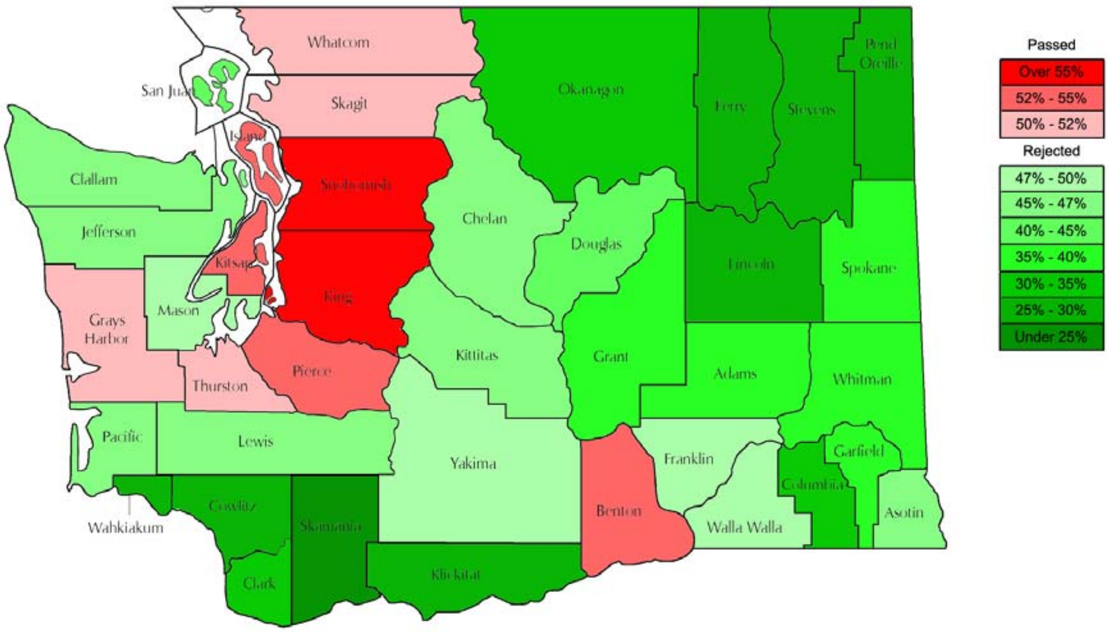


Table 4.1: County Population and Median Household Income, Washington State, 1997

	Population	Rank	Income	Rank		Population	Rank	Income	Rank
State	5,663,763		42,399						
Adams	15,989	31	31,795	30	Lewis	67,350	16	33,610	26
Asotin	20,273	29	31,499	31	Lincoln	9,883	34	35,838	17
Benton*	136,503	10	44,057	5	Mason	47,277	20	36,524	16
Chelan	65,936	17	35,662	19	Okanogan	40,277	23	28,047	39
Clallam	62,889	18	34,770	22	Pacific	20,813	28	28,974	37
Clark	317,324	5	45,705	3	Pend Oreille	11,816	33	31,223	33
Columbia	4,527	37	30,820	34	Pierce*	668,103	2	42,596	7
Cowlitz	90,728	12	36,738	14	San Juan	12,906	32	41,134	9
Douglas	31,252	26	36,855	13	Skagit*	97,848	11	38,449	10
Ferry	7,127	36	30,489	35	Skamania	9,559	35	37,409	12
Franklin	47,206	21	35,770	18	Snohomish*	557,016	3	50,680	1
Garfield	2,252	39	34,792	21	Spokane	409,553	4	34,920	20
Grays Harbor*	68,188	15	31,368	32	Stevens	37,609	24	32,435	29
Grant	70,433	13	33,977	25	Thurston*	199,081	8	43,748	6
Island*	68,967	14	41,901	8	Wahkiakum	3,883	38	36,566	15
Jefferson	25,116	27	34,282	23	Walla Walla	55,238	19	34,094	24
King*	1,679,516	1	48,271	2	Whatcom*	157,460	9	37,553	11
Kitsap*	228,181	6	44,098	4	Whitman	40,815	22	28,697	38
Kittitas	32,325	25	29,775	36	Yakima	223,917	7	32,946	28
Klickitat	18,627	30	33,543	27					

Source: State of Washington (2008a, 2008b).
Note: * denotes that Referendum 48 passed in the county.

Table 4.2: Population and Income, Counties that Passed Referendum 48, 1997

	Population	Rank	Income	Rank
Benton	136,503	10	44,057	5
Grays Harbor	68,188	15	31,368	32
Island	68,967	14	41,901	8
King	1,679,516	1	48,271	2
Kitsap	228,181	6	44,098	4
Pierce	668,103	2	42,596	7
Skagit	97,848	11	38,449	10
Snohomish	557,016	3	50,680	1
Thurston	199,081	8	43,748	6
Whatcom	157,460	9	37,553	11

Source: See Table 1.

Table 4.3: Variables and Descriptive Statistics

Variable	Description	Min	Max	Mean	S.D.
YES	Number of yes votes	315	275,368	21,035	48,410
NO	Number of no votes	634	213,092	20,092	37,891
DIST	Driving distance to Seattle	0	350	164	95
UNEMP	Unemployment percent of labor force	2.2	13.7	7.2	2.7
RETIRED	Population 65 years and older per capita	0.084	0.224	0.140	0.038
CHILDREN	Public school enrollment per capita (1998-1999)	0.124	0.230	0.179	0.027
POP	Total population of county	2,397	1,737,034	151,131	305,164
CLOSEST	Smaller of driving distance to Seattle or Pullman	0	212	105	55
BS	Bachelor Degrees per 1000	10.4	42.6	17.1	7.0
DEMO	Votes cast for Clinton divided by Votes cast for Dole, 1996	0.689	1.96	1.12	0.344

Sources: Voting data are from the Secretary of the State of Washington (2007).

Demographic data are from U.S. Census Bureau (2009).

Table 4.4: Correlation Matrix

	DIST	UNEMP	RETIRED	CHILDREN	POP	CLOSEST	BS	DEMO	INCOME
DIST	1								
UNEMP	-0.342	1							
RETIRED	-0.041	0.309	1						
CHILDREN	-0.194	0.479	-0.096	1					
POP	-0.436	-0.362	-0.373	-0.175	1				
CLOSEST	0.403	0.662	-0.005	0.320	-0.448	1			
BS	-0.217	-0.681	-0.221	-0.554	0.393	-0.563	1		
DEMO	-0.502	-0.259	0.168	-0.433	0.354	-0.259	0.171	1	
INCOME	-0.614	-0.630	0.020	-0.480	0.557	-0.437	0.590	0.468	1

Table 4.5: Regression Results

Variable	Traditional OLS	Group Logit
Constant	-0.343	0.022
	(1.005)	(0.850)
DIST	-0.003*	-0.004*
	(.001)	(0.001)
UMEMP	-0.016	0.076***
	(.040)	(0.039)
RETIRED	0.499	0.025
	(1.699)	(2.02)
CHILDREN	2.225	-3.54
	(3.298)	(3.50)
POP	7.06E-07	1.03E-06*
	(4.84E-07)	(3.80E-07)
POP^2	-3.89E-13	-6.38E-13*
	(2.65E-13)	(2.07E-13)
CLOSEST	0.001	0.002***
	(.001)	(0.001)
BS	0.016	0.031**
	(.014)	(0.013)
DEMO	-0.170	-0.233
	(.225)	(0.185)
#OBS	39	39
DF	29	29
R^2	0.584	0.798
$AdjR^2$	0.462	0.735

Notes: Standard errors are in parentheses. *Significant at 99% level. **Significant at 95% level. ***Significant at 90% level.