Measuring Racial Profiling

Katherine Barnes[1]

James E. Rogers College of Law, University of Arizona

preliminary and incomplete

do not cite or quote without permission

1 Introduction

What if racial profiling were useful? Even, perhaps, very useful? Are the costs of racial profiling so significant that racial profiling should be banned? Courts and commentators do not ask these questions; indeed, they rarely, if ever, acknowledge that racial profiling may be useful. This paper explores what costs police would have to justify, and how they might do so in order to use racial profiling legally.

Using race as a part of a profile of criminal has some efficiency value; it is not solely (or even at all) about racial animus. Race is a marker of criminal behavior; it would be miraculous if it where not. Height, weight, gender, hair style, ear piercings, all of these characteristics almost surely delineate some difference in the commission rate of at least one crime. This is, essentially, a mathematical truism: race and crime are not independent. Because race is "useful" when "useful" is defined as "information worth more than zero," it is important to move beyond the initial question — is it useful — to the more important question — how useful is it? What do we give up in allowing or banning racial profiling? What are courts missing when they discuss racial profiling? This paper is an attempt to get at least some of the empirics correct regarding the direct costs and benefits of racial

[1]This project has been support by NSF Law & Social Science and Measurement, Methodology & Statistics Sections Grant no. 1024389. I would also like to acknowledge the assistance of Professor John Hobbins at the McGill Law Library, for kindly providing work space and access to materials in Summer 2010.

profiling; it leaves the important research of mapping the collateral consequences of racial profiling to others.

But even when one puts aside the collateral consequences of focusing the might of the criminal justice system on one specific race, there are many complicating factors in measuring how useful racial profiling. First, how do you measure the gap in crime rates (that is, the different crime rates for different racial groups and different crimes)? Without this, no one could calibrate their behavior: police would not know how much to profile, and the courts would not know how much reliance on race in police decision-making is too much. This is also important as a policy matter, to determine what society loses if racial profiling were not allowed. To be more concrete, what if whites were twice as likely to sell cocaine as Blacks were? Does this justify profiling them? Perhaps the answer to this depends on the base rate: how much more cocaine do white people sell? Is this a difference between one and two percent, or fifteen and thirty percent of individuals who could be targeted. One needs to know more than just a single statistic. A second complicating factor, beyond the initial measurement question, is whether race is too easy to over-use. People see race quickly, and believe it is a highly salient part of a persons physical description; police might over-rely on this characteristic because it is easy to determine (or at least perceived to be so). A third complication is that race is rarely the only available information about a potential suspect, and a better question than whether race is correlated with crime rates is whether information about a suspects race adds any useful information to a profile. But, again, the chance that these two probabilities are always equal, for every type of crime and any other possible set of information is zero. Race matters. This bears emphasis not because it isnt obvious, but because many opposed to racial profiling deem it inefficient automatically; that is, they argue that racial profiling does not work because race is not a predictor of criminality.

Answering these questions is a difficult, but not insurmountable task. But these questions also suggest that a simple focus on whether race is a predictor of criminality is misplaced; the answer to that first question is yes. The more important question, which legislatures, courts, legal and economic researchers alike have generally ignored, is when and how useful racial profiling is in different situations. To answer this question, one needs to investigate the costs and benefits of racial profiling, which the above complicating factors suggest can be done, but needs more attention to the detail of the reality of the relationship of race and crime to answer. This paper seeks to quantify the costs and benefits of racial profiling, and in doing so creates new methods to control for the problem of selection bias in the data that are generally available to evaluate this question.

2 The Problem of Selection Bias in Racial Profiling

Statisticians have long recognized the problem of selection into samples that selection into a sample, if non-random, can create significant bias. The prototypical example is an examination of womens wages in the 1970s: women may choose not to work, meaning that they self-select to remain out of any sample of womens wages. A sample of employed women by definition includes only those women who do work. Performing inference on this non-random sample incorrectly estimates the importance of various factors on womens wages. (Heckman 1979).

The same selection bias problem occurs in a wide variety of circumstances, and is particularly acute in many legal applications where researchers must rely upon criminal justice data. Indeed, scholars have long recognized the threat of selection bias in data related to the criminal justice system, and have focused on racial profiling as one context in which selection bias is a significant concern. Racial profiling involves acting on the belief — be it based upon racial animus, rational, or a combination of the two — that certain crimes are committed disproportionately by a particular race. Unfortunately, data on criminal activity — necessary to disentangle whether and why racial profiling occurs, are not free of society's assumptions and stereotypes. Not only does society shape views on race including stereotypes and assumptions — but society's views also feedback to shape the society we live in and how one describes it empirically. Look for drug couriers only among Blacks and you will likely find some; what you wont find is a white drug courier. Then look at the people youve arrested and jailed, and it is easy to see how the initial perceptions of criminality get reinforced. Over decades, a stereotype that Blacks are more likely to engage in criminal activity can transform itself into large and statistically significant differences in arrest and incarceration rates for Blacks and whites.

There are some data suggesting that American society has experienced this form of self-fulfilling racial bias in drug prosecutions. For example, Black youth are seen as more likely than white youth to engage in drug use and violent behavior; arrest and incarceration rates reinforce this perception. Yet self-reported data consistently reveals that except for homicide, Black and white youth have similar propensities to violent crime and Blacks are, if anything, less likely than whites to use illegal drugs. Because biases can be reinforced through action, an initial focus on Blacks as drug couriers will be validated and reinforced by the data: police will believe that their profile is accurate, because they arrest more Blacks for drug possession. Put simply, the data on stop and search rates and the data on arrest and incarceration rates alone cannot demonstrate that Blacks are more likely to be guilty of drug-related offenses. Without an answer to this empirical question, one cannot separate fact from stereotype in the debate surrounding racial profiling as a tool of drug interdiction.

This project creates methods to control for selection bias and other data problems. In doing so, it seeks to separate fact from stereotype by estimating empirically the extent to which the stereotypes used to justify racial profiling are based on actual propensities to commit drug crimes. In particular, the project investigates the counterfactual: what would police find if they did not use race in deciding whom to search during highway traffic stops? By doing so, the methods allow quantification of the potential benefits and costs of using race as a factor in the decision to stop and search a vehicle. Quantifying these costs and benefits brings empirical reality, rather than stereotype and supposition, to the normative debate regarding the efficacy and justifications of racial profiling. The new methods, of course, are not limited to the racial profiling context, and provide an important tool to researchers who seek to control for selection bias in the wide variety of contexts where individual-level data on the non-selected population is unavailable.

3 Literature Review

3.1 Legal and Policy Motivation

Racial profiling has attracted a great deal of interest from researchers interested in understanding whether racial profiling is an efficient use of police resources, or simply a product of racial animus. Much of the research has been conducted by economists, with Knowles, Persico and Todd (2001) (KPT) being the leading example. KPT posits a straightforward game theoretic model of discrimination in police stops and searches. Using data from Maryland, KPT determines that any use the state police make of race in their decision making is rational; that is, it furthers the polices legitimate goal of drug interdiction, rather than being based on racial animus. Several extensions to this model have relaxed assumptions and used different identification strategies. (Antonovics and Knight 2005; Anwar and Fang 2005; Hernandez-Murillo and Knowles 2004; Persico and Todd 2005). All of these projects assume, contrary to existing law, that a rational use of race in decision making is unproblematic; in fact, the use of race by state actors such as police must be justified by a much higher standard to pass constitutional muster. See Grutter v. Bollinger, 539 U.S. 306 (2003). Persico (2008) notes this disparity, but concludes that it is unimportant, conflating the legal question of whether race was used with the legal justification of the use of race. Instead, KPT seek only to differentiate between types of racial profiling — the "good" kind, which is rational and based on Blacks having an a priori higher payoff to carrying drugs, and the "bad" kind, which is irrational, inefficient, and solely based on racial animus. Antonovics and Knight suggest that this is the legally salient question. On a practical level, Antonovics and Knight may simply be suggesting that causation is difficult to prove in racial profiling cases; a rational use of race might, in fact, be a rational use of some unobserved characteristic of the motorists that is simply correlated with race.[2] Nonetheless, whether the police engage in racial profiling is an important threshold question, as it determines what level of scrutiny is applied in an Equal Protection challenge. [3]

[2] Some courts do this as well.

[3] See Grutter v. Bollinger, 539 U.S. 306, 326 (2003) (reaffirming that use of race in government decision-making triggers strict scrutiny under the Equal Protection Clause).

At best, the KPT model is agnostic with respect to whether race is an explicit factor in the troopers decision to search.

The equal protection standard articulated in Grutter requires the state to articulate a compelling interest (e.g., drug interdiction), and that the means by which the state realizes its goal be narrowly tailored to minimize the extraneous racial effect of the policy. At a minimum, this requires some knowledge of how the policy allowing the use of race in decision-making affects minorities, and what the benefits of such a policy are — that is, a cost/benefit analysis. Instead of performing this cost/benefit analysis, economics researchers conflate the question of whether police use race in decision-making with the question of whether that use is justified.

A second strand of research focus solely on the threshold question of whether racial profiling occurs. (Harris 2002; Lovrich 2003). Harris (2002) is a leading example of this strand of scholarship, but the analysis neglects to model the complex process by which police make search decisions, thereby provided too simple a portrayal of the process in general, and racial profiling in particular. Lovrich et al. engage in sophisticated modeling of the decision to stop or search, but do not obtain outcome data on the results of those decisions, and therefore are unable to perform . (Lovrich et al. 2003) This strand of research, however, accurately separates the two legal questions regarding racial profiling (first, whether it occurs, and second, whether it is justified if it does); the research simply stops too soon in the analysis, both for lack of data and lack of appropriate methods to fully utilize the data available.

In prior work, I investigated whether racial profiling of stopped vehicles using race as a factor in the decision to search a stopped vehicle was productive (Barnes 2005). The results suggested that the police did engage in racial profiling, and that its effects were ambiguous: while disproportionately searching vehicles stopped with Black drivers increases the amount of drugs seized slightly, it also increases the number of innocent motorists whose vehicles are searched. This cost suggests that the policy may not be narrowly tailored. But the universe of stopped vehicles was clearly not a random sample of drivers on the highway; based on a study of the race of drivers on the highway, Black motorists were 1.7 times more likely to be stopped than white motorists. Without a random sample of drivers on the highway, the full picture of racial profiling is unknown. Because the Heckman selection model I use only controls for the second step of the selection, from stop to search, the conclusions I draw are valid only if one assumes that the police will not change their selection of whom to stop. The selection model cannot control for the first step: the decision to stop a vehicle driving down the highway, and therefore must assume that the universe of stopped vehicles is equivalent to the universe of vehicles driving down the highway. This lack of individual-level data on highway drivers motivates the statistical model that follows.

3.2 Statistical Motivation

Currently, there are several statistical methods to control for selection bias. The best known is the Heckman selection model, created by economist James Heckman. (Heckman 1979). While a significant leap in the ability to control for selection bias, the Heckman selection model has two primary disadvantages. First, its parametric assumptions identify the model without any information from the data, meaning that the assumptions of the model, rather than the data, may drive the results. (Heckman and Honore 1990). Second, the Heckman selection model requires the researcher to obtain individual-level data on the non-selected population in addition to individual-level data on the selected population. In many cases, these data are unavailable. In the racial profiling context, one question of interest is whether police use the race of the driver in their decision to search a vehicle. In order to answer this question, one would need to control for the mix of drivers that a police officer could stop, along with the mix of suspicious behavior in which different drivers engage. There are several selection mechanisms involved here: a driver must decide to drive down the highway (and decide whether to carry contraband); a police officer must decide whether to stop a vehicle; a police officer must decide the level of scrutiny to apply to the driver and car to determine whether to search the car; and, finally, the police officer must decide whether to search the car, and with what intensity. All of these decisions to drive (with or without possessing drugs), stop or search a carare non-random, and depend on the individual propensity of the driver to carry drugs. Without controlling for these decisions, it is impossible to answer many questions about racial profiling. Finally, without individual-level data on individuals who did not get stopped, or searched, or did not drive down the highway, even the Heckman selection model cannot be estimated.

With respect to the first disadvantage, several researchers, including Heckman himself, have worked to create nonparametric selection models that allow the data, rather than the parametric model assumptions, to determine the results. (See, e.g., Heckman and Vytlacil 2006). This project focuses on the second disadvantage of the Heckman selection model, that it requires individual-level data on the non-selected population. This is a significant limitation because such data can be very difficult to obtain.

A second general method that corrects for confounding of variables, and to some extent selection bias, is propensity score matching. This method has been used successfully in legal empirical applications (Ho 2005), and is widely used in other disciplines. It too, however, requires individual-level data.

4 A Bayesian Model of Selection

This project tackles the second disadvantage of the Heckman selection model: the need for individual-level data. I develop a method by which a researcher can control for selection bias using aggregate data: means, standard deviations, and correlations of variables rather than the actual data from individuals. As an illustration of the method, I estimate the model using drug interdiction data to determine whether police engage in racial profiling.

To create a Bayesian model of selection, I modify the rubric of Geweke et al. (2003) (GGT), and model the selection as a two-stage process: first, I model the police decision to stop and search an individual traveling on the highway as a Bayesian probit regression, and second, for those that they search, I model the quantity of drugs seized as a standard Bayesian regression. As with the standard Heckman model, the two equations are related by correlated errors. The model differs from GGT because the exogenous variables of the outcome equation (X_i) are not observable if the individual was not searched. The full model is given by:

$$\begin{bmatrix} y_{1i}^* = X_i\beta + \epsilon_{1i} \\ y_{2i}^* = Z_i\beta + \epsilon_{2i} \end{bmatrix} \quad \text{with} \quad \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \sim N\begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}$$

and the following definitions:

$$\begin{aligned} y_{2i} &= 1 \text{ if } y_{2i}^* > 0 \\ y_{2i} &= 0 \text{ if } y_{2i}^* \leq 0 \\ y_{1i} &= 1 \text{ if } y_{1i}^* > 0 \,\&\, y_{2i} = 1 \\ y_{1i} &= 0 \text{ if } y_{1i}^* \leq 0 \,\&\, y_{2i} = 1 \\ y_{1i} & \quad \text{is undefined if } y_{2i} = 0 \end{aligned}$$

In this model, y_{1i}^* and $y_{2i}^*$are latent (unobserved) variables that determine the observed dependent variables y_{1i} and y_{2i}. The first model, with $y_{1i}^* = X_i\beta + \epsilon_{1i}$, is the latent outcome equation, modeling the whether an individual i carried drugs for all individuals. If $y_{1i}^* > 0$, then the police would have found some drugs if they searched, and $y_{1i} = 1$ if the search occurred (i.e., if $y_{2i}^* > 0$). Otherwise, when $y_{1i}^* \leq 0$, the police would not have found any drugs, and therefore $y_{1i} = 0$ if it is observed. The second model, $y_{2i}^* = Z_i\beta + \epsilon_{2i}$, is the selection equation modeling the decision to search, and has the exact same structure. Thus, $y_{2i}^* > 0$ implies that the police searched car i, and $y_{2i}^* \leq 0$ implies that no search occurred. The correlation between the two error terms, ρ defines the extent of selection; a larger ρ (in absolute value) means greater selection and selection bias. In practice, for a Bayesian model, I re-parametrize the error terms and their correlation as follows: $\epsilon_{1i} = \delta\epsilon_{2i} + \eta_i$ where $\eta_i \sim N(0, 1)$. Operationally, δ is easier to estimate directly, and ρ can easily be recovered by the formula: $\rho = \frac{\delta}{\sqrt{1+\delta^2}}$.

In a standard Bayesian selection model, β λ, and δ are parameters that require prior distributions to estimate; y^*_{1i}, y^*_{2i} are latent variables that are treated as parameters to be estimated because they are unobserved; and y_{1i}, y_{2i}, X_i, Z_i are known data used to estimate the model. The model is computed using Bayes Law (familiar to statisticians and Evidence scholars both):

$$Pr(\beta, \delta, \lambda, y^*_{1i}, y^*_{2i}|Y_1, Y_2, X_i, Z_i) = \frac{Pr(Y_1, Y_2|\beta, \delta, \lambda, X_i, Z_i) \times Pr(\beta, \delta, \lambda, y^*_{1i}, y^*_{2i})}{Pr(Y_1, Y_2|X_i, Z_i)}$$

$$Pr(\beta, \delta, \lambda, y^*_{1i}, y^*_{2i}|Y_1, Y_2, X_i, Z_i) \propto Pr(Y_1, Y_2|\beta, \delta, \lambda, Z_i, X_i) \times Pr(\beta, \delta, \lambda, y^*_{1i}, y^*_{2i})$$

With appropriate prior guesses that are vague and uninformative, the data drives the resulting final estimate I assume a prior distribution for each of the three sets of parameters, β, λ, δ, to be normally distributed, and independent of each other. Specifically,

$$\begin{aligned} Pr(\beta) &\sim N(\beta_0, \Sigma_\beta), \text{ where } \Sigma_\beta = \sigma^2_\beta I \\ Pr(\lambda) &\sim N(\lambda_0, \Sigma_\lambda), \text{ where } \Sigma_\lambda = \sigma^2_\lambda I \\ Pr(\delta) &\sim N(\delta_0, \sigma^2_\delta) \end{aligned}$$

The hyperparameters $\beta_0, \lambda_0, \delta_0$ are fixed to 0 a priori in the model, but could theoretically be another parameter to estimate, allowing for more flexibility. There are also priors on the variances of ($\sigma^2_\beta, \sigma^2_\lambda, \sigma^2_\delta$, respectively). As is standard in Bayesian regression, these variances are assumed to be distributed inverse gamma, which allows for all values from 0 to ∞. Specifically,

$$\begin{aligned} Pr(\sigma^2_\beta) &\sim IG(A_\beta, S^2_\beta) \\ Pr(\sigma^2_\lambda) &\sim IG(A_\lambda, S^2_\lambda) \\ Pr(\sigma^2_\delta) &\sim IG(A_\delta, S^2_\delta) \end{aligned}$$

where $A_\beta, A_\lambda, A_\delta, S^2_\beta, S^2_\lambda, S^2_\delta$ are all fixed a priori to create minimally informative priors. Together Equations **??** and **??** define the prior of the model. The likelihood is described completely by Equations **??**, **??** and the fact that both ϵ_{2i} and η_i are distributed as standard normals.

4.1 An Initial Comparison

Recent work by Strnad has suggested that legal empirical scholarship incorporate more Bayesian techniques. (Strnad 2007). An initial comparison of the Bayesian approach to the more traditional Heckman model suggests that simply using a Bayesian selection model may provide more accurate results. As with all but the simplest Bayesian models, there is no closed-form solution to the model, and therefore, the estimation is iterative. Using markov-chain monte carlo methods; the model is straightforward to estimate. [4]

Table 1 provides results comparing the Bayesian selection model to the Heckman selection model and a standard probit model that does not control for selection. This initial comparison suggests that a Bayesian approach may be practical, as well as theoretically more attractive. In particular, because the models are estimated with simulated data, one can compare the true values of the parameters to their estimates in the model. For all parameters, the Bayesian selection model performs significantly better; none of the 95%

[4]Carlin & Louis provide an overview of how to estimate this and other Bayesian models. (Carlin & Louis 1996). Each complete iteration creates one estimate of all of the parameters, essentially moving the parameter estimates around the possible support of the parameters (where they are likely to be); the model is run until there are sufficient iterations that the researcher is confident that the sample is representative of the true probability density. There are a series of test statistics that help one make this decision, but there is no definitive answer of when stop the model running. Theoretically, it could run for infinite time.

Table 1: Comparison of Bayesian and Heckman Selection Models

Parameter			Bayesian Model	Heckman Model	No Selection Model
		True Value	Mean $(95\%CS$	Mean $95\%CI$	Mean $95\%CI$
Outcome Equation	Black	3.0	3.03 (2.68, 3.37)	1.98 (1.95, 2.01)	1.72 (1.69, 1.75)
	Hispanic	-2.0	-2.01 (-2.10, -1.92)	-1.23 (-1.41, -1.05)	-1.64 (-1.82, -1.45)
	Year	-0.5	-0.49 (-0.53, -0.45)	-0.33 (-0.34, -0.32)	-0.32 (-0.33, -0.31)
	Constant	-0.5	-0.49 (-0.49,-0.48)	-0.32 (-0.37, -0.29)	0.11 (0.09, 0.14)
Selection Equation	Black	2.0	2.15 (2.05, 2.28)	2.01 (1.97, 2.04)	
	Hispanic	2.5	2.63 (2.55, 2.78)	2.53 (2.41, 2.64)	
	Speed	-0.25	-0.22 (-0.28, -0.19)	-0.25 (-0.26, -0.25)	
	Constant	-0.5	-0.53 (-0.68, -0.34)	-0.50) (-0.51,-0.49)	
Both	Rho	0.75	0.72 (0.69, 0.77)	0.73 (0.70, 0.76)	

confidence intervals for the Heckman selection model include the true value, for example. In addition, the naive model that does not control for selection bias at all (despite a high selection bias parameter of 0.75) is not often not worse than the Heckman model at estimating the true value of the parameters. In contrast, the Bayesian model provides 95% credible sets (akin to confidence intervals) that contain the true value for all of the outcome equation parameters, and the estimates are significantly closer, on average, to the true values. This suggests that the Bayesian model has some significant advantages over the standard Heckman model.

4.2 Incorporating Aggregate-Level Data

The model I describe above is simply a Bayesian version of a standard Heckman selection model; the Bayesian model also requires individual-level data. Assume for the moment that one did not have individual-level data on the initial selection decision. This means that the Z_i would be missing data in the model. Importantly, y_{2i} is not missing data; for every individual, we know whether the individual is in the selected sample or not. What we do not know are the exogenous characteristics of the non-selected individuals for the selection regression equation. This is the key theoretically benefit of using a Bayesian approach. Missing data is straightforward to model in a Bayesian framework; it is simply another parameter that must be estimated. Thus, the new model would simply treat Z_i as a parameter in the model, putting on the parameter side of the line, instead of the data side of the line:

$$\begin{aligned} Pr(\beta, \delta, \lambda, y_{1i}^*, y_{2i}^*, Z_i | Y_1, Y_2, X_i) \quad &\propto \quad Pr(Y_1, Y_2 | \beta, \delta, \lambda, Z_i, X_i) \\ &\times Pr(\beta, \delta, \lambda, y_{1i}^*, y_{2i}^*, Z_i) \end{aligned}$$

The question simply becomes how to create a prior on Z_i that incorporates ones prior knowledge about the individuals driving down the highway; after that, it is technically complicated, but conceptually straightforward to include a step updating the Z_i in the model estimation. The approach is similar to Greenland (2003), which uses a Bayesian model that treats unmeasured confounding variables as missing data. (Greenland 2003; Steenland and Greenland 2004). The key to this model is to estimate the Z_i accurately, despite the lack of individual-level data. Thus, I must use aggregate data means, standard deviations and correlations to recreate the possible Z_i. If the entire moment structure were known in other words, if the researcher knew all of the means, correlations, and higher moments among and between different characteristics of the underlying population of motorists then simulating Z_i would be straightforward: one could create the dataset that contains all of the correct marginal values. This would mean require the researcher to know, for example, the percentage of cars driven by men, the percentage of cars driven by Blacks, and the interaction between race and sex in the composition of drivers, as well as the higher moments for these variables. Unfortunately, some of these data are also hard to find. Greenland (2003) suggests using prior elicitation or other direct external evidence to create an appropriate prior on the Z_i.

The strategy for this project, however, deviates slightly from the standard Bayesian rubric, and uses simulations of potential Z_i variables, averaging across simulations to obtain estimates of the posterior distributions of the parameters. To simulate the Z_i variables, I use aggregate data on drivers, including the percentage of old cars on the road, distribution of the race and sex of drivers, and when cars are driven; I also include the correlations between these variables to accurately simulate the Z_i. These data are culled from highway usage reports as well as census data from the area.

5 Data

I estimate the model using data from the Maryland State Police (MSP). The data are in two separate files: data on stops which records every stop on the 49.5-mile stretch of I95 north of the City of Baltimore to the Delaware state line, and data on searches, which includes searches throughout the state. The stops data begin in May 1997, and record the location, date and time of the stop, the make of car, any traffic code violation, whether a ticket, warning or safety violation was issued, and the race and sex of the driver. In a subset of the cases, the data also record the posted speed limit and the speed of the vehicle. The search dataset contains additional information, including the stated reason for the search, and whether the search was justified by consent or probable cause. After combining the two datasets, there are 59,070 stops in the dataset from May 1,1999 to December 31, 2005. Narrowing the search dataset to the time period and geographic area covered by the stop dataset yields approximately 525 traffic stops for which search data is also available. Table 2 provides the details of the descriptive statistics for the combined dataset.

Table 2: Descriptive Statistics of MSP Dataset

Variable	Population	Population Percentage	Searched Population	Search Rate	CDS Found	Large Haul Found
Black	16377	27.7%	262	1.6%	40.5%	1.5%
White	38960	66.0%	241	0.6%	37.8%	1.2%
Other race	3733	6.3%	22	0.6%	13.6%	4.5%
Male	43525	74.7%	480	1.1%	39.2%	1.7%
Female	15545	26.3%	45	0.3%	26.7%	0%
Old car	12527	21.2%	186	1.5%	38.2%	1.1%
Night driver	8948	15.2%	79	0.9%	30.3%	1.3%
Instate plate	22746	38.5%	147	0.7%	42.9%	0.7%
NY plate	5503	9.3%	40	0.7%	42.5%	2.5%
Consent Search	–	–	180	34.3%	17.8%	2.2%
Totals	59070	100%	525	0.9%	38.1%	1.5%

In addition to these demographic variables, the dataset also contains the ID number of each police officer; I construct dummy variables for each police officer who conducted at least ten searches, and use these in the selection equation as psuedo-instrumental variables. As Table 2 demonstrates, Black drivers, male drivers, and drivers of old cars are targeted for searches more often. Of targeted cars, those with drivers who are not white or Black are least likely to be found with a controlled substance (CDS); however, these individuals were also most likely to be found with a large amount of drugs (defined as the equivalent of over 500 grams of CDS). Black drivers are more likely to be searched than average, and slightly more likely to be carrying drugs. Men make up the vast majority of the searched individuals (about 75%), and are more likely, according to these descriptive statistics, to be carrying drugs. But, of course, these descriptive statistics suffer from selection bias; one cannot rely on them to determine what the true probability of an individual carrying drugs (or carrying a large amount of drugs) is. The results of the Bayesian model, discussed below, remedy this bias.

6 Results

The results of the Bayesian selection model, using individual-level data, suggest that the police clearly have a profile that they rely upon, and are therefore "selecting" individuals to be search. The results do not suggest, however, that much selection bias occurs: selection bias only occurs when unobserved (rather than observed) selection is present, and there is little evidence of this in the data. In particular, the distribution of the selection parameter, ρ, significantly overlaps with 0; its mean value is $\rho = -0.20$ with a standard deviation of $\sigma = 0.13$ This means that the selection model is not absolutely required here; still, the model is what determined that no unobserved selection was present.

Table 3 provides the specific results with respect to whether or not any amount of contraband was found. The lower left quadrant of the table provides the results of the selection equation, that is, the observed selection criteria. Police officers are significantly more likely to search Black and white motorists (as compared with those of a different race); Black and white motorists are about equally likely to be searched. Women are also more likely to be searched, after controlling for other characteristics. Drivers from instate (Maryland, in this case) are much more likely to be searched, as compared to drivers from New York, who, in turn, are somewhat more likely to be searched than drivers from the other 48 states. [5] The

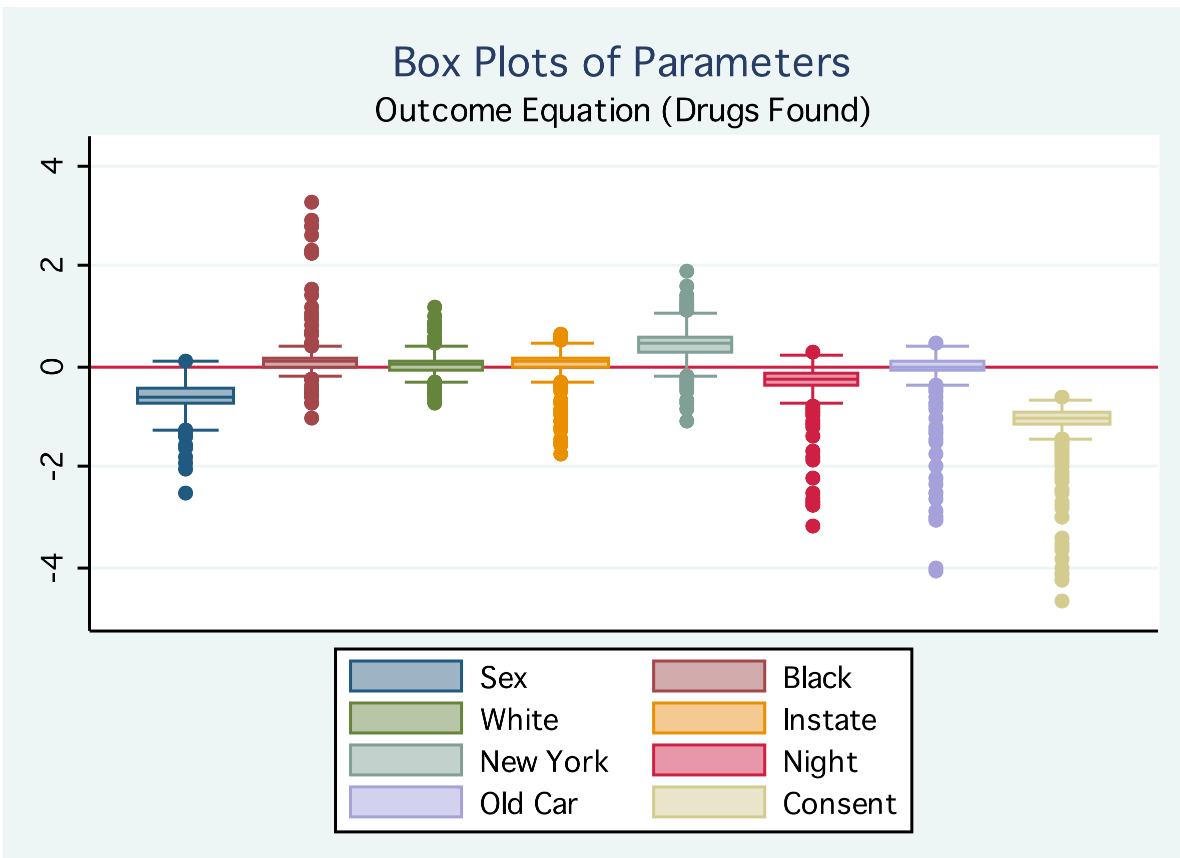
[5] New York drivers are thought to be part of a profile of drug couriers from New York City to Washington D.C.

Table 3: Results for Aggregate & Individual Selection Models: Drugs Found

Equation	Parameter	Individual Data Model Mean	St. Dev.	Aggregate Data Model Mean	St. Dev.
Outcome	Black	0.11	0.29	0.08	0.11
CDS Found	White	0.02	0.17	0.02	0.12
	Female	-0.62	0.28	-0.55	0.22
	Old car	-0.4	0.39	0.02	0.12
	Night driver	-0.22	0.17	-0.29	0.32
	Instate plate	0.10	0.13	0.08	0.21
	NY plate	1.14	0.39	0.45	0.27
	Consent search	-1.00	0.13	-1.11	0.40
Selection	Black	15.4	5.4		
Equation	White	16.6	3.4		
	Female	4.0	1.3		
	Old car	14.7	4.9		
	Night driver	5.1	1.4		
	Instate plate	12.0	1.8		
	NY plate	2.1	0.5		
	Cop Dummies			No Cop Dummies	
	ρ	-0.20	0.13	-0.06	0.08

top left quadrant of Table 3 provides results relevant to the probability that a random sample (having controlled for unobserved and observed selection) of drivers would be carrying drugs. Here, only a few variables are quite significantly different from zero. Specifically, the drivers who consented to a search were significantly less likely to be found with drugs, despite the apocryphal stories of criminal procedure professors. When one considers, however, that the alternative reason for a search was probable cause, this simply suggests that police officer "hunches" that lead to the request for consent (in the absence of probable cause) are not very useful. [6] The mean and standard deviations of the parameters masks

Figure 1: Bayesian Selection Model: Drugs Found



the nuance of the actual results; Bayesian researchers can directly ask the question how likely a parameter is to be zero (unlike classical statistician). Figure 1 provides box plots of the entire distribution of each β parameter. Only one, the parameter on consent, is never

[6]Note that only one person refused to give consent in the dataset.

equal to zero. Several others, however, are rarely close to zero; the probability that female drivers a good target for a profile is very low. In contrast, more than 75% of the time, New York drivers are a good target (being more likely than average to carry drugs). This is also true for Black drivers, although there is more variability, suggesting that it is a harder parameter to measure accurately. On average, drivers at night are also a bad target; they are less likely to carry drugs over 75% of the time.

Table 4 provides the results for the model that estimates the probability of finding a large amount of drugs. Here, again, there is very little evidence of unobserved selection, with $\rho = -0.08$. There is, again, significant observed selection; the bottom left quadrant of Table 4 provides the details, which are quite similar to the selection mechanism estimated for the previous model. [7] There is also a more obvious profile of the optimal target: it is not a Black or white driver, who are quite likely, in comparison to drivers of other races, to carry large amounts of drugs. While most of the means are quite close to zero, Figure 2 provides the complete distribution of the parameters. Most other parameters are also generally negative, suggesting that these characteristics should form a profile of an individual who should not be targeted for search.

6.1 Calibration of the Loss in Precision

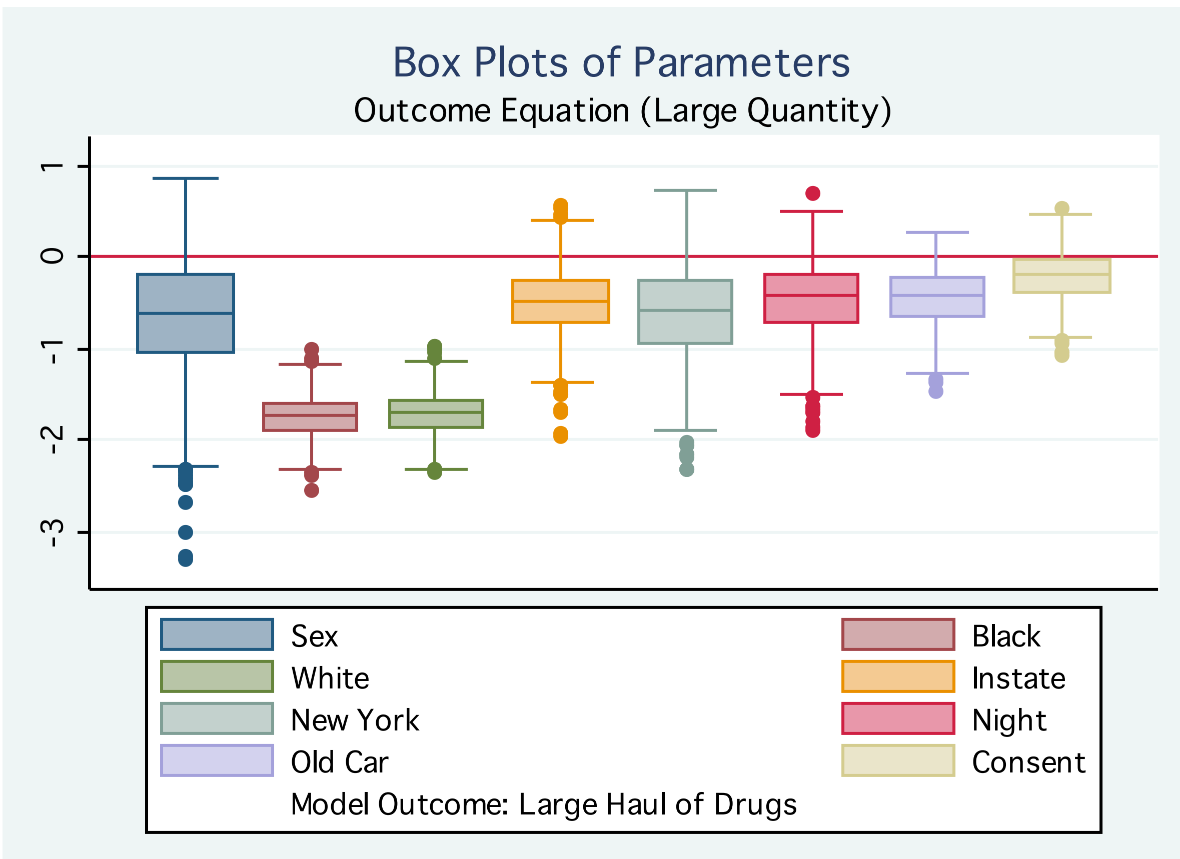
Here I compare the individual-level data bayesian model with the aggregate level data model. Tables 3 and 4 provide the results. The aggregate-level model does quite well in comparison using this data, providing very similar results. This may partially be due to the fact that there is not significant unobserved selection in these data; the aggregate model only has to control for observed selection, which is (statistically speaking) an easier task. However, with that caveat, the results are strikingly similar. In Table 3, almost all the values are within 10% of the individual-level data model, with the exception of New York plates (which moderated from 1.14 to 0.45), and having an old car (which flipped from -.4 to 0.02, both with large standard deviations). In essence, while there is a loss of precision,

[7] As a theoretical matter, there is no link between these two models which both model the decision to search a vehicle; however, as they are estimated using the same data, they should be quite similar.

Table 4: Bayesian Selection Model: Large Haul Found

Equation	Parameter	Individual Data Model Mean	St. Dev.	Aggregate Data Model Mean	St. Dev.
Outcome	Black	-1.74	0.22	-1.74	0.20
Large Haul	White	-1.71	0.24	-1.74	0.24
Found	Female	-0.66	0.66	-0.73	0.65
	Old car	-0.44	0.29	-0.32	0.27
	Night driver	-0.46	0.41	-0.32	0.37
	Instate plate	-0.50	0.35	-0.37	0.32
	NY plate	-0.62	0.50	-0.59	0.49
	Consent search	-0.20	0.26	-0.18	0.25
Selection	Black	13.5	2.9		
Equation	White	14.7	1.8		
	Female	3.2	0.6		
	Old car	13.4	2.3		
	Night driver	4.7	0.5		
	Instate plate	10.6	0.8		
	NY plate	1.8	0.3		
	Cop Dummies			No Cop Dummies	
	ρ	-0.08	0.13	-0.06	0.07

Figure 2: Bayesian Selection Model: Large Haul Found

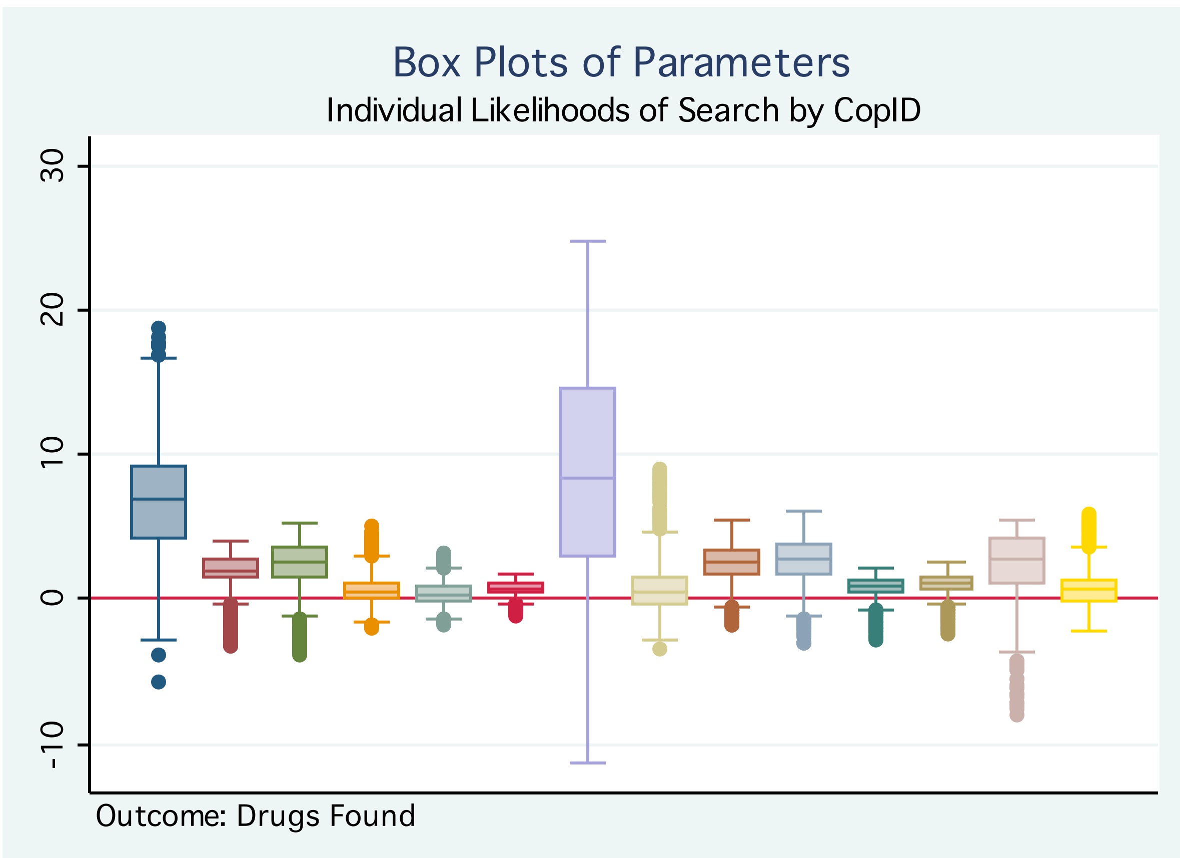


the aggregate model is able to control for the observed selection. (Selection parameters, while not reported, were generally about half the size, with similar standard deviations).

Table 4 provides a similar comparison for the estimation of the large haul outcome. Here, there is almost no significant differences in results; Black and white drivers still should not be targeted, and others should be targeted less often.

6.2 Application to Maryland State Police

Figure 3: Individual Police Officer Selection Parameters



One element of the estimation strategy that the model exploits is that individual police officers may have very different search rates (and search critieria). The individual-level model is capable of estimating the specific search probabilities of different individual police officers; indeed, the identification strategy is based upon the assumption that individual officers have different profiles, but once searching, they all have the same intensity of search (i.e., the error term for the outcome equation is uncorrelated with the cop dummies). The model also estimates how likely an officer is to search a vehicle. Figure 3 provides the distributions of the individual cop dummies; these are best viewed as relative likelihoods.

One can see that there are significant differences in selection thresholds for these individuals. Further work could include determining whether the individual officers have different profiles, rather than simply different thresholds for the decision to search.

7 Extensions

Note: These extensions have not yet been implemented, but hopefully they provide a sense of the flexibility of the statistical approach I am taking.

7.1 Modeling Additional Uncertainty

A second advantage of a Bayesian approach to this selection problem is that it is possible to directly model additional uncertainty. This uncertainty comes from two main sources: first, unmeasured confounding variables, and second, data errors. Consider a set of unmeasured confounding variables . Our model of interest expands from $p(\beta, \lambda, \delta | X_i, Z_i, y_{1i}, y_{2i})$ to a similar model indexed by nuisance parameter ν: $p(\beta, \lambda, \delta; \nu | X_i, Z_i, y_{1i}, y_{2i})$. The ν are not parameters of direct interest, but instead are nuisance parameters that serve only to complicate the analysis. Theoretically, we are interested in the marginal distribution $p(\beta, \lambda, \delta | X_i, Z_i, y_{1i}, y_{2i})$ which integrates out potential values of ν; in practice, we may only be able to estimate $p(\beta, \lambda, \delta; \nu | X_i, Z_i, y_{1i}, y_{2i})$, which estimates the parameters of interest for specific (unknown) values of ν. In a Bayesian setting, we incorporate a prior on ν in order to integrate out the set of nuisance parameters ν and obtain the marginal distribution of interest, $p(\beta, \lambda, \delta | X_i, Z_i, y_{1i}, y_{2i})$. In practice, modeling ν as a single dichotomous variable that incorporates all of the confounding unmeasured variables (akin in some ways to a propensity score) controls for a broad range of unmeasured confounding. (Greenland 2002).

The same approach applies for data errors. In many racial profiling cases, there is a strong suggestion that the data that police collect and report are incomplete and perhaps purposefully biased. More specifically, there is evidence that the practice of ghosting, or recording the race of the driver of a stopped or unsuccessfully searched car incorrectly as white instead of Black or Hispanic in order to make the data look less racist. (Gross and Barnes 2002). The ACLU of Maryland has documented this practice, and specifically sued regarding several individual Black motorists who have been stopped and searched, but who do not appear in the database of searches. (Gross and Barnes 2002). This was a major impetus for the collection of stop data as well.

The Bayesian model easily incorporates, and thereby controls for, ghosting as well, with a parameter that estimates the probability of ghosting. Let be the race-specific probability of ghosting. Incorporating this probability is a straightforward application of Bayesian law: effectively, it converts the dichotomous race variable (minority/white) in that is an element in and into a probability of being white, specific to each individual. Prior elicitation (through expert witness reports in the court cases in Maryland and Arizona) allows straightforward inclusion of this source of model uncertainty into the complete Bayesian model.

7.2 Modeling Simultaneous Choices

The results in Section **??** demonstrate that there are significant tradeoffs in deciding whom to search: picking the individual most likely to carry a large amount of drugs is more likely to be a completely unfruitful search than any other. However, one cannot simply look at the results of two separate models - the model of whether CDS was recovered, and the model of whether a large haul was recovered - to determine what the optimal strategy is. This is because the decision is simultaneous; the model needs to be as well. There are two reasons why the two separate models may not provide the same results as one joint model: first, the two outcomes - large haul and any CDS recovered - are correlated, and two separate models would assume that they are not; and second, the two separate models estimate two separate search decisions, even though there is, in reality, only one decision to search. The Bayesian approach, however, makes modeling several outcomes simultaneously straightforward: instead of estimating two correlated probit equations, one would model 3 (or more) equations: a single selection equation, and multiple outcome equations. Each outcome equation would have its own selection coefficient (allowing different levels of selection bias for different outcomes) and the outcomes would be correlated with each other.

7.3 Modeling Additional Levels of Selection

The model posited above controls for one aggregate selection mechanism. In the racial profiling context, however, there are at least two separable mechanisms: the decision to stop a vehicle, and the decision to search a stopped vehicle. Modeling these processes separately allows one to answer questions regarding each separate decision: what factors did the police use in deciding whether to stop a vehicle? Was the race of the driver salient? Separately, what factors did the police use in deciding whether to search a vehicle? Was the race of the driver salient? While not necessarily completely separate legal questions the use of race in either case triggers an equal protection analysis separating these two decisions separates the equal protection analysis as well. In turn, this separation allows one to focus on the specific actions of the police that may be problematic, and sharpens the normative debate over the appropriate use of race in governmental decision-making in the context of drug interdiction and immigration enforcement. Knowing exactly how race is used in decision-making allows for a rich discussion of why race is used, how it is beneficial in that context (or not), and whether racialized decision-making is justified in that context. Because the justifications for the use of race in decision-making, and racial profiling in particular, are so context specific, it is important to separate the two analytically distinct decisions that are made in both the empirical and the normative analysis. To model a second level of selection, one incorporates a third equation into the full model. Similar to modeling simultaneous outcomes, this structure models sequential selection decisions; the limit is (as always) the limit of the data. Again, one re-parametrizes the correlation structure across the three models for ease of estimation. This extension allows the researcher to determine which selection decision creates the most bias, and, from a legal perspective, which decision is most troublesome (for example, the evidence for the use of race is strongest).

8 Conclusion

This paper develops new methods to control for selection bias without relying on individual-level data of the non-selected population. The project tests this model against simulations and new data on racial profiling in the context of drug interdiction on the Maryland highway. The estimation finds that Black and white drivers are more likely to be targeted, despite lower probabilities of carrying large amounts of drugs (and about even probabilities of carrying some drugs). This suggests, as an initial matter, that the racial profiling engaged in by the Maryland State Police was not efficient, and that a more efficient profile would focus on other minorities than Black drivers and men. The consequences of this more efficient profile are, when looking only at direct costs, win-win: less innocent individuals would be targeted, and more drug dealers (as opposed to simple users) would be found. Further investigation would provide additional information about the possible tradeoffs to be made, as well as quantify how uncertainty or misspecification affects the results.

While focusing on selection bias in the racial profiling context, this project has wide applicably to many empirical questions, and provides a significant step in controlling for selection bias when individual-level data is unavailable.