THE UNIVERSITY OF CHICAGO

ESSAYS ON ECONOMETRICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
JOSHUA SHEA

CHICAGO, ILLINOIS
DECEMBER 2022

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I am deeply grateful to my advisors Alexander Torgovitsky, Stéphane Bonhomme, and Peter Hull, who have provided invaluable guidance and support and shown immeasurable patience. I would also like to thank Yuehao Bai, Jeffrey Grogger, Jack Mountjoy, Derek Neal, Guillaume Pouliot, Evan Rose, Azeem Shaikh, Max Tabord-Meehan, Hung Ho, Christine Blandhol, John Bonney, Jiaying Gu, Lee Lockwood, Francesca Molinari, Alexandre Poirier, participants at the Becker Applied Economics Workshop, and members of the Econometrics Advising Group. A special thank you to Laura Sale, Francisco del Villar, Dan Kashner, Nadav Kunievsky, and the Metrics Student Group for the regular and thoughtful discussions of my research.

Finally, I am immensely grateful for the continual love and support of my parents, Esther and Steve, and my brothers, Graham, Kyle, and John. I do not know what I would accomplish without them.

ABSTRACT

This thesis contains three essays. The first essay develops methodology for detecting and measuring racial bias in police traffic searches. The second essay develops methodology for using instrumental variable strategies to extrapolate beyond complier subpopulations, and is coauthored with Alexander Torgovitsky. The third essay provides large sample distribution theory for support vector regression with ℓ_1-norm regularization, and is coauthored with Yuehao Bai, Hung Ho, and Guillaume Pouliot.

CHAPTER 1

DETECTING RACIAL BIAS IN POLICE TRAFFIC SEARCHES

1.1 Introduction

Since the police killings of Eric Garner, Michael Brown, and Tamir Rice in 2014, the potentially fatal cost of encountering police for Black Americans has become a central theme in public and political discourse. The cost of these events is not only the welfare and lives of Blacks, but also the credibility and authority of the police (Manski & Nagin, 2017; Owens, 2020).[1] Over the last two decades, confidence in the police among African Americans has fallen from 37% to 19%, with 84% believing they are treated unfairly by the police (Horowitz, Brown, & Cox, 2019; Jones, 2021).[2] This decline in confidence has led to growing demand for police accountability, a lack of which has historically been the norm.[3] In the cases of Eric Garner, Michael Brown, and Tamir Rice, none of the officers responsible for the deaths were indicted by a grand jury, an outcome that sparked protests across the country against racial bias in policing.

But holding officers accountable for their actions by establishing misconduct is difficult for several reasons.[4] In particular, we do not know the thoughts of an officer, making it difficult to determine if he has abused his discretion. This is exemplified by a comment from a juror involved in the case against officer Jeronimo Yanez, who was acquitted for shooting and killing Philando Castile in front of his girlfriend and her four-year-old daughter during a traffic stop in 2016:[5]

[1] Trinkner, Kerrison, and Goff (2019) find when officers are perceived as being racist, they feel their authority is lessened. This correlates with their condoning of greater use of force to establish control.

[2] Confidence in the police among whites has fallen from 63% to 56% in the last two decades, and 63% of whites believe Blacks are treated unfairly by the police (Horowitz et al., 2019; Jones, 2021).

[3] See Morin and Stepler (2016) and https://www.bloomberg.com/news/articles/2020-06-09/a-history-of-protests-against-police-brutality

[4] For example, officers refuse to testify against their colleagues ('Blue Wall of Silence'), and prosecutors typically refuse to go after officers in fear of jeopardizing their working relationship with the police department. See https://fivethirtyeight.com/features/why-its-still-so-rare-for-police-officers-to-face-legal-consequences-for-misconduct/ and https://www.vox.com/21497089/derek-chauvin-george-floyd-trial-police-prosecutions-black-lives-matter.

[5] https://www.mprnews.org/story/2017/06/23/74-seconds-yanez-juror

> "It just came down to us not being able to see what was going on in the car. Some of us were saying that there was some recklessness there, but that didn't stick because we didn't know what escalated the situation: was [Yanez] really seeing a gun? We felt [Yanez] was an honest guy…and in the end, we had to go on his word, and that's what it came down to."

In this paper, I develop a test for racial bias that recognizes the disparity in the information available to the officer versus the researcher. Specifically, I use a partial identification approach to test whether individual officers have different preferences for searching white and minority drivers who are stopped. These preferences govern an economic choice model for searching drivers that depends on the officer's beliefs about how likely a driver carries contraband (e.g., drugs or weapons), which are unobserved by the researcher. Using this approach, I am able to make sharp inferences on how the officer's decisions depend on his beliefs. Restrictions on the officer's preferences and the probability that drivers carry contraband may be layered in a flexible and transparent manner. The test does not require officers to be randomly assigned to drivers and may be performed for each officer separately.

The test for bias checks whether the sharp identified set for the officer's search preferences (i.e., the smallest set of preferences consistent with the model and data) includes an equivalent pair of preferences for white and minority drivers. If not, then the officer's preferences must differ by race, implying he is biased. The partial identification approach permits the test to be valid even when officers have dissimilar beliefs for white and minority drivers, which can occur for several reasons. One possibility is statistical discrimination, where the quality of the signals used by officers to form their beliefs differ for white and minority drivers (Aigner & Cain, 1977). Another possibility is sample selection, which occurs because officers choose to stop different types of drivers depending on their race (e.g., bias in traffic stops). A third possibility is that white and minority drivers in population are simply different. Implementing the test entails solving a bilinear program, a type of non-convex problem that can be solved to global optimality. Bilinear programs are not only novel in the context of discrimination, but also in the context of partial identification.

A distinguishing feature of my test is how I model an officer's search decision. Similar to earlier papers, the officer is modeled to search drivers only if their probability of carrying contraband ('risk') exceeds a threshold, where the threshold represents the officer's search preference. However, whereas recent papers have required or assumed fixed thresholds (see Canay, Mogstad, and Mountjoy (2020) and Hull (2021) for a discussion on this restriction), I use a random threshold. Consequently, there is no longer a single driver at the margin of search, but a 'marginal driver' at every level of risk. This means a biased officer is not restricted to searching all drivers of one race with a given level of risk, while searching none of the equally risky drivers of the other race, as implied by a fixed threshold. Instead, the officer can search both groups of drivers at different intensities, e.g., whites with 10% risk are searched 10% of the time, whereas equally risky minorities are searched 20% of the time. Officers can even change direction of bias depending on the level of risk, e.g., whites with 10% risk are half as likely to be searched compared to equally risky minorities, but whites with 20% risk are twice as likely to be searched compared to equally risky minorities. Using a bilinear program, it is feasible to estimate sharp bounds on these differences conditional on the risk of drivers. The random threshold therefore permits a richer and more refined analysis of racial bias, where the direction and intensity of bias may depend on unobserved (to the researcher) characteristics of the driver.

Identification stems from an instrumental variable (IV) that shifts the distribution of risk among drivers stopped without shifting the officer's preferences. This variation allows the researcher to 'trace' out the preferences of the officer, similar to how an IV is used to overcome simultaneity when estimating demand curves. More intuitively, the identification argument is simply that an officer's preferences may be learned by observing him make search decisions for many types of drivers. Since it is possible to vary the risk of drivers stopped for each officer separately, it is possible to apply the proposed methods on each officer separately.

I apply the test using data collected by the Metropolitan Nashville Police Department

(MNPD) between 2010 and 2019. As with all police traffic data, the data are limited to drivers whom officers have chosen to stop. Due to the scarcity of traffic searches, I restrict my attention to the 50 officers with the most number of searches, each of whom have made at least 270 stops and 26 searches for each race of drivers.[6] I find that 8 (14) officers fail the test at the 5% (10%) significance level. Among these officers, I find that at least 28% of searches of minority drivers are attributable to racial bias. Moreover, if these biased officers treat white drivers as minority drivers, then their search rate for white drivers would increase by 92% on average. The estimates also suggest that the intensity of bias varies with the risk of the driver.

The paper proceeds as follows. Section 1.2 reviews the literature on testing for racial bias; Section 2.2 presents the model of an individual officer's search decision; Section 1.4 formalizes how bias may be detected and measured; Section 1.5 presents the data; Section 1.6 discusses the estimation procedure; Section 1.7 presents the estimates; and Section 1.8 concludes.

1.2 Literature review

It is well documented that Black civilians are disproportionately impacted by policing across the US. Compared to their white counterparts, Black pedestrians are stopped up to six times as often (Gelman, Fagan, & Kiss, 2007); Black motorists who are stopped are searched twice as often (Pierson et al., 2020); and Black civilians are almost seven times more likely to be killed by an officer.[7] Whether these disparities are the outcome of racial bias is a difficult question because officers are given a lot of discretion in making their decisions, but the researcher does not know what the officer is thinking at the time of the decision. In this

[6] The original data contains over 2,200 officers. These 50 officers are responsible for a third of all searches in the data.

[7] Source: Fatal Force, Washington Post. Since 2015, the Washington Post has recorded every fatal shooting by on-duty officers in the US. The collection of such data was a response to their investigation in the previous year revealing that the FBI undercounted fatal police shootings since police departments are not required to report such incidents and only do so voluntarily. Similarly, Collaborators et al. (2021) found over half of all deaths in the US due to police violence between 1980 and 2018 were unreported in the National Vital Statistics System.

section, I summarize earlier approaches to answering this question.

The seminal paper by Knowles, Persico, and Todd (2001) lays the foundation for detecting racial bias in traffic searches through its use of the outcome test proposed by Becker (1957, 1993). In Knowles et al. (2001), officers are homogeneous and only search drivers whose perceived risk of carrying contraband exceeds a fixed threshold. The authors assume officers have accurate beliefs, meaning that the risk they perceive is equal to the true risk. If the thresholds differ for white and minority drivers, then officers are racially biased, so the researcher's objective is to recover these thresholds. If risk is observed by the researcher and continuously distributed over the unit interval, then the thresholds are identified from the risk of the white and minority drivers at the margin of search.

However, because risk is not actually observed, the researcher must use a different strategy. Developing an econometric solution was considered to be difficult at the time and has only recently been accomplished (Arnold, Dobbie, & Yang, 2018; Ayres, 2002). So Knowles et al. (2001) instead form a game-theoretic argument that drivers of the same race have the same risk in equilibrium, placing every driver at the margin of search.[8] In addition, the authors show that if officers are unbiased, then all white and minority drivers carry contraband with equal probability. This results in a straightforward test for bias: if officers have different success ('hit') rates when searching white and minority drivers, then officers are biased. However, the model's assumption of homogeneous officers, as well as its implication of homogeneous drivers (within race), may both be rejected using officer-level data. For instance, Ba, Knox, Mummolo, and Rivera (2021) find that the rate at which officers stop, arrest, and use force against civilians varies with the race and sex of the officer.[9] Also, the variation across MNPD officers in the success rate of a search reveals that drivers are not homogeneous in risk.

[8]The argument is that drivers who are more likely to carry contraband will be searched more frequently. These drivers are therefore discouraged from carrying contraband. So in equilibrium, all drivers carry contraband with equal probability and officers search at random.

[9]From surveys conducted on officers, Morin, Parker, Stepler, and Mercer (2017) find that men are three times more likely than women to have discharged their service weapon while on duty (30% versus 11%). White officers are also 80% more likely than Black officers to have been in an altercation with a civilian within a month prior to the interview (36% versus 20%).

Anwar and Fang (2006) propose an alternative test that allows for heterogeneity in officer decisions and driver risk.[10] Using a similar model for officers as Knowles et al. (2001) (i.e., officers only search drivers whose risk exceed a fixed threshold), but allowing different officers to have different thresholds, Anwar and Fang (2006) test for bias using pairwise comparisons of search decisions across groups of officers (e.g., white officers versus Black officers). If both groups of officers are unbiased, then the ranking of their search rates should be the same regardless of the race of the driver. While this approach can detect bias, it cannot determine which group is biased, nor which group of drivers is being discriminated against.

Recently, Arnold et al. (2018) made an important contribution to the literature by using random assignment of decision makers as an instrument to detect bias. However, their method is not directly applicable to the setting of this paper since officers select who to stop and are therefore non-randomly assigned to drivers. Nevertheless, Arnold et al. (2018) provide an empirical strategy to recover the thresholds of decision makers without trivializing the distribution of risk for each race (as in Knowles et al., 2001). The authors use the marginal treatment effect framework of J. J. Heckman and Vytlacil (2005a) to compare the decisions of a continuum of decision makers and show that this point-identifies the thresholds of all decision makers. This method is referred to as the marginal outcome test, and is valid as long as all decision makers have common distributions of risk (hence the importance of random assignment) and fixed thresholds (Canay et al., 2020). To see whether this approach extends to police traffic searches, Gelbach (2021) tests three implications of the model on police traffic data from Florida and Texas. These implications are not satisfied in the data and the author points to different distributions of risk across officers as the potential reason. Papers using the marginal outcome test to study bias in policing therefore require restrictions on the distributions of risk. For example, Marx (2021) requires the distributions to either be common across officers, or that the distribution for one officer second-order stochastically dominates that of another. Feigenberg and Miller (2021) do not restrict how the distributions of risk

[10] Antonovics and Knight (2009) independently developed the same test as Anwar and Fang (2006).

differ across officers, but assume that they are independent of the officer's threshold and rule out sample selection on unobservables.[11] While Arnold, Dobbie, and Hull (2020) extend the method of Arnold et al. (2018) to allow decision makers to face different distributions of risk, it comes at the expense of parametric assumptions on the joint distribution of thresholds and risk.[12]

Other papers have used statistical approaches to test whether civilian race has an effect on police decisions, including stop-and-frisk and use of force (Fryer Jr, 2019; Gaebler et al., 2020; Gelman et al., 2007; Goel, Rao, & Shroff, 2016a, 2016b; Grogger & Ridgeway, 2006; Knox, Lowe, & Mummolo, 2020; MacDonald & Fagan, 2019; Ridgeway, 2006; Ridgeway & MacDonald, 2009). These papers either assume that the distribution of risk may be balanced across races, or cannot attribute the effect of race to racial bias. Knox et al. (2020) is noteworthy for emphasizing how difficult it is to identify the effect of race on post-stop decisions alone (e.g., use of force, traffic searches) because of sample selection. The authors show that, under a principal strata framework, this is only possible in the knife-edge scenario where the biases from sample selection and omitted variables cancel each other out. So the effect of race on post-stop decisions considered in their paper includes the effect of race on stop decisions as well. The reason I am able to measure bias in the search decision alone in spite of sample selection is because I impose an economic choice model.

An area in the literature that has received increasing attention is inaccurate beliefs (see Bohren, Haggag, Imas, & Pope, 2020; Bohren, Imas, & Rosenberg, 2019; Bordalo, Coffman, Gennaioli, & Shleifer, 2016). In my setting, this corresponds to an officer incorrectly assessing the driver's risk, which can generate patterns of searches that resemble bias even when the officer is unbiased. Hence, the concern is that tests for bias conflate inaccurate beliefs with racial bias. Given the difficulty of distinguishing between the two without knowing the

[11] The difference-in-differences strategy used by Goncalves and Mello (2021) to study racial bias among officers writing speeding tickets also rules out sample selection on unobservables.

[12] See also Simoiu, Corbett-Davies, and Goel (2017), Pierson, Corbett-Davies, and Goel (2018), Pierson et al. (2020), and Chan, Gentzkow, and Yu (2019), who impose similar parametric restrictions to identify thresholds of decision makers.

decision maker's beliefs, researchers have begun using experiments to elicit beliefs of decision makers when studying discrimination (Bohren et al., 2020). In the case of observational data, only Hull (2021) provides sufficient conditions to reject the hypothesis of accurate beliefs when using the marginal outcome test, although the framework in this paper permits a similar test for inaccurate beliefs.

1.3 Model

In this section, I model the search decision of a single officer (he) for drivers who are stopped (she). The analysis allows for heterogeneity across officers, but I suppress the officer indexes for brevity. I also suppress the notation indicating the analysis is conditional on drivers who are stopped. This conditioning is important to keep in mind, though, as it affects the interpretation of the model's assumptions.

1.3.1 Setup and notation

For each stop i, the officer observes the driver's race $R_i \in \{w, m\}$ (white or minority), and a set of non-race characteristics $V_i \in \mathcal{V}$ that may include the driver's demeanor, the direction of travel, and any other details the officer notices. Some of the components in V_i may be observed by the officer prior to the stop; some components may also be observable to one officer but not another so that officers vary in their perceptiveness and form different beliefs about the driver's risk. The researcher only observes R_i but not V_i; any other observed characteristics are implicitly conditioned on throughout.

The driver may carry contraband (e.g., drugs, weapons), denoted by $Guilty_i \in \{0, 1\}$, but the officer does not know this unless he searches the driver, denoted by $Search_i \in \{0, 1\}$. At the end of each traffic stop, the officer reports whether or not contraband was found, referred to as a 'hit',

$$Hit_i \equiv Search_i \times Guilty_i,$$

which is observed by the researcher. I therefore assume that the officer finds contraband if and only if he searches a guilty driver.

Finally, I assume drivers are drawn from a distribution that depends on the setting of the stop, $Z_i \in \mathcal{Z}$. For example, Z_i may consist of the hour and day of the stop, and the interpretation is that different types of drivers are stopped at different times. This may be because the composition of drivers on the road changes with time, or because the officer's stop decision changes with time.[13] The setting is observed by both the officer and researcher, and will play the role of an instrument.

When deciding whether to search, the officer considers the four possible outcomes of his decision: (i) searching an innocent driver; (ii) searching a guilty driver; (iii) not searching an innocent driver; and (iv) not searching a guilty driver. Associated with each outcome is an *ex post* utility that the officer learns after interacting with the driver and observing all of her characteristics, but prior to making his search decision. Let $\mathcal{U}_i^s(g;r)$ denote this utility when $Search_i = s$ and $Guilty_i = g$ for drivers with race $R_i = r$. Note that these utilities are random and can vary across drivers who are observationally equivalent to the officer. The distributions of these utilities represent the officer's search preferences, and the objective of the test is to detect whether race has a direct effect on these distributions. To do this, I make the following assumption about the utilities $\{\mathcal{U}_i^0(g;r), \mathcal{U}_i^1(g;r)\}_{(g,r)\in\{0,1\}\times\{w,m\}}$, which I denote by $\{\mathcal{U}_i^s\}$ for brevity.

Assumption 1.

[13] If there are variables that inform the officer's stop decision and are visible for one value of Z_i but not another (e.g., race is visible during the day before stopping a driver, but is not visible at night), then the distribution of drivers stopped will vary with Z_i even if the composition of drivers on the road do not. This type of variation is used in the Veil of Darkness test by Grogger and Ridgeway (2006) to test whether race affects the stop decision.

(i) $\mathcal{U}_i^1(1; R_i) - \mathcal{U}_i^1(0; R_i) > 0$ *and* $\mathcal{U}_i^0(1; R_i) - \mathcal{U}_i^0(0; R_i) < 0$.

(ii) $\{\mathcal{U}_i^s\}$ *are identically distributed across stops* i.

(iii) $\{\mathcal{U}_i^s\} \perp\!\!\!\perp (Z_i, Guilty_i, V_i)$.

Assumption 1(i) states that the officer prefers to make the correct decision by searching guilty drivers and not searching innocent drivers. This implies that officers are more likely to search drivers who are more likely to carry contraband. This would be violated if the officer instead prefers to make the wrong decision by searching innocent drivers and releasing guilty drivers.

Assumption 1(ii) states that the utilities are drawn from a common distribution, which allows me to pool the drivers of the same race together to infer the officer's preferences. Conditioning the analysis on variables that affect the distribution of utilities (e.g., age and sex of the driver) helps to satisfy this assumption. If instead $\{\mathcal{U}_i^s\}$ and $\{\mathcal{U}_{i'}^s\}$ were drawn from different distributions for every $i \neq i'$, then the researcher must infer the search preference faced by each driver using one observation only.

Assumption 1(iii) is the key assumption of the model and determines how racial bias is defined and how it may be detected, so I spend more time discussing its parts. The independence between the utilities $\{\mathcal{U}_i^s\}$ and the setting Z_i is the exogeneity assumption for the instrument, whose role is to shift the distribution of risk (by changing the drivers stopped) without shifting the officer's preferences. This variation is what allows me to learn about the officer's preference. Moreover, this instrument enables the test to be applied to each officer separately since the variation in risk may be generated within officer. This instrument separates my approach from those using random assignment of decision makers as the instrument (Arnold et al., 2020, 2018), which instead aim to shift the officer's preferences without shifting the distribution of risk. Such an instrument is harder to justify in the setting of police searches since officers choose their distribution of risk by choosing who to stop. Randomly assigning officers to roads does not resolve this problem, as it does not guarantee that officers stop the same drivers.

The independence between the utilities and whether the driver is guilty means that the officer can only infer how likely a driver carries contraband using the characteristics of the driver and setting of the stop, but not his utilities. This rules out clairvoyance, where the officer infers the driver's guilt using information beyond what is provided by the driver.

Finally, the independence between the utilities and the unobserved characteristics of the driver allows the researcher to make a direct link between officer preferences and race. That is, any dependence between the officer's preferences and the driver is through the race of the driver. This part of Assumption 1(iii) is crucial to any test of racial bias and has generated discussion among researchers. I elaborate on this point after defining the officer's search decision rule.

Under Assumption 1, any dependence between the officer's preferences and the driver's race can only be through race, leading to the following definition of racial bias.

Definition 1. *The officer is* racially biased in traffic searches *if* $(\mathcal{U}_i^0(0;w), \mathcal{U}_i^0(1;w), \mathcal{U}_i^1(0;w), \mathcal{U}_i^1(1;w))$ *and* $(\mathcal{U}_i^0(0;m), \mathcal{U}_i^0(1;m), \mathcal{U}_i^1(0;m), \mathcal{U}_i^1(1;m))$ *do not share the same distribution.*

The objective of the test is thus to determine whether the distribution of utilities depends on race.

1.3.2 Search decision

To map the preferences of the officer to the data, I assume the officer chooses the search decision that maximizes his utility. Since the driver's guilt is not known to him, he chooses the decision that maximizes his expected utility,

$$\begin{aligned}
&\mathbb{E}[\mathcal{U}_i^s(\mathit{Guilty}_i; R_i) \mid R_i = r, Z_i = z, V_i = v] \\
&= G(r,z,v)\, \mathcal{U}_i^s(1; R_i) \ + (1 - G(r,z,v))\, \mathcal{U}_i^s(0; R_i),
\end{aligned}$$

where

$$G(r,z,v) \equiv \mathbb{P}\{\mathit{Guilty}_i = 1 \mid R_i = r, Z_i = z, V_i = v\}$$

is the officer's belief of how likely the driver caries contraband, which I refer to as the 'risk' of the driver. His search decision may then be written as

$$\begin{aligned} S_i &\equiv \operatorname*{arg\,max}_{s\in\{0,1\}} \mathbb{E}[\mathcal{U}_i^s(\mathit{Guilty}_i; R_i) \mid R_i, Z_i, V_i] \\ &= \mathbb{1}\left\{G(R_i, Z_i, V_i) \geq T_i\right\}, \end{aligned} \tag{1.1}$$

where

$$T_i \equiv \frac{\mathcal{U}_i^0(0;R_i) - \mathcal{U}_i^1(0;R_i)}{\left[\mathcal{U}_i^1(1;R_i) - \mathcal{U}_i^1(0;R_i)\right] - \left[\mathcal{U}_i^0(1;R_i) - \mathcal{U}_i^0(0;R_i)\right]}$$

is a random utility threshold representing the officer's preferences. See Appendix A.1 for the full derivation. The officer thus searches a driver if and only if her risk is sufficiently large, and how large that risk must be may vary across stops. The researcher observes neither $G(R_i, Z_i, V_i)$ nor T_i.

From its definition, T_i inherits the properties of $\{\mathcal{U}_i^s\}$ stated in Assumption 1 and may be used to detect racial bias.

Corollary 1.

(i) $T_i \mid R_i = r$ *is identically distributed across stops* i *for* $r \in \{w, m\}$.

(ii) $T_i \perp\!\!\!\perp (Z_i, \mathit{Guilty}_i, V_i) \mid R_i = r$ *for* $r \in \{w, m\}$.

(iii) *The officer is racially biased in traffic searches if* $T_i \not\perp\!\!\!\perp R_i$.

Proof. See Appendix A.1. ■

So instead of comparing distribution of $\{\mathcal{U}_i^s\}$ across races to detect bias, it suffices to compare T_i across races.[14]

This approach of comparing thresholds to detect bias is standard in the literature. Many papers model the choice of the decision maker as in (1.1), except their thresholds are deterministic functions of race (Anwar & Fang, 2006; Arnold et al., 2018; Knowles et al., 2001), which corresponds to the special case where $T_i \mid R_i = r$ is degenerate for $r \in \{w, m\}$.[15] But in all the papers, the threshold is independent of the unobserved characteristics of the driver after conditioning on race. This independence property is crucial to any test for racial bias, as it allows the researcher to make a direct link between an officer's preferences and a driver's race. Without it, any dependence between the utilities and race may be argued to be a result of confounding variables in V_i (Canay et al., 2020).

To provide some justification for this independence assumption, I refer to the MNPD manual, which states that "individuals shall only be subjected to stops, seizures, or detentions upon reasonable suspicions" based on "specific and articulable facts, which taken together with rational inferences from those facts, reasonably warrant an officer to believe that criminal activity is afoot."[16] So conditional on the risk of the driver, the officer should be impartial to drivers with different characteristics. However, the researcher may expect this mandate to be violated by officers letting their decisions be influenced by their preferences toward certain characteristics. For example, an officer may feel differently about searching a young male driver versus an elderly female driver. But for condition (ii) in Corollary 1 to hold, an officer may *only* be partial to the race of the driver. The analysis must therefore be conditional on all non-race characteristics of the driver that the officer is partial to and may be correlated

[14] Types of biases that generate identical thresholds will be impossible to detect. For instance, let the constant $\overline{u}$ denote the bias, and suppose $\mathcal{U}_i^s(g; w) = \mathcal{U}_i^s(g; m) + \overline{u}$ for $(s, g) \in \{0, 1\}^2$. Then $\overline{u}$ drops out of T_i, and the officer has the same decision rule for both groups of drivers. I ignore these cases since the bias neither affects the search decisions nor the impact on drivers.

[15] A threshold that is a deterministic function of race can be obtained by assuming $\{\mathcal{U}_i^s\}$ are degenerate random variables. Another way is to begin with decision rule (1.1) and then assume that the threshold is a deterministic function of race. Hull (2021) provides conditions where such a decision rule is a normalization of a more general decision rule ranking drivers in the order they will be searched.

[16] Section 4.40 of Metropolitan Nashville Police Department Manual (2018).

with race.

The unobserved characteristics V_i then only affect the search decision through the risk of the driver. This in turn implies that omitted variables, sample selection, and statistical discrimination—the usual confounders of bias—only affect the search decision through $G(R_i, Z_i, V_i)$. The econometric challenge of detecting bias is thus to separately infer the distributions of T_i and $G(R_i, Z_i, V_i)$; I defer this discussion to Section 1.4.

To elaborate on how the three confounders affect the distribution of risk, consider an example where V_i is the color of the car, and red cars are more likely to contain contraband. Omitted variable bias pertains to differences in the distribution of V_i across races in population, e.g., whites are twice as likely to drive red cars than minorities in population. So even if the officer stopped drivers at random, the distribution of risk may differ across race since the underlying determinant V_i differs across race. Sample selection pertains to differences in the distribution of V_i across races for drivers who are stopped, e.g., the officer may prefer to stop minority drivers in red cars, so conditional on being stopped, whites are only half as likely to be in red cars than minorities, despite how whites are twice as likely to drive red cars in population. Finally, statistical discrimination (in the sense of Aigner and Cain (1977)) pertains to how V_i maps to risk differently for white and minority drivers, e.g., red cars are correlated with possessing contraband for whites, but not for minorities. This notion of statistical discrimination also extends to other officers, where different officers observe different components of V_i (Arnold et al., 2020; Hull, 2021). For example, an experienced officer may know to consider the direction of travel along a highway when assessing the driver's risk (Barnes, 2004), whereas an inexperienced officer may not. Since the test may be applied to each officer separately, I place no restrictions on how different officers infer the risk of the driver.

The instrument Z_i also affects the search decision exclusively through the risk of the driver. But unlike the confounders of bias, the variation in risk generated by Z_i is helpful in partially identifying the distribution of T_i. The intuition for this is that, by seeing how an officer makes his search decisions in a variety of settings, I can build a profile for the types of white and minority drivers he likes to search and then compare these profiles. In Section 1.4, I show it is possible to detect bias without Z_i by only using the variation in search decisions generated by R_i. However, such a test will be weak.

Notice that, conditional on race, Z_i may shift $G(R_i, Z_i, V_i)$ in two ways. The first is through shifting the distribution of V_i, e.g., $G(R_i, Z_i, V_i) = G(R_i, V_i)$ but $Z_i \not\perp\!\!\!\perp V_i \mid R_i$. An example of this is if the time of the traffic stop contains no information on whether the driver is guilty, but criminals and drug dealers tend to drive at night. The second way Z_i may shift risk is to have a direct effect on $G(R_i, Z_i, V_i)$, i.e., $G(R_i, z_1, V_i) \neq G(R_i, z_2, V_i)$ for $z_1 \neq z_2$. This reflects how the same signals can be interpreted differently depending on the setting of the stop (Engel & Johnson, 2006; Novak & Chamlin, 2012). For example, stopping a white driver in a predominantly white suburb may not arouse much suspicion, whereas stopping the same driver in a predominantly Black neighborhood may lead to more questions. Similarly, stopping a high school student in the afternoon shortly after school has ended is less suspicious than stopping the same student late into the night.

Under Assumption 1 and decision rule (1.1), the probability that a driver is searched only depends on the race and the risk of the driver,

$$
\begin{aligned}
&\mathbb{P}\{Search_i = 1 \mid R_i = r, Z_i = z, V_i = v\} \\
&= \mathbb{P}\{G(R_i, Z_i, V_i) \geq T_i \mid R_i = r, Z_i = z, V_i = v\} \\
&= \mathbb{P}\{G(r, z, v) \geq T_i \mid R_i = r, Z_i = z, V_i = v\} \\
&= \mathbb{P}\{G(r, z, v) \geq T_i \mid R_i = r\} \\
&= F_{T|R}(G(r, z, v) \mid r),
\end{aligned}
$$

where the third equality follows from Assumption 1(iii), and F_X denotes the CDF of random variable X. For each level of risk $g \in [0,1]$, the officer's bias may be measured by

$$\beta(g) \equiv F_{T|R}(g \mid m) - F_{T|R}(g \mid w),$$

with $\beta(g) > 0$ indicating bias against minority drivers with risk g. Since $\beta(g)$ can vary with g and even change sign, the intensity and direction of bias can vary with the unobserved (to the researcher) risk of the driver. This feature of the model arises from the random threshold and distinguishes it from earlier models, under which the officer searches all drivers with a given level of risk or none at all (Anwar & Fang, 2006; Arnold et al., 2018; Hull, 2021; Knowles et al., 2001). A random threshold thus extends the notion of the marginal driver to every level of risk and permits a more nuanced analysis of bias.[17] I show in Section 1.4 how sharp bounds on $\beta(\cdot)$ may be derived.

There are two concerns with all tests of racial bias, including the one I present. The first is that the model restricts *Guilty* to be binary, when in reality officers consider factors beyond whether the driver is or is not carrying contraband. For example, an officer may be more willing to search a driver if he suspects there is 0.1 probability of finding a kilogram of heroin versus a 0.1 probability of finding a few grams of marijuana, even though both drivers are guilty of possessing contraband.[18] These additional dimensions of contraband (i.e., type and quantity) can be thought of as components of V_i, and if these components impact an officer's willingness to search a driver, then Assumption 1(iii) is violated. So if white and minority drivers carry different types and quantities of contraband, then racial differences in T_i may be a result of the racial differences in the nature of contraband carried rather than the race of the driver. If the researcher has a prior on how these omitted payoffs correlate with driver race and officer preferences, then it is possible to sign the (statistical) bias of the test. The model can also be generalized to accommodate continuous outcomes, such as when the officer considers only the quantity of contraband and not the type of contraband when making his

[17] But if the researcher wishes to maintain a fixed threshold, the methods I propose can also accommodate this.

[18] Marijuana is not authorized for recreational or medical use in Tennessee.

search decision.

The second concern pertains to the accuracy of the decision maker's beliefs and whether it is possible to distinguish between inaccurate beliefs and racial bias. To illustrate the problem, suppose an unbiased officer incorrectly believes minority drivers are twice as risky as they truly are. His search decision may then be written as

$$\begin{aligned} S_i &= \mathbb{1}\left\{(1 + \mathbb{1}\{R_i = m\})\; G(R_i, Z_i, V_i) \geq T_i\right\} \\ &= \mathbb{1}\left\{G(R_i, Z_i, V_i) \geq \widetilde{T}_i\right\} \end{aligned}$$

where $\widetilde{T}_i \equiv T_i/(1 + \mathbb{1}\{R_i = m\})$. So in this example, the effect of inaccurate beliefs is observationally equivalent to the officer drawing thresholds that are half as large for minorities compared to whites. The test may then incorrectly detect bias since $\widetilde{T}_i \not\perp R_i$, despite how T_i is the true object of interest. Nevertheless, these tests for bias are still valuable since the effects of inaccurate beliefs and bias are the same for drivers. The test may serve as a preliminary check to determine which officers ought to be reviewed, and further investigation may reveal whether differences in search behavior stem from bias or inaccurate beliefs. In the next section, I also show how certain cases of inaccurate beliefs can be detected using the proposed methods.

1.4 Testing for racial bias

In line with Becker's (1957, 1993) outcome test, the test I propose checks whether an officer's search decisions are consistent with him being unbiased. If they are not, then the officer is biased. To avoid conflating bias with omitted variable bias, sample selection, and statistical discrimination, I use a partial identification approach to make inferences on the officer's preferences separately from the distribution of risk.

1.4.1 Defining the test

For each traffic stop, I observe the driver's race R_i; the setting of the stop Z_i; the search decision Search_i; and whether contraband is found, Hit_i. From this, I am able to construct the officer's search and hit rates for each race $r \in \{w, m\}$ and setting $z \in \mathcal{Z}$,

$$\mathbb{P}\{\mathit{Search}_i = 1 \mid R_i = r, Z_i = z\} = \int_{\mathcal{V}} F_{T|R}(G(r, z, v) \mid r)\ dF_{V|R,Z}(v \mid r, z), \tag{1.2}$$

$$\mathbb{P}\{\mathit{Hit}_i = 1 \mid R_i = r, Z_i = z\} = \int_{\mathcal{V}} G(r, z, v)\ F_{T|R}(G(r, z, v) \mid r)\ dF_{V|R,Z}(v \mid r, z). \tag{1.3}$$

See Appendix A.1 for the derivations. The instrument Z_i varies the search and hit rates by varying the distributions of risk. From the ratio of these rates, I also obtain the conditional hit rate, which is the probability that contraband is found conditional on a traffic search,

$$\mathbb{P}\{\mathit{Hit}_i = 1 \mid \mathit{Search}_i = 1, R_i = r, Z_i = z\} = \frac{\mathbb{P}\{\mathit{Hit}_i = 1 \mid R_i = r, Z_i = z\}}{\mathbb{P}\{\mathit{Search}_i = 1 \mid R_i = r, Z_i = z\}}.$$

To define the identified set of the model, let $\mathcal{F}$ denote the space of distributions of $(R_i, Z_i, V_i, T_i, \mathit{Guilty}_i)$ satisfying Assumption 1. The sharp identified set is then

$$\{F \in \mathcal{F} : (1.2) \text{ and } (1.3) \text{ are satisfied for all } (r, z) \in \{w, m\} \times \mathcal{Z}\}.$$

However, in testing for racial bias, the parameters of interest are only $F_{T|R}(\cdot \mid w)$ and $F_{T|R}(\cdot \mid m)$. So I instead consider a projection of the identified set when testing for bias. To define this projection, I introduce the following notation,

$$G_i \equiv G(R_i, Z_i, V_i),$$

$$\sigma(\cdot; r) \equiv F_{T|R}(\cdot \mid r),$$

where G_i denotes the risk in stop i; $\sigma(\cdot; r)$ denotes the probability a driver with of race r is searched conditional on her risk, and represents the officer's search preference for race r. The parameters of interest are then $\sigma(\cdot;w)$ and $\sigma(\cdot;m)$, and the distribution of risk conditional on race and setting is

$$F_{G|R,Z}(g \mid r, z) \equiv \int_{\mathcal{V}} \mathbb{1}\{G(r,z,v) \leq g\} \; dF_{V|R,Z}(v \mid r, z).$$

Equations (1.2)–(1.3) may then be written as

$$\mathbb{P}\{Search_i = 1 \mid R_i = r, Z_i = z\} = \int_0^1 \sigma(g;r) \; dF_{G|R,Z}(g \mid r, z), \tag{1.4}$$

$$\mathbb{P}\{Hit_i = 1 \mid R_i = r, Z_i = z\} = \int_0^1 g \; \sigma(g;r) \; dF_{G|R,Z}(g \mid r, z). \tag{1.5}$$

Let Σ denote the space of non-decreasing, right-continuous functions with domain and codomain equal to $[0,1]$; and let $\mathcal{F}_G$ denote the space of distributions for scalar random variables with support $[0,1]$. Then the sharp identified set for the parameters of interest is

$$\Sigma^\star \equiv \left\{ (\sigma(\cdot;w), \sigma(\cdot;m)) \in \Sigma^2 : \begin{array}{c} \exists F_{G|R,Z}(\cdot \mid r, z) \in \mathcal{F}_G \text{ s.t. (1.4) and (1.5) are} \\ \text{satisfied } \forall (r,z) \in \{w,m\} \times \mathcal{Z} \end{array} \right\}. \tag{1.6}$$

A test for racial bias immediately follows from (1.6).

Corollary 2. *Under* (1.1) *and Assumption 1, if there does not exist* $\sigma^\star \in \Sigma$ *such that* $(\sigma^\star, \sigma^\star) \in \Sigma^\star$*, then the officer is biased.*

Proof. By the definition of an unbiased officer (see Definition 1 and property (iii) of Corollary 1), the contrapositive of the corollary is true. ■

Since $\Sigma^\star$ is sharp, Corollary 2 is the strongest testable implication of the model for unbiasedness. As discussed below, having variation in the search and hit rates across Z_i strengthens the test by reducing the size of $\Sigma^\star$.

1.4.2 Building intuition for the test

When $|\text{supp}\,(G)\,| = 2$

In this section, I show how the variation in search and hit rates can reveal whether an officer is biased. But instead of inferring whether $\sigma(\cdot; r)$ is the same for $r \in \{w, m\}$, it is easier and equivalent to infer whether

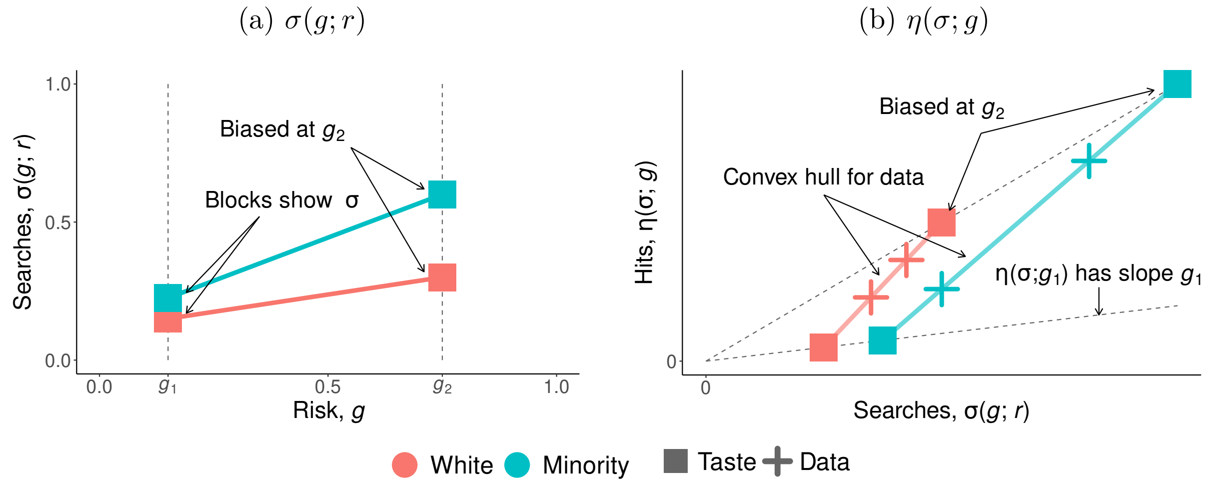
$$\eta(\sigma(\cdot; r); g) \equiv g\ \sigma(g; r)$$

depends on r, where η maps the probability a driver is searched to the probability of a hit, conditional on the risk of the driver. More intuitively, η maps the volume of traffic searches to the volume of hits for drivers with a certain level of risk, and is akin to a constant-returns-to-scale production function, e.g., searching drivers with risk g twice as often will double the amount of contraband recovered from them. The advantage of comparing η across races instead of σ is that η has a convenient geometric relationship with the data that lessens the computational burden of the test. I show in Appendix A.1 that $\sigma(\cdot; r)$ depends on r if and only if $\eta(\sigma(\cdot, r); \cdot)$ depends on r.

To provide a simple illustration of how Z_i can reveal the officer's preferences, suppose that drivers stopped are either low- or high-risk so that $\text{supp}\,(G) = \{g_1, g_2\}$ with $g_1 < g_2$. The left panel of Figure 1.1 displays $\sigma(\cdot; w)$ and $\sigma(\cdot; m)$ for an officer, and the blocks indicate how often the officer searches a driver conditional on her race and risk. Since the configurations of the blocks differ for white and minority drivers, it follows that $\sigma(\cdot; w) \neq \sigma(\cdot; m)$ and the officer is biased.

The right panel displays $\eta(\sigma(\cdot; w); \cdot)$ and $\eta(\sigma(\cdot; m); \cdot)$. The dashed lines depict $\eta(\sigma(\cdot; w); g)$ for $g \in \{g_1, g_2\}$, and show how much contraband an officer will recover if he exclusively searches drivers with a given level of risk. Both of the dashed lines intersect the origin because no contraband will be found if no searches occur; and both lines are linear since every additional search is equally likely to uncover contraband, with the slopes being equal

Figure 1.1: Testing for bias when $|\operatorname{supp}(G)| = 2$



to the risk of the driver. How far the blocks lie along the dashed lines indicate how often the low- and high-risk drivers are searched. If the positions of the blocks differ by race in the left panel, then they will also differ in the right panel, which is why bias may be detected by comparing $\eta(\sigma(\cdot; r); \cdot)$ across races.

To see how Z_i is informative of preferences, consider the data points

$$\mathcal{D}(r) \equiv \{(\mathbb{P}\{Search_i = 1 \mid R_i = r, Z_i = z\}, \mathbb{P}\{Hit_i = 1 \mid R_i = r, Z_i = z\})\}_{z \in \mathcal{Z}}$$

for $r \in \{w, m\}$, depicted by the crosses in the right panel of Figure 1.1. As implied by (1.4)–(1.5), for each race, the data points are convex combinations of the blocks and must therefore lie inside the convex hull of the blocks. Since there are only two levels of risk/blocks for each race, the convex hulls are simply the colored lines in the right panel. The higher up a data point lies along the line, the greater the proportion of high-risk drivers stopped at that setting.

If the officer is unbiased, then the position of the blocks in both panels of Figure 1.1 should be the same across races. This means that $\mathcal{D}(w)$ and $\mathcal{D}(m)$ should lie inside the same convex hull/along the same line. To show this, let

$$p_{r,z}(g) \equiv \mathbb{P}\{G_i = g \mid R_i = r, Z_i = z\}$$

denote the distribution of risk conditional on the race of the driver and setting of the stop. Then the search and hit rates for an unbiased officer are

$$\begin{aligned}
\mathbb{P}\{Search_i \mid R_i = r, Z_i = z\} &= \sigma^\star(g_1)\ p_{r,z}(g_1) + \sigma^\star(g_2)\ (1 - p_{r,z}(g_1)), \\
\mathbb{P}\{Hit_i \mid R_i = r, Z_i = z\} &= g_1\ \sigma^\star(g_1)\ p_{r,z}(g_1) + g_2\ \sigma^\star(g_2)\ (1 - p_{r,z}(g_1)),
\end{aligned}$$

for some race-neutral $\sigma^\star \in \Sigma$. The linear relationship that the search and hit rates have with $p_{r,z}(g_1)$ indicate that the data lie on a line, and that variation in the data only stems from differences in the composition of drivers. Since the data in Figure 1.1 do not lie on a line, the officer is revealed to be biased.

In this special case where $|\text{supp}\,(G)\,| = 2$, the officer's preferences are summarized by the colored lines containing the data and the IV enables us to 'trace' out these preferences, similar to how an IV may be used to trace out a demand curve. This result can also be viewed as a control function approach, since the instrument provides a way to condition on the unobserved risk of the drivers. Regardless, when $|\text{supp}\,(G)\,| = 2$, a simple IV regression is able to detect bias.

Proposition 1. *Suppose* $|\text{supp}\,(G)\,| = 2$ *and* $Var[Search_i \mid R_i = r, Z_i] > 0$ *for* $r \in \{w, m\}$. *Then*

$$\mathbb{E}[Hit_i \mid R_i = r, Z_i = z] = \alpha_0(r) + \alpha_1(r)\ \mathbb{E}[Search_i \mid R_i = r, Z_i = z],$$

where $\alpha_0(r) \leq 0$ *and* $\alpha_1(r) > 0$ *for* $r \in \{w, m\}$. *The coefficients* $\alpha_0(r)$, $\alpha_1(r)$ *are identified by an IV regression of* Hit_i *on* $Search_i$, *using* Z_i *as an instrument.*

(i) *If* $\alpha_0(w) \neq \alpha_0(m)$ *or* $\alpha_1(w) \neq \alpha_1(m)$, *then the officer is biased.*

(ii) *If* $\alpha_0(r_1) > \alpha_0(r_2)$ *and* $\alpha_1(r_1) > \alpha_1(r_2)$*, then the officer is biased against race* r_2*.*

Proof. See Appendix A.1. ■

Notice how condition (ii) in Proposition 1 is required to infer the direction of bias, since bias may now vary with unobserved risk. I show later what may be learned about the direction and intensity of bias at each level of risk.

If the researcher does not have an instrument, Proposition 1 still offers a way to test for bias.

Corollary 3. *Suppose* $|\text{supp}\,(G)\,| = 2$ *and* $Var[Search_i \mid R_i] > 0$ *for* $r \in \{w, m\}$*. Consider the regression of* $\mathbb{E}[Hit_i \mid R_i]$ *on* $\mathbb{E}[Search_i \mid R_i]$ *and an intercept so that*

$$\mathbb{E}[Hit_i \mid R_i = r] = \alpha_0 + \alpha_1 \mathbb{E}[Search_i \mid R_i = r].$$

If $\alpha_0 > 0$ *or* $\alpha_1 \leq 0$*, then the officer is biased.*

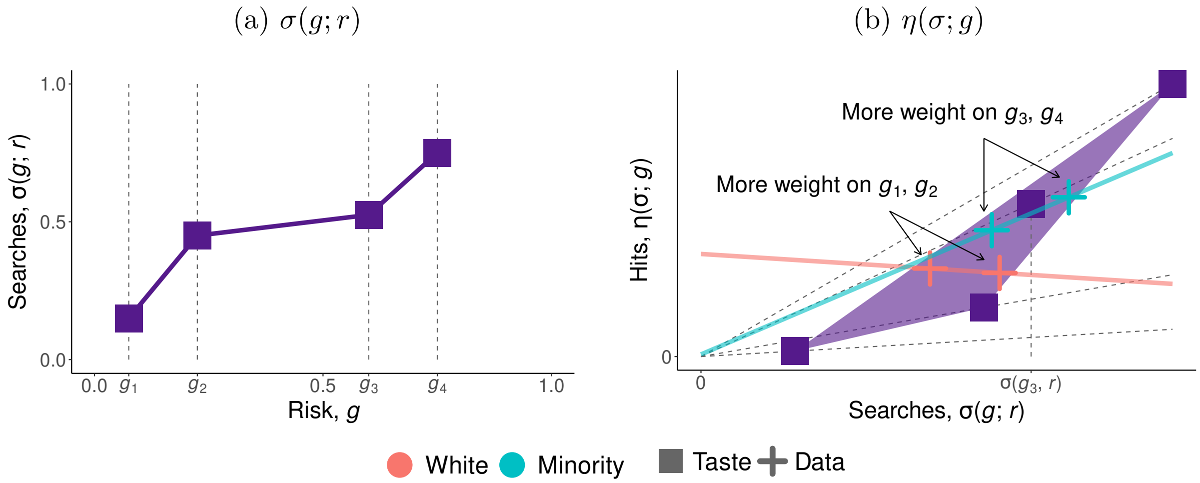
Proof. See Appendix A.1. ■

The intuition behind Corollary 3 is that an unbiased officer should search minorities more often than whites only if minorities have more high-risk drivers. If this is indeed the case, then the officer should experience more hits with minorities and there should be a positive relationship between the search and hit rates. If instead the relationship is weak or negative, then the higher search rates cannot be due to there being more high-risk drivers and must be due to bias.

When $|\text{supp}\,(G)\,| > 2$

Testing for bias when $|\text{supp}\,(G)\,| > 2$ is much more difficult. The IV result in Proposition 1 does not extend to this setting because the convex hulls containing the data need not be lines, so $\mathcal{D}(w)$ and $\mathcal{D}(m)$ need not lie on the same line for an unbiased officer. An example of this is provided in Figure 1.2, which displays the preferences and data for an unbiased

Figure 1.2: An unbiased officer when $|\mathrm{supp}\,(C)\,| > 2$



officer, where the purple triangle in the right panel represents the convex hull of the blocks containing the data.

To highlight another challenge, as well as illustrate how officer preferences may only be partially identified, notice that the convex hull in Figure 1.2 contains the block corresponding to $\sigma(g_3; r)$. While the vertices of the convex hull may be identified at infinity, the block inside of the convex hull cannot. This means that the officer's preference for searching drivers with $G_i = g_3$ may never be recovered, even with infinite values of Z_i. Bias among drivers with $G_i = g_3$ may therefore remain undetected.

To show how Corollary 2 may be used to detect bias in this general setting, define

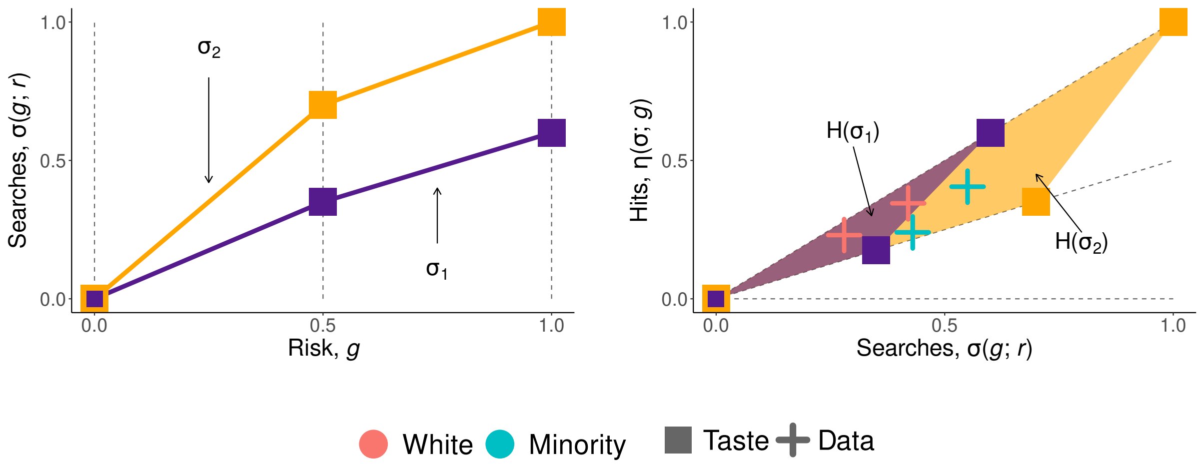
$$\mathcal{H}(\sigma(\cdot; r)) \equiv \mathrm{conv}\left(\{(\sigma(g; r), g\ \sigma(g; r))\}_{g\in \mathrm{supp}(G)}\right)$$

to be the convex hull generated by preference $\sigma(\cdot; r)$ that is supposed to contain the data $\mathcal{D}(r)$ for $r \in \{w, m\}$. In Figure 1.2, $\mathcal{H}(\sigma(\cdot; r))$ is depicted by the purple triangle. Corollary 2 asks whether there exists a race-neutral $\sigma^\star \in \Sigma$ such that

$$\mathcal{D}(w), \mathcal{D}(m) \subseteq \mathcal{H}(\sigma^\star). \tag{1.7}$$

Figure 1.3: Applying Corollary 2 when $|\text{supp}\,(C)| = \{0, 0.5, 1\}$

(a) $\sigma(g; r)$ (b) $\eta(\sigma; g)$



If so, the officer may be unbiased; and if not, the officer must be biased. Figure 1.3 provides an example of how this test may be performed by iteratively checking whether $\sigma \in \Sigma$ satisfy (3.11). In practice, iterating over the elements in Σ is an inefficient way to test for bias, but it is helpful at conveying the idea of the test; I show below how Corollary 2 may be implemented by solving a bilinear programming (BP) problem instead. Figure 1.3 shows that $\sigma_1 \in \Sigma$ (purple triangle) is not a suitable candidate for $\sigma^\star$ since $\mathcal{D}(m) \nsubseteq \mathcal{H}(\sigma_1)$. In contrast, $\sigma_2 \in \Sigma$ (orange triangle) is a possible candidate for $\sigma^\star$ since $\mathcal{D}(w), \mathcal{D}(m) \subseteq \mathcal{H}(\sigma_2)$. The officer in this example thus passes the test as it is feasible for him to be unbiased.

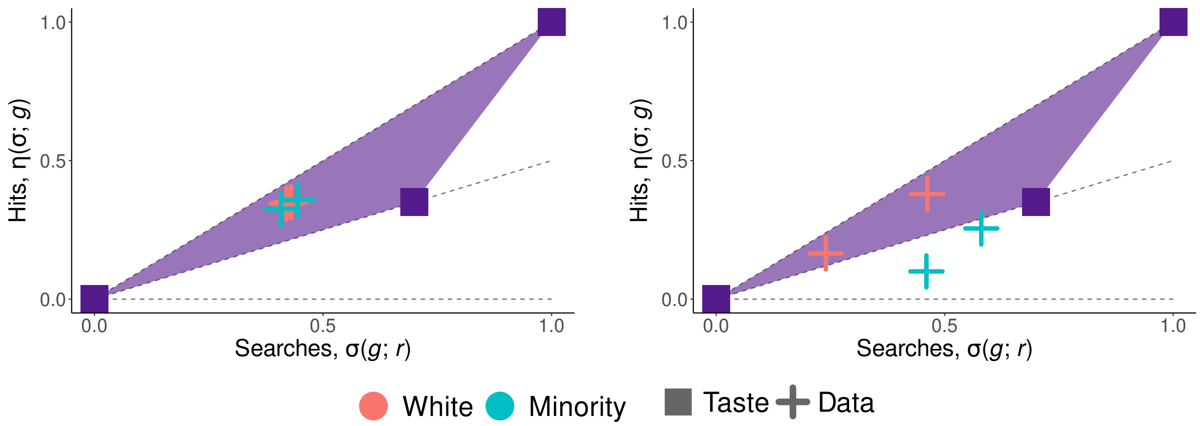
Figure 1.4 demonstrates how the strength of the test depends on the variation in $\mathcal{D}(w)$ and $\mathcal{D}(m)$. If there is little variation and the data are clustered around a point, then it is easy to find a $\sigma^\star \in \Sigma$ such that $\mathcal{D}(w), \mathcal{D}(m) \subseteq \mathcal{H}(\sigma^\star)$ and hard to detect bias. If instead there is a lot of variation in the data, then it becomes more difficult to find such a $\sigma^\star \in \Sigma$ and therefore easier to detect bias.

1.4.3 Applying Corollary 2 using a bilinear program

In this section, I show how Corollary 2 may be implemented as a bilinear program. To limit the computational cost of the program, I continue to discretize G_i so that $\text{supp}\,(G) = \{g_0, \ldots, g_K\}$

Figure 1.4: How variation in $\mathcal{D}(w)$, $\mathcal{D}(m)$ affects the strength of the test

(a) $\sigma \in \Sigma^\star$ (b) $\sigma \notin \Sigma^\star$



for $K < \infty$ and

$$\mathbb{E}[Search_i \mid R_i = r, Z_i = z] = \sum_{k=0}^{K} \sigma(g_k; r)\; p_{r,z}(g_k),$$
$$\mathbb{E}[Hit_i \mid R_i = r, Z_i = z] = \sum_{k=0}^{K} g_k\; \sigma(g_k; r)\; p_{r,z}(g_k).$$

However, under appropriate restrictions on the model, the BP problem is feasible even when G_i is continuous; I discuss this in Section 1.4.4.

To state the BP problem, I introduce the following notation.

$$\begin{aligned}
\mathbf{m}^S_{r,z} &\equiv \mathbb{P}\{Search_i = 1 \mid R_i = r, Z_i = z\} \\
\mathbf{m}^H_{r,z} &\equiv \mathbb{P}\{Hit_i = 1 \mid R_i = r, Z_i = z\} \\
\mathbf{g} &\equiv \{g_0, \ldots, g_K\} \\
\varsigma &\equiv (\sigma^\star(g_0), \ldots, \sigma^\star(g_K)) \\
\mathbf{p}_{r,z} &\equiv (p_{r,z}(g_0), \ldots, p_{r,z}(g_K))
\end{aligned}$$

The objects $\mathbf{m}^S_{r,z}$, $\mathbf{m}^H_{r,z}$ are the search and hit rates for each race r and setting z; and the vector $\mathbf{g}$ is the support of G_i. I assume these three objects are known to the researcher.

The unknown parameters of the BP problem are ς, which are the values of $\sigma^\star(\cdot)$ evaluated at each point of $\mathbf{g}$; and $\{\mathbf{p}_{r,z}\}_{(r,z)\in\{w,m\}\times\mathcal{Z}}$, which are the distributions of risk conditional on race and setting. For notational brevity, I refer to the distributions of risk by $\{\mathbf{p}_{r,z}\}$. The objective of the problem is to optimize over these parameters to match the conditional moments $\mathbf{m}^S_{r,z}$, $\mathbf{m}^H_{r,z}$.

To ensure that the parameters of the model are consistent with their definitions, I impose two baseline constraints. The first is that

$$0 \leq \varsigma_k \leq \varsigma_{k+1} \leq 1 \text{ for } k = 0, \ldots, K-1,$$

where ς_k denotes the k^{th} component of ς. This ensures $\sigma^\star \in \Sigma$, as required by Corollary 2. The second is that

$$\mathbf{p}_{r,z,k} \in [0,1] \text{ and } \sum_{k=0}^{K} \mathbf{p}_{r,z,k} = 1 \text{ for all } (r,z) \in \{w,m\} \times \mathcal{Z},$$

where $\mathbf{p}_{r,z,k}$ denotes the k^{th} component of $\mathbf{p}_{r,z}$. This ensures $\mathbf{p}_{r,z} \in \mathcal{F}_G$ for all $(r,z) \in \{w,m\} \times \mathcal{Z}$, as required by the definition of $\Sigma^\star$. These restrictions are linear and may be written as

$$\mathbf{A} \begin{bmatrix} \varsigma \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b},$$

where matrix $\mathbf{A}$ and vector $\mathbf{b}$ characterize the restrictions (see Appendix A.2 for more details). To simplify the discussion, I assume that $\operatorname{supp}(Z_i \mid R_i = w) = \operatorname{supp}(Z_i \mid R_i = m)$, but this assumption is not necessary.

Corollary 2 may then be implemented as follows.

Proposition 2. *Define the criterion $Q^\star$ as the solution to the following BP program,*

$$Q^\star \equiv \min_{\varsigma, \{\mathbf{p}_{r,z}\}} \sum_{r,z} \left| \varsigma' \mathbf{p}_{r,z} - \mathbf{m}^S_{r,z} \right| + \sum_{r,z} \left| (\mathbf{g} \odot \varsigma)' \mathbf{p}_{r,z} - \mathbf{m}^H_{r,z} \right| \tag{1.8}$$

$$\text{s.t.} \quad \mathbf{A} \begin{bmatrix} \varsigma \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b},$$

where $\odot$ denotes the Hadamard (element-wise) product. The officer is biased if $Q^\star > 0$.

Proof. If $Q^\star > 0$, then there exists an $(r,z) \in \{w,m\} \times \mathcal{Z}$ such that $|\varsigma' \mathbf{p}_{r,z} - \mathbf{m}^S_{r,z}| > 0$ or $|(\mathbf{g} \odot \varsigma)' \mathbf{p}_{r,z} - \mathbf{m}^H_{r,z}| > 0$. Then there does not exist a $\sigma^\star \in \Sigma$ such that (1.4)–(1.5) are satisfied for all $(r,z) \in \{w,m\} \times \mathcal{Z}$. Then by Corollary 2, the officer is biased. ∎

The criterion $Q^\star$ in Proposition 2 is the minimum ℓ_1-norm between the moments of the model and the moments of the data. In theory, other norms may be used, but I choose the ℓ_1-norm because it leaves (1.8) quadratic, which is easier to solve.[19]

Note that any of the constraints in (1.8) may be tested. To do this, define

$$\varsigma_r \equiv (\sigma(g_0; r), \cdots, \sigma(g_K; r))$$

to be the officer's preference for race r, and consider the BP problem of matching only the

[19] If I instead use the ℓ_2-norm, then (1.8) becomes quartic.

moments for race r when the constraints are defined by $(\mathbf{A}, \mathbf{b})$,

$$Q^{\star}_{C,r}(\mathbf{A}, \mathbf{b}) \equiv \min_{\varsigma_r, \{\mathbf{p}_{r,z}\}} \sum_z \left| \varsigma_r' \mathbf{p}_{r,z} - \mathbf{m}^S_{r,z} \right| + \sum_z \left| (\mathbf{g} \odot \varsigma_r)' \mathbf{p}_{r,z} - \mathbf{m}^H_{r,z} \right| \tag{1.9}$$

$$\text{s.t.} \quad \mathbf{A} \begin{bmatrix} \varsigma_r \\ \mathbf{p}_{r,1} \\ \vdots \\ \mathbf{p}_{r,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b}.$$

Suppose that the set of constraints $(\mathbf{A}_1, \mathbf{b}_1)$ is a strict subset of $(\mathbf{A}_2, \mathbf{b}_2)$, i.e., the rows of $\mathbf{A}_1$ are a strict subset of those of $\mathbf{A}_2$, and likewise for $\mathbf{b}_1$ and $\mathbf{b}_2$. Then if $Q^{\star}_{C,r}(\mathbf{A}_1, \mathbf{b}_1) = 0 < Q^{\star}_{C,r}(\mathbf{A}_2, \mathbf{b}_2)$, the additional constraints in $(\mathbf{A}_2, \mathbf{b}_2)$ may be rejected.

For example, suppose $(\mathbf{A}_1, \mathbf{b}_1)$ and $(\mathbf{A}_2, \mathbf{b}_2)$ contain all the restrictions on ς_r and $(\mathbf{p}_{r,z})$ as $(\mathbf{A}, \mathbf{b})$ in (1.8), except $(\mathbf{A}_1, \mathbf{b}_2)$ excludes the monotonicity restriction on ς. Then if $Q^{\star}_{C,r}(\mathbf{A}_1, \mathbf{b}_1) = 0 < Q^{\star}_{C,r}(\mathbf{A}_2, \mathbf{b}_2)$, it means the monotonicity restriction cannot be satisfied and may be rejected. That is, the data reveals the officer is searching certain drivers with greater probability compared to other drivers with lower risk. There are two interpretations of this result. The first is that Assumption 1 is violated and the model is misspecified. The second is that Assumption 1 holds but the officer has inaccurate beliefs, which is similar to the interpretation used by Hull (2021) to detect inaccurate beliefs using the marginal outcome test.

1.4.4 Adding restrictions

The framework allows the researcher to strengthen the test by adding restrictions to Σ and $\mathcal{F}_G$ in a transparent, modular fashion. All additional restrictions may also be easily tested, as just described.

For example, if the researcher believes $\sigma^{\star}(\cdot)$ is smooth, then it can be modeled as a Bernstein polynomial, which has several convenient properties. First, it is highly flexible.

Second, it is linear in its parameters, so the test remains as a BP problem. Third, restrictions on its range and derivatives take the form of linear constraints and are easy to impose.[20]

The framework also nests the earlier models in the literature where $Search_i = \mathbb{1}\{G_i \geq t(R_i)\}$ for some deterministic function t. These models effectively impose an integrality constraint on ς so that

$$\sigma^{\star}(g_k) \in \{0,1\} \text{ for } k = 0, \dots K. \tag{1.10}$$

The researcher may also impose restrictions on the distributions of risk, $\{\mathbf{p}_{r,z}\}$. For example, restrictions such as (ranking across settings)

$$\mathbb{E}[G_i \mid R_i = r, Z_i = z_1] \leq \cdots \leq \mathbb{E}[G_i \mid R_i = r, Z_i = z_{|\mathcal{Z}|}].$$

and (ranking across races)

$$\mathbb{E}[G_i \mid R_i = w, Z_i = z] \lesseqgtr \mathbb{E}[G_i \mid R_i = m, Z_i = z].$$

are linear constraints and easy to impose.[21] It is also possible to model $\{\mathbf{p}_{r,z}\}$ as Bernstein polynomials.[22] If both ς and $\{\mathbf{p}_{r,z}\}$ are modeled as Bernstein polynomials, then the BP

[20]The Bernstein basis of degree L is defined by

$$\mathbf{b}_l^L(g) \equiv \binom{L}{l}(1-g)^{L-l}g^l$$

for $l = 0, \dots L$ and $g \in [0,1]$. So $\sigma^{\star}$ can be modeled as

$$\sigma^{\star}(g) = \sum_{l=0}^{L} \theta_l \mathbf{b}_l^L(g)$$

for some $\theta \equiv (\theta_0, \dots, \theta_L)$. See Appendix A.2 for a summary of Bernstein polynomials and their properties. See Farouki (2012) for a more detailed summary.

[21]The inequality constraint $\mathbb{E}[G_i \mid R_i = r_1, Z_i = z_1] \leq \mathbb{E}[G_i \mid R_i = r_2, Z_i = z_2]$ may be written as

$$\mathbf{g}'\mathbf{p}_{r_1,z_1} \leq \mathbf{g}'\mathbf{p}_{r_2,z_2}.$$

[22]This is not an unreasonable assumption since the Beta density function, which is used to model random

problem is feasible even when G_i is continuous (see Appendix A.3 for details).[23] But when making restrictions on $\{\mathbf{p}_{r,z}\}$, the researcher must bear in mind that the restrictions pertain to the risk of drivers stopped. So these restrictions should be justified by evidence in the data, or reasonable priors on the distribution of risk in the population and how sample selection interacts with this distribution.

To provide a visual example of how these restrictions strengthen the test, consider restricting the mass of drivers to be decreasing in risk,

$$\mathbf{p}_{r,z,k} \geq \mathbf{p}_{r,z,k+1} \text{ for } k = 0, \ldots K-1 \text{ and } (r,z) \in \{w,m\} \times \mathcal{Z}. \tag{1.11}$$

This assumption is plausible as long as the mass of low-risk drivers in population is sufficiently large. See Appendix A.2 for a simulated example of this restriction holding true even when the officer is much more likely to stop high-risk drivers.[24] Figure 1.5 shows how this restriction shrinks $\mathcal{H}(\sigma_2)$ from the example in Figure 1.3. So while it was previously feasible for the officer to be unbiased and have search preference σ_2, this is no longer true after imposing (1.11). In fact, there do not exist any race-neutral $\sigma^\star \in \Sigma$ capable of generating the data for both races while satisfying (1.11). The test thus detects bias under this restriction.

1.4.5 Determining the direction and intensity of bias

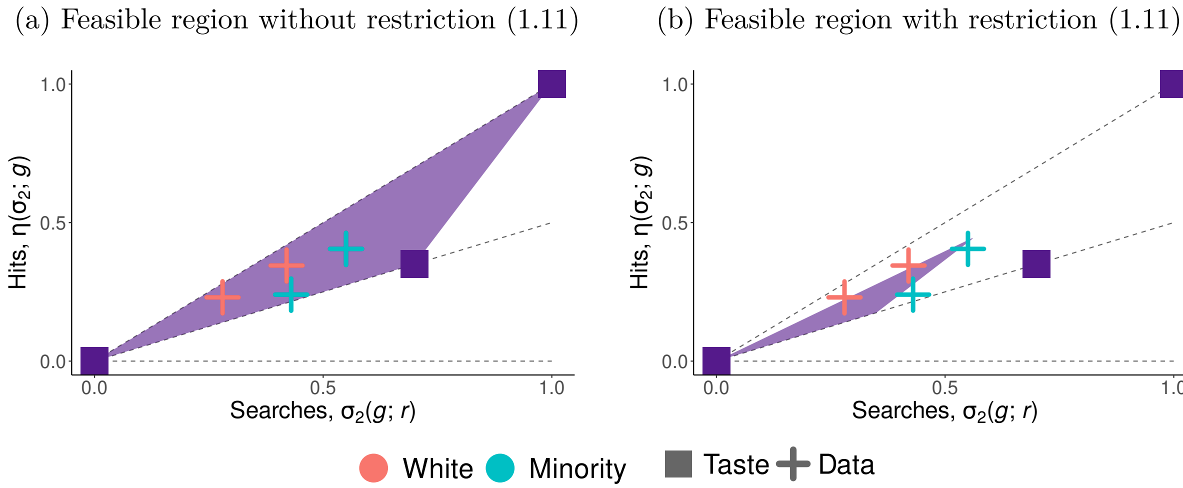
If the test detects bias, the next step is to determine *how* the officer is biased. This may be done in two ways. The first is to derive bounds on $\beta(g_k) \equiv \sigma(g_k; m) - \sigma(g_k; w)$ for

probabilities, is itself a Bernstein polynomial. This restricts the density function of risk to be smooth and allows the researcher to impose various shape restrictions.

[23] If ς and $\mathbf{p}_{r,z}$ are Bernstein polynomials, then their product—which is used to construct $\varsigma'\mathbf{p}_{r,z}$ in the objective function in the BP problem—is also a Bernstein polynomial. The unknown coefficients of this polynomial are bilinear functions of the unknown coefficients of ς and $\mathbf{p}_{r,z}$. Since the integral of any Bernstein basis polynomial of degree L over the unit interval is $(1+L)^{-1}$, Proposition 2 may still be applied by solving a BP problem that is feasible even when G_i is continuous.

[24] It is not unreasonable to assume that most drivers in the population are low-risk. If the mass of low-risk drivers in population is sufficiently large compared to high-risk drivers, then the officer may still primarily stop low-risk drivers, even if he prefers to stop high-risk drivers.

Figure 1.5: Strengthening the test by restricting $\{\mathbf{p}_{r,z}\}$



$k = 0, \ldots, K$, whose sharp identified set is

$$\mathcal{B}_k \equiv \left\{ b \in \mathbb{R} : b = \sigma(g_k; m) - \sigma(g_k; w) \text{ for } (\sigma(\cdot; w), \sigma(\cdot; m)) \in \Sigma^\star \right\}.$$

The sharp bounds on $\beta(g_k)$ may then be obtained as follows.

Proposition 3. *For $g_k \in \operatorname{supp}(G)$, the sharp bounds on $\beta(g_k)$ are obtained by solving the following BP problem,*

$$\begin{aligned} \beta_{\text{lb}}(g_k), \beta_{\text{ub}}(g_k) \equiv & \min_{\varsigma_w, \varsigma_m, \{\mathbf{p}_{w,z}\}, \{\mathbf{p}_{m,z}\}}/\max \quad \varsigma_{m,k} - \varsigma_{w,k} \\ \text{s.t.} \quad & \sum_{r,z} \left| \varsigma_r' \mathbf{p}_{r,z} - \mathbf{m}^S_{r,z} \right| + \sum_{r,z} \left| (\mathbf{g} \odot \varsigma_r)' \mathbf{p}_{r,z} - \mathbf{m}^H_{r,z} \right| = 0 \\ & \mathbf{A} \begin{bmatrix} \varsigma_w \\ \varsigma_m \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b}, \end{aligned} \tag{1.12}$$

where $\varsigma_{r,k}$ is the k^{th} element of ς_r.

Proof. The constraints in (1.12) characterize the sharp identified set $\Sigma^\star$, so the bounds are sharp by definition. ■

The bounds in Proposition 3 are sharp in the sense that they equal the smallest and largest values in the identified set $\mathcal{B}_k$. However, because bilinear programs are non-convex, $\mathcal{B}_k$ need not be the full interval $[\beta_{\text{lb}}(g_k), \beta_{\text{ub}}(g_k)]$, and may instead be a union of disjoint intervals contained in $[\beta_{\text{lb}}(g_k), \beta_{\text{ub}}(g_k)]$. I focus the discussion on the simpler bounds in Proposition 3, although $\mathcal{B}_k$ may be identified by 'inverting' (1.12), i.e., minimizing the criterion subject to the constraint that $\beta(g_k) = b$ for some $b \in \mathbb{R}$, where $b \in \mathcal{B}_k$ if and only if the criterion is zero. This procedure is similar to inverting a statistical test to construct a confidence interval. See Appendix A.2 for more details.

If $\beta_{\text{lb}}(g_k) > 0$, then the officer is biased against minorities with risk g_k; and if $\beta_{\text{ub}}(g_k) < 0$, then the officer is biased against whites. So it is possible that the officer is biased against one race when risk equals g_1, but biased against the other when risk equals g_2. It is also possible for the officer to fail the test in Proposition 2, but have $\beta_{\text{lb}}(g) < 0 < \beta_{\text{ub}}(g)$ for all $g \in \operatorname{supp}(G)$. For such an officer, I can detect that he is biased, but cannot determine the direction of bias. If the researcher has a strong prior on the direction of bias, then this prior may be imposed via a sign restriction on β, e.g., $\beta(g_k) \geq 0$ for all $k = 0, \ldots, K$ so that bias is always against minorities.[25]

The second way to measure the direction and intensity of bias is to average $\beta(g)$ across $g \in \operatorname{supp}(G)$. To do this, the researcher must first choose a weight $\omega = (\omega_0, \ldots, \omega_K)$ such that $\omega_k \in [0, 1]$ for $k = 0, \ldots, K$ and $\sum_{k=0}^{K} \omega_k = 1$. The average bias is then defined as

$$\mathbb{E}[\beta(G_i); \omega] \equiv \sum_{k=0}^{K} \omega_k \ \beta(g_k),$$

[25]The direction of bias can be fixed for a subset of values of risk instead.

and its sharp identified set is

$$\mathcal{E} \equiv \left\{ b \in \mathbb{R} : b = \sum_{k=0}^{K} \omega_k \ (\sigma(g_k; m) - \sigma(g_k; w) \text{ for } (\sigma(\cdot; w), \sigma(\cdot; m)) \in \Sigma^\star \right\}.$$

What $\mathbb{E}[\beta(G_i); \omega]$ measures is the average difference in the probability that equally risky white and minority drivers are searched, under the counterfactual where $\mathbb{P}\{G_i = g_k \mid R_i = r\} = \omega_k$ for $r \in \{w, m\}$.

Proposition 4. *For a choice of ω, the sharp bounds on $\mathbb{E}[\beta(G_i); \omega]$ are obtained by solving the following BP problem,*

$$\begin{aligned} \mathbb{E}[\beta(G_i); \omega]_{\text{lb}}, \mathbb{E}[\beta(G_i); \omega]_{\text{ub}} \equiv \quad & \min_{\varsigma_w, \varsigma_m, \{\mathbf{p}_{w,z}\}, \{\mathbf{p}_{m,z}\}}/\max \quad \omega' \left(\varsigma_m - \varsigma_w\right) \\ \text{s.t.} \quad & \sum_{r,z} \left| \varsigma_r' \mathbf{p}_{r,z} - \mathbf{m}^S_{r,z} \right| + \sum_{r,z} \left| (\mathbf{g} \odot \varsigma_r)' \mathbf{p}_{r,z} - \mathbf{m}^H_{r,z} \right| = 0 \\ & \mathbf{A} \begin{bmatrix} \varsigma_w \\ \varsigma_m \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b}. \end{aligned} \tag{1.13}$$

Proof. The constraints in (1.13) characterize the sharp identified set $\Sigma^\star$, so the bounds are sharp by definition. ■

Similar to the bounds in Proposition 3, the bounds in Proposition 4 are sharp in the sense that they equal the smallest and largest values in the identified set $\mathcal{E}$. See Appendix A.2 for how $\mathcal{E}$ maybe fully recovered by inverting (1.13). These bounds may also be seen as a non-parametric, partial identification approach to the decomposition methods of Oaxaca (1973), Blinder (1973), and DiNardo, Fortin, and Lemieux (1996).[26]

[26]These methods decompose average outcomes into structural and composition effects. By reweighting the structural effects, the authors are able to form counterfactuals. This is similar to how I decompose the search

For many interesting counterfactuals, the weights ω may be functions of $\{\mathbf{p}_{r,z}\}$ and therefore be unknown. For example, the researcher may wish to know what the average bias is if minority drivers are equally risky as white drivers in the data. This corresponds to choosing

$$\begin{aligned}\omega_k &= \mathbb{P}\{G_i = g_k \mid R_i = w\} \\ &= \sum_{z \in \mathcal{Z}} \mathbb{P}\{G_i = g_k \mid R_i = w, Z_i = z\}\ \mathbb{P}\{Z_i = z \mid R_i = w\} \\ &= \sum_{z \in \mathcal{Z}} p_{w,z}(g_k)\ q_{w,z}, \end{aligned} \tag{1.14}$$

for $k = 0, \ldots, K$, where the second equality follows from law of iterated expectations, and $q_{w,z} = \mathbb{P}\{Z_i = z \mid R_i = w\}$ is observed in the data. More generally, the researcher can choose ω to be

$$\omega = \mathbf{P}_w \mathbf{q}_w + \mathbf{P}_m \mathbf{q}_m,$$

where $\mathbf{P}_r$ is an unknown $K \times |\mathcal{Z}|$ matrix whose l^{th} column is $\mathbf{p}_{r,z_l}$; and $\mathbf{q}_r$ is a known vector whose purpose is to weight the different settings Z_i for race r. Proposition 4 continues to hold under this choice of ω since the objective function

$$\begin{aligned}\omega'(\varsigma_m - \varsigma_w) &= (\mathbf{P}_w \mathbf{q}_w + \mathbf{P}_m \mathbf{q}_m)'\ (\varsigma_m - \varsigma_w) \\ &= \sum_{r \in \{w,m\}} \sum_{l=1}^{|\mathcal{Z}|} \mathbf{q}_{r,l} \sum_{k=0}^{K} \underbrace{\mathbf{p}_{r,z,k}\ (\varsigma_{m,k} - \varsigma_{w,k})}_{\text{Bilinear terms}}\end{aligned}$$

is still bilinear.

and hit rates into σ and $p_{r,z}$, and reweight σ to construct counterfactuals. See Fortin, Lemieux, and Firpo (2011) for a summary of decomposition methods in economics.

1.4.6 Testing the model

Model misspecification test for each race

Before testing for racial bias, the researcher may first want to test whether the model is misspecified for each race of drivers. To see how this can be done, let $\Sigma_r^\star$ denote the sharp identified set of preferences for race r, which is simply a projection of $\Sigma^\star$. Then testing for model misspecification for race r amounts to testing whether $\Sigma_r^\star$ is empty. This can be done using the test in Bugni, Canay, and Shi (2015a), which outlines a misspecification test for partially identified models defined by moment inequalities.

Overidentifying restrictions test

If there are multiple instruments available, the researcher may also test the instrumental exogeneity condition in Assumption 1(iii). To see how, let $\Sigma^\star(Z_i)$ denote the identified set of white and minority preferences obtained using instrument Z_i. If $\Sigma^\star(Z_{i,1}) \cap \Sigma^\star(Z_{i,2}) = \emptyset$ for instruments $Z_{i,1}$ and $Z_{i,2}$, then the officer's preferences must be distinct for $Z_{i,1}$ and $Z_{i,2}$, which violates Assumption 1(iii).

1.5 Data

I apply the test to police traffic data from the Metropolitan Nashville Police Department (MNPD). The data contain records of traffic stops for over 2,200 MNPD officers between 2010 and 2019 and is made available by the Stanford Open Policing Project (Pierson et al., 2020).

1.5.1 MNPD traffic stop data

Each observation in the data is a traffic stop made by an officer. The researcher observes the driver's race, age, sex, and state of registration, but does not observe sensitive information such as her license number and vehicle tag number. All information on the officer is hidden, other than an anonymized identifier. Logistic details of the traffic stop are observed and include the date, time, address, and geocoordinates of the stop. The researcher observes the reason for the traffic stop, whether a search occurred, why the search occurred, whether any contraband was found, and the outcome of the stop (i.e., arrest, citation, warning). However, there are no detailed descriptions of the interaction between the officer and driver.[27]

Although though the data categorize contraband into weapons and drugs, I treat all forms of contraband as being the same. This is because traffic searches are infrequent and typically unsuccessful, so there are relatively few traffic stops that uncover contraband. This makes it infeasible to evaluate the officer's search decision for weapons and drugs separately. It is also unknown whether the officer was searching for weapons or drugs to begin with.

I supplement the MNPD police traffic data with additional MNPD data on criminal incidents and calls for services.[28] The purpose of these data are to control for environmental variables that may correlate with the setting of the stop but also affect the officer's preference toward searching a driver. For the same reasons, I also include local measures of racial composition and median household income from the American Community Survey of the US Census Bureau.

1.5.2 Restricting the sample

To study bias in traffic searches, the search decision must be discretionary. So traffic searches motivated by rules or mandates are excluded from the study. This includes searches that are incidental to an arrest, inventory searches, and searches based on warrants.[29] In total, 28% of the traffic searches in the data must be discarded for potentially being non-discretionary.

I restrict my attention to the 50 officers with the highest number of traffic searches. This

[27]For a small number of stops, a short note written by the officer summarizing the stop is available.

[28]I restrict both criminal incidents and calls for services to those related to violent crimes, theft, or drugs, as these may affect an officer's decision to search for contraband.

[29]Searches incidental to an arrest occur after a driver has been arrested. Hernández-Murillo and Knowles (2004) propose a methodology to incorporate non-discretionary searches into the analysis. Inventory searches are required whenever a vehicle is impounded by the police. Warrants to search a driver are typically obtained before the traffic stop, implying that warrant-based searches are predetermined.

Table 1.1: Summary of stops, searches, and hits for select 50 officers

	Full sample		Avg. by officer	
	White	Minority	White	Minority
Stops	109,023	113,405	2,180	2,268
Searches	12,622	15,732	252	315
Hits	1,831	2,741	37	55
Search rate	0.1158	0.1387	0.1546	0.1884
Uncon. hit rate	0.0168	0.0242	0.0277	0.0297
Con. hit rate	0.1451	0.1742	0.2431	0.2135

is because the methods discussed in Section 1.4.3 are performed on each officer separately, and in order to reasonably estimate their search and hit rates, I require each of them to have made a large number of traffic stops and searches. On average, these officers make 2,180 stops and 250 searches for white drivers, and 2,268 stops and 314 searches for minority drivers. Remarkably, this small fraction of officers make up a third of all searches in the data.

Finally, I focus on comparing the officer's preferences for searching white drivers against that of Black and Hispanic drivers. 'Minority' therefore exclusively refers to Black and Hispanic drivers.

Table 1.1 summarizes the number of traffic stops, searches, and hits in the restricted sample.

1.5.3 Context variable Z_i

The officer's search preferences are realized after he observes driver characteristics R_i, V_i and before he searches the driver. The options for Z_i depend on the determinants of the officer's preferences, which I assume include basic demographic variables (e.g., race, age, sex of the driver), the reason for the stop, the interaction with the driver, and the surrounding environment. For example, the officer may feel less comfortable searching female drivers than male drivers, suggesting that female drivers face larger draws of T_i. A motorist stopped for reckless driving may earn the ire of an officer and thereby face lower draws of T_i. A charismatic driver may be able to dissuade the officer from searching, or at least discourage him, implying larger draws of T_i. Finally, patrolling a dangerous neighborhood may put the officer on edge, resulting in lower draws of T_i. For Z_i to satisfy the independence conditions in Assumption 1, it is necessary for me to control for the determinants of officer preferences that may be correlated with Z_i, as they may induce a correlation between Z_i and T_i.

In the application, I choose Z_i to be combinations of the day of the week and the patrol shift. I divide the days into weekdays and weekends, and patrol shifts are either in the morning (7am–3pm), evening (3pm–11pm), or night (11pm–7am), giving me up to six values of Z_i for each officer. Below, I discuss the controls I use to support this assumption. Table 1.2 provides summary statistics for these controls, and Tables 1.3–1.4 show how they vary with Z_i.

The first set of controls are the observable (to the researcher) characteristics of the driver, i.e., race, age, sex, state of registration. To see why this is necessary, imagine that officers prefer not to inconvenience elderly female drivers by searching them, but do not feel such reservations toward college-age male drivers. Tables 1.3–1.4 show that drivers who are stopped late at night are younger and more likely to be male compared to drivers stopped earlier in the day. So if elderly females primarily drive in the mornings, and college-age males primarily drive late at night when there is no school, then Z_i may violate the independence assumption.

The second set of controls include the details of the traffic encounter, namely the reason for the stop and, if a search took place, the reason for the search. I categorize the reason for stop into three groups: driving-related reasons, non-driving related reasons, and investigative reasons.[30] If officers have different search preferences for drivers depending on the (reported) reason they are stopped, and the (reported) reason for stops varies with the day or shift, then

[30] Driving-related reasons correspond to how the driver maneuvers her vehicle and how she interacts with other drivers on the road. This include moving traffic violations, safety violations, and vehicle equipment violations. Non-driving reasons correspond to reasons unrelated to how the vehicle is driven, and include seat belt violations, parking violations, registration violations, and issues with child restraints. Investigative stops are its own category and not an aggregate of other reasons.

Table 1.2: Summary of control variables

	Drivers stopped		Drivers searched	
	White	Minority	White	Minority
Driver characteristics				
Male	0.6032	0.6007	0.6613	0.7722
Age	37.28	34.64	32.31	30.49
Out of state	0.0638	0.0330	0.0490	0.0340
Reason for stop				
Driving	0.8803	0.8776	0.8668	0.8687
Non-driving	0.1070	0.1065	0.1072	0.1031
Investigation	0.0127	0.0159	0.0260	0.0282
Reason for search				
Plain view			0.4978	0.2606
Consent			0.4336	0.5938
Probable Cause			0.0686	0.1456
Location				
Highway	0.1228	0.0644	0.0759	0.0495
Precinct 1	0.0763	0.0509	0.0640	0.0521
Precinct 2	0.1190	0.1760	0.0882	0.1920
Precinct 3	0.1042	0.1446	0.0913	0.1377
Precinct 4	0.0395	0.0249	0.0789	0.0381
Precinct 5	0.3618	0.2567	0.2573	0.2227
Precinct 6	0.0400	0.1100	0.0257	0.0774
Precinct 7	0.1366	0.1528	0.1469	0.1540
Precinct 8	0.1225	0.0842	0.2477	0.1260
Census tract demographics				
Percent white	0.5901	0.4523	0.6028	0.4580
Median household income	49038	41170	48642	40029
Crime incident rate	0.0256	0.0369	0.0305	0.0400
Calls for MNPD services	0.0207	0.0216	0.0212	0.0227

Notes: Crime and call rates are per capita and are restricted to those pertaining to violent crimes, theft, or drugs.

Table 1.3: Controls by Z_i, white drivers

	Weekday			Weekend		
	Morning	Evening	Night	Morning	Evening	Night
Driver characteristics						
Male	0.5792	0.6072	0.6521	0.5912	0.6230	0.6286
Age	39.46	36.37	34.43	40.98	35.89	32.01
Out of state	0.0684	0.0533	0.0615	0.0738	0.0654	0.0805
Reason for stop						
Driving	0.8686	0.8502	0.9413	0.8920	0.8948	0.9376
Non-driving	0.1212	0.1403	0.0395	0.0999	0.0862	0.0353
Investigation	0.0103	0.0094	0.0192	0.0081	0.0190	0.0271
Reason for search						
Plain view	0.1252	0.3814	0.6302	0.0781	0.5954	0.7899
Consent	0.7318	0.5345	0.3297	0.8438	0.3505	0.1759
Probable Cause	0.1430	0.0842	0.0401	0.0781	0.0541	0.0342
Location						
Highway	0.1445	0.0877	0.1279	0.0854	0.1049	0.1313
Precinct 1	0.0503	0.0568	0.1505	0.0517	0.1205	0.1449
Precinct 2	0.1429	0.1324	0.0550	0.0505	0.1179	0.0548
Precinct 3	0.0846	0.1011	0.1226	0.2909	0.1631	0.1197
Precinct 4	0.0300	0.0312	0.0527	0.0331	0.0478	0.1170
Precinct 5	0.4180	0.3815	0.2656	0.1545	0.2913	0.1969
Precinct 6	0.0572	0.0331	0.0149	0.0314	0.0289	0.0156
Precinct 7	0.1219	0.1736	0.1527	0.0935	0.0918	0.0846
Precinct 8	0.0949	0.0904	0.1859	0.2944	0.1388	0.2664
Census tract demographics						
Percent white	0.6044	0.5610	0.5986	0.5733	0.5689	0.6174
Median household income	51915	45154	48417	49006	45648	49590
Crime incident rate	0.0193	0.0434	0.0208	0.0049	0.0128	0.0205
Calls for MNPD services	0.0232	0.0251	0.0096	0.0087	0.0131	0.0197

Notes: Crime and call rates are per capita and are restricted to those pertaining to violent crimes, theft, or drugs. Rates for reasons for search are calculated using only stops involving searches. All other rates are estimated using all stops in the data.

Table 1.4: Controls by Z_i, minority drivers

	Weekday			Weekend		
	Morning	Evening	Night	Morning	Evening	Night
Driver characteristics						
Male	0.5606	0.5963	0.6663	0.6127	0.6111	0.6540
Age	36.12	34.56	32.89	37.84	34.06	31.45
Out of state	0.0350	0.0272	0.0368	0.0421	0.0323	0.0480
Reason for stop						
Driving	0.8630	0.8547	0.9305	0.8755	0.8984	0.9375
Non-driving	0.1229	0.1323	0.0472	0.1067	0.0874	0.0348
Investigation	0.0141	0.0130	0.0224	0.0177	0.0142	0.0276
Reason for search						
Plain view	0.1126	0.2034	0.3451	0.0439	0.2728	0.5067
Consent	0.6723	0.6319	0.5501	0.8772	0.5754	0.4083
Probable Cause	0.2151	0.1647	0.1048	0.0789	0.1518	0.0850
Location						
Highway	0.0780	0.0388	0.0930	0.0570	0.0483	0.0987
Precinct 1	0.0348	0.0272	0.1215	0.0306	0.0456	0.1123
Precinct 2	0.1623	0.2362	0.1008	0.0598	0.1764	0.1047
Precinct 3	0.1186	0.1216	0.1734	0.3652	0.2204	0.1963
Precinct 4	0.0227	0.0169	0.0271	0.0124	0.0265	0.0864
Precinct 5	0.2948	0.2509	0.2398	0.1063	0.2505	0.1842
Precinct 6	0.1508	0.1120	0.0443	0.1364	0.1046	0.0490
Precinct 7	0.1480	0.1642	0.1909	0.0857	0.0865	0.1012
Precinct 8	0.0679	0.0711	0.1023	0.2034	0.0895	0.1659
Census tract demographics						
Percent white	0.4814	0.4078	0.4925	0.4424	0.4147	0.5200
Median household income	46113	37043	42050	42478	36647	43331
Crime incident rate	0.0237	0.0619	0.0232	0.0060	0.0176	0.0187
Calls for MNPD services	0.0221	0.0293	0.0090	0.0089	0.0139	0.0164

Notes: Crime and call rates are per capita and are restricted to those pertaining to violent crimes, theft, or drugs. Rates for reasons for search are calculated using only stops involving searches. All other rates are estimated using all stops in the data.

it becomes necessary to condition on why the driver is stopped.[31] For example, Makofske (2020) shows that officers in Louisville arrest 40% of drivers stopped for failing to signal, compared to 1% of drivers stopped for any other reason. This suggests that certain stops are pretextual, and that the reason for stop may indicate the officer's search preference. In addition, the data show a 10% increase in the proportion of stops being attributed to driving-related reasons between the evening and night shifts.[32] If the reason for stop is indeed correlated with both search preferences and setting, then it is necessary to condition on it to satisfy Assumption 1.

Reasons for traffic searches include driver consent, probable cause, and plain view of contraband, and provide some insight into the interaction between the officer and driver. To see why it is necessary to condition on the reason for search, consider a traffic stop where the driver behaves belligerently. Not only may her behavior raise the officer's suspicion that she is hiding contraband and there is probable cause to search her, but it may also frustrate the officer and result in a lower draw of T_i. In contrast, a respectful driver may be disarming, and the officer may instead ask for consent to search, or forgo the search entirely. Consequently, the search basis and officer preference may be correlated. If this type of behavior is also correlated with Z_i—e.g., belligerent drivers are more common during the weekend night shifts—then it becomes necessary to condition on the search basis. Tables 1.3–1.4 show strong correlation between the setting Z_i and reason for searches.

The final set of controls pertains to the environment where the stop takes place. This includes whether the stop is on a highway, the police precinct, the racial composition and income level of the census tract, the crime rate of the census tract, and the frequency of calls for MNPD services from the census tract. The concern is that an officer may be more cautious and mindful of his safety in some neighborhoods, but more carefree in others, and his mindset may influence his search preferences. For example, Roh and Robinson (2009) find there to be

[31] Durlauf and Heckman (2020) raise concerns about the credibility of self-reported police data. While the concern is valid, there is currently not a good solution.

[32] In a study on endogenous driving behavior, Kalinowski, Ross, and Ross (2020) find that minority drivers adjust their driving behavior during the day, when their race is more visible to the officer.

spatial correlation in traffic search decisions even after controlling for driver characteristics. The authors attribute the correlation to similarities in environmental variables, such as the racial composition of the neighborhood and the volume of police allocated nearby. Novak and Chamlin (2012) also find that the police workload (measured via calls for services) and degree of 'social disorganization' (e.g., percentage of single parent households, percentage of residents in poverty) are predictive of officer behavior. If officers patrol different locations depending on the time and day, then this may induce a correlation between the officer's preferences and the setting of the stop, making these controls necessary.

A concern may be that officers are not randomly assigned to shifts, but instead choose which shifts to work. However, this is a threat to identification only if an officer's shift preference correlates with his search preferences. For example, an officer may reasonably wish to avoid the night shift between 3pm and 11pm since that limits the time he has with his family. But as long as he is equally willing to search drivers on each shift, then the exogeneity condition will hold.

Another concern may be that there are ticket quotas enforced by the MNPD that affect an officer's stop decisions. While Tennessee has explicit laws banning quotas on traffic citations, it has not stopped departments from implementing such targets [33] However, ticket quotas are only a concern if they affect the officer's search preferences. If quotas only induce the officer to stop drivers more often, then quotas simply impact the distribution of risk among drivers who have been stopped, similar to sample selection. So even if the effect of quotas on stop decisions differ for white and minority drivers, the methods discussed still allow me to rule out racial disparities in risk as being the cause for the racial disparities in police search decisions.

There is also the possibility that officers are instructed to search more intensely during certain times. For example, Spartanburg County in South Carolina carries out Operation

[33] See Tennessee Code §39-16-516 (2014) for the law. The mayor of Ridgetop, TN tried to have the city's police department enforce a ticket quota to raise money for the city, only to be turned in by the city's police chief; see https://fox17.com/news/local/ferrier-files-ridgetop-disbands-police-department-after-illegal-ticket-quotas-exposed.

Rolling Thunder each October. This entails a week of aggressive traffic stops and searches performed by officers from different agencies.[34] So a concern may be that MNPD policy requires officers to have different search standards depending on the time of the day or day of the week. To the best of my knowledge, there were no such policies during time frame of the data being analyzed. Such policies were only implemented beginning in July of 2019.[35]

1.6 Estimation

Estimation is performed in several steps and done separately for each officer. First, I estimate an individual officer's search and hit rates. Then I test for bias using a bilinear program. If the officer is biased, I then construct bounds on $\beta(g)$ for each g, as well as $\mathbb{E}[\beta(g);\omega]$ for some ω.

1.6.1 Estimating search and hit rates

Let X_i denote the vector of control variables. I condition on $X_i = \overline{x}$ throughout, where $\overline{x}$ is defined as follows. For continuous controls, $\overline{x}_j = (1/n)\sum_i X_{i,j}$, i.e., the sample average, where j indexes components in the vector of controls. For categorical controls, $\overline{x}_j = \arg\max_{x_j} \widehat{\mathbb{P}}\{X_{i,j} = x_j\}$, i.e., the sample mode, where $\widehat{\mathbb{P}}$ denotes the empirical distribution.

To construct the search rate $\widehat{\mathbb{E}}[\mathit{Search}_i \mid R_i, Z_i, X_i]$, I use a logistic regression. To construct

[34] See https://www.wspa.com/news/crime/spartanburg-co-sheriffs-office-to-give-update-on-operation-rolling-thunder/

[35] In July of 2019, the MNPD introduced the Entertainment District Initiative, which assigned 17 additional officers to the Entertainment District on Fridays and Saturdays between 6 p.m. and 4 a.m. to improve public safety. These officers performed high-visibility patrols on foot, bike, and utility task vehicles, and would make unannounced visits to local establishments. In February of 2021, the MNPD introduced the Office of Alternative Policing Strategies to address an increase in violent crime in Nashville. A new shift of 80 officers working between 5:30pm–3:30am was added across all precincts to perform high-visibility patrols to deter and detect violent crimes. See https://tennesseelookout.com/2021/02/01/in-nashville-mayor-cooper-chief-drake-announce-policing-reforms-to-address-murders-gun-crimes/, https://www.newschannel5.com/news/special-report-mnpd-puts-60-extra-officers-in-nashvilles-entertainment-district-on-weekends-why, and https://www.nashville.gov/departments/police/news/new-initiative-further-enhance-public-safety-nashvilles-entertainment.

the hit rate $\widehat{\mathbb{E}}[Hit_i \mid R_i, Z_i, X_i]$, I use the relation

$$\mathbb{E}[Hit_i \mid R_i, Z_i, X_i] = \mathbb{E}[Hit_i \mid Search_i = 1, R_i, Z_i, X_i]\ \mathbb{E}[Search_i \mid R_i, Z_i, X_i].$$

So I first estimate the conditional hit rate $\widehat{\mathbb{E}}[Hit_i \mid Search_i = 1, R_i, Z_i, X_i]$ using a logistic regression on the subsample of drivers who are searched. I then scale these estimates by the estimated search rates, $\widehat{\mathbb{E}}[Search_i \mid R_i, Z_i, X_i]$.

An alternative approach to estimating the hit rate is to simply regress Hit_i on R_i, Z_i, and X_i. However, for some officers, this results in nonsensical estimates of the hit rate that exceed their search rate. This would imply a conditional hit rate greater than 1, which is not possible.

To summarize the variation in search and hit rates generated by Z_i, Tables 1.5–1.6 present logistic regressions of the search and hit indicators on Z_i, conditional on race, controls, and officer fixed effects. For ease of interpretation, the estimates presented are the exponentiated logit coefficients and they reflect the multiplicative impact that a change of setting has on the odds of being searched or finding contraband. The estimates suggest that, relative to stops during weekday evenings, the average odds of being searched can fall by 36% and rise by 68% across the settings. Conditional on being searched, the average odds of finding contraband can fall by 20% and increase by up to 168% across settings.

Figure 1.6 provides examples of the search and hit rates for specific officers after conditioning on controls. For officer 49, the data are fairly similar across races, which may be consistent with the officer being unbiased. But the same cannot be said of the other officers. For instance, officer 6 is an example where bias is certainly going to be detected, since the absence of any hits for both white and minority drivers suggests both groups of drivers stopped are low-risk, yet minorities are searched three times more often than white drivers. Also, to see that the conditional hit rate varies with setting at the officer level, simply note that the data for each race do not lie along a ray extending from the origin.[36]

[36]The conditional hit rate for each data point is equal to the ratio of the y-coordinate and x-coordinate of

Table 1.5: Pooled logistic regression of $Search_i$ on Z_i, X_i

Contraband found	Estimate	Std. Err.	p-value	C.I. lower	C.I. upper
White					
Weekend	**0.7855**	0.0498	0.0001	0.6937	0.8895
Morning	**0.6356**	0.0271	0	0.5846	0.6909
Night	**1.4779**	0.0666	0	1.353	1.6144
Weekend × Morning	0.9252	0.1382	0.6029	0.6903	1.24
Weekend × Night	**0.7845**	0.0638	0.0028	0.6689	0.9201
N	109,023				
Minority					
Weekend	**0.8027**	0.0389	0	0.7299	0.8828
Morning	**0.5737**	0.0203	0	0.5353	0.6148
Night	**1.4698**	0.0528	0	1.3698	1.5771
Weekend × Morning	**1.6803**	0.1525	0	1.4065	2.0074
Weekend × Night	**0.8504**	0.0535	0.01	0.7517	0.9621
N	113,405				

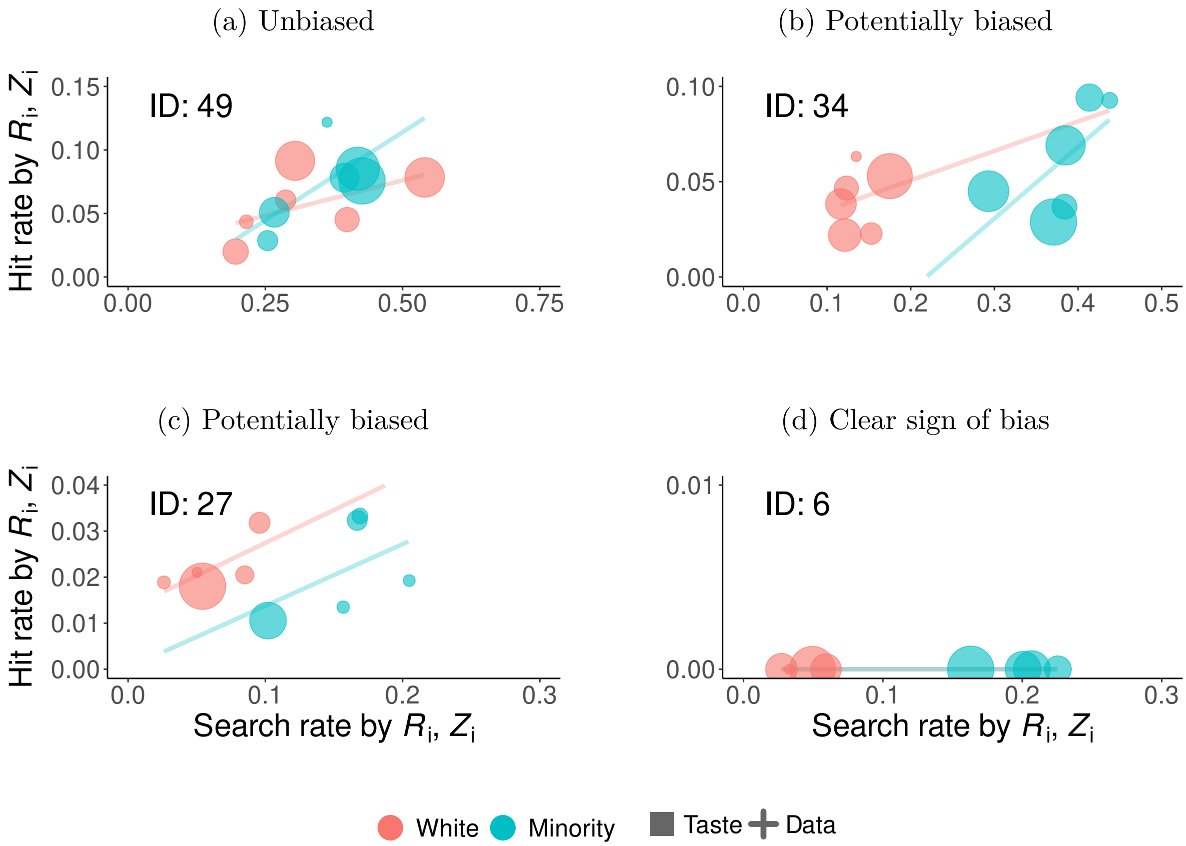
Notes: Estimates presented are the multiplicative impact a change of setting has on the odds ratio, and are relative to weekday evening stops. Estimates condition on officer fixed effects. Confidence intervals are at the 95% confidence level.

Table 1.6: Pooled logistic regression of Hit_i on Z_i, X_i, conditional on being searched

Contraband found	Estimate	Std. Err.	p-value	C.I. lower	C.I. upper
White					
Weekend	0.7987	0.1382	0.1938	0.569	1.1211
Morning	0.9345	0.0952	0.5063	0.7653	1.1411
Night	1.1581	0.122	0.1633	0.9421	1.4237
Weekend × Morning	**2.3021**	0.9085	0.0346	1.0622	4.9894
Weekend × Night	1.0303	0.2224	0.8899	0.6749	1.573
N	12,622				
Minority					
Weekend	0.8047	0.0968	0.0708	0.6357	1.0186
Morning	1.0007	0.0872	0.9938	0.8436	1.1871
Night	**1.4871**	0.1269	0	1.258	1.7578
Weekend × Morning	1.6823	0.4608	0.0575	0.9835	2.8777
Weekend × Night	0.9221	0.1445	0.6049	0.6783	1.2537
N	15,732				

Notes: Estimates presented are the multiplicative impact a change of setting has on the odds ratio, and are relative to weekday evening stops. Estimates condition on officer fixed effects. Confidence intervals are at the 95% confidence level.

Figure 1.6: Example of officer-level data



Note: Size of points correspond to number of traffic stops in setting Z_i. Search and hit rates are conditional on X_i.

1.6.2 Restrictions $\operatorname{supp}(G)$ and $\{\mathbf{p}_{r,z}\}$

In order for the test to be computationally feasible, I discretize the support of risk to be

$$\mathbf{g} = \{\underbrace{0, 0.025, 0.05, 0.075}_{\text{Increments of } 0.025}, \underbrace{0.1, 0.15, 0.20, 0.25}_{\text{Increments of } 0.05}, \underbrace{0.3, 0.4, 0.5, 0.6}_{\text{Increments of } 0.1}, 0.75, 1\}.$$

The grid $\mathbf{g}$ is deliberately chosen to be finer at lower levels of risk since the average conditional hit rate across officers is not particularly high, between 20% and 25% (see Table 1.1). This suggests most drivers searched are relatively low risk. In order to distinguish between these drivers, I allocate more points in the grid to lower-levels of risk.

Furthermore, since the drivers searched represent a riskier subset of the driver stopped, the low conditional hit rates suggest that most drivers stopped are relatively low risk as well. So I also impose the monotonicity restriction in (1.11), requiring that the PMF $\mathbf{p}_{r,z}$ is decreasing in risk for all $(r, z) \in \{w, m\} \times \mathcal{Z}$

I do not impose any restrictions on σ other than that it is non-decreasing in risk and bounded in the unit interval. But with more computational time, I can impose additional restrictions (e.g., model σ using Bernstein polynomials) in future versions of the draft.

1.6.3 Testing for bias

To test for bias, I solve the empirical counterpart to (1.8) in Proposition 2. I reweight the moments to improve efficiency, although an optimal weighting scheme has yet to be

the point. If the conditional hit rates are the same across all settings, then the data points will lie along a line extending from the origin.

determined. The problem I solve is then

$$\widehat{Q}^{\star} \equiv \min_{\varsigma, \{\mathbf{p}_{r,z}\}} \sum_{r,z} \widehat{\mathbf{w}}^{S}_{r,z} \left| \varsigma' \mathbf{p}_{r,z} - \widehat{\mathbf{m}}^{S}_{r,z} \right| + \sum_{r,z} \widehat{\mathbf{w}}^{H}_{r,z} \left| (\mathbf{g} \odot \varsigma)' \mathbf{p}_{r,z} - \widehat{\mathbf{m}}^{H}_{r,z} \right| \tag{1.15}$$

$$\text{s.t.} \quad \mathbf{A} \begin{bmatrix} \varsigma \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b},$$

where

$$\widehat{\mathbf{m}}^{S}_{r,z} \equiv \widehat{\mathbb{P}}\{Search_i = 1 \mid R_i = r, Z_i = z, X_i = \overline{x}\}$$

$$\widehat{\mathbf{m}}^{H}_{r,z} \equiv \widehat{\mathbb{P}}\{Hit_i = 1 \mid R_i = r, Z_i = z, X_i = \overline{x}\}$$

and the weights are[37]

$$\widehat{\mathbf{w}}^{S}_{r,z} \equiv \frac{\sqrt{\sum_{i:R_i=r} \mathbb{1}\{R_i = r, Z_i = z\}}}{\widehat{\text{s.e.}}(\widehat{\mathbb{P}}\{Search_i = 1 \mid R_i = r, Z_i = z, X_i = \overline{x}\})},$$

$$\widehat{\mathbf{w}}^{H}_{r,z} \equiv \min\left\{ \frac{\sqrt{\sum_{i:R_i=r} \mathbb{1}\{R_i = r, Z_i = z\}}}{\widehat{\text{s.e.}}(\widehat{\mathbb{P}}\{Hit_i = 1 \mid R_i = r, Z_i = z, X_i = \overline{x}\})},\ 20\ \widehat{\mathbf{w}}^{S}_{r,z} \right\}.$$

The standard errors in the denominators of the weights adjust for how well the search and hit rates are estimated. These standard errors are estimated using a stratified bootstrap, where the number of stops drawn for a given R_i and Z_i is equal to that of the original sample. The numerator in the weights is the square-root of the number of traffic stops for a given race and setting. Its purpose is to account for how the standard errors may be artificially low for settings where the officer has made only a few stops. For example, if an officer makes five stops for $(R_i, Z_i) = (w, z)$ and happens to search the driver every time, then $\widehat{\text{s.e.}}(\{\widehat{\mathbb{P}}\{Search_i = 1 \mid R_i = w, Z_i = z, X_i = \overline{x}\})$ will be small. As a result, these search

[37] In practice, I also scale the weights by a constant to improve numerical stability when optimizing.

and hit rates will be weighted too heavily, and the test will primarily target these moments despite how they make up a small fraction of the officer's stops.

In addition, the weight assigned to the hit rate is limited to twenty times that of the search rate. This is to prevent excessive weight being placed on the hit rates for officers who almost never find contraband, for whom $\widehat{\text{s.e.}}(\{\widehat{\mathbb{P}}\{Hit_i = 1 \mid R_i = w, Z_i = z, X_i = \overline{x}\})$ is small.

To conduct statistical inference, I make a modification to the test described in Proposition 2. In particular, I do not use $\widehat{Q}^\star$ as the test statistic to detect bias. Doing so not only tests whether the officer is biased, but also tests whether the model in general is misspecified. As a result, the distribution of $\widehat{Q}^\star$ may be overly dispersed and the test becomes very weak. Instead, I solve a second BP problem where the officer is allowed to have different preferences ς_w, ς_m for each race of drivers,

$$\widehat{Q}^\star_B \equiv \min_{\varsigma_w, \varsigma_m, \{\mathbf{p}_{r,z}\}} \sum_{r,z} \widehat{\mathbf{w}}^S_{r,z} \left| \varsigma_r' \mathbf{p}_{r,z} - \widehat{\mathbf{m}}^S_{r,z} \right| + \sum_{r,z} \widehat{\mathbf{w}}^H_{r,z} \left| (\mathbf{g} \odot \varsigma_r)' \mathbf{p}_{r,z} - \widehat{\mathbf{m}}^H_{r,z} \right| \tag{1.16}$$

$$\text{s.t.} \quad \mathbf{A} \begin{bmatrix} \varsigma_w \\ \varsigma_m \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b}.$$

I then construct the test statistic as

$$\widehat{\tau} \equiv \mathbb{1}\{\widehat{Q}^\star > 0\} \frac{\widehat{Q}^\star - \widehat{Q}^\star_B}{\widehat{Q}^\star_B},$$

which compares the fit of the model when the officer is restricted to being unbiased against the fit without the restriction. For example, if $\widehat{\tau} = 0.05$, that means the fit of the model under the unbiasedness restriction is 5% worse relative to the fit without the restriction. The test therefore only tests the unbiasedness restriction.[38]

[38] See Bugni et al. (2015a); Bugni, Canay, and Shi (2017) and Chernozhukov, Newey, and Santos (2020)

To obtain the distribution of $\widehat{\tau}$, I use a stratified bootstrap, where the number of stops drawn for a given R_i and Z_i is equal to that of the original sample. I then reject that an officer is unbiased if the estimated α-quantile of τ exceeds threshold $\overline{\tau}$, for some choice of α and $\overline{\tau}$. This heuristic approach is not guaranteed to control the size of the test; a more formal approach is under development.

Due to the computational demands of (1.15)–(1.16), inference is performed using 200 bootstrap samples and BP programs are terminated after five minutes. I construct $\widehat{\tau}$ conservatively whenever the BP program is terminated before being solved to global optimality. Specifically, when solving (1.15), a lower and upper bound on $\widehat{Q}^\star$ is obtained, and an optimal solution is reached when the lower and upper bounds coincide.[39] If (1.15) is terminated before this occurs, then the lower bound of $\widehat{Q}^\star$ is used to construct $\widehat{\tau}$. In contrast, if (1.16) is terminated before optimality is achieved, then the upper bound of $\widehat{Q}^\star_B$ is used to construct $\widehat{\tau}$. Together, this minimizes $\widehat{\tau}$, resulting in a more conservative test.

1.6.4 Bounding intensity of bias

If bias is detected, then the direction and intensity of the bias may be estimated by solving the empirical counterparts to (1.12)–(1.13) in Propositions 3–4. The bounds on $\beta(g_k)$ for

for a discussion on inference for partially identified models defined by moment restrictions.

[39]In practice, optimality is achieved once the difference between the lower and upper bounds fall below some tolerance. In the application, I set the tolerance to be 1%.

$k = 0, \ldots, K$ are estimated by

$$\widehat{\beta}_{\text{lb}}(g_k), \widehat{\beta}_{\text{ub}}(g_k) \equiv \min_{\varsigma_w, \varsigma_m, \{\mathbf{p}_{w,z}\}, \{\mathbf{p}_{m,z}\}}/\max \; \varsigma_{m,k} - \varsigma_{w,k} \tag{1.17}$$

$$\text{s.t.} \quad \sum_{w,z} \widehat{\mathbf{w}}^S_{w,z} \left| \varsigma_w' \mathbf{p}_{w,z} - \widehat{\mathbf{m}}^S_{w,z} \right| + \sum_{w,z} \widehat{\mathbf{w}}^H_{w,z} \left| (\mathbf{g} \odot \varsigma_w)' \mathbf{p}_{w,z} - \widehat{\mathbf{m}}^H_{w,z} \right| \leq \widehat{Q}^\star_{B,w}(1+\kappa)$$

$$\sum_{m,z} \widehat{\mathbf{w}}^S_{m,z} \left| \varsigma_m' \mathbf{p}_{m,z} - \widehat{\mathbf{m}}^S_{m,z} \right| + \sum_{m,z} \widehat{\mathbf{w}}^H_{m,z} \left| (\mathbf{g} \odot \varsigma_m)' \mathbf{p}_{m,z} - \widehat{\mathbf{m}}^H_{m,z} \right| \leq \widehat{Q}^\star_{B,m}(1+\kappa)$$

$$\mathbf{A} \begin{bmatrix} \varsigma_w \\ \varsigma_m \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b},$$

where the moments from the data have been replaced by their sample counterparts; $\widehat{Q}^\star_{B,r}$ is the minimized criterion for race r obtained in (1.16), i.e.,

$$\widehat{Q}^\star_{B,r} \equiv \sum_z \widehat{\mathbf{w}}^S_{r,z} \left| \widehat{\varsigma}_r' \mathbf{p}_{r,z} - \widehat{\mathbf{m}}^S_{r,z} \right| + \sum_z \widehat{\mathbf{w}}^H_{r,z} \left| (\mathbf{g} \odot \widehat{\varsigma}_r)' \mathbf{p}_{r,z} - \widehat{\mathbf{m}}^H_{r,z} \right|,$$

where $\widehat{\varsigma}_r$ is part of the solution to (1.16); and $\kappa \geq 0$ is a tuning parameter controlling the slackness in the moment matching criterion, where the slackness ensures the optimization problem is always feasible.[40] In the application, I set $\kappa = 0.001$. The bounds I present in Section 1.7 may be tightened by choosing a smaller value of κ.

To estimate the average bias, I define ω as in (1.14) and set $q_{w,z} = \widehat{\mathbb{P}}\{Z_i = z \mid R_i = w\}$. That is, the average bias corresponds to the average difference in the probability of being searched when minority drivers have the same distribution of risk as white drivers in the

[40] The tuning parameter κ converges to zero as the number of traffic stops grows. See Mogstad, Santos, and Torgovitsky (2018a) for another example of such a tuning parameter.

data. The bounds on $\mathbb{E}[\beta(G_i);\omega]$ are estimated by

$$\widehat{\mathbb{E}}[\beta(G_i);\omega]_{\mathrm{lb}}, \widehat{\mathbb{E}}[\beta(G_i);\omega]_{\mathrm{ub}} \equiv \min_{\varsigma_w,\varsigma_m,\{\mathbf{p}_{w,z}\},\{\mathbf{p}_{m,z}\}}/\max \; \mathbf{q}_w' \mathbf{P}_w \left(\varsigma_m - \varsigma_w\right) \tag{1.18}$$

$$\text{s.t.} \quad \sum_z \widehat{\mathbf{w}}_{w,z}^S \left| \varsigma_w' \mathbf{p}_{w,z} - \widehat{\mathbf{m}}_{w,z}^S \right| + \sum_z \widehat{\mathbf{w}}_{w,z}^H \left| (\mathbf{g} \odot \varsigma_w)' \mathbf{p}_{w,z} - \widehat{\mathbf{m}}_{w,z}^H \right| \leq \widehat{Q}_{B,w}^{\star}(1+\kappa)$$

$$\sum_z \widehat{\mathbf{w}}_{m,z}^S \left| \varsigma_m' \mathbf{p}_{m,z} - \widehat{\mathbf{m}}_{m,z}^S \right| + \sum_z \widehat{\mathbf{w}}_{m,z}^H \left| (\mathbf{g} \odot \varsigma_m)' \mathbf{p}_{m,z} - \widehat{\mathbf{m}}_{m,z}^H \right| \leq \widehat{Q}_{B,m}^{\star}(1+\kappa)$$

$$\mathbf{A} \begin{bmatrix} \varsigma_w \\ \varsigma_m \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b}.$$

The BP problems (1.17)–(1.18) are terminated after five minutes. If optimality is not yet reached, then the lower bound of the BP objective value is used to estimate the lower bound on the measure of bias, and the upper bound of the BP objective value is used to estimate the upper bound on the measure of bias.

To conduct inference, the researcher may construct the confidence interval for each measure of bias by inverting the test in Proposition 2. Specifically, for $b \in [-1, 1]$ and $k = 1, \ldots, K$, the researcher can test the restriction $\beta(g_k) = b$. If the test does not reject the restriction, then b is contained in the confidence interval for $\beta(g_k)$. The confidence interval for $\mathbb{E}[\beta(G_i);\omega]$ may be constructed in the same way. See Appendix A.2.4 for a full description of this procedure. This is a computationally demanding approach since it involves iterating over a grid of values for $b \in [-1, 1]$, and each iteration involves bootstrapping a BP problem.

1.7 Results

Prior to testing for racial bias, I test whether the model is misspecified for each officer and race using the test described in Bugni et al. (2015a). For white drivers, the model is rejected for one officer.[41] For minority drivers, the model is not rejected for any officer.

Table 1.7 displays the number of officers who fail the test under various specifications of the test. Each column indicates the α-quantile of the empirical distribution of the test statistic $\widehat{\tau}$ used to detect bias, and each row indicates the threshold $\overline{\tau}$ for how much the model fit must worsen under the unbiasedness restriction before the officer is flagged as biased. Each entry indicates the number of officers who fail the test for a choice of α and $\overline{\tau}$. The counts are not adjusted for multiple hypothesis testing.

The size of the test increases as α increases and $\overline{\tau}$ decreases. For instance, consider the entry corresponding to $\alpha = \overline{\tau} = 0.05$ (third row, third column). The entry indicates that, for 8 officers, the unbiasedness restriction worsens the fit of the model by at least 5% (relative to the fit without the restriction) in 95% of the bootstrap samples. The entry to the right shows that, for 14 officers, the restriction worsens the fit by at least 5% in 90% of the bootstrap samples; and the entry above shows that, for 13 officers, the restriction worsens the fit by at least 2.5% for 95% of the bootstrap samples. I focus on the test results for $\alpha = 0.05$ and $\tau = 0.05$, and reject the null hypothesis that the officer is biased for 8 officers. The tuning parameter τ is chosen based on Monte Carlo simulations to control the size of the test.[42]

Figure 1.7 shows the relationship between the racial disparities in search and hit rates, and whether an officer is flagged as being racially biased. Positive disparities in search and hit rates indicate that minority drivers have higher search and hit rates, respectively, compared to whites. Not surprisingly, we see officers with large disparities in search rates or hit rates being flagged as racially biased. However, we also see that the test is able to detect bias even among officers with relatively similar search and hit rates across white and minority drivers. This suggests the proposed methodology is able to pick up subtleties in the data that may escape earlier tests making simple comparisons of search and hit rates across groups of

[41] The model is rejected for Officer 19 at the 2.5% significance level. This officer is not among those who fail the test for racial bias. See Appendix A.4.2 for the search and hit rates of each officer.

[42] The Monte Carlo simulations are based on an unbiased officer who has made 2,000 stops for each race of drivers with a search rate of approximately 15%, similar to the average officer in Table 1.1. With more computational time, τ may be chosen separately for each officer based on his traffic stop data.

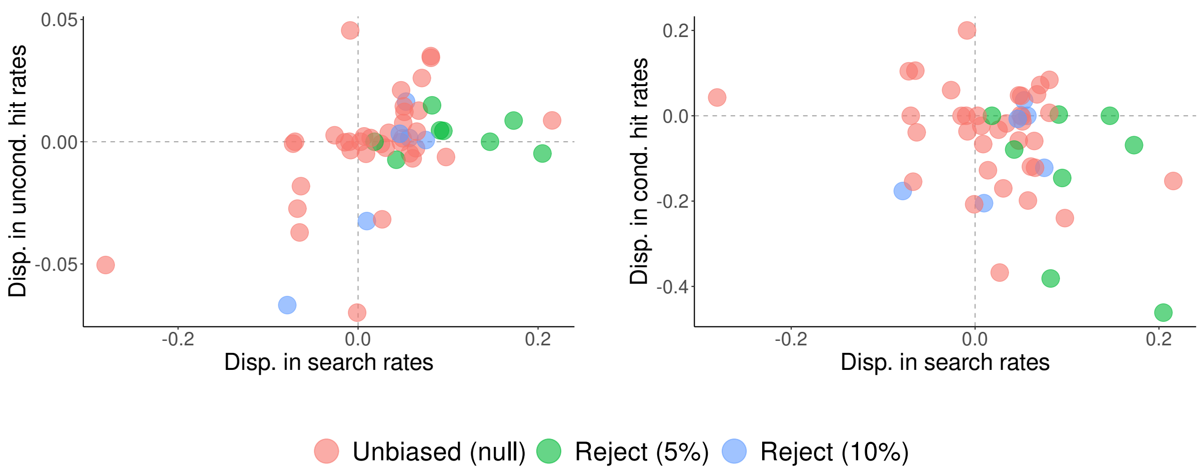
Table 1.7: Number of biased officers

	α-percentile		
$\overline{\tau}$	0.01	0.05	0.10
0.000	16	31	42
0.025	7	13	24
0.050	6	8	14
0.100	5	6	9
0.200	4	6	6

Notes: Estimates are based on 200 bootstrap samples. An officer is biased if the α-percentile of the test statistic $\widehat{\tau}$ strictly exceeds threshold $\overline{\tau}$. The counts above do not adjust for multiple hypothesis testing.

Figure 1.7: Racial disparities in search and hit rates by officer

(a) Unconditional hit rate (b) Conditional hit rate



Note: Each point corresponds to an individual officer. For each race of drivers, the search and hit rates of each officer are averaged across the different settings. Positive disparities indicate that minority drivers have higher rates compared to white drivers. Red points indicate officers for whom the null hypothesis of being unbiased is not rejected. Green (blue) points indicate officers for whom the null is rejected when $\alpha = 0.05$ (0.1).

drivers (Anwar & Fang, 2006; Knowles et al., 2001).

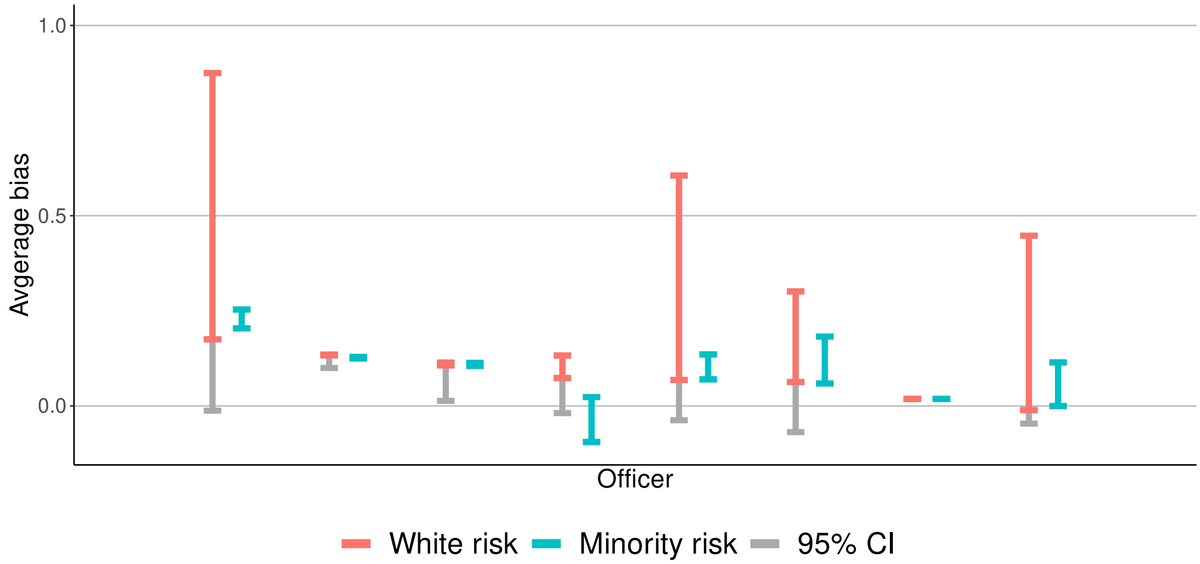
Figure 1.8 presents the estimated bounds on the average bias for the 8 officers who fail the test. The red bounds correspond to the bias being averaged over the distribution of risk of white drivers and indicate how much more white drivers would be searched if they were treated as minorities. The gray bounds correspond to the lower bound of the 95% confidence interval. Weighting the officers by the number of stops they have made for white drivers, the average of the red lower bounds is 6.2 percentage points. That is, the search rate of white drivers across these 8 officers is expected to increase by at least 6.2 percentage points if these officers treated white drivers the same way they treat equally risky minorities. This is a large difference considering these officers on average search white drivers 6.7% of the time.

The blue bounds correspond to the bias being averaged over the distribution of risk of minority drivers and indicate how much more minority drivers are being searched compared to if they were treated as whites. The corresponding confidence intervals are being computed at the time of this draft and will be included in future drafts. Weighting the officers by the number of stops they have made for minority drivers, the average of the blue lower bounds is 4.5 percentage points. That is, the search rate of minority drivers across these 8 officers is expected to decrease by at least 4.5 percentage points if these officers treated minority drivers the same way they treat equally risky white drivers. This is over a quarter of the searches performed by these officers, who on average search minority drivers 16% of the time.

Figure 1.8 also allows me to answer the question of whether minority drivers should be searched more frequently than white drivers. This can be done by subtracting the average of the blue upper bounds—8.6 percentage points—from the observed search rate of minority drivers. This difference implies that these 8 officers will search minority drivers at least 7.4% of the time on average, even if they treat minority drivers as white drivers. Since this search rate exceeds the observed search rate white drivers (6.7%), the distribution of risk for minority drivers warrants them a higher search rate than for white drivers.

The estimates in Figure 1.8 allow the officer to change their direction of bias. See Appendix A.4.1 for additional results where β is constrained to either be non-negative (anti-minority) or non-positive (anti-white). Appendix A.4.1 also includes estimates when risk

Figure 1.8: Bounds on average bias $\mathbb{E}[\beta(G_i);\omega]$ for biased officers



Note: Positive average bias indicates that the officer searches minority drivers more often than equally risky white drivers on average. Red (blue) bounds indicate the average bias when ω is equal to the distribution of risk for white (minority) drivers. Gray bounds indicate the lower bound of the 95% confidence interval.

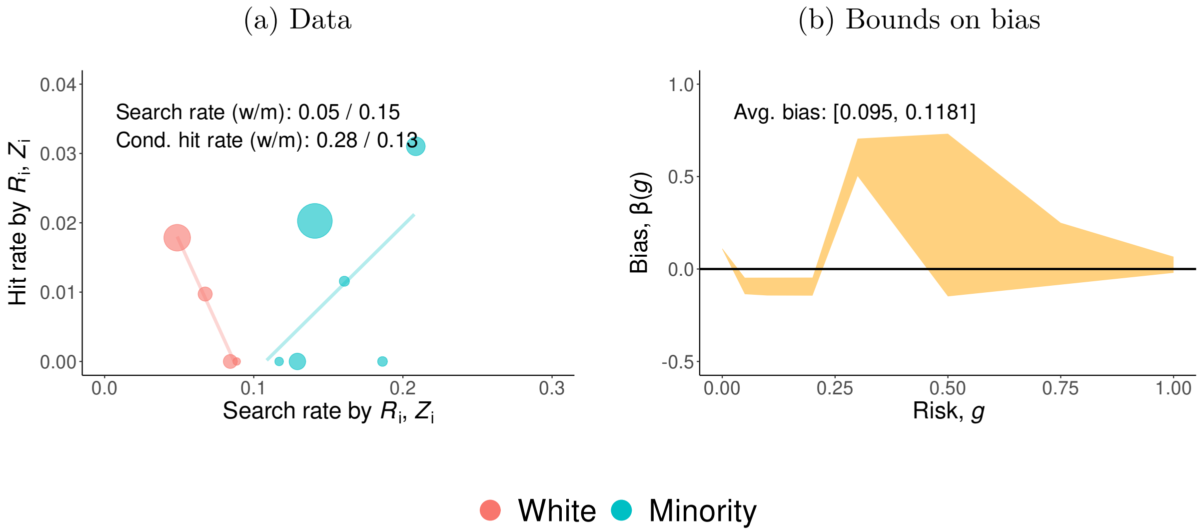
modeled as being continuous.

Figures 1.9–1.12 take a closer look at individual officers and their bias conditional on risk. To reduce computational cost, I estimate the bounds on $\beta(g_k)$ for only a subset of $g_k \in \text{supp}\,(G)$.[43] In Appendix A.4.2, I show the estimated bounds for all 50 officers. Inference will be performed on these bounds in future versions of the paper.

Figure 1.9 presents the data and bounds for officer 8, who fails the test and is estimated to change his direction of bias. For zero-risk drivers, he is biased against minorities; as risk increases to 0.05 and 0.1, he becomes biased against white drivers; as risk increases to 0.3, he is again biased against minority drivers. Once the risk exceeds 0.5, the direction of bias is unknown, but the bounds on $\beta(\cdot)$ shrink towards zero, a common pattern in the estimates (see Appendix A.4.2). Intuitively, this makes sense, as it suggests that the officer's preferences have less of an impact on the search decision as it becomes increasingly apparent that the

[43] I estimate the bounds on $\beta(g)$ for roughly every other point in $\mathbf{g}$, i.e., for g in $\{0, 0.05, 1, 0.2, 0.3, 0.5, 0.75, 1\}$.

Figure 1.9: Officer 8 switches direction of bias



driver carries contraband.

Figure 1.10 presents the data and bounds for officer 6, who also fails the test. As his hit rates are approximately zero, it is implied that both groups of drivers stopped primarily have zero risk. Yet, the officer searches minority drivers 15 percentage points more than white drivers, indicating that the officer is biased against zero-risk minority drivers. Moreover, since the data suggests both groups of drivers have similar distribution of risk, the differences in search rates must stem from differences in preferences, resulting in very tight bounds on the average bias.

Figure 1.11 shows an example where the officer appears to be biased at first glance. Specifically, the data alone shows that officer 34 is more than twice as likely to search minority drivers as white drivers, despite how he is half as likely to find contraband on minority drivers compared to white drivers. Nevertheless, he passes the test and is not flagged as biased.

Finally, Figure 1.12 presents an example where the data for white and minority drivers are similar. This suggests the identified sets of $\sigma(\cdot; w)$ and $\sigma(\cdot; m)$ are also similar, and the data may be generated using the same search preference for both groups of drivers. As expected, the officer passes the test.

Figure 1.10: Officer 6 is clearly biased against zero-risk drivers

(a) Data (b) Bounds on bias

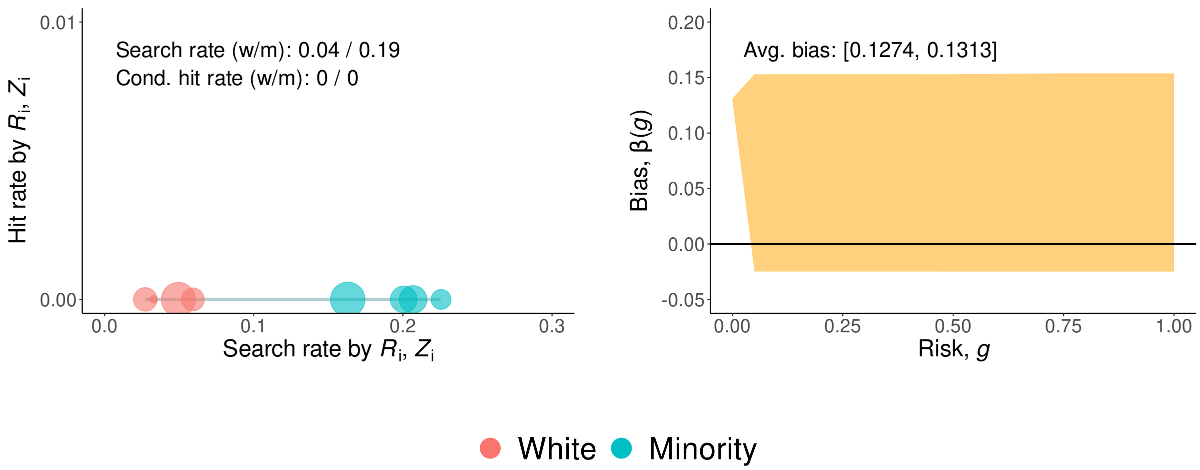
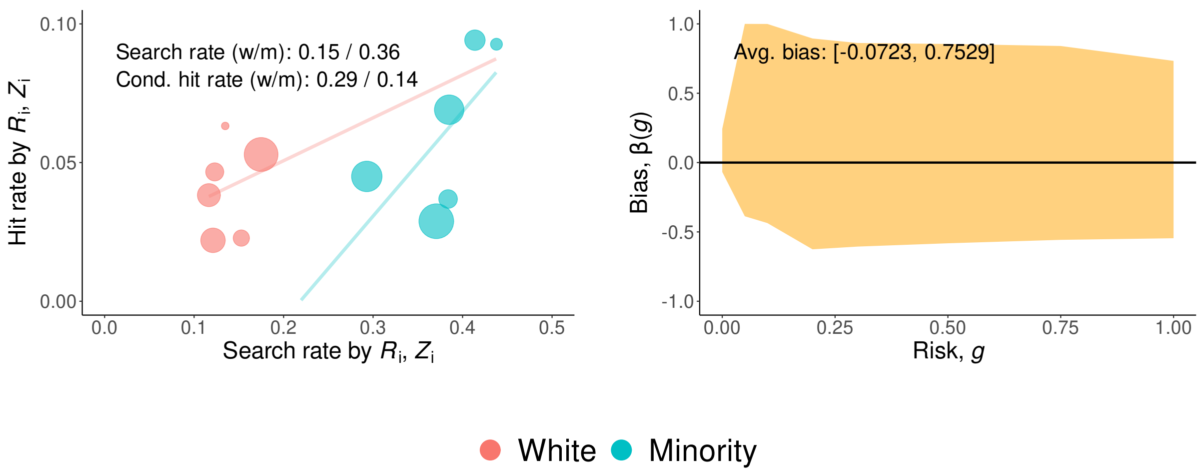


Figure 1.11: Officer 34 passes the test

(a) Data (b) Bounds on bias

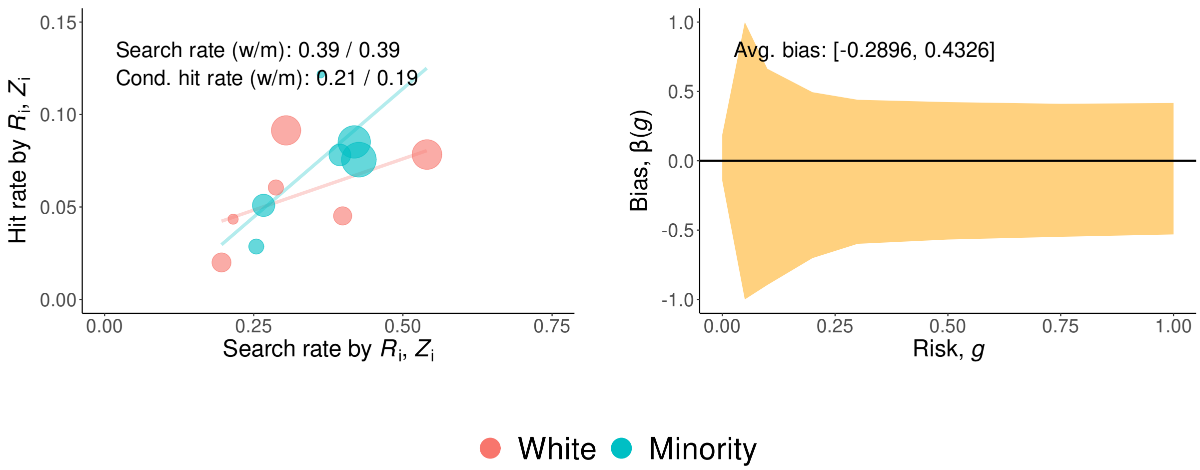


1.8 Conclusion

In this paper, I provide a flexible approach to detecting and measuring racial bias in police traffic searches. The partial identification framework enables the test to be applied even amid sample selection on unobservables and statistical discrimination. In addition, by using an IV to vary the risk among drivers stopped, the methods I propose may be applied to individual officers, allowing for unrestricted heterogeneity in preferences and beliefs across officers.

Figure 1.12: Officer 49 passes the test

(a) Data (b) Bounds on bias



This paper also contributes to the literature from a modeling standpoint, as earlier papers studying racial bias have either assumed or required choice models with deterministic thresholds, whereas I allow the threshold to be random. This relaxation permits a richer notion of bias, where the direction and intensity of bias may depend on the unobserved (to the researcher) risk of the driver. Moreover, sharp bounds on these measures immediately follow from the econometric model. Additional restrictions to tighten these bounds, as well as strengthen the test, may be imposed in a transparent and modular fashion.

Implementing these methods involves solving several bilinear programs, which is novel in the literature on discrimination. There is commercial software freely available to academic institutions capable of solving these problems to global optimality, making it feasible to estimate the sharp bounds discussed. Bilinear programs also have the potential to be used more generally to study mixture models, and is a possible area of future research. Another topic that requires further study is statistical inference for bilinear programs, for which there is currently no formal procedure.

I apply the proposed methods on police traffic data from the Metropolitan Nashville Police Department, and find evidence to suggest 8 officers are biased. The estimates also suggest that officers are more likely to be biased against low-risk minority drivers, and the bias disappears as the risk of the driver increases. A convenient feature of these methods is that they may be performed using fairly standard police traffic data sets. The assumptions of the model are better satisfied when the police data are supplemented with local demographic data, such as household incomes and crime rates, and such data is often public or available upon request. So a natural extension of the paper is to apply these methods to other police data sets from across the US.

Another avenue for future research is to extend these methods to study bias in traffic stops. Although the framework in this paper was intended to circumvent the challenges that bias in traffic stops imposed on measuring bias in traffic searches, it would be interesting to decompose the *total* effect of bias on searches into the biases that occur before and after a driver is stopped. There are now commercial data sets on driver demographics collected by tracking smartphones in vehicles, and the availability of such data provides an opportunity to tackle this long-standing question in new ways.

CHAPTER 2

IVMTE: AN R PACKAGE FOR IMPLEMENTING MARGINAL TREATMENT EFFECT METHODS

COAUTHORED WITH ALEXANDER TORGOVITSKY*

2.1 Introduction

A central task in many empirical fields is to determine the effect (the *causal* effect) of one variable on another. The task is often complicated by the fact that the effecting variable (the treatment) is not only not randomly assigned, it is *chosen* by an agent with information unavailable to the researcher. For example, in the application discussed later, the treatment is the number of children a family decides to have, and the empirical question is the effect that bearing more children has on the mother's labor force participation. Since having a child and working are joint decisions a family makes using their own private information, strategies such as propensity score matching are unlikely to eliminate systematic unobserved differences between families with more children and those with fewer children. Different empirical strategies are needed to credibly identify a causal effect.

Instrumental variables (IVs) are one extremely popular strategy (e.g. Baiocchi, Cheng, & Small, 2014; Bollen, 2012; J. J. Heckman & Robb, 1985; Imbens, 2014). An IV (or instrument) is an observed variable that is correlated with the treatment variable, but uncorrelated with confounding unobservable differences. A well-known example of an instrument for fertility is the same-sex instrument introduced by Angrist and Evans (1998). This instrument is a binary variable that is equal to 1 if a family's first two children had the same sex (female-female or male-male) and is 0 otherwise.[1] The key assumption of an IV model is that the sex of the second child is as good as randomly assigned—and therefore independent of any confounding

*Kenneth C. Griffin Department of Economics, University of Chicago

[1]Using the same-sex instrument requires restricting the analysis to families with two or more children.

unobservable differences across families—while still impacting a family's decision to have a third child due to a preference for having both a male and female child. The intuition is that by comparing the labor supply decisions of families whose first two children were the same sex to families whose children were mixed sex, one picks up only the differences that are caused by the decision to have additional children.

While IV strategies have been widely studied and applied for many decades (see Stock & Trebbi, 2003, for a history), it wasn't until the 1980s that researchers started to focus on IV models with unobserved heterogeneity in treatment effects.[2] In an influential paper, Imbens and Angrist (1994) provided nonparametric conditions under which a simple linear IV estimator can be interpreted as estimating the average causal effect (the "local average treatment effect," or LATE) among a subpopulation described as the compliers. The compliers are the individuals whose treatment choice would have been different had their instrument been different. In the fertility application, they are the families who would have had a third child if and only if their first two children had the same sex.

An important implication is that the interpretation of a linear IV estimator depends on the instrument used. If there is unobserved treatment effect heterogeneity, then linear IV estimators cannot in general be interpreted as providing estimates of conventional parameters such as the average treatment effect (ATE) or the average treatment effect on the treated (ATT). One response to this finding is to continue to use linear IV estimators and change the research question to serve the definition of the complier group, a practice espoused by Angrist and Krueger (2001) and Angrist and Pischke (2009, 2010). A number of authors in multiple disciplines have criticized this practice (e.g. Deaton, 2010; J. Heckman, 1997; Pearl, 2011; Robins & Greenland, 1996; Swanson & Hernán, 2014, among many others). Another response is to change the estimator and extrapolate from the compliers to the subpopulation that better answers the researchers' empirical question (see Mogstad & Torgovitsky, 2018,

[2] An early example is J. J. Heckman (1976). See also J. J. Heckman and Robb (1985), Björklund and Moffitt (1987), and Manski (1990). Taking a more expansive view of unobserved heterogeneity in "treatment effects" to also encompass random coefficient models, one can trace interest back to the Cowles Foundation (Hurwicz, 1950; Rubin, 1950) as well as foundational economic analyses like Becker and Chiswick (1966).

for a discussion of different approaches).

In a series of papers, J. J. Heckman and Vytlacil (2005b, 1999, 2007a, 2007b) developed the concept of the marginal treatment effect (MTE) and showed how it can be used to nonparametrically model this type of extrapolation under the same "monotonicity" condition used by Imbens and Angrist (1994). Carneiro, Heckman, and Vytlacil (2011) and Brinch, Mogstad, and Wiswall (2012, 2017) showed how to apply their idea to identify and estimate semiparametric MTE models. The method is now widely applied in empirical economics, see e.g. Maestas, Mullen, and Strand (2013), Carneiro, Lokshin, and Umapathi (2016), and Bhuller, Dahl, Løken, and Mogstad (2020). Mogstad, Santos, and Torgovitsky (2018b) extended these approaches to provide a general moment-based framework that also accommodates partial identification (bounds) in cases when the researcher's assumptions are not strong enough (or the data is not rich enough) to pin down a unique conclusion.

In this paper, we expand on the moment-based framework in Mogstad et al. (2018b) by considering a related regression-based framework. Instead of focusing on fitting specific moments, the regression framework minimizes a least squares criterion. As we discuss, this has both benefits and drawbacks that depend on the researcher's empirical setting and goals. We then describe the `R` package `ivmte`, which can be used to implement both the moment and regression frameworks. The package provides a flexible environment for using IV methods for rigorous policy evaluation in the presence of unobserved heterogeneity.

2.2 Model and identification

2.2.1 Potential outcomes and choices

The model is about the impact of a binary treatment $D_i \in \{0, 1\}$ on individual i's observed outcome variable, Y_i. Let $Y_i(0)$ and $Y_i(1)$ denote the unobserved potential outcomes for Y_i if individual i had received $D_i = 0$ or 1, respectively, so that $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$. The researcher is interested in features of the distribution of the causal effect, $Y_i(1) -$

$Y_i(0)$. The researcher has access to some observable covariates, X_i, but they are concerned that D_i is still dependent with $Y_i(0)$ or $Y_i(1)$ even after conditioning on X_i, so that the unconfoundedness (selection on observables) assumption (e.g. Barnow, Cain, Goldberger, et al. (1980), Rosenbaum and Rubin (1983), J. J. Heckman, Ichimura, Smith, and Todd (1996)) does not hold.

However, the researcher also has access to an instrumental variable, Z_i. The instrument influences individual i's treatment choice with $D_i(z)$ denoting their unobserved potential treatment choice if Z_i were set to z. Their observed treatment choice is related to these potential choices via $D_i = \sum_{z \in \mathcal{Z}} \mathbb{1}\left[Z_i = z\right] D_i(z)$, where $\mathcal{Z}$ is the support of Z_i. In contrast to D_i, the instrument is assumed to be as good as randomly assigned, conditional on X_i, in the sense that Z_i is independent of $(Y_i(0), Y_i(1), \{D_i(z)\}_{z \in \mathcal{Z}})$, conditional on X_i.

2.2.2 The selection model

Imbens and Angrist (1994) introduced an additional assumption that they described as monotonicity. The monotonicity assumption says that for any pair of instrument values z and z', either $D_i(z) \geq D_i(z')$ for all individuals i, or else $D_i(z') \geq D_i(z)$ for all individuals i. That is, a shift from z to z' either pushes every individual towards treatment, or else pushes every individual away from treatment. Whichever direction holds, the monotonicity condition maintains that there are no individuals who deviate from this ordering, a requirement sometimes described as "no defiers."[3]

Vytlacil (2002) showed that the monotonicity condition is equivalent to the latent variable selection model

$$D_i = \mathbb{1}\left[U_i \leq p(X_i, Z_i)\right], \tag{2.1}$$

[3]Despite the name "monotonicity," the condition would be more accurately described as "uniformity," since it restricts heterogeneity in how the instrument impacts treatment choice (J. J. Heckman, Urzua, & Vytlacil, 2006).

where U_i is a continuously distributed unobserved random variable, and $p(x, z) \equiv \mathbb{P}[D_i = 1|X_i = x, Z_i = z]$ is the propensity score. The latent variable U_i is independent of Z_i, conditional on X_i, and customarily normalized to be uniformly distributed on $[0, 1]$.[4] It can be interpreted as the individual i's rank (quantile) of latent willingness to choose $D_i = 1$, with smaller values of U_i corresponding to more-willing individuals. When D_i is a variable chosen by an agent, such as in the fertility example, we expect that U_i will be dependent with $Y_i(0)$ and $Y_i(1)$ if these potential outcomes themselves either directly influence the agent's choice or are correlated with other factors that do.

2.2.3 Marginal treatment response and effect functions

The advantage of the latent variable model (2.1) is that it facilitates modeling unobserved heterogeneity in the effect of D_i on Y_i. The key object for this purpose is the marginal treatment response (MTR) function

$$m(d|u, x) \equiv \mathbb{E}[Y_i(d)|U_i = u, X_i = x] \tag{2.2}$$

The MTR function describes how expected treated and untreated outcomes vary conditional on both observed covariates, X_i, and the unobserved latent propensity to take treatment, U_i. The marginal treatment effect (MTE) of J. J. Heckman and Vytlacil (2005b, 1999, 2007a, 2007b) is the difference of the MTR function between treatment states: $m(1|u, x) - m(0|u, x)$. For example, if the MTE is declining in u, then individuals who are less likely to choose treatment (larger u) would experience smaller treatment effects than those who are more likely to choose treatment. Thus, the MTE captures the idea of selection on *unobservables*, where the unobservable in question is an individual's latent propensity to take treatment, U_i.

[4] See J. J. Heckman and Vytlacil (2005b), Matzkin (2007), or Mogstad and Torgovitsky (2018) for a detailed discussion of the normalization.

2.2.4 Target parameters

Many treatment effect parameters can be written as weighted averages of the MTR functions. For example, the average treatment effect (ATE) can be written as

$$\begin{aligned}\underbrace{\mathbb{E}[Y_i(1) - Y_i(0)]}_{\equiv \text{ATE}} &= \mathbb{E}\left[\mathbb{E}[Y_i(1)|U_i, X_i] - \mathbb{E}[Y_i(0)|U_i, X_i]\right] \\ &= \mathbb{E}[m(1|U_i, X_i) - m(0|U_i, X_i)] = \mathbb{E}\left[\int_0^1 m(1|u, X_i) - m(0|u, X_i)\, du\right],\end{aligned} \tag{2.3}$$

where the final equality used the normalization on the distribution of U_i to be uniform and independent of X_i. Similarly, the average treatment effect on the treated (ATT) can be written as

$$\underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]}_{\equiv \text{ATT}} = \mathbb{E}\left[\int_0^1 (m(1|u, X_i) - m(0|u, X_i)) \times \frac{\mathbb{1}\left[u \leq p(X_i, Z_i)\right]}{\mathbb{P}[D_i = 1]}\, du\right], \tag{2.4}$$

see e.g. J. J. Heckman and Vytlacil (2005b). As in Mogstad et al. (2018b), we view both (2.3) and (2.4) as examples of *target parameters* τ with the general form

$$\tau(m) \equiv \sum_{d \in \{0,1\}} \mathbb{E}\left[\int_0^1 m(d|u, X_i)\omega_\tau(d|u, X_i, Z_i)\, du\right], \tag{2.5}$$

where ω_τ is a weighting function that is either known to the researcher (as in (2.3)) or point identified from the distribution of (D_i, X_i, Z_i) (as in (2.4)). J. J. Heckman and Vytlacil (2005b) and Mogstad et al. (2018b) provide extensive discussions and many examples of target parameters, along with their weighting functions, ω_τ.

2.2.5 Implied observable quantities

The model implies a relationship between the MTR function and moments of the observed outcome, Y_i. In particular, Mogstad et al. (2018b, Proposition 1) show that for any (measurable) function s of (D_i, X_i, Z_i)

$$\begin{aligned}\mathbb{E}[Y_i s(D_i, X_i, Z_i)] = \quad & \mathbb{E}\left[s(0, X_i, Z_i) \int_0^1 m(0|u, X_i)\mathbb{1}[u > p(X_i, Z_i)]\, du\right] \\ & + \mathbb{E}\left[s(1, X_i, Z_i) \int_0^1 m(1|u, X_i)\mathbb{1}[u \leq p(X_i, Z_i)]\, du\right] \equiv \gamma_s(m). \quad (2.6)\end{aligned}$$

A similar expression can be derived for the conditional moments of Y_i:

$$\begin{aligned}&\mathbb{E}[Y_i|D_i = 1, X_i = x, Z_i = z] \\ &\quad = \mathbb{E}[Y_i(1)|U_i \leq p(x, z), X_i = x] = \int_0^1 m(1|u, x)\frac{\mathbb{1}[u \leq p(x, z)]}{p(x, z)}\, du,\end{aligned}$$

and, symmetrically,

$$\mathbb{E}[Y_i|D_i = 0, X_i = x, Z_i = z] = \int_0^1 m(0|u, x)\frac{\mathbb{1}[u > p(x, z)]}{(1 - p(x, z))},\ du.$$

We combine the right-hand side of these two relationships using the notation

$$\gamma_{\text{cm}}(m|d, x, z) = \int_0^1 m(0|u, x)(1 - d)\frac{\mathbb{1}[u > p(x, z)]}{(1 - p(x, z))} + m(1|u, x)d\frac{\mathbb{1}[u \leq p(x, z)]}{p(x, z)}\, du. \quad (2.7)$$

2.2.6 Identification

We use expressions (2.6) and (2.7) to define two identified sets for the target parameter, τ. To do this, we first define identified sets for the MTR function. We assume that m lives in some set $\mathcal{M}$ contained in a vector space, where $\mathcal{M}$ encodes any additional assumptions we might want to place on m, such as parameterizations or shape restrictions.

One identified set matches a collection of unconditional moments (2.6) formed by a collection of functions $s \in \mathcal{S}$:

$$\mathcal{M}^{\star}_{\mathcal{S}} = \{m \in \mathcal{M} : \gamma_s(m) = \mathbb{E}[Y_i s(D_i, X_i, Z_i)] \text{ for all } s \in \mathcal{S}\}. \quad (2.8)$$

The moment approach is based on $\mathcal{M}^\star_{\mathcal{S}}$. Another identified set matches the conditional mean of the observed outcome:

$$\mathcal{M}^\star_{\text{cm}} = \{m \in \mathcal{M} : \gamma_{\text{cm}}(m|d,x,z) = \mathbb{E}[Y_i|D_i = d, X_i = x, Z_i = z] \text{ for almost every } d,x,z\}\,. \tag{2.9}$$

The regression approach is based on $\mathcal{M}^\star_{\text{cm}}$. While $\mathcal{M}^\star_{\text{cm}} \subseteq \mathcal{M}^\star_{\mathcal{S}}$ for any choice of $\mathcal{S}$, there are some practical and conceptual considerations that may nevertheless favor the moment approach (see Section 2.3.4).

Our object of interest is not an identified set for the MTR function, but rather an identified set for the target parameter, τ. An identified set for the target parameter can be formed by taking the image of either $\mathcal{M}^\star_{\mathcal{S}}$ or $\mathcal{M}^\star_{\text{cm}}$ under τ:

$$\mathcal{T}^\star_{\mathcal{S}} \equiv \{\tau(m) : m \in \mathcal{M}^\star_{\mathcal{S}}\} \quad \text{and} \quad \mathcal{T}_{\text{cm}} \equiv \{\tau(m) : m \in \mathcal{M}^\star_{\text{cm}}\}\,. \tag{2.10}$$

The set $\mathcal{T}^\star_{\mathcal{S}}$ gives the values of the target parameter that are consistent with the assumptions of the model and the unconditional cross-moments (2.6) for $s \in \mathcal{S}$. The set $\mathcal{T}^\star_{\text{cm}}$ is the subset of $\mathcal{T}^\star_{\mathcal{S}}$ that is consistent with the entire conditional mean of the observed outcome.

2.2.7 Point identification vs. partial identification

The formulation in the previous section allows the identified sets $\mathcal{M}^\star_{\mathcal{S}}, \mathcal{M}^\star_{\text{cm}}$, $\mathcal{T}^\star_{\mathcal{S}}$, and $\mathcal{T}^\star_{\text{cm}}$ to be either singletons or proper non-singleton sets. In the first case, we say that m or τ is point identified, while in the second case we say that they are partially identified. Point identification of m implies point identification of τ. When τ is not point identified, its identified sets $\mathcal{T}^\star_{\mathcal{S}}$ and $\mathcal{T}^\star_{\text{cm}}$ will still be closed intervals under weak conditions (see Mogstad et al., 2018b, for a precise statement). One can thus describe the partial identification case as providing bounds on the target parameter. It is also possible for the identified sets to be empty, in which case the model is said to be misspecified.

The size and cardinality of the identified sets depend on a few factors. Having a smaller parameter space $\mathcal{M}$—that is, maintaining more restrictive assumptions—mechanically shrinks the identified sets. Making $\mathcal{S}$ a larger set of functions also mechanically shrinks the moment-based identified set. The number of distinct functions one can potentially include in $\mathcal{S}$ is determined by the supports of Z_i and X_i. Richer supports of Z_i allow for smaller identified sets and thus tighter conclusions; richer supports of X_i can also be helpful if $\mathcal{M}$ is such that $m(d|u, x)$ depends on x in a restricted way. These richer supports get automatically incorporated into the regression-based identified set, so that it necessarily shrinks with additional support points.

2.2.8 Criterion functions

For implementation, it is useful to have a scalar function that determines if a candidate MTR function m is in either $\mathcal{M}^\star_{\mathcal{S}}$ or $\mathcal{M}^\star_{\text{cm}}$. For the moment approach, we let $c_s \equiv \mathbb{E}[Y_i s(D_i, X_i, Z_i)]$, and stack both c_s and $\gamma_s(m)$ across $s \in \mathcal{S}$ into vectors $c_{\mathcal{S}}$ and $\gamma_{\mathcal{S}}(m)$. Define

$$Q_{\mathcal{S}}(m) = \left\| \gamma_{\mathcal{S}}(m) - c_{\mathcal{S}} \right\|, \tag{2.11}$$

for some choice of norm $\|\cdot\|$. Then $m \in \mathcal{M}^\star_{\mathcal{S}}$ if and only if $Q_{\mathcal{S}}(m) = 0$, so that

$$\mathcal{T}^\star_{\mathcal{S}} = \left\{ \tau(m) : Q_{\mathcal{S}}(m) = 0 \right\}.$$

For the regression approach, we define the least squares criterion:

$$Q_{\text{cm}}(m) \equiv \mathbb{E}\left[(Y_i - \gamma_{\text{cm}}(m|D_i, X_i, Z_i))^2 \right]. \tag{2.12}$$

Then $m \in \mathcal{M}^\star_{\text{cm}}$ if and only if $m \in \arg\min_{m' \in \mathcal{M}} Q_{\text{cm}}(m')$, so that

$$\mathcal{T}^\star_{\text{cm}} = \left\{ \tau(m) : Q_{\text{cm}}(m) = \min_{m' \in \mathcal{M}} Q_{\text{cm}}(m') \right\}.$$

2.3 Estimation and computation

2.3.1 Linear basis representation

Implementation requires evaluating the functions τ and γ_s or γ_{cm} at candidate choices of the MTR functions. Some dimension reduction is needed for computation. In particular, let

$$m_\theta(d|u,x) = \sum_{k=1}^{K} \theta_k b_k(d|u,x) \quad \text{for some } \theta \in \mathbb{R}^K, \tag{2.13}$$

where b_k are known basis functions and θ_k are unknown parameters. Then we assume that the parameter space is $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$ for some subset Θ of $\mathbb{R}^K$. That is, the MTR function is assumed to be a member of the class of functions formed by taking linear combinations of the basis functions. This reduces the dimension of the function m to a K–dimensional real vector θ.

Linear-in-parameters specifications like (2.13) are commonplace in statistical models. For example, if x is scalar, one could specify

$$\begin{aligned} m_\theta(d|u,x) = (1-d)\,&\overbrace{(\theta_1 + \theta_2 u + \theta_3 x + \theta_4 ux)}^{m_\theta(0|u,x)} \\ &+ d\underbrace{\left(\theta_5 + \theta_6 u + \theta_7 u^2 + \theta_8 x + \theta_9 x^2\right)}_{m_\theta(1|u,x)} \end{aligned} \tag{2.14}$$

which corresponds to $K = 9$ parameters with basis functions (e.g.) $b_3(d|u,x) = (1-d)x$ and $b_7(d|u,x) = du^2$. The assumption used in the `ivmte` package is that $\mathcal{M}$ contains only MTR functions such that both $m_\theta(0|u,x)$ and $m_\theta(1|u,x)$ are either polynomials or polynomial B–splines in u. This is certainly a special case of (2.13), but one that is popular both as a parametric restriction (e.g. a polynomial, like (2.14)) and as an approximating basis for nonparametric sieve estimation (e.g. Chen, 2007).

2.3.2 Sample analogs

The benefit of using the linear-in-parameters specification (2.13) is that it preserves the linearity of τ, γ_s, and γ_{cm} as functions of m. In particular,

$$\tau(m_\theta) = \theta' T, \tag{2.15}$$

where T is a K–dimensional vector with kth element

$$\tau(b_k) = \sum_{d \in \{0,1\}} \mathbb{E}\left[\int_0^1 b_k(d|u, X_i)\omega_\tau(d|u, X_i, Z_i)\, du\right]. \tag{2.16}$$

Given a sample of data $\{(Y_i, D_i, X_i, Z_i)\}_{i=1}^n$, each component of T can be estimated by its sample analog

$$\hat{\tau}(b_k) \equiv \frac{1}{n}\sum_{i=1}^n \sum_{d \in \{0,1\}} \int_0^1 b_k(d|u, X_i)\hat{\omega}_\tau(d|u, X_i, Z_i)\, du, \tag{2.17}$$

where $\hat{\omega}_\tau$ is an estimate of ω_τ. Requiring $m_\theta(d|u, x)$ to be a polynomial or B–spline in u means that the integral in $\hat{\tau}(b_k)$ can be computed analytically as long as $\hat{\omega}_\tau(u, x, z)$ has a tractable form, which it does for all conventional target parameters.[5] By the same reasoning,

$$\gamma_s(m_\theta) = \theta' \Gamma_s \quad \text{and} \quad \gamma_{\text{cm}}(m_\theta|d, x, z) = \theta' \Gamma_{\text{cm}}(d, x, z),$$

where Γ_s and $\Gamma_{\text{cm}}(d, x, z)$ are K–dimensional vectors with kth elements given by $\gamma_s(b_k)$ and $\gamma_{\text{cm}}(b_k|d, x, z)$. A sample analog estimator of $\gamma_s(b_k)$ is

$$\begin{aligned}\hat{\gamma}_s(b_k) \equiv \quad & \frac{1}{n}\sum_{i=1}^n \hat{s}(0, X_i, Z_i)\int_0^1 b_k(0|u, X_i)\mathbb{1}[u > \hat{p}(X_i, Z_i)]\, du \\ & + \frac{1}{n}\sum_{i=1}^n \hat{s}(1, X_i, Z_i)\int_0^1 b_k(1|u, X_i)\mathbb{1}[u \le \hat{p}(X_i, Z_i)]\, du,\end{aligned}$$

[5] `ivmte` allows for any target parameter for which $\hat{\omega}_\tau(u, x, z)$ can be written as a constant spline in u—see Section 2.4.3.

where $\hat{s}$ is an estimator of s, and $\hat{p}$ is an estimator of p. A sample analog estimator of $\gamma_{\text{cm}}(b_k|d,x,z)$ is

$$\hat{\gamma}_{\text{cm}}(b_k|d,x,z) = \int_0^1 b_k(0|u,x)(1-d)\frac{\mathbb{1}[u > \hat{p}(x,z)]}{(1-\hat{p}(x,z))} + b_k(1|u,x)d\frac{\mathbb{1}[u \leq \hat{p}(x,z)]}{\hat{p}(x,z)}\, du.$$

We use these sample analogs to define sample criterion functions. The moment-based sample criterion is

$$\hat{Q}_{\mathcal{S}}(m_\theta) \equiv \left\| \hat{\Gamma}_{\mathcal{S}}\theta - \hat{c}_{\mathcal{S}} \right\|,$$

where $\hat{\Gamma}_{\mathcal{S}}$ is an $|\mathcal{S}| \times K$ matrix with rows $\hat{\Gamma}'_s \equiv [\hat{\gamma}_s(b_1), \ldots, \hat{\gamma}_s(b_K)]$, and $\hat{c}_{\mathcal{S}}$ is a vector with elements

$$\hat{c}_s \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i \hat{s}(D_i, X_i, Z_i).$$

The regression-based sample criterion is

$$\hat{Q}_{\text{cm}}(m_\theta) \equiv \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \theta'\hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)\right)^2,$$

where $\hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)$ is a K–dimensional vector with kth element $\hat{\gamma}_{\text{cm}}(b_k|D_i, X_i, Z_i)$.

2.3.3 Estimation

Estimation differs for point and partially identified cases.

Point identification

In the point identified case, we assume that the parameter space is $\Theta = \mathbb{R}^K$. This simplification allows for closed-form estimators.

For the moment criterion, we use the generalized method of moments (Hansen, 1982, "GMM") estimator of θ:

$$\hat{\theta} = \text{argmin}_{\theta \in \mathbb{R}^K} \left(\hat{\Gamma}_{\mathcal{S}}\theta - \hat{c}_{\mathcal{S}}\right)' \hat{\Omega} \left(\hat{\Gamma}_{\mathcal{S}}\theta - \hat{c}_{\mathcal{S}}\right), \tag{2.18}$$

where $\hat{\Omega}$ is a positive semi-definite weighting matrix. The minimizer $\hat{\theta}$ of (2.18) is the minimizer of $\hat{Q}_{\mathcal{S}}$ when $\|\cdot\|$ is taken to be the Euclidean norm weighted by $\hat{\Omega}$. In a point identified case, one would expect that (2.18) has a unique solution, in which case it can be solved for analytically.

The regression sample criterion $\hat{Q}_{\text{cm}}$ is simply the ordinary least squares criterion for a linear regression of Y_i onto the vector of generated regressors $\hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)$. It has a unique minimizer if and only if the matrix

$$\sum_{i=1}^{n} \hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)\hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)' \tag{2.19}$$

is invertible. If the researcher believes that point identification holds, then a simple estimator $\hat{\theta}$ of θ is the ordinary least squares estimator from a regression of Y_i on $\hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)$.

For both the moment and regression approaches, we then set

$$\hat{\tau}^{\star} \equiv \hat{\theta}'\hat{T} \tag{2.20}$$

where $\hat{T}$ is the K–dimensional vector with kth element $\hat{\tau}(b_k)$. Then $\hat{\tau}^{\star}$ is our point estimator of the (assumed singleton) identified set for the target parameter, $\mathcal{T}^{\star}_{\mathcal{S}}$ or $\mathcal{T}^{\star}_{\text{cm}}$, depending on the criterion function used.

Partial identification

For partially identified cases we use a two-step estimator developed by Mogstad et al. (2018b) for both the moment and regression approaches. In the first step, we minimize the criterion function to find

$$\hat{Q}^{\star} \equiv \min_{\theta \in \Theta} \hat{Q}(m_\theta), \tag{2.21}$$

where $\hat{Q}$ could be either $\hat{Q}_{\mathcal{S}}$ or $\hat{Q}_{\text{cm}}$, and now the parameter space Θ is allowed to be a proper subset of $\mathbb{R}^K$. In the second step, we then minimize and maximize the target parameter over the set of $\theta \in \Theta$ that produce sample criteria close to the best possible value, $\hat{Q}^{\star}$. That is, we solve for

$$\hat{\tau}^{\star}_{\text{lb}}/\hat{\tau}^{\star}_{\text{ub}} \equiv \min/\max_{\theta \in \Theta} \theta' \hat{T} \qquad \text{subject to} \quad \hat{Q}(m_\theta) \leq (1+\sigma)\hat{Q}^{\star}, \tag{2.22}$$

where $\sigma \geq 0$ is a tuning parameter used in the asymptotic theory (see Mogstad et al., 2018b, for more detail). The feasible set in (2.22) is always non-empty due to the definition of $\hat{Q}^{\star}$, so that both $\hat{\tau}^{\star}_{\text{lb}}$ and $\hat{\tau}^{\star}_{\text{ub}}$ are always well-defined. Mogstad et al. (2018b) provide conditions under which $[\hat{\tau}^{\star}_{\text{lb}}, \hat{\tau}^{\star}_{\text{ub}}]$ is a consistent set estimator of $\mathcal{T}^{\star}_{\mathcal{S}}$ if $\hat{Q} = \hat{Q}_{\mathcal{S}}$, and of $\mathcal{T}^{\star}_{\text{cm}}$ if $\hat{Q} = \hat{Q}_{\text{cm}}$.

To facilitate computation, we assume that the constraint set Θ can be written in terms of linear inequality constraints:

$$\Theta = \left\{ \theta \in \mathbb{R}^K : r_{\text{lb}} \leq R\theta \leq r_{\text{ub}} \right\}, \tag{2.23}$$

for vectors $r_{\text{lb}}, r_{\text{ub}}$, and a conformable matrix R. In practice, these constraints typically represent bounds on levels and/or derivatives of $m(0|\cdot, x)$ and $m(1|\cdot, x)$ and/or $m(1|\cdot, x) - m(0|\cdot, x)$ on a large grid of evaluation points. We discuss shape constraints in more detail in Sections 2.4.5 and 2.5.

For the moment approach, the structure of the first and second step programs depends on the choice of norm $\|\cdot\|$. In the `ivmte` module, we take $\|\cdot\|$ to be the ℓ_1 norm so that

$$\hat{Q}_{\mathcal{S}}(m_\theta) = \sum_{s \in \mathcal{S}} \left| \hat{\Gamma}_s \theta - \hat{c}_s \right|. \tag{2.24}$$

This choice is attractive given (2.23) because one can then reformulate the first and second step problems (2.21) and (2.22) as linear programs by replacing absolute values with appropriate slack variables. Linear programs scale quite well with the number of parameters and constraints.

Given (2.23), the program defining $\hat{Q}^\star$ in the regression approach is a convex quadratic program. The second step programs in (2.22) are convex quadratically-constrained quadratic programs (QCQPs). Mature algorithms exist for solving both types of programs to global optimality. As one might expect, QCQPs do not tend to scale as well as LPs, and can be more sensitive to numerical issues.

2.3.4 Trade-offs between the moment and regression approaches

As already noted, the identified set for the regression approach is always weakly smaller than in the moment approach: $\mathcal{T}^\star_{\text{cm}} \subseteq \mathcal{T}^\star_{\mathcal{S}}$. Not only that, but the researcher does not need to specify the set $\mathcal{S}$, as they would in the moment approach. In point identified cases, the regression approach has the additional benefit of being implementable through ordinary least squares, which is computationally trivial and can be expected to have good statistical properties. These are certainly strong points in favor of the regression approach.

There are, however, also some benefits to the moment approach. Being able to choose the set of moments $\mathcal{S}$ that are fit can be attractive since it draws a clear line between the portions of the observed data that are used in inference and the portions that are not. For example, Mogstad and Torgovitsky (2018) suggest reporting common linear IV model estimates such as various two-stage least squares specifications—which do not in general estimate an interesting target parameter—together with bounds on the target parameter that incorporate the same information by using the same linear IV estimands as functions in $\mathcal{S}$. The moment-based criterion can also be easier to interpret; it simply measures the distance to satisfying the moments, so if the number of moments is small and the MTR function flexible, it can be exactly zero indicating that all the moments can be reproduced. The other primary benefit of the moment approach is computation in partially identified cases, where it produces an LP implementation that can usually be expected to be easier to compute than the QCQPs required in the regression approach.

2.4 The `ivmte` package

2.4.1 Installation and requirements

The `ivmte` package is available in CRAN, and can be installed and loaded as usual:

```
install.packages("ivmte")
library("ivmte")
```

The most up-to-date version can be installed directly from the GitHub repository:

```
devtools::install_github("jkcshea/ivmte")
```

No additional packages are required for implementing the point estimators discussed in Section 2.3.3.

For the partially identified cases, `ivmte` requires a solver package. If using the moment approach, the options are `gurobi`, `Rmosek`, `cplexAPI` (Roettger, Gelius-Dietrich, & Fritzemeier, 2019), or `lpSolveAPI` (Konis, 2019). The first package requires a Gurobi (Gurobi Optimization, Inc., 2015) license, the second requires a MOSEK (MOSEK ApS, 2021) license, while the third requires a CPLEX (IBM, 2010) license. These are available at no cost for academic researchers. Alternatively, `lpSolveAPI` is freely available through CRAN and does not require a license. For the regression approach with partial identification, `ivmte` requires either `gurobi` or `Rmosek`, since the other solvers cannot solve QCQPs.[6]

All of the examples shown ahead in Section 2.5 were computed using `gurobi`.

[6] CPLEX can solve QCQPs, but its `R` API does not appear to allow for it.

2.4.2 Basic syntax

The main command in `ivmte` is called `ivmte`. It requires the following arguments

```
ivmte(data, target, m0, m1, ivlike, propensity)
```

where `data` is the usual `dataframe` and

- `target` specifies the target parameter, τ.
- `m0` and `m1` are formulas indicating the specification for the MTR function m broken up into treatment arms $m(0|u,x)$ and $m(1|u,x)$.
- `ivlike` indicates whether to use the moment or regression criterion. For the moment criterion, it is a list of formulas that determine the set of functions $\mathcal{S}$ that define the moment conditions.[7]
- `propensity` is a formula that specifies how the propensity score is estimated.

In the remainder of this section we discuss how to specify these arguments to implement the methodology previously described. Along the way, we cover additional options that provide extra functionality.

2.4.3 Specifying the target parameter

The `target` option can be set to one of `ate`, `att`, `atu`, `late`, or `genlate`, which correspond respectively to the average treatment effect (ATE), the average treatment on the treated (ATT), the average treatment on the untreated (ATU), the local average treatment effect (LATE; Imbens and Angrist, 1994) and the generalized LATE (J. J. Heckman and Vytlacil, 2005b; Mogstad et al., 2018b). The choice of this argument specifies the form of the target parameter τ via its weighting function $\omega_\mathcal{T}$ in (2.5). Nothing else has to be specified for `ate`, `att`, and `atu`. It is also possible to specify a custom parameter by specifying the weights $\omega_\mathcal{T}$.

[7]The terminology comes from Mogstad et al. (2018b), who described the class of cross-moments $c_s \equiv \mathbb{E}[Y_i s(D_i, X_i, Z_i)]$ as "IV–like" estimands because they nest standard linear IV estimands via particular choices of s.

LATE

The local average treatment effect (LATE) from shifting the instrument Z_i from z_0 to z_1 is defined as

$$\text{LATE}(z_0 \to z_1) \equiv \mathbb{E}[Y_i(1) - Y_i(0) | D_i(z_0) = 0, D_i(z_1) = 1].$$

In terms of the equivalent selection model (2.1),

$$\text{LATE}(z_0 \to z_1) = \mathbb{E}\left[\int_{p(X_i,z_0)}^{p(X_i,z_1)} (m(1|u, X_i) - m(0|u, X_i)) \left(\frac{1}{p(X_i, z_1) - p(X_i, z_0)}\right) du\right],$$

which takes the form (2.5) with weighting function

$$\omega_{\mathcal{T}}(d|x, z) = (-1)^{d+1} \left(\frac{\mathbb{1}[p(x, z_0) < u \leq p(x, z_1)]}{p(x, z_1) - p(x, z_0)}\right).$$

This is the form of $\omega_{\mathcal{T}}$ used if `target = late`. The user must pass `late.from` and `late.to`, which should be named lists indicating the identity and value of z_0 and z_1, respectively.

As defined, the LATE parameter averages over all covariates. The `ivmte` package also allows for "effect modification," where the LATE is computed conditional on $V_i = v$, for some function V_i of the covariate vector X_i (e.g. Kennedy, Lorch, & Small, 2019; Ogburn, Rotnitzky, & Robins, 2015):

$$\text{LATE}(z_0 \to z_1 | v) \equiv \mathbb{E}\left[\int_{p(X_i,z_0)}^{p(X_i,z_1)} (m(1|u, X_i) - m(0|u, X_i)) \left(\frac{1}{p(X_i, z_1) - p(X_i, z_0)}\right) du | V_i = v\right].$$

To do this, set `target = late` and `late.from`, `late.to` as above, but also pass a named list `late.X` to indicate the variable V_i and value v. Note that no smoothing is done for the conditional expectation, so V_i should be a discrete variable.

Generalized LATE

The selection model (2.1) allows conceptualizing a generalized LATE where instead of choosing instrument values z_0 and z_1, we choose values u_0 and u_1 for the latent propensity variable U_i (J. J. Heckman & Vytlacil, 2005b). This can be useful for diagnosing the robustness of a standard LATE to broadening the complier subpopulation (Mogstad & Torgovitsky, 2018). The formal definition is

$$\text{GenLATE}(u_0, u_1) \equiv \mathbb{E}\left[\int_{u_0}^{u_1} (m(1|u, X_i) - m(0|u, X_i)) \frac{1}{(u_1 - u_0)} du\right], \tag{2.25}$$

for values $u_0, u_1 \in [0, 1]$ with $u_0 < u_1$. To set the target parameter to (2.25) in `ivmte`, pass `target = genlate`, `genlate.lb` and `genlate.ub`, where the latter two parameters correspond to u_0 and u_1. Effect modification can also be incorporated by passing `late.X`, in the same way as for the usual LATE discussed in the previous section.

Custom target parameters

The user can define their own target parameters by directly specifying the weight function $\omega_\mathcal{T}$ in (2.5). To facilitate computation, these weight functions are required to be constant splines in u, i.e.

$$\omega_\mathcal{T}(d|u, X_i, Z_i) = \sum_{j=1}^{J_d} \mathbb{1}\left[\kappa_{j-1}(d|X_i, Z_i) < u \leq \kappa_j(d|X_i, Z_i)\right] \bar{\omega}_{\mathcal{T},j}(d|X_i, Z_i),$$

where $\kappa_0(d|X_i, Z_i) \equiv 0$, and $k_{J_d}(d|X_i, Z_i) \equiv 1$. The user sets these weights by passing the $J_0 - 1$ knot functions $(\kappa_1(0|\cdot,\cdot), \ldots, \kappa_{J_0-1}(0|\cdot,\cdot))$ as a list via `target.knots0` and the J_0 weight functions $(\bar{\omega}_{\mathcal{T},1}(0|\cdot,\cdot), \ldots, \bar{\omega}_{\mathcal{T},J_0}(0|\cdot,\cdot))$ as a list via `target.weight0`. The analogous options `target.knots1` and `target.weight1` for the treated $(d = 1)$ weights also need to be specified. For any component of these lists, a constant (scalar numeric) can be passed instead of a function to indicate a function that does not vary with (x, z). Note that the option `target` is ignored when any of the custom `target.*` options are passed.

2.4.4 Specifying the MTR functions

The required `m0` and `m1` arguments accept specifications for two treatment arms of the MTR function using the standard `R` formula syntax familiar from functions like `lm` or `glm`. However, these formulas involve an unobservable variable whose default name is `u`.[8] Typical specifications will involve combinations of `u` and other covariates. For example,

```
m0 <- ~ var1 + u + I(var1 * u) + I(u^2)
```

specifies $m(0|u, x)$ to be quadratic in the unobservable `u` (u) and linear in `var1` (a subcomponent of x), with a first order interaction between `u` and `var1`. Note that the left-hand side of these formulas is empty. Also note the use of `I()` to inhibit the interpretation of `*` and `^` as formula operators.

Currently, `ivmte` requires specifications of `m0` and `m1` to either be polynomials or B-splines in `u`. B-splines are incorporated using the function `uSplines`, which is an interpreter that utilizes the `splines2` package (Wang & Yan, 2018). An example of the syntax is

```
m1 <- ~ var1 + uSplines(degree = 0, knots = c(0.2, 0.5, 0.8))
```

which would specify $m(1|u, x)$ to be linear in `var1` and piecewise constant in `u` with jumps at the specified knot points.[9] Splines can be interacted with other variables and intermingled with other polynomials, for example

```
m0 <- ~ u + I(u^2) + var1:uSplines(degree = 2, knots = c(0.3, 0.4, 0.5, 0.7))
```

would specify a quadratic function of `u` and a linear function of `var1` whose slope varies with u according to a quadratic B-spline with knot points at `.3`, `.4`, `.5`, and `.7`.

[8] The name can be changed with the `uname` option.

[9] As Wang and Yan (2018) describe in their vignette, the only difference between the `bSpline` function in `splines2` and the `bs` function in the core package `splines` is that `bSpline` allows for degree 0 splines, i.e. piecewise constant functions. This turns out to be particularly useful for our purposes because piecewise constant functions have a special place in the theory developed by Mogstad et al. (2018b); see their Proposition 4.

2.4.5 Imposing shape constraints

For partial identification cases, `ivmte` also allows the user to require the MTR and/or MTE functions to be bounded and/or monotone in u. The bounds are imposed through the arguments `m0.lb`, `m0.ub`, `m1.lb`, `m1.ub`, `mte.lb`, and `mte.ub`. Note that the default action of `ivmte` is to set the upper and lower bounds on `m0` and `m1` to the largest and smallest values of the response variable observed in the data, which also implies values for `mte.lb` and `mte.ub`. Monotonicity in u, in either an increasing or decreasing sense, is set through the boolean arguments `m0.dec`, `m0.inc`, `m1.dec`, `m1.inc`, `mte.dec`, and `mte.inc`. These arguments are set to `FALSE` by default.

These shape constraints (boundedness and monotonicity) are enforced through an "auditing" procedure. The procedure is designed to circumvent the difficulty of determining whether a polynomial function is bounded or monotone on its domain. It starts by imposing the desired shape constraints on the MTR function at all points on a well-spaced, relatively coarse *constraint grid*. After producing the bound estimates $\hat{\tau}^{\star}_{\text{lb}}$ and $\hat{\tau}^{\star}_{\text{ub}}$, the solution MTR functions at these bounds are checked ("audited") for shape restrictions on a much finer *audit grid*. If the solutions satisfy the shape restricts across the entire audit grid, then the process ends. Otherwise, the estimator is recomputed with an expanded constraint grid that contains some of the points in the audit grid where the restrictions were violated. The procedure repeats until the solutions pass the audit, or until a maximum number of iterations are reached.

The user can adjust the size of the initial constraint grid through the arguments `initgrid.nu` and `initgrid.nx`. These arguments control the initial number of points at which to impose the constraints for `u`, via `initgrid.nu`, and all other variables included in the specification of `m0` and `m1`, via `initgrid.nx`. For the latter, the points are drawn randomly from the empirical distribution in `data`. The default values of `initgrid.nu` and `initgrid.nx` are both 20, so that the total initial constraint grid size is 400.

The user can also adjust the size of audit grid through the arguments `audit.nu` and `audit.nx`. The default for `audit.nu` is 25, while the default for `audit.nx` is set at 2500. By default then, the solution MTR function must satisfy the shape constraints on an audit grid with 62,500 points.[10] When a solution MTR function fails an audit, the number of violating points that are added to the constraint grid from the audit grid (for each shape constraint) is given by `audit.add`, which has a default of 100.

The audit is terminated after the solution MTR functions satisfy the constraints on the entire audit grid, or after `audit.max` rounds of the audit procedure, which has a default of 25 rounds. If `audit.max` is hit, the user should investigate the `audit.grid$violations` field of the list that `ivmte` returns. This reports the points of the audit grid at which the shape restrictions are violated. Small regions of violation can likely be ignored without seriously affecting the estimated bounds $\hat{\tau}^{\star}_{\text{lb}}$ and $\hat{\tau}^{\star}_{\text{ub}}$. If the violations occur on a large region, the user can let the audit procedure run for more rounds by increasing `audit.max`.

2.4.6 Specifying the criterion function

To use the regression approach, simply leave the `ivlike` input empty and indicate the outcome variable Y_i as `outcome = y`.

To use the moment approach, one needs to specify the collection of functions $\mathcal{S}$ via `ivlike` by passing a vector of formulas, each of which has the same outcome variable (Y_i) on the left-hand side. For example,

```
ivlike <- c(
  y ~ d,
  y ~ d | z,
  y ~ d + x | z + I(z^2) + x
)
```

has three formulas, with the second two specified using the `|` syntax familiar from the `ivreg` command in the `AER` package (Kleiber & Zeileis, 2018). The first formula is an OLS

[10] Assuming of course that there are at least 2500 unique values of X_i in the data. Otherwise, the entire empirical support of X_i is used.

regression of y on d and a constant. The second formula uses z as an instrument for d, as in just-identified IV regressions. The third formula uses z and z^2 as instruments for d, with x serving as a covariate that instruments for itself.

The default behavior of ivmte is to include all of the estimated coefficients from each specification as functions $s \in \mathcal{S}$. In the example above, this would mean the coefficients on the constant and d in the first and second specifications, and the coefficients on the constant, d and x in the third, for a total of 7 moments to match. The user can change this behavior with the components argument. This argument expects a list of the same length as ivlike, with the jth component of the list being a vector that indicates which coefficients should be included from the jth IV–like specification in ivlike. In the example above, we could have used

```
components <- l(intercept, d, c(d, x))
```

to indicate that we want only the coefficient on the constant (intercept) from the first specification, only the coefficient on d in the second, and both the coefficients on d and x in the third, for a total of 4 moments. Note that intercept is used to refer to the implied constant term in the formula specifications, and so should be viewed as a reserved word when it comes to naming data columns.[11]

Conditioning subsets for the IV–like specification can be set through the optional subset argument. This option expects a list of the same length as ivlike, with each component of the list representing a logical statement. For example,

```
subset <- l(z == 1, , x %in% c(2, 3))
```

would estimate the first IV–like specification only on the subset with z == 1, the second for all observations, and the third only for the subset for which either x == 2 or x == 3. This

[11]The l function is a generalization of the list function, and allows the user to list variables and expressions without having to enclose them by quotation marks.

provides an easy way to specify conditional moments as components of $\mathcal{S}$, e.g. via

$$\mathbb{E}[Y_i|Z_i = 1] = \mathbb{E}\left[Y_i \underbrace{\frac{\mathbb{1}[Z_i = 1]}{\mathbb{P}[Z_i = 1]}}_{\text{example of } s(D_i, X_i, Z_i)}\right]. \tag{2.26}$$

2.4.7 Propensity score estimation

Estimating $\hat{\gamma}_s$ and $\hat{\gamma}_{\text{cm}}$, as well as $\hat{\tau}$ for many choices of target parameter requires first estimating the propensity score, $p(x, z) \equiv \mathbb{P}[D_i = 1|X_i = x, Z_i = z]$. This is communicated through the `propensity` argument. Typically, the user will pass a formula for `propensity` in which the treatment variable appears on the left-hand side, for example

```
propensity <- d ~ x + z
```

By default, this estimates a logit model using `glm` with the specified right-hand side variables, but the user can change this to probit or linear by passing `link = "probit"` or `link = "linear"`.

Alternatively, the user can estimate the propensity score before running `ivmte`, save estimates of $p(X_i, Z_i)$ in their dataframe as a new column, say `p`, and then pass `propensity = p`. When this is done, the user must also indicate the name of the treatment variable through the argument `treat`. When a formula is passed for `propensity`, the treatment variable is inferred to be the response variable of the formula, and the `treat` argument is ignored unless it doesn't match the inferred variable, in which case an error is thrown.

2.4.8 Solving

By default, `ivmte` attempts to determine whether there is a unique solution to either the moment-based or regression-based criteria (depending on the user's specification of `ivlike`) by checking the rank of their first-order equations. If it determines that there is a unique solution, and `point` is either not passed, or passed as `point = TRUE`, then it proceeds to solve for the unique solution and form a point estimate of the target parameter as described in Section 2.3.3. For the moment criterion, the default behavior is to use the optimal two-step weighting for $\hat{\Omega}$, but this can be changed to the identity weighting by passing `point.eyeweight = TRUE`.

If `ivmte` determines there is not a unique solution, or if `point = FALSE` is passed, then it proceeds with the two-step bounds estimator described in Section 2.3.3 The solver package for these problems is set using the option `solver`, which currently accepts the following values: `gurobi`, `Rmosek`, `cplexAPI`, and `lpSolveAPI`.[12] If no value is passed for `solver`, then `ivmte` searches for a solver in the order given above and uses the first one that is found. The value of the tuning parameter, σ, in (2.22) is set to 10^{-4} by default, and can be changed with the `criterion.tol` argument.

2.4.9 Confidence intervals

The `ivmte` command can construct confidence intervals by resampling (bootstrapping or subsampling). The number of replications is determined by the argument `bootstraps`, which is set to 0 by default so that confidence intervals are not computed. The size of the resampled dataset is determined by `bootstraps.m`, which is set to the sample size of `data` by default. The default behavior is to draw the resampled data with replacement from `data`, but this can be toggled with the boolean argument `bootstraps.replace`. Confidence intervals are reported for all levels in `levels`, which has the default of `c(.99, .95, .90)`.

For the point-identified case, the reported intervals are formed from the resampled distribution of (2.20).[13] Conducting statistical inference on bounds in the partially identified case is more delicate due to their potentially non-standard asymptotic distributions.[14] There does not currently exist a solution for the MTE framework that is both theoretically

[12] The `gurobi` package is included with Gurobi, while `cplexAPI` and `lpSolveAPI` are available from CRAN. The version requirements of `ivmte` as of this writing are: `gurobi` 7.5-1 or later, `Rmosek` 9.2.38 or later, `cplexAPI` 1.3.3 or later, and `lpSolveAPI` 5.5.2 or later.

[13] The moment-based criterion uses re-centered moment conditions (Brown & Newey, 2002; Hall & Horowitz, 1996).

[14] See Canay and Shaikh (2017) for a recent survey on inference in partially identified models.

satisfactory and computationally tractable. Instead, `ivmte` implements the forward and reverse bootstrap procedures discussed by Andrews and Han (2009).[15] While these are known to *not* be valid in general, they may still provide a reasonable indication of statistical uncertainty for the user. In addition to confidence intervals for each level in `levels`, `imvte` also returns a p-value, computed as the smallest level a such that a $1 - a$ confidence interval would not contain 0.

2.4.10 Specification tests

If using the moment-based criterion function, `ivmte` will also conduct a bootstrap test of the null hypothesis that the model is correctly specified (i.e. of the null hypothesis that the minimum value of the population criterion is zero) whenever `bootstraps` is a positive number. In the point-identified case, the test used is the well-known Hansen (1982) overidentification test for GMM using the adjustment for bootstrapping discussed by Hall and Horowitz (1996). In the partially-identified case, the test used is the "re-sampling" test of Bugni, Canay, and Shi (2015b). In either case, `ivmte` returns a p-value for the null hypothesis of correct specification. The user can turn off the specification test by passing `specification.test = FALSE`.

2.4.11 Output

The return of `ivmte` is a named list with a large number of fields.[16] The most important fields are `pointestimate` and `bounds`, which return (2.20) or (2.22), depending on whether `point` is `TRUE` or `FALSE`. If confidence intervals are being computed, these are returned in the fields `pointestimate.ci` or `bounds.ci`, with the p-value returned in the field `pvalue`. Other fields that may be useful for diagnostics or debugging are `s.set`, which contains the results of running the IV–like specifications, `propensity`, which contains the results of

[15] The default is to compute and report the results from both backward and forward procedures. This behavior can be changed by passing `ci.type = "backward"` or `ci.type = "forward"`.

[16] In case memory usage is an important issue to the user, we have included an option `smallreturnlist` that can be set to `TRUE` to limit the number of objects that are returned.

the propensity score estimation, `audit.criterion`, which gives the value $\hat{Q}^\star$ in (2.21), and `audit.grid$violations`, which reports points at which the audit procedure failed to secure compliance with the desired shape restrictions.

2.5 Empirical illustration

2.5.1 Data and motivation

We illustrate the motivation and usage of `ivmte` by revisiting Angrist and Evans's (1998) analysis of the relationship between fertility on labor supply. The data comes from the 1980 Census Public Use Micro Samples (PUMS); a detailed description can be found in Angrist and Evans (1998).[17] Our illustration uses three main variables: `worked` is an indicator for whether a woman worked for pay in the year prior to the survey, `morekids` is an indicator for whether a woman has exactly two children (`morekids = 0`) or three or more children (`morekids = 1`), and `samesex` is an indicator that is 1 if the first two children had the same sex. Later, we will also use the woman's year of birth (`yob`) and indicators for her race (`hisp`, `black`, `other`) to demonstrate specifications with covariates. Our interest is in the effect of having more than two children (`morekids`) on labor supply (`worked`).

A simple linear regression of `worked` on `morekids` shows that 58% of women with two children work, compared to only 44% of those with three or more children:

```
lm(data = AE, worked ~ morekids)
```

```
##
## Call:
## lm(formula = worked ~ morekids, data = AE)
##
## Coefficients:
```

[17] The original data can be downloaded from https://economics.mit.edu/files/1199 or from http://sites.bu.edu/ivanf/files/2014/03/m_d_806.dta_.zip. The data we use is restricted to women who were at least 20 years old at their first birth. The cleaned subsample data with only the variables relevant to the current analysis is included as data with ‘ivmte‘.

```
## (Intercept)     morekids
##      0.5822      -0.1423
```

The coefficient on `morekids` of $-.14$ probably overstates the causal impact of fertility on labor supply, since women who choose to have more children likely do so in part because their labor market prospects are weaker. An IV regression using `samesex` as an instrument for `morekids` returns a coefficient on `morekids` that is substantially smaller in magnitude:

```
library("AER")
ivreg(data = AE, worked ~ morekids | samesex)$coeff["morekids"]
```

```
##    morekids
## -0.08484221
```

Moreover, if there is heterogeneity in the effect of fertility on working, then this latter estimate only reflects the same-sex compliers, that is, those women who would have a third child if and only if their first two had the same sex. The "first stage" regression of `morekids` on `samesex` shows that this group is rather small, comprising less than 6% of the population.

```
lm(data = AE, morekids ~ samesex)$coeff["samesex"]
```

```
##    samesex
## 0.05886826
```

If our research question requires knowing a quantity involving the entire population, such as the ATE or the ATT, then this linear IV estimate is not particularly helpful.

2.5.2 Extrapolation to the ATE under different assumptions

The `ivmte` package can be used to extrapolate from the small complier group to larger groups by providing a coherent framework under which additional assumptions can be imposed. Suppose for example that we assume that the MTR functions are both quadratic in u, so that the pair is characterized by six parameters. Since both `morekids` and `samesex` are binary, we only have four moments at our disposal to identify these six parameters, so the model is not point identified. However, we can use `ivmte` to estimate bounds on the ATE:

```
ivmte(
  data = AE,
  ivlike = c(worked ~ morekids + samesex + morekids * samesex),
  target = "ate",
  m0 = ~ u + I(u^2),
  m1 = ~ u + I(u^2),
  propensity = morekids ~ samesex
)
```

```
##
## Bounds on the target parameter: [-0.2862919, 0.1050867]
## Audit terminated successfully after 1 round
```

As a comparison, Manski's (1990) nonparametric IV bounds on the ATE are $[-.548, .393]$. The bounds produced by `ivmte` are much tighter because they impose a parametric assumption on the model primitives which smooths out the extreme cases at which Manski's bounds are obtained. The parametric assumption says that if we line up families by their latent propensity to have a third child, then families who are close to having the same propensity (similar u) are, on average, not too dissimilar in their potential work outcomes. A weaker parameterization for the MTR, such as a polynomial of higher order than 2, would allow families with similar fertility propensities to be more different, since such a function could "wiggle" more quickly between the natural bounds of 0 and 1 for Y_i.

While narrower than Manski's nonparametric bounds, the bounds under a quadratic parameterization are still quite wide; they are consistent with both large negative and modest positive causal effects. However, because the outcome is binary and all four potential s functions have been incorporated into the saturated specification `ivlike = c(worked ~ morekids + samesex + morekids*samesex)`, we know from Proposition 3 of Mogstad et al. (2018b) that the bounds are sharp (best possible) in the sense of fully exhausting the information contained in the model and the data. Thus, if the researcher is unsatisfied with the width of the bounds, they have two paths to satisfaction: (i) make stronger (or different) assumptions, or (ii) ask a less ambitious question by changing the target parameter.

A natural way to strengthen the assumptions is to eliminate the quadratic terms in the MTR specifications, so that there are only four parameters:

```r
ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~u,
  m1 = ~u,
  propensity = morekids ~ samesex
)
```

```
## Warning: MTR is point identified via GMM. Shape constraints are ignored.

##
## Point estimate of the target parameter: -0.07791036
```

The bounds have collapsed to a point. This makes sense since we have not changed `ivlike`, so we still have four moments, but relative to the quadratic case we have reduced the number of parameters from six to four (Brinch et al., 2012, 2017). If we had done this moment-counting exercise ahead of time, we could have added `point = TRUE` to the call:

```r
ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~u,
  m1 = ~u,
  propensity = morekids ~ samesex,
```

```
  point = TRUE
)
```

```
##
## Point estimate of the target parameter: -0.07791036
```

Linearity is a restrictive parameterization, and one might be uncomfortable with the fact that it allows for complete extrapolation from the 6% of the population represented in the LATE to the entire population represented in the ATE. As an alternative, consider combining the quadratic case with shape restrictions. For example, we could assume that the MTRs must generate an MTE curve that is negative and increasing:

```
ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~ u + I(u^2),
  m1 = ~ u + I(u^2),
  mte.inc = TRUE,
  mte.ub = 0,
  propensity = morekids ~ samesex
)
```

```
##
## Bounds on the target parameter: [-0.08484221, -0.06323574]
## Audit terminated successfully after 1 round
```

The assumption behind this shape restriction is that the effect of having another child on working is negative ($m(1|u,x)-m(0|u,x) \leq 0$, imposed via `mte.ub = 0`), and is more negative for women who are more likely to have more children ($m(1|u,x) - m(0|u,x)$ increasing as a function of u, imposed via `mte.inc = TRUE`). Adding the assumption narrows the bounds considerably, from $[-.286, .105]$ to $[-.085, -.063]$. In this case, the resulting bounds happen to be similar to the original linear IV estimate for compliers, but they are the product of a formally-justified theoretical framework for extrapolation, rather than verbal extrapolation and wishful thinking.

2.5.3 Easier extrapolation problems

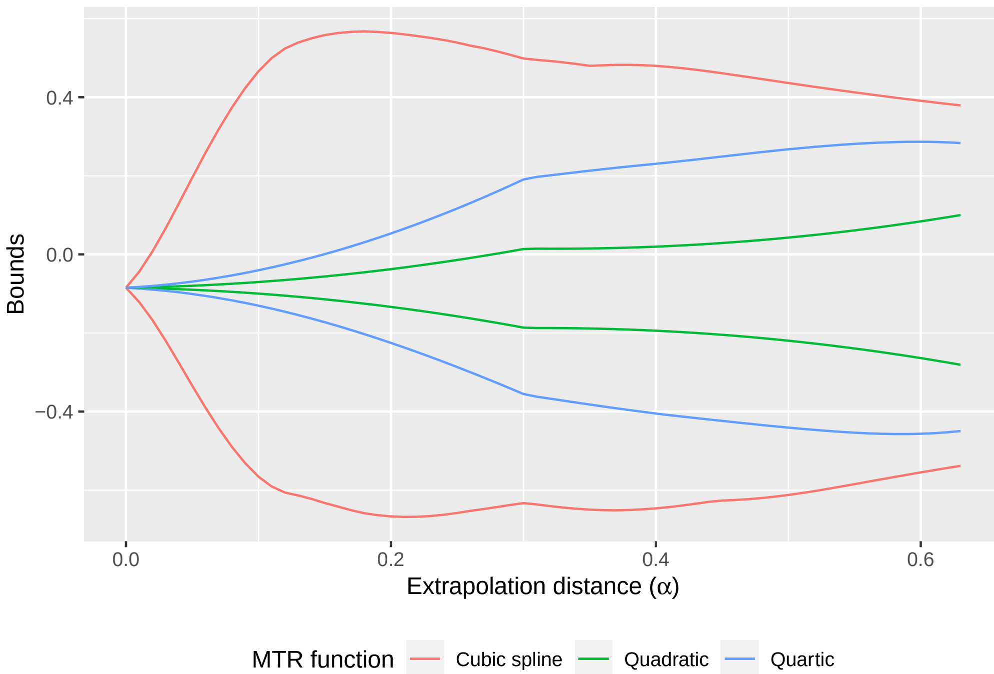
Extrapolating from a small complier group represented in a LATE (6%) to the entire population represented in the ATE is a heroic challenge. Changing the target parameter to something less ambitious makes the extrapolation problem easier. For example, one could consider extrapolated LATEs, i.e. generalized LATEs (2.25) with $u_{\text{lb}} = \max\{p(0) - \alpha, 0\}$ and $u_{\text{ub}} = \min\{p(1) + \alpha, 1\}$ for different non-negative values of α (Mogstad et al., 2018b, Section 4.2). For $\alpha = 0$, the extrapolated LATE is equivalent to the usual LATE, while as $\alpha \to \max\{p(0), 1 - p(1)\}$, it returns to the ATE.

```
# Set up ivmte arguments as a list so they can be easily changed
args <- list(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "genlate",
  m0 = ~ u + I(u^2),
  m1 = ~ u + I(u^2),
  propensity = morekids ~ samesex,
  audit.nu = 200
)

# Get propensity score and construct alpha list
p <- predict(lm(data = AE, morekids ~ samesex),
  newdata = data.frame(samesex = c(0, 1)),
  type = "response"
)
alphalist <- seq(from = 0, to = max(p[1], (1 - p[2])), by = .01)
```

```r
# Function for computing genlate bounds at different values
loopivmte <- function(args, alphalist) {
  df.lb <- data.frame(alpha = alphalist, value = NA, type = "lb")
  df.ub <- data.frame(alpha = alphalist, value = NA, type = "ub")
  for (i in 1:length(alphalist)) {
    args[["genlate.lb"]] <- max(p[1] - alphalist[i], 0)
    args[["genlate.ub"]] <- min(p[2] + alphalist[i], 1)
    r <- do.call(ivmte, args)
    df.lb$value[i] <- r$bound[1]
    df.ub$value[i] <- r$bound[2]
  }
  return(rbind(df.lb, df.ub))
}

# Run the quadratic case
plotquadratic <- loopivmte(args, alphalist)
plotquadratic$name <- "Quadratic"

# Run the quartic case
args[["m0"]] <- ~ u + I(u^2) + I(u^3) + I(u^4)
args[["m1"]] <- args[["m0"]]
plotquartic <- loopivmte(args, alphalist)
plotquartic$name <- "Quartic"

# Run the spline case
args[["m0"]] <- ~ uSplines(degree = 3, knots = seq(from = .1, to = .9, by = .1))
args[["m1"]] <- args[["m0"]]
plotspline <- loopivmte(args, alphalist)
plotspline$name <- "Cubic spline"
```

```r
library("ggplot2")
plotdf <- rbind(plotquadratic, plotquartic, plotspline)
ggplot(plotdf, aes(x = alpha, y = value, color = name)) +
```

Figure 2.1: Bounds as a function of assumptions and extrapolation difficulty




```
geom_line(data = subset(plotdf, type == "lb")) +
geom_line(data = subset(plotdf, type == "ub")) +
labs(
  x = expression(paste("Extrapolation distance (", alpha, ")")),
  y = "Bounds",
  color = "MTR function"
) +
theme(legend.position = "bottom")
```

Figure 2.1 reports bounds on extrapolated LATEs as a function of α for three different specifications of the MTR functions. The tightest specification is the unconstrained quadratic used above. The quartic specification takes the quadratic and adds third and fourth order terms to both MTR functions. The spline specification is a flexible cubic spline with nine knots.

As expected, the bounds are always ordered in width with quadratic being narrowest and the cubic spline being widest. For all specifications, the bounds start as a point at $\alpha = 0$ (the

LATE) and tend towards the ATE bounds as $\alpha \to 1$. This shows how the MTE framework allows the researcher to achieve bounds of any width they desire, while still being constrained by the reality that stronger conclusions require stronger assumptions. Given this freedom, it is unlikely that the researcher's preferred trade-off between assumptions and conclusions is the corner solution of reporting only nonparametrically point-identified parameters such as the LATE, which reflect both the weakest assumptions and the weakest conclusions.

2.5.4 Covariates

Covariates (X_i) serve two roles in all IV strategies. First, they can increase the credibility of the assumption that the instrument is as good as randomly assigned by making that assumption conditional on other observables. Second, covariates can reduce sampling uncertainty to the extent that they soak up residual variation in the outcome and/or treatment variables. In the MTE framework, covariates can also be used in a third role to provide identifying content through separability (e.g. Brinch et al., 2012, 2017; Carneiro et al., 2011).

To demonstrate this, we return to the quadratic specification with the ATE as the target parameter, but now we fully interact the MTE specification in `yob`, viewed here as X_i:

```
set.seed(1234) # the covariate part of the audit grid is stochastic
ivmte(
  data = AE,
  ivlike = worked ~ (morekids + samesex + morekids * samesex) * yob,
  target = "ate",
  m0 = ~ u + yob + u * yob + I(u^2) + I(u^2) * yob,
  m1 = ~ u + yob + u * yob + I(u^2) + I(u^2) * yob,
  propensity = morekids ~ yob + samesex + samesex * yob
)
```

```
##
## Bounds on the target parameter: [-0.2790478, 0.09365855]
## Audit terminated successfully after 1 round
```

The bounds are quite similar to the previous bounds that we obtained without covariates. This is expected because for each new interacted moment that is being matched we are adding an interaction term in the MTR that must be fit. Eliminating one or more of these interaction terms imposes *separability*, that is, the assumption that unobserved heterogeneity in potential outcomes operates similarly for different values of the covariate. Here we eliminate the quadratic interaction and see that the bounds narrow considerably.

```
set.seed(1234)
ivmte(
  data = AE,
  ivlike = worked ~ (morekids + samesex + morekids * samesex) * yob,
  target = "ate",
  m0 = ~ u + yob + u * yob + I(u^2),
  m1 = ~ u + yob + u * yob + I(u^2),
  propensity = morekids ~ yob + samesex + samesex * yob
)
```

```
##
## Bounds on the target parameter: [-0.1206799, 0.03139476]
## Audit terminated successfully after 1 round
```

With multiple types of assumptions to impose there is naturally a trade-off. For example, we might want to use the information we obtain with separability to buy a more flexible functional form.

```
set.seed(1234)
ivmte(
  data = AE,
  ivlike = worked ~ (morekids + samesex + morekids * samesex) * yob,
  target = "ate",
  m0 = ~ uSplines(degree = 3, knots = seq(from = .25, to = .75, by = .25)) + yob,
  m1 = ~ uSplines(degree = 3, knots = seq(from = .2, to = .75, by = .25)) + yob,
```

```
  propensity = morekids ~ yob + samesex + samesex * yob
)
```

```
##
## Bounds on the target parameter: [-0.2993263, 0.1696851]
## Audit terminated successfully after 2 rounds
```

The regression approach starts to become particularly attractive in rich specifications with multiple different covariates. This is because the number of possible moments that could be matched blows up; using all of them is potentially unwise due to small-sample bias, but it is also not necessarily clear how to choose which ones to use. The regression approach removes this choice through the usual least squares weighting. For example:

```
set.seed(1234)
ivmte(
  data = AE,
  outcome = worked,
  target = "ate",
  m0 = ~ uSplines(degree = 3, knots = seq(from = .1, to = .9, by = .1)) + yob +
    black + hisp + other,
  m1 = ~ uSplines(degree = 3, knots = seq(from = .1, to = .9, by = .1)) + yob +
    black + hisp + other,
  propensity = morekids ~ samesex + yob + black + hisp + other,
  audit.nu = 50
)
```

```
##
## Bounds on the target parameter: [-0.3097528, 0.0173328]
## Audit terminated successfully after 7 rounds
```

2.5.5 Confidence intervals

To conclude, we demonstrate how `ivmte` constructs confidence intervals. If we estimate the model assuming point identification, as with a linear specification, `ivmte` returns:

```
set.seed(1234) # the bootstrap is stochastic
r <- ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~u,
  m1 = ~u,
  point = TRUE,
  bootstraps = 100,
  propensity = morekids ~ samesex
)
summary(r)
```

```
##
## Point estimate of the target parameter: -0.07791036
## MTR coefficients: 4
## Independent/total moments: 4/4
##
## Bootstrapped confidence intervals (nonparametric):
##     90%: [-0.150156, -0.006348619]
##     95%: [-0.161036, 0.003467422]
##     99%: [-0.1650715, 0.02561266]
## p-value: 0.08
## Number of bootstraps: 100
```

While for the general case of bound estimation, `ivmte` returns:

```
set.seed(1234)
r <- ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~ u + I(u^2),
```

```
  m1 = ~ u + I(u^2),
  mte.inc = TRUE,
  mte.ub = 0,
  bootstraps = 100,
  propensity = morekids ~ samesex
)
summary(r)
```

```
##
## Bounds on the target parameter: [-0.08484221, -0.06323574]
## Audit terminated successfully after 1 round
## MTR coefficients: 6
## Independent/total moments: 4/4
## Minimum criterion: 0
## Solver: Gurobi ('gurobi')
##
## Bootstrapped confidence intervals (backward):
##     90%: [-0.1409956, -0.01898701]
##     95%: [-0.14766, -0.009421106]
##     99%: [-0.1587216, 2.775558e-17]
## p-value: 0.02
## Number of bootstraps: 100
```

2.6 Acknowledgments

This research was supported in part by National Science Foundation grants SES-1530538 and SES-1846832. We thank Christine Blandhol and John Bonney for testing the package during its development.

CHAPTER 3

INFERENCE FOR SUPPORT VECTOR REGRESSION UNDER ℓ_1 REGULARIZATION

COAUTHORED WITH YUEHAO BAI,* HUNG HO,† GUILLAUME A. POULIOT‡

3.1 Introduction

This paper studies inference for support vector regression (SVR) with ℓ_1-norm regularization (ℓ_1-SVR). SVR is the extension of the support vector machine (SVM) classification method (Vapnik, 1998) to the regression problem (Basak, Pal, & Patranabis, 2007; Cherkassky & Ma, 2004; Drucker, Burges, Kaufman, Smola, & Vapnik, 1997; Smola & Schölkopf, 2004). SVR is designed to reproduce the good out-of-sample performance of SVM classification in the regression setting. It has been frequently used in regression analysis across fields such as geophysical sciences (Ghorbani, Zargar, & Jazayeri-Rad, 2016; P. Li, Tan, Yan, & Deng, 2011), environmental science (Sánchez, Nieto, Fernández, del Coz Díaz, & Iglesias-Rodríguez, 2011), engineering (J. Li, West, & Platt, 2012; Pelossof, Miller, Allen, & Jebara, 2004; Xi, Poo, & Chou, 2007), and image compression (Jiao, Li, Wang, & Li, 2005).

However, methodology for inference is lacking. To the best of our knowledge, theory and closed form expressions for the asymptotic variance of the regression coefficient estimates or of tests which may be inverted for inference are not available. Some non-asymptotic methodology is currently available for producing error bars for support vector machines regression, but it necessitates distributional assumptions. For instance, Gao, Gunn, Harris, and Brown (2002) produce a Gaussian process formulation which delivers exact, small sample inference. Law and Kwok (2001) cast support vector machines regression in a Bayesian framework, and likewise produce exact, small sample inference. Because distributional assumptions are not

*Department of Economics, University of Michigan
†Booth School of Business, University of Chicago
‡Harris School of Public Policy, University of Chicago

satisfied in practice, and the choice of prior distributions places an additional burden on the user, we find these methods impractical.

In many standard inference problems, such distributional assumptions are typically circumvented by resorting to either asymptotic approximations of the distribution of the estimated coefficients, or by inverting asymptotically valid test procedures.

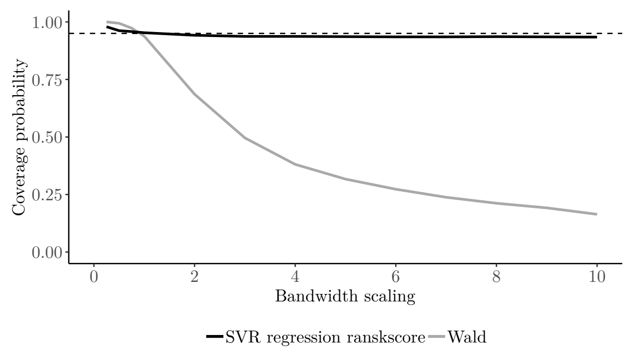
Meanwhile, theory for asymptotic and exact inference with quantile regression is well developed (Bai, Pouliot, & Shaikh, 2019; Koenker, 2005), and analogy with quantile regression (see Figure 3.2) suggests that similar ideas and methods may work for support vector machines regression with ℓ_1-norm regularization.

Indeed, it has been shown that calculations akin to those used to derive asymptotic distributions of quantile regression coefficient estimates may be used to produce asymptotic approximations of conditional (on features) probabilities of classification for SVM (G. Pouliot, 2018). These derivations produce asymptotic distributions for the coefficients of the SVM classification problem, and are readily extended to produce asymptotic distributions for the regression coefficients of SVR.

However, regression estimators with support vectors[1] share the undesirable feature that the asymptotic variance of the estimated regression coefficients depends on the density of the regression errors (Koenker & Machado, 1999). This is intuitive since the stability of the fitted hyperplane resting on support vectors ought to depend on the stability of the support vectors. It is indeed an unfortunate property. Any estimate of the asymptotic variance, a key ingredient for producing the error bars, will rely on a nonparametric density estimate. Density estimation itself requires the somewhat arbitrary choice of a bandwidth parameter, which will induce an arbitrarily scaling of the width of error bars. This may allow users to present deceptively narrow confidence intervals whose coverage probabilities fall well below the nominal level, as illustrated in Figure 3.1.

[1]There is no general, agreed upon definition of support vectors. For both quantile regression and SVM regression with ℓ_1-norm, they may be rigorously defined as the active basis in the dual program, when written as a linear program in standard form (Bertsimas & Tsitsiklis, 1997) with box constraints (Koenker, 2005).

Figure 3.1: Coverage Probability of ℓ_1-SVR Regression Rankscore Confidence Interval versus Wald Confidence Interval



Note: Coverage probabilities of the 95% confidence interval constructed using the SVR regression rankscore test and Wald test when the errors are heteroskedastic and Laplacian. The SVR regression rankscore confidence interval achieves the correct coverage probability independently of the bandwidth parameter for density estimation. In contrast, the Wald confidence interval achieves the correct coverage only at a sufficiently small bandwidth, as the theoretical test statistic corresponds to the limiting case where the bandwidth is 0. These results extend to all other error distributions considered in Table 3.1.

Here again, the analogy with quantile regression is fruitful. Although the asymptotic distribution of the regression coefficients inherently depends on the density of the regression errors, there are powerful statistical tests whose own asymptotic distribution does not depend on the density of the regression errors. By the duality between testing and inference, these tests may be inverted to produce confidence intervals for regression coefficients which do not require plugging in an estimate of a nonparametric density estimate or selecting a bandwidth parameter.

Contribution Our main contribution is to deliver what is, to the best of our knowledge, the first derivation of error bars for SVR which does not require distributional assumptions. We give the first rigorous treatment of large sample inference for SVR. We further improve on this by developing a bandwidth-free procedure based on the inversion of a novel regression rankscore test statistic.

Outline The remainder of the paper is organized as follows. Section 3.2 introduces the setup and notation. Section 3.3 introduces the inference procedure and theory. Section 3.4 presents a simulation study on the properties of the ℓ_1-SVR regression rankscore test, and compares the test against its natural competitor, the median regression rankscore test. Section 3.5 concludes. Technical proofs are deferred to the supplementary appendix.

3.2 Setup and Notation

Let $W_i = (Y_i, X_i, Z_i) \in \mathbb{R} \times \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$, $1 \le i \le n$ be i.i.d. random vectors. We assume the first element of X_i is 1. Let P denote the distribution of W_i. For a random variable (vector) A, define the vector (matrix) $\mathbf{A}_n = (A_1, \dots, A_n)'$. Let $Q_Y(x,z)$ denote the conditional median of Y given $X = x, Z = z$. We assume that this regression function is linear, i.e.,

$$Q_Y(x,z) = x'\beta(P) + z'\gamma(P) \ , \tag{3.1}$$

where $\beta(P) \in \mathbf{R}^{d_x}$ and $\gamma(P) \in \mathbf{R}^{d_z}$ are unknown parameters. We omit the dependence of β and γ on P whenever it is clear from the context.

The covariates are (X_i, Z_i). We distinguish X_i and Z_i to make transparent that the covariate Z_i is the one for which we conduct inference.

Consider the following ℓ_1-SVR:

$$\begin{aligned} \min_{(b,r)\in\mathbb{R}^{d_x}\times\mathbb{R}^{d_z}} \quad n^{-1} \sum_{1\le i\le n} &\max\{0, |Y_i - X_i'b - Z_i'r| - \epsilon\} + \\ &\lambda_n(\|b\|_1 + \|r\|_1) \ , \end{aligned} \tag{3.2}$$

where

$$\|b\|_1 = \sum_{1\le j\le d_x} |b_j|$$

and similarly for $\|r\|_1$.

As can be seen in Figure 3.2, the loss functions for median regression and support vector regression are similar, which suggests a close analogy between the methods and means for inference. In support vector regression, the errors are also penalized linearly, but only if they are bigger than ϵ in absolute value.

Define $F_Y(y|x,z)$ as the conditional distribution at $Y=y$ given $X=x$ and $Z=z$ and $f_Y(y|x,z)$ as the corresponding conditional density. We impose the following conditions on the distribution P.

Assumption 2. *The distribution P is such that*

(a) $0 < E\left[\begin{pmatrix} X_iX_i' & X_iZ_i' \\ Z_iX_i' & Z_iZ_i' \end{pmatrix} f_Y(X_i'\beta + Z_i'\gamma - \epsilon|X_i,Z_i)\right] < \infty.$

(b) $f_Y(y|x,z)$ *exists for all* $(y,x,z) \in \mathbb{R} \times \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$.

(c) $f_Y(\cdot|x,z)$ *is symmetric around* $x'\beta + z'\gamma$ *for all* $(x,z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$.

(d) $f_Y(x'\beta + z'\gamma - \epsilon|x,z) > 0$ *for all* $(x,z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$.

(e) *Define*

$$\Gamma = \big\{ (x,z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z} : \\ y \in [x'\beta + z'\gamma - c, x'\beta + z'\gamma + c] \big\}.$$

There exists $c > 0$ such that

$$\sup_{(x,z)\in\Gamma} \frac{|f_Y(y|x,z) - f_Y(x'\beta + z'\gamma|x,z)|}{|y - x'\beta - z'\gamma|} < \infty \ .$$

Assumption 2(a), (b) and (e) are commonly imposed in the quantile regression literature (Bai et al., 2019; Koenker, 2005) in order to establish the asymptotic distributions of estimators. Assumption 2(c)–(d) are imposed so that the estimators from ℓ_1-SVR are consistent for the coefficients of linear conditional medians. See Theorem 9.8 of Steinwart and Christmann (2008) for more details.

At times we will require the following homoskedasticity assumption on P. This strong but powerful assumption delivers the pivotal procedure.

Assumption 3. *The distribution P is such that*

(a) $f_Y(x'\beta + z'\gamma - \epsilon | x, z) = g(\epsilon)$ *for all* $(x, z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$, *for some function* g.

(b) $F_Y(x'\beta + z'\gamma - \epsilon | x, z) \equiv p_\epsilon$ *across all* $(x, z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$, *where the constant* p_ϵ *may depend on* ϵ.

Conditions similar to Assumption 3(a) are imposed, often implicitly, in the study of regression rankscore tests in quantile regression, so that the density terms cancel in the expression of the limiting variances of test statistics. As Bai et al. (2019) point out, the assumptions are tacitly imposed in the framework of fixed regressors, such as in Gutenbrunner and Jurecková (1992) and Gutenbrunner, Jurečková, Koenker, and Portnoy (1993). Assumption 3(b) is imposed so that the test statistic is simpler, but is not required in general. See Remark 3.3.4 for more details.

We impose the following condition on the tuning parameter λ_n. It is satisfied when $\lambda_n = \lambda/n$, where λ is a constant.

Assumption 4. $\lambda_n \to 0$ *as* $n \to \infty$.

Linear programming duality is at the core of the design of our proposed test statistic. Let $\mathbf{1}_d$ denote the $d \times 1$ vector of 1's. Note that the ℓ_1-SVR problem (3.2) has the following primal linear programming formulation.

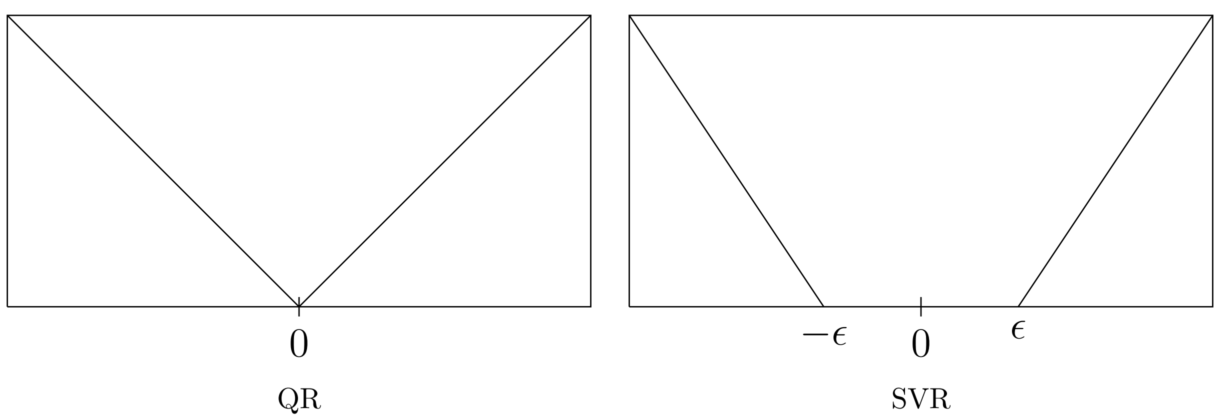
$$\max \mathbf{1}_n'\sigma + \lambda_n(\mathbf{1}_{d_x}'b^+ + \mathbf{1}_{d_x}'b^- + \mathbf{1}_{d_z}'r^+ + \mathbf{1}_{d_z}'r^-) \tag{3.3}$$

$$\begin{aligned} \text{subject to } u - v &= \mathbf{Y}_n - \mathbf{Z}_n r - \mathbf{X}_n b \\ \sigma - s &= u + v - \epsilon \mathbf{1}_n \\ b^+, b^-, r^+, r^-, u, v, \sigma, s &\geq 0 \ , \end{aligned}$$

where the optimization is over $b^+, b^-, r^+, r^-, u, v, \sigma, s$. See the supplementary appendix for the derivation of (3.3). From standard duality arguments, the dual of (3.3) is:

$$
\begin{aligned}
\max_{a^+,a^-} \quad & \mathbf{Y}_n' a^+ + \epsilon \mathbf{1}_n' a^- && (3.4)\\
\text{subject to} \quad & -\lambda_n \mathbf{1}_{d_x} \leq \mathbf{X}_n' a^+ \leq \lambda_n \mathbf{1}_{d_x} \\
& -\lambda_n \mathbf{1}_{d_z} \leq \mathbf{Z}_n' a^+ \leq \lambda_n \mathbf{1}_{d_z} \\
& a^- \leq a^+ \leq -a^- \\
& a^- \in [-1, 0]^n \, .
\end{aligned}
$$

Figure 3.2: Loss functions for median regression (QR) and for support vector regression (SVR).



3.3 Inference

Our construction of confidence regions builds on the duality between hypothesis testing and inference. The duality relies on the tautological statement that the confidence region made of all γ_0's for which the null $H_0 : \gamma = \gamma_0$ fails to be rejected at critical level α has coverage probability $1 - \alpha$.

Our error bars are obtained by test inversion as follows. Suppose, for simplicity of exposition, that $d_z = 1$. With a test statistic T_n and its null distribution in hand, we may find a critical value $c_{1-\alpha}$ such that

$$\lim_{n\to\infty} P\{T_n(\mathbf{W}_n, \gamma_0) \leq c_{1-\alpha}\} = 1-\alpha$$

$\forall P$ such that $\gamma(P) = \gamma_0$ and $\forall \gamma_0 \in \mathbb{R}$. Define the confidence region as

$$C_n = \{\gamma_0 \in \mathbb{R} : T_n(\mathbf{W}_n, \gamma_0) \leq c_{1-\alpha}\} \ .$$

We of course only obtain well-defined error bars if the test inverts to produce an interval. Importantly, Theorem 2 below ensures that the confidence region C_n is indeed an interval. This guarantees that the formula for the error bars for γ is given by

$$[\underline{\gamma}, \overline{\gamma}] := \left[\min_{\gamma_0 \in C_n} \gamma_0, \max_{\gamma_0 \in C_n} \gamma_0\right] \ .$$

For instance, Theorem 3.1 below justifies using $c_{0.95} = 1.96$ for error bars with 95% coverage, with a specific choice of $T_n(\mathbf{W}_n, \gamma_0)$.

We now introduce the test statistic for general values of d_z. For a prespecified $\gamma_0 \in \mathbb{R}^{d_z}$, we are interested in inverting tests of

$$H_0 : \gamma(P) = \gamma_0 \text{ versus } H_1 : \gamma(P) \neq \gamma_0 \tag{3.5}$$

at level $\alpha \in (0, 1)$.

For that purpose, consider the following "short" ℓ_1-SVR problem constructed by replacing

$\mathbf{Y}_n$ with $\mathbf{Y}_n - \mathbf{Z}_n\gamma_0$ and omitting $\mathbf{Z}_n$ from the regressors.

$$\begin{aligned} \max_{b^+,b^-,u,v,\sigma,s} \quad & \mathbf{1}_n'\sigma + \lambda_n(\mathbf{1}_{d_x}'b^+ + \mathbf{1}_{d_x}'b^-) \\ \text{subject to } & u - v = \mathbf{Y}_n - \mathbf{Z}_n\gamma_0 - \mathbf{X}_n(b^+ - b^-) \\ & \sigma - s = u + v - \epsilon\mathbf{1}_n \\ & b^+, b^-, u, v, \sigma, s \geq 0 \ . \end{aligned} \tag{3.6}$$

Define $\hat{\beta}_n$ as $b^+ - b^-$ where b^+ and b^- are part of the solution to (3.6). From standard duality arguments, the dual of (3.6) is

$$\begin{aligned} \max_{a^+,a^-} \quad & (\mathbf{Y}_n - \mathbf{Z}_n\gamma_0)'a^+ + \epsilon\mathbf{1}_n'a^- \\ \text{subject to } & -\lambda_n\mathbf{1}_{d_x} \leq \mathbf{X}_n'a^+ \leq \lambda_n\mathbf{1}_{d_x} \\ & a^- \leq a^+ \leq -a^- \\ & a^- \in [-1, 0]^n \ . \end{aligned} \tag{3.7}$$

Denote the solution to (3.7) by $\hat{a}^+$ and $\hat{a}^-$.

We collect a characterization of the complementary slackness condition for the ℓ_1-SVR problem which will be key in the design of our proposed test statistic.

Lemma 1. *The solution to* (3.7) *satisfies:*

(a) *If* $Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n > \epsilon$*, then* $\hat{a}_i^- = -1$ *and* $\hat{a}_i^+ = 1$*.*

(b) *If* $Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n < -\epsilon$*, then* $\hat{a}_i^- = -1$ *and* $\hat{a}_i^+ = -1$*.*

(c) *If* $0 < |Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n| < \epsilon$*, then* $\hat{a}_i^+ = \hat{a}_i^- = 0$*.*

Proof of Lemma 1. By complementary slackness,

$$|Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n| < \epsilon \Leftrightarrow s_i > 0 \Rightarrow \hat{a}_i^- = 0$$

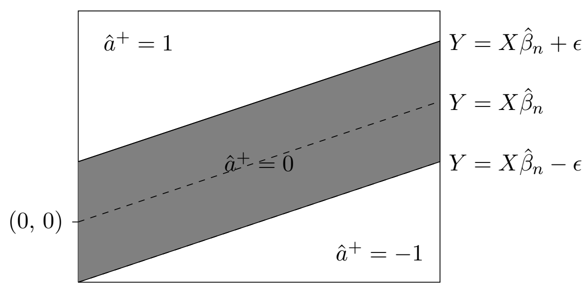
$$|Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n| > \epsilon \Leftrightarrow \sigma_i > 0 \Rightarrow \hat{a}_i^- = -1$$

$$Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n > 0 \Leftrightarrow u_i > 0 \Rightarrow \hat{a}_i^+ + \hat{a}_i^- = 0$$

$$Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n < 0 \Leftrightarrow v_i > 0 \Rightarrow \hat{a}_i^- - \hat{a}_i^+ = 0 \ ,$$

and the result follows. ■

Figure 3.3: $\hat{a}^+$ in different regions of regression residuals $Y - X\hat{\beta}_n$.



Following Bai et al. (2019), we construct the ℓ_1-*SVR regression rankscore test statistic* as

$$T_n(\mathbf{W}_n, \gamma_0) = \frac{n^{-1/2}\mathbf{Z}_n'\hat{a}^+}{\sqrt{n^{-1}\mathbf{Z}_n'\mathbf{M}_n\mathbf{Z}_n\hat{p}_n}} \ , \tag{3.8}$$

$$\mathbf{M}_n = I - \mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n' \tag{3.9}$$

and

$$\hat{p}_n = \frac{1}{n}\sum_{1\le i\le n} I\{|Y_i - X_i'\hat{\beta}_n - Z_i'\gamma_0| \ge \epsilon\}. \tag{3.10}$$

We define the ℓ_1-*SVR regression rankscore test* as

$$\phi_n(\mathbf{W}_n, \gamma_0) = I\{|T_n(\mathbf{W}_n, \gamma_0)| > z_{1-\frac{\alpha}{2}}\} \ , \tag{3.11}$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-th quantile of the standard normal distribution.

We give two intuitive interpretations for $T_n(\mathbf{W}_n, \gamma_0)$ in (3.8). Since $\lambda_n \to 0$, we intuitively view it as 0. First, as shown in Lemma 1 and Figure 3.3, $\hat{a}_i^+$ is a monotonic transformation of the residuals $\hat{u}_i = Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n$. If H_0 holds, i.e., $\gamma(P) = \gamma_0$, then we expect a low correlation between Z and $\hat{u}$, and any monotonic transformation of $\hat{u}$. Therefore $\mathbf{Z}_n'\hat{a}^+$ should be small under the null, but larger the more $\gamma(P)$ differs from γ_0. Second, consider running the SVR of $Y - Z'\gamma_0$ on X, with $\hat{a}^+$ as the solution to the dual problem. If H_0 holds, i.e., $\gamma(P) = \gamma_0$, then running a regression of $Y - Z'\gamma_0$ on X and Z should result in an estimated coefficient "close" to 0 on Z. Hence, including Z or not in the regression should have "close" to zero effect on the primal or dual results. Equivalently, adding the constraint $\mathbf{Z}_n'\hat{a}^+ = 0$ to (3.7) should not change the solution very much, so that $\mathbf{Z}_n'\hat{a}^+ = 0$ holds approximately when the null hypothesis holds, but may be large otherwise.

Remark 3.3.1. The complementary slackness conditions in (3.6) and (3.7) are the key ingredients in the construction of the test statistic in (3.8). These conditions are summarized in Lemma 1. Figure 3.3 displays the regions defined by regression residuals and the corresponding values of $\hat{a}^+$. For simplicity, we assume $d_x = 2$, the intercept is 0, and $\gamma_0 = 0$. Note that unit i contributes to $T_n(\mathbf{W}_n, \gamma_0)$ only when $|Y_i - X_i\hat{\beta}_n| > \epsilon$. The graph is different from that under the median regression, where the shaded region in which $\hat{a}_i^+ = 0$ collapses to a single line. ■

Remark 3.3.2. The ℓ_1-SVR regression rankscore test is equivalent to the median regression rankscore test when $\epsilon = \lambda = 0$ and $\hat{p}_n$ is set to $\frac{1}{2}$. In Section 3.4, we compare the finite sample performances of the two tests. ■

The following theorem is our main result. It establishes the asymptotic distribution of the test statistic defined in (3.8) under the null. As a consequence, it guarantees asymptotic exactness of the test defined in (3.11) in the sense that the limiting rejection probability under the null is equal to the nominal level.

Theorem 1. *Suppose P satisfies Assumption 2, λ_n satisfies Assumption 4, and P additionally satisfies the null hypothesis, i.e., $\gamma(P) = \gamma_0$. Then,*

$$\begin{aligned} & n^{-1/2}\mathbf{Z}_n'\hat{a}^+ \xrightarrow{d} \\ & \quad N(0, 2E[\tilde{Z}_i\tilde{Z}_i' F_Y(X_i'\beta + Z_i'\gamma_0 - \epsilon|X_i, Z_i)]) \ , \end{aligned} \tag{3.12}$$

where

$$\begin{aligned} \tilde{Z}_i = Z_i - E[Z_i X_i' f_Y(X_i'\beta + Z_i'\gamma_0 - \epsilon|X_i, Z_i)] \times \\ E[X_i X_i' f_Y(X_i'\beta + Z_i'\gamma_0 - \epsilon|X_i, Z_i)]^{-1} X_i \ . \end{aligned}$$

If P additionally satisfies Assumption 3, then

$$T_n(\mathbf{W}_n, \gamma_0) \xrightarrow{d} N(0, 1) \ ,$$

and therefore, for the problem of testing (3.5) at level $\alpha \in (0, 1)$, $\phi_n(\mathbf{W}_n)$ defined in (3.11) satisfies

$$\lim_{n \to \infty} E[\phi_n(\mathbf{W}_n, \gamma_0)] = \alpha \ .$$

Corollary 4. *Suppose P satisfies Assumptions 2–3 and the null hypothesis, and λ_n satisfies Assumption 4. Then the asymptotic variance in (3.12) can be consistently estimated without density estimation by*

$$\frac{1}{n}\mathbf{Z}_n'\mathbf{M}_n\mathbf{Z}_n\hat{p}_n$$

where $\mathbf{M}_n$ and $\hat{p}_n$ are defined in (3.9) and (3.10), respectively.

Remark 3.3.3. Let $\theta = (\beta, \gamma)$ and $V = (X, Z)$. Then

$$
\begin{aligned}
n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} N\Big(0, \frac{1}{2}\ & E[V_i V_i' f_Y(V_i'\theta - \epsilon \mid V_i)]^{-1} \times \\
& E[V_i V_i' F_Y(V_i'\theta - \epsilon \mid V_i)] \times \\
& E[V_i V_i' f_Y(V_i'\theta - \epsilon \mid V_i)]^{-1}\Big),
\end{aligned}
$$

from which the Wald confidence interval obtains. See the supplementary appendix for the derivation. However, as shown in Figure 3.1, Wald confidence intervals are highly sensitive to the bandwidth used for density estimation. ■

Remark 3.3.4. If P violates Assumption 3(b) but satisfies Assumption 3(a), it is straightforward to construct a consistent estimator of the asymptotic variance in (3.12) by using the law of iterated expectations. This estimator involves conditional distributions rather than densities, thus does not require a choice of bandwidth. The statistic studentized by the estimate of the asymptotic variance will be asymptotically exact. If P violates Assumption 3(a), it is also possible to construct consistent estimators of the asymptotic variance in (3.12). The statistic studentized by any consistent estimator of the asymptotic variance will also be asymptotically exact. However, the studentization involves density estimation, which is why Assumption 3(a) is important for our purposes. See, for example, Powell (1991). ■

According to Theorem 1, we could construct confidence regions by inverting the test $\phi_n(\mathbf{W}_n, \gamma_0)$ in (3.11). The following corollary shows that the limiting coverage probability of the confidence region is indeed correct.

Corollary 5. *Let $\phi_n(\mathbf{W}_n, \gamma_0)$ denote the test in* (3.11) *with level α. Define*

$$
C_n = \{\gamma_0 \in \mathbb{R} : \phi_n(\mathbf{W}_n, \gamma_0) = 0\} \ . \tag{3.13}
$$

Suppose P satisfied Assumptions 2 and 3, and λ_n satisfies Assumption 4. Then,

$$
\lim_{n \to \infty} P\{\gamma \in C_n\} = 1 - \alpha \ .
$$

Monotonicity of the test statistic is essential to the tractability of the inversion procedure, as it limits the procedure to a search for the two points where the test statistic, as a function of the posited null parameter, crosses the critical value. Note that without monotonicity, we cannot construct the confidence region as an interval. Error bars are obtained by inverting the test for each individual covariate. As discussed above, these are well-defined only insofar as the individual tests invert to produce intervals—as opposed to disconnected regions. Such a guarantee is given in Theorem 2.

Theorem 2. *If $d_z = 1$, then $T_n(\mathbf{W}_n, \gamma_0)$ in (3.8) is monotonically decreasing in γ_0.*

PROOF OF THEOREM 2. First note that the denominator of $T_n(\mathbf{W}_n, \gamma_0)$ is monotonically increasing in γ_0 because $\hat{p}_n$ in (3.10) is so. Next, we show the numerator is monotonically decreasing in γ_0. Denote by $\hat{a}^+(r)$ the solution to (3.7) with $\gamma_0 = r$. Given $r_1 > r_2$, by definition of the optimization problem,

$$(\mathbf{Y}_n - r_1\mathbf{Z}_n)'\hat{a}^+(r_1) - (\mathbf{Y}_n - r_1\mathbf{Z}_n)'\hat{a}^+(r_2) > 0 \ ,$$

while

$$(\mathbf{Y}_n - r_2\mathbf{Z}_n)'\hat{a}^+(r_1) - (\mathbf{Y}_n - r_2\mathbf{Z}_n)'\hat{a}^+(r_2) < 0 \ .$$

The two observations indicate that

$$r_1\mathbf{Z}_n'(\hat{a}^+(r_2) - \hat{a}^+(r_1)) > 0 > r_2\mathbf{Z}_n'(\hat{a}^+(r_2) - \hat{a}^+(r_1)) \ ,$$

so that

$$(r_1 - r_2)\mathbf{Z}_n'(\hat{a}^+(r_2) - \hat{a}^+(r_1)) > 0 \ ,$$

and the result follows since $r_1 > r_2$. ∎

3.4 Simulation

In this section, we investigate the finite sample performance of the ℓ_1-SVR regression rankscore test. As ℓ_1-SVR provides a robust estimate of the median, a natural benchmark for inference is the quantile regression rankscore test, which is the default method for inference for quantile regression (Koenker et al., 2019). Through a simulation study, we characterize the size, power, width of the error bars, and the robustness to misspecification of the homoskedasticity hypothesis for both the ℓ_1-SVR regression rankscore test and the median regression rankscore test. The results suggest that our test not only meets the standard of the default inference methodology for quantile regression, but in fact improves upon it under a wide variety of error distributions.

The data is generated according to the following equation,

$$Y = -0.8 + 2X + \gamma Z + G^{-1}(\tau),$$

where

$$(X, Z) \sim N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 10 & 4 \\ 4 & 8 \end{pmatrix}\right),$$

G^{-1} is the inverse-CDF for the error, τ is uniformly distributed over a subset of the $[0, 1]$ interval, and $(X, Z) \perp\!\!\!\perp \tau$. In all simulations, the sample size is 500. The results of the simulation carry over to cases of higher dimensional vectors of covariates, so long as the sample size is sufficiently large.

We consider five distributions for the error terms: Gaussian, Laplace, a symmetric mixture of Gaussian distributions, Student's t, and χ^2. The latter three distributions permit us to measure the performance of the test when the error distribution exhibits either multiple modes, fat tails, or asymmetry. For each distribution, we also consider two models which differ in the support restrictions on the error terms.

$$\textbf{Unrestricted support}: \ \tau \sim \text{Unif}[0,1].$$

$$\textbf{Restricted support}: \ \tau \sim \text{Unif}([0,0.4] \cup [0.6,1]).$$

The restricted support model may be thought of as an extreme version of the data generating processes which SVR is meant for when conceived as a regression extension of SVM classification.

Throughout the simulation, we set $\lambda_n = 0$. The parameter ϵ is adjusted according to the distribution of the error so that $G(\epsilon) - G(-\epsilon) = 0.2$. For symmetric distributions of the error term, ϵ thus excludes the quantiles in the interval $(0.4, 0.6)$. We set the true parameter $\gamma = 0$ under the null to study size properties, and $\gamma = 0.5$ under the alternative to study power properties at critical level $\alpha = 0.05$.

Columns 1–4 of Table 3.1 present the simulations in which the errors are homoskedastic. In all the cases of unrestricted support for the errors, the distributions are centered and scaled to be mean zero with a standard deviation of 15.[2] Upon restricting the support, the errors are normalized to be mean zero, but their standard deviation may change. Columns 1 and 3 indicate that the size properties of the SVR and median regression rankscore tests are about equal under homoskedasticity. However, columns 2 and 4 suggest that the SVR regression rankscore test has better power properties than median regression, the former outperforming the latter in 7 of the 10 settings.

Columns 5–8 of Table 3.1 present the simulations in which the errors are heteroskedastic, their variance being determined by the covariate X.[3] To account for the heteroskedastic errors,

[2] The mixture consists of evenly weighted Gaussian distributions centered around 10 and −10, with the same scale parameter. Errors under the Student's t distribution are drawn from the counterpart with 10 degrees of freedom, and then rescaled. Errors under the χ^2 distribution are drawn from the counterpart with 3 degrees of freedom, and then rescaled and recentered.

[3] The scale parameters of the distributions are set equal to a normalized value of X. For the Gaussian, Laplace, and mixture distributions, X is normalized to have a standard deviation of 1 and is then recentered so its mean is equal to the scale parameter required for the error distribution to have a standard deviation of 15. For the Student's t and χ^2 distributions, X is instead recentered around the degrees of freedom stated in

Table 3.1: Rejection Probabilities for Different Distributions of the Errors, Under Homo/Heteroskedasticity

	Homoskedastic				Heteroskedastic				Homo. stat, hetero. data			
	SVR		QR		SVR		QR		SVR		QR	
Distribution	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$
Gaussian	5.7	34.4	5.4	31.3	4.5	34.9	4.7	30.3	5.5	37.0	5.5	32.4
Gaussian, restricted	5.6	15.7	5.7	13.1	4.7	15.5	4.1	10.9	5.5	17.8	5.4	12.6
Laplace	5.6	64.8	5.3	65.4	4.2	63.6	4.7	67.4	5.5	67.9	5.7	70.4
Laplace, restricted	5.7	23.6	5.6	17.8	4.2	25.2	3.9	17.1	5.3	30.1	5.1	20.3
Mixture	4.6	37.9	4.3	30.6	4.5	37.7	4.4	32.1	4.7	38.8	4.7	32.7
Mixture, restricted	5.2	16.4	5.0	14.5	4.8	18.4	4.6	11.7	5.1	19.1	4.8	13.5
Student's t	5.7	40.1	5.4	35.9	4.4	37.0	4.6	31.7	5.7	40.2	5.3	35.9
Student's t, restricted	5.9	17.1	5.7	14.0	4.8	14.1	4.7	11.7	5.4	16.8	5.6	14.0
χ^2	5.1	39.2	5.4	39.6	4.7	16.3	3.7	15.4	4.7	17.0	4.4	17.3
χ^2, restricted	5.4	14.3	5.4	14.8	4.8	16.1	4.0	15.6	5.5	16.6	4.7	16.4

we obtain a consistent estimate of the variance expression in (3.12) using the methodology of Powell (1991). This requires density estimation of the errors, for which the tuning parameters are shown in Table 3.2.[4] The test statistic is then studentized using the variance estimate. We keep ϵ the same as in the homoskedastic simulations. As before, columns 5–8 of Table 3.1 suggest that the size properties of the two tests are roughly equal under heteroskedasticity, whereas the SVR regression rankscore test exhibits greater power in 9 of the 10 cases.

Table 3.2: Tuning Parameters for Heteroskedastic Test Statistic

	Unrestricted		Restricted	
Distribution	h	κ	h	κ
Gaussian	1	1.75	2.5	2
Laplace	1	0.75	1.5	1.75
Mixture	2	2	2.75	2.5
Student's t	1	1	2	2.75
χ^2	1.5	1.5	2.75	1.5

To gauge the robustness of the SVR regression rankscore test, columns 9–12 present simulations in which the homoskedastic test statistic is used with heteroskedastic errors. Similar to the earlier results, columns 9 and 11 suggest that both tests have similar size properties. However, columns 10 and 12 reveal that the SVR regression rankscore test demonstrates greater statistical power, having higher rejection rates in 8 of the 10 simulations. Similar robust behavior in heteroskedastic environments of regression rankscore tests constructed under homoskedasticity assumptions are documented using real and simulated data in Bai et al. (2019) and G. A. Pouliot (2019).

In each iteration of the simulations in Table 3.1, we can construct the confidence interval by inverting (3.11). The results in Table 3.1 suggest the SVR regression rankscore test has greater power against the alternatives considered when inverting the test, as compared to the

Footnote 2. In the rare event that the normalized X falls below 0, its absolute value is taken. While the distributions of parameters determining the standard deviation of the errors are centered around the value that would correspond to a standard deviation of 15, the standard deviations of the actual error terms need not be 15.

[4]The bandwidth used in density estimation is $\kappa\left(\Phi^{-1}(0.5 + hn^{-\frac{1}{3}}) - \Phi^{-1}(0.5 - hn^{-\frac{1}{3}})\right)$, where h and κ are constants, and Φ is the standard normal CDF.

Figure 3.4: Confidence Intervals Under Homoskedasticity

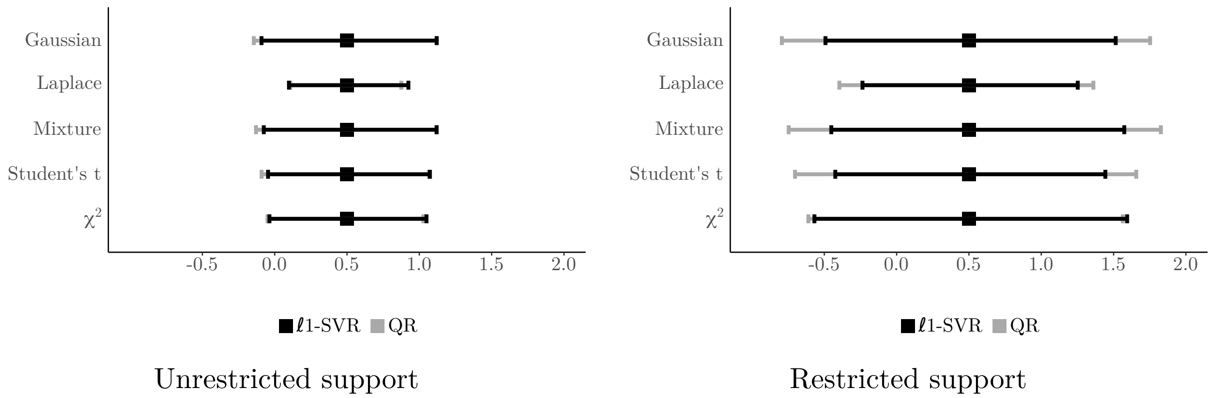
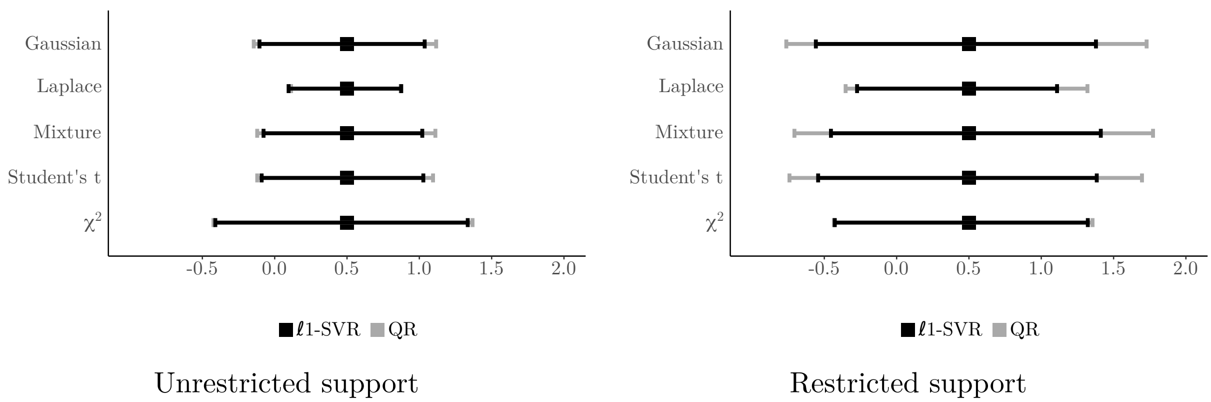


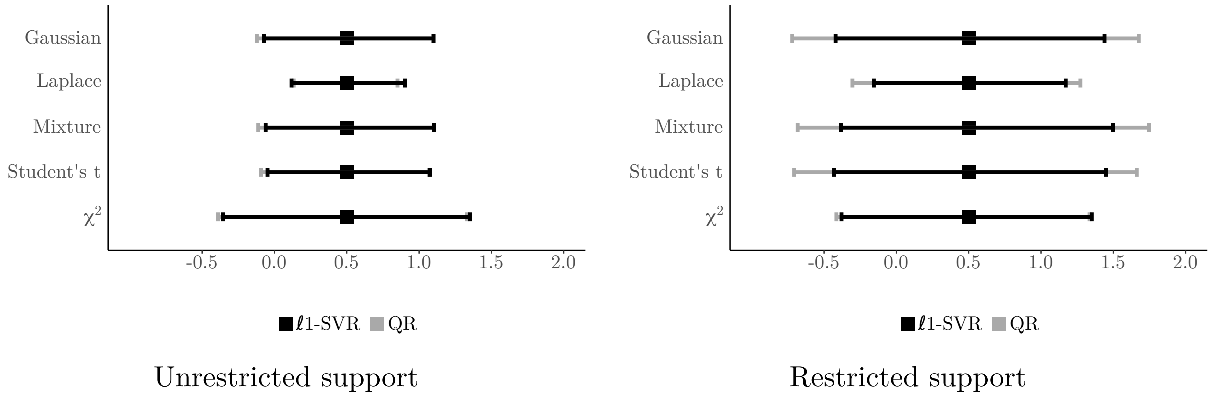
Figure 3.5: Confidence Intervals Under Heteroskedasticity



median regression rankscore test. From the duality between hypothesis testing and inference, this implies tighter confidence intervals under the former test procedure. Indeed, we find this to be the case. Figures 3.4–3.6 present the average confidence interval for each simulation where $\gamma = 0.5$. In all three figures, the results closely align with those of Table 3.1. That is, for the data generating process where the SVR regression rankscore test has greater power than the median regression rankscore test, the confidence interval of the former is narrower than that of the latter. The reduction in the error bars becomes rather substantial when we restrict the support of the error terms. This is to be expected, as the SVR loss function is able to account for the restricted support of the error term, whereas the median regression loss function cannot (see Figure 3.2).

It is rather remarkable that modifications to the quantile regression procedure intended

Figure 3.6: Confidence Intervals Under Heteroskedasticity Using Homoskedastic Test Statistic



for robustness deliver greater inference accuracy. This naturally suggests using the SVR regression rankscore test for inference in standard quantile regression analysis, even if the point estimate is obtained using quantile regression.

3.5 Conclusion

While SVR is largely used in practice, the lack of large sample inference methodology limits its practicality. In this article, we developed classical large sample inference, and furthermore delivered methodology producing asymptotically valid error bars while circumventing the need to select a bandwidth parameter. The ℓ_1-SVR regression rankscore test is the natural analog of the default procedure for inference in quantile regression. The asymptotic theory developed to establish the validity of the error bars is novel for SVR, and may be of independent interest. Remarkably, simulation evidence suggests that the regression rankscore test with our proposed regression rankscore test statistic may outperform the standard median regression rankscore test in inference for the regression parameters of the linear median regression function.

REFERENCES

Aigner, D. J., & Cain, G. G. (1977). Statistical Theories of Discrimination in Labor Markets. *Indutrial and Labor Relations Review*, *30*(2), 175–187.

Andrews, D. W. K., & Han, S. (2009). Invalidity of the bootstrap and the m out of n bootstrap for confidence interval endpoints defined by moment inequalities. *Econometrics Journal*, *12*, S172-S199.

Angrist, J. D., & Evans, W. N. (1998, June). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review*, *88*(3), 450-477.

Angrist, J. D., & Krueger, A. B. (2001, October). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives*, *15*(4), 69–85.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Angrist, J. D., & Pischke, J.-S. (2010, May). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, *24*(2), 3–30. doi: 10.1257/jep.24.2.3

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

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

Arnold, D., Dobbie, W., & Hull, P. (2020). *Measuring Racial Discrimination in Bail Decisions* (Tech. Rep.).

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, *133*(4), 1885–1932.

Ayres, I. (2002). Outcome Tests of Racial Disparities in Police Practices. *Justice Research and Policy*, *4*(1-2), 131–142.

Ba, B. A., Knox, D., Mummolo, J., & Rivera, R. (2021). The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago. *Science*, *371*(6530), 696–702.

Bai, Y., Pouliot, G. A., & Shaikh, A. M. (2019). *On regression rankscore inference* (Working Paper).

Baiocchi, M., Cheng, J., & Small, D. S. (2014). Instrumental variable methods for causal inference. *Statistics in Medicine*, *33*(13), 2297–2340. doi: 10.1002/sim.6128

Barnes, K. Y. (2004). Assessing the Counterfactual: The Efficacy of Drug Interdiction Absent Racial Profiling. *Duke LJ*, *54*, 1089.

Barnow, B. S., Cain, G. G., Goldberger, A. S., et al. (1980). *Issues in the analysis of selectivity bias*. University of Wisconsin, Inst. for Research on Poverty.

Basak, D., Pal, S., & Patranabis, D. C. (2007). Support vector regression. *Neural Information Processing-Letters and Reviews*, *11*(10), 203–224.

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

Becker, G. S. (1993). Nobel lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy*, *101*(3), 385–409.

Becker, G. S., & Chiswick, B. R. (1966). Education and the Distribution of Earnings. *The American Economic Review*, *56*(1/2), 358–369.

Bertsimas, D., & Tsitsiklis, J. N. (1997). *Introduction to linear optimization* (Vol. 6). Athena Scientific Belmont, MA.

Bhuller, M., Dahl, G. B., Løken, K. V., & Mogstad, M. (2020, April). Incarceration, Recidivism, and Employment. *Journal of Political Economy*, *128*(4), 1269–1324. doi: 10.1086/705330

Björklund, A., & Moffitt, R. (1987, February). The Estimation of Wage Gains and Welfare Gains in Self-Selection Models. *The Review of Economics and Statistics*, *69*(1), 42-49.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 436–455.

Bohren, J. A., Haggag, K., Imas, A., & Pope, D. G. (2020). *Inaccurate Statistical Discrimination: An Identification Problem* (Tech. Rep.). National Bureau of Economic Research.

Bohren, J. A., Imas, A., & Rosenberg, M. (2019). The Dynamics of Discrimination: Theory and Evidence. *American Economic Review*, *109*(10), 3395–3436.

Bollen, K. A. (2012, August). Instrumental Variables in Sociology and the Social Sciences. *Annual Review of Sociology*, *38*(1), 37–72. doi: 10.1146/annurev-soc-081309-150141

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, *131*(4), 1753–1794.

Brinch, C. N., Mogstad, M., & Wiswall, M. (2012). Beyond LATE with a Discrete Instrument. *Working paper*.

Brinch, C. N., Mogstad, M., & Wiswall, M. (2017, August). Beyond LATE with a Discrete Instrument. *Journal of Political Economy*, *125*(4), 985-1039. doi: 10.1086/692712

Brown, B. W., & Newey, W. K. (2002, October). Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference. *Journal of Business & Economic Statistics*, *20*(4), 507-517. doi: 10.1198/073500102288618649

Bugni, F. A., Canay, I. A., & Shi, X. (2015a). Specification Tests for Partially Identified Models Defined by Moment Inequalities. *Journal of Econometrics*, *185*(1), 259–282.

Bugni, F. A., Canay, I. A., & Shi, X. (2015b, March). Specification tests for partially identified models defined by moment inequalities. *Journal of Econometrics*, *185*(1), 259-282.

Bugni, F. A., Canay, I. A., & Shi, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, *8*(1), 1–38.

Canay, I. A., Mogstad, M., & Mountjoy, J. (2020). *On the Use of Outcome Tests for Detecing Bias in Decision Making* (Tech. Rep.).

Canay, I. A., & Shaikh, A. M. (2017). Practical and Theoretical Advances in Inference for Partially Identified Models. In B. Honore, A. Pakes, M. Piazzesi, & L. Samuelson (Eds.), *Advances in Economics and Econometrics* (p. 271-306). Cambridge University Press. doi: 10.1017/9781108227223.009

Carneiro, P., Heckman, J. J., & Vytlacil, E. J. (2011). Estimating Marginal Returns to Education. *American Economic Review*, *101*(6), 2754-81. doi: 10.1257/aer.101.6.2754

Carneiro, P., Lokshin, M., & Umapathi, N. (2016, May). Average and Marginal Returns to Upper Secondary Schooling in Indonesia. *Journal of Applied Econometrics*, *32*(1), 16–36. doi: 10.1002/jae.2523

Chan, D. C., Gentzkow, M., & Yu, C. (2019). *Selection with Variation in Diagnostic Skill: Evidence from Radiologists* (Tech. Rep.). National Bureau of Economic Research.

Chen, X. (2007). Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics* (Vols. Volume 6, Part 2, p. 5549-5632). Elsevier.

Cherkassky, V., & Ma, Y. (2004). Practical selection of svm parameters and noise estimation for svm regression. *Neural networks*, *17*(1), 113–126.

Chernozhukov, V., Newey, W. K., & Santos, A. (2020). *Constrained Conditional Moment Restriction Models* (Tech. Rep.).

Collaborators, G. . P. V. U. S., et al. (2021). Fatal Police Violence by Race and State in the USA, 1980–2019: A Network Meta-Regression. *The Lancet*, *398*(10307), 1239–1255.

Deaton, A. (2010, June). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, *48*(2), 424–455. doi: 10.1257/jel.48.2.424

DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica: Journal of the Econometric Society*, 1001–1044.

Doha, E., Bhrawy, A., & Saker, M. (2011). On the Derivatives of Bernstein Polynomials: An Application for the Solution of High Even-order Differential Equations. *Boundary Value Problems*, *2011*, 1–16.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. In *Advances in neural information processing systems* (pp. 155–161).

Durlauf, S. N., & Heckman, J. J. (2020). An Empirical Analysis of Racial Differences in Police Use of Force: A Comment. *Journal of Political Economy*, *128*(10), 000–000.

Engel, R. S., & Johnson, R. (2006). Toward a Better Understanding of Racial and Ethnic Disparities in Search and Seizure Rates. *Journal of Criminal Justice*, *34*(6), 605–617.

Farouki, R. T. (2012). The Bernstein Polynomial Basis: A Centennial Retrospective. *Computer Aided Geometric Design*, *29*(6), 379–419.

Farouki, R. T., & Rajan, V. (1988). Algorithms for polynomials in bernstein form. *Computer Aided Geometric Design*, *5*(1), 1–26.

Feigenberg, B., & Miller, C. (2021). *Would Eliminating Racial Disparities in Motor Vehicle Searches Impose Efficiency Costs?* (Tech. Rep.).

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition Methods in Economics. In *Handbook of labor economics* (Vol. 4, pp. 1–102). Elsevier.

Fryer Jr, R. G. (2019). An Empirical Analysis of Racial Differences in Police Use of Force. *Journal of Political Economy*, *127*(3), 1210–1261.

Gaebler, J., Cai, W., Basse, G., Shroff, R., Goel, S., & Hill, J. (2020). Deconstructing Claims of Post-treatment Bias in Observational Studies of Discrimination. *arXiv preprint arXiv:2006.12460*.

Gao, J. B., Gunn, S. R., Harris, C. J., & Brown, M. (2002). A probabilistic framework for svm regression and error bar estimation. *Machine Learning*, *46*(1-3), 71–89.

Gelbach, J. B. (2021). *Testing Economic Models of Discrimination in Criminal Justice* (Tech. Rep.).

Gelman, A., Fagan, J., & Kiss, A. (2007). An Analysis of the New York City Police Department's "Stop-and-frisk" Policy in the Context of Claims of Racial Bias. *Journal of the American statistical association*, *102*(479), 813–823.

Ghorbani, M., Zargar, G., & Jazayeri-Rad, H. (2016). Prediction of asphaltene precipitation using support vector regression tuned with genetic algorithms. *Petroleum*, *2*(3), 301–306.

Goel, S., Rao, J. M., & Shroff, R. (2016a). Personalized Risk Assessments in the Criminal Justice System. *American Economic Review*, *106*(5), 119–23.

Goel, S., Rao, J. M., & Shroff, R. (2016b). Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy. *The Annals of Applied Statistics*, *10*(1), 365–394.

Goncalves, F., & Mello, S. (2021). A Few Bad Apples? Racial Bias in Policing. *American Economic Review*, *111*(5), 1406–1441.

Grogger, J., & Ridgeway, G. (2006). Testing for Racial Profiling in Traffic Stops from Behind a Veil of Darkness. *Journal of the American Statistical Association*, *101*(475), 878–887.

Gurobi Optimization, Inc. (2015). Gurobi Optimizer Reference Manual.

Gutenbrunner, C., & Jurecková, J. (1992). Regression rank scores and regression quantiles. *The Annals of Statistics*, 305–330.

Gutenbrunner, C., Jurečková, J., Koenker, R., & Portnoy, S. (1993). Tests of linear hypotheses based on regression rank scores. *Nonparametric Statistics*, *2*(4), 307–331.

Hall, P., & Horowitz, J. L. (1996, July). Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. *Econometrica*, *64*(4), 891-916.

Hansen, L. P. (1982, July). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*(4), 1029-1054.

Heckman, J. (1997). Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *The Journal of Human Resources*, *32*(3), 441–462.

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*.

Heckman, J. J., Ichimura, H., Smith, J., & Todd, P. (1996). Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation. *Proceedings of the National Academy of Sciences*, *93*(23), 13416-13420.

Heckman, J. J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions. In J. J. Heckman & B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data.* Cambridge University Press.

Heckman, J. J., Urzua, S., & Vytlacil, E. (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics*, *88*(3), 389-432. doi: 10.1162/rest.88.3.389

Heckman, J. J., & Vytlacil, E. (2005a). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, *73*(3), 669–738.

Heckman, J. J., & Vytlacil, E. (2005b). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, *73*(3), 669-738.

Heckman, J. J., & Vytlacil, E. J. (1999, April). Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(8), 4730-4734.

Heckman, J. J., & Vytlacil, E. J. (2007a). Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics* (Vols. Volume 6, Part 2, p. 4779-4874). Elsevier.

Heckman, J. J., & Vytlacil, E. J. (2007b). Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics* (Vols. Volume 6, Part 2, p. 4875-5143). Elsevier.

Hernández-Murillo, R., & Knowles, J. (2004). Racial Profiling or Racist Policing? Bounds

Tests in Aggregate Data. *International economic review*, *45*(3), 959–989.

Horowitz, J. M., Brown, A., & Cox, K. (2019). *Race in America 2019.* Pew Research Center.

Hull, P. (2021). *What Marginal Outcome Tests Can Tell Us About Racially Biased Decision-Making* (Tech. Rep.).

Hurwicz, L. (1950). Systems with Nonadditive Disturbances. In T. Koopmans (Ed.), *Cowles 10* (pp. 410–418).

IBM. (2010). *IBM ILOG AMPL Version 12.2.* International Business Machines Corporation.

Imbens, G. W. (2014, August). Instrumental Variables: An Econometrician's Perspective. *Statistical Science*, *29*(3), 323–358. doi: 10.1214/14-STS480

Imbens, G. W., & Angrist, J. D. (1994, March). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467-475.

Jiao, R., Li, Y., Wang, Q., & Li, B. (2005). Svm regression and its application to image compression. In *International conference on intelligent computing* (pp. 747–756).

Jones, J. M. (2021). *In U.S., Black Confidence in Police Recovers from 2020 Low.* Gallup.

Kalinowski, J., Ross, M., & Ross, S. L. (2020). *Endogenous Driving Behavior in Tests of Racial Profiling in Police Traffic Stops* (Tech. Rep.). Working Paper.

Kennedy, E. H., Lorch, S., & Small, D. S. (2019). Robust causal inference with continuous instruments using the local instrumental variable curve. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *81*(1), 121–143. doi: 10.1111/rssb.12300

Kleiber, C., & Zeileis, A. (2018). Aer: Applied econometrics with r [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=AER (R package version 1.2-6)

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

Knox, D., Lowe, W., & Mummolo, J. (2020). Administrative Records Mask Racially Biased Policing. *American Political Science Review*, 1–19.

Koenker, R. (2005). *Quantile regression.* Cambridge University Press. Retrieved from https://books.google.com/books?id=WjOdAgAAQBAJ

Koenker, R., & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, *94*(448), 1296–1310.

Koenker, R., Portnoy, S., Ng, P. T., Zeileis, A., Grosjean, P., & Ripley, B. D. (2019). *R package* ***quantreg****.*

Konis, K. (2019). lpSolveAPI: R Interface to 'lp_solve' [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=lpSolveAPI (R package version 5.5.2.0-17.4)

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer.

Law, M. H., & Kwok, J. T.-Y. (2001). Bayesian support vector regression. In *Aistats.*

Li, J., West, S., & Platt, G. (2012). Power decomposition based on svm regression. In *2012 proceedings of international conference on modelling, identification and control* (pp. 1195–1199).

Li, P., Tan, Z., Yan, L., & Deng, K. (2011). Time series prediction of mining subsidence based on a svm. *Mining Science and Technology (China)*, *21*(4), 557–562.

MacDonald, J. M., & Fagan, J. (2019). Using Shifts in Deployment and Operations to Test for Racial Bias in Police Stops. In *Aea papers and proceedings* (Vol. 109, pp. 148–51).

Maestas, N., Mullen, K. J., & Strand, A. (2013). Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. *The American Economic Review*, *103*(5), 1797–1829.

Makofske, M. (2020). *Pretextual traffic stops and racial disparities in their use* (Tech. Rep.).

Manski, C. F. (1990, May). Nonparametric Bounds on Treatment Effects. *The American Economic Review*, *80*(2), 319-323.

Manski, C. F., & Nagin, D. S. (2017). Assessing Benefits, Costs, and Disparate Racial Impacts of Confrontational Proactive Policing. *Proceedings of the National Academy of Sciences*, *114*(35), 9308–9313.

Marx, P. (2021). An absolute Test of Racial Prejudice. *The Journal of Law, Economics, and Organization*.

Matzkin, R. L. (2007). Chapter 73 Nonparametric identification. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics* (Vols. Volume 6, Part 2, p. 5307-5368). Elsevier.

Metropolitan Nashville Police Department Manual. (2018). https://www.nashville.gov/Police-Department/Department-Manual.aspx. Nashville, Davidson County, TN.

Mogstad, M., Santos, A., & Torgovitsky, A. (2018a). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica*, *86*(5), 1589–1619.

Mogstad, M., Santos, A., & Torgovitsky, A. (2018b). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica*, *86*(5), 1589-1619. doi: 10.3982/ecta15463

Mogstad, M., & Torgovitsky, A. (2018, May). Identification and Extrapolation of Causal Effects with Instrumental Variables. *Annual Review of Economics*, *10*(1). doi: 10.1146/annurev-economics-101617-041813

Morin, R., Parker, K., Stepler, R., & Mercer, A. (2017). *Behind the Badge* (Tech. Rep.). Pew Research Center.

Morin, R., & Stepler, R. (2016). *The Racial Confidence Gap in Police Performance* (Tech. Rep.). Pew Research Center.

MOSEK ApS. (2021). MOSEK Licensing Guide.

Novak, K. J., & Chamlin, M. B. (2012). Racial threat, suspicion, and police behavior: The Impact of Race and Place in Traffic Enforcement. *Crime & Delinquency*, *58*(2), 275–300.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International economic review*, 693–709.

Ogburn, E. L., Rotnitzky, A., & Robins, J. M. (2015). Doubly robust estimation of the local average treatment effect curve. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *77*(2), 373–396. doi: 10.1111/rssb.12078

Owens, E. (2020). The Economics of Policing. *Handbook of Labor, Human Resources and Population Economics*, 1–30.

Pearl, J. (2011, March). Principal Stratification – a Goal or a Tool? *The International Journal of Biostatistics*, *7*(1), 1–13. doi: 10.2202/1557-4679.1322

Pelossof, R., Miller, A., Allen, P., & Jebara, T. (2004). An svm learning approach to robotic grasping. In *Ieee international conference on robotics and automation, 2004. proceedings. icra'04. 2004* (Vol. 4, pp. 3512–3518).

Pierson, E., Corbett-Davies, S., & Goel, S. (2018). Fast Threshold Tests for Detecting Discrimination. In *International conference on artificial intelligence and statistics* (pp. 96–105).

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., ... others (2020). A Large-scale Analysis of Racial Disparities in Police Stops Across the United States. *Nature human behaviour*, 1–10.

Pouliot, G. (2018). Equivalence of multicategory svm and simplex cone svm: Fast computations and statistical theory. In *International conference on machine learning* (pp. 4130–4137).

Pouliot, G. A. (2019). *Instrumental variables quantile regression with multivariate endogenous variable* (Working Paper).

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. *Nonparametric and semiparametric methods in Econometrics*, 357–384.

Ridgeway, G. (2006). Assessing the Effect of Race Bias in Post-Traffic Stop Outcomes Using Propensity Scores. *Journal of quantitative criminology*, *22*(1), 1–29.

Ridgeway, G., & MacDonald, J. M. (2009). Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops. *Journal of the American Statistical Association*, *104*(486), 661–668.

Robins, J. M., & Greenland, S. (1996). Identification of Causal Effects Using Instrumental Variables: Comment. *Journal of the American Statistical Association*, *91*(434), 456–458. doi: 10.2307/2291630

Roettger, M., Gelius-Dietrich, G., & Fritzemeier, J. C. (2019). cplexAPI: R Interface to C API of IBM ILOG CPLEX [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=cplexAPI (R package version 1.3.6)

Roh, S., & Robinson, M. (2009). A geographic approach to racial profiling: The microanalysis and macroanalysis of racial disparity in traffic stops. *Police quarterly*, *12*(2), 137–169.

Rosenbaum, P. R., & Rubin, D. B. (1983, April). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*(1), 41-55.

Rubin, H. (1950). Note on Random Coefficients. In T. Koopmans (Ed.), *Cowles 10* (pp. 419–421).

Sánchez, A. S., Nieto, P. G., Fernández, P. R., del Coz Díaz, J., & Iglesias-Rodríguez, F. J. (2011). Application of an svm-based regression model to the air quality study at local scale in the avilés urban area (spain). *Mathematical and Computer Modelling*, *54*(5-6), 1453–1466.

Simoiu, C., Corbett-Davies, S., & Goel, S. (2017). The Problem of Infra-Marginality in Outcome Tests for Discrimination. *The Annals of Applied Statistics*, *11*(3), 1193–1216.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199–222.

Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.

Stock, J. H., & Trebbi, F. (2003, August). Retrospectives: Who Invented Instrumental Variable Regression? *Journal of Economic Perspectives*, *17*(3), 177-194. doi: 10.1257/089533003769204416

Swanson, S. A., & Hernán, M. A. (2014, August). Think Globally, Act Globally: An Epidemiologist's Perspective on Instrumental Variable Estimation. *Statistical Science*, *29*(3), 371–374. doi: 10.1214/14-STS491

Trinkner, R., Kerrison, E. M., & Goff, P. A. (2019). The Force of Fear: Police Stereotype Threat, Self-Legitimacy, and Support for Excessive Force. *Law and human behavior*, *43*(5), 421.

Van Der Vaart, A. W., & Wellner, J. A. (1996). Weak convergence. In *Weak convergence and empirical processes* (pp. 16–28). Springer.

Vapnik, V. (1998). *Statistical learning theory*. John Wiley & Sons Inc., New York.

Vytlacil, E. (2002, January). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica*, *70*(1), 331-341.

Wang, W., & Yan, J. (2018). splines2: Regression spline functions and classes [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=splines2 (R package version 0.2.8)

Xi, X.-C., Poo, A.-N., & Chou, S.-K. (2007). Support vector regression model predictive control on a hvac plant. *Control Engineering Practice*, *15*(8), 897–908.

APPENDIX A

APPENDIX TO CHAPTER 1

A.1 Proofs

A.1.1 Deriving the random threshold in (1.1)

The officer wishes to maximize his expected utility. As shown in the main paper, the expected utility for decision $Search_i = s$ is

$$
\begin{aligned}
&\mathbb{E}[\mathcal{U}_i^s(Guilty_i; R_i) \mid R_i = r, Z_i = z, V_i = v] \\
&= G(r,z,v)\, \mathcal{U}_i^s(1; R_i) \; + (1 - G(r,z,v))\, \mathcal{U}_i^s(0; R_i) \\
&= \mathcal{U}_i^s(0; R_i) + G(r,z,v)\, (\mathcal{U}_i^s(1; R_i) - \mathcal{U}_i^s(0; R_i))
\end{aligned}
$$

So the officer chooses to search the driver if the expected utility from searching is at least as great as that of not searching, which is equivalent to

$$
\begin{aligned}
&\mathbb{E}[\mathcal{U}_i^1(Guilty_i; R_i) \mid R_i = r, Z_i = z, V_i = v] \geq \mathbb{E}[\mathcal{U}_i^0(Guilty_i; R_i) \mid R_i = r, Z_i = z, V_i = v] \\
\Longleftrightarrow\; &\mathcal{U}_i^1(0; R_i) + G(r,z,v) \left(\mathcal{U}_i^1(1; R_i) - \mathcal{U}_i^1(0; R_i)\right) \geq \mathcal{U}_i^0(0; R_i) + G(r,z,v) \left(\mathcal{U}_i^0(1; R_i) - \mathcal{U}_i^0(0; R_i)\right) \\
\Longleftrightarrow\; &G(r,z,v) \begin{bmatrix} \left(\mathcal{U}_i^1(1; R_i) - \mathcal{U}_i^1(0; R_i)\right) \\ - \left(\mathcal{U}_i^0(1; R_i) - \mathcal{U}_i^0(0; R_i)\right) \end{bmatrix} \geq \mathcal{U}_i^0(0; R_i) - \mathcal{U}_i^1(0; R_i) \\
\Longleftrightarrow\; &G(r,z,v) \geq \underbrace{\frac{\mathcal{U}_i^0(0; R_i) - \mathcal{U}_i^1(0; R_i)}{[\mathcal{U}_i^1(1; R_i) - \mathcal{U}_i^1(0; R_i)] - [\mathcal{U}_i^0(1; R_i) - \mathcal{U}_i^0(0; R_i)]}}_{\text{Random utility threshold } T_i} .
\end{aligned}
$$

The final line follows from Assumption 1(i), which ensures the denominator in the expression for T_i is strictly positive.

A.1.2 Proof of Corollary 1

Proof. The random threshold T_i is a deterministic function of the utilities $\{\mathcal{U}_i\}$. Properties (i)–(ii) of the corollary follow immediately from Assumptions 1(ii)–1(iii). Property (iii) follows immediately from Definition 1. ■

A.1.3 Deriving the search and hit rates

The search rate is derived as follows.

$$
\begin{aligned}
&\mathbb{E}[Search_i \mid R_i = r, Z_i = z] \\
&= \mathbb{E}[\mathbb{E}[Search_i \mid R_i = r, Z_i = z, V_i] \mid R_i = r, Z_i = z] && \text{(A.1)} \\
&= \mathbb{E}[\mathbb{E}[\mathbb{1}\{G(R_i, Z_i, V_i) \geq T_i\} \mid R_i = r, Z_i = z, V_i] \mid R_i = r, Z_i = z] && \text{(A.2)} \\
&= \mathbb{E}[F_{T|R}(G(r, z, V_i) \mid r) \mid R_i = r, Z_i = z] && \text{(A.3)} \\
&= \int_{\mathcal{V}} F_{T|R}(G(r, z, v) \mid r) \; dF_{V|R,Z}(v \mid r, z),
\end{aligned}
$$

where the first equality is by law of iterated expectations; the second equality is by substituting the definition of $Search_i$; the third equality follows from $T_i \perp\!\!\!\perp (Z_i, V_i) \mid R_i$ imposed by property (ii) in Corollary 1; the final equality follows by definition of conditional expectations.

The hit rate is derived as follows.

$$
\begin{aligned}
&\mathbb{E}[Hit_i \mid R_i = r, Z_i = z] \\
&= \mathbb{E}[\mathbb{E}[Hit_i \mid R_i = r, Z_i = z, V_i] \mid R_i = r, Z_i = z] \\
&= \int_{\mathcal{V}} \mathbb{E}[Hit_i \mid R_i = r, Z_i = z, V_i = v] \; dF_{V|R,Z}(v \mid r, z), && \text{(A.4)}
\end{aligned}
$$

where the first equality is by law of iterated expectations; and the second equality is by definition of conditional expectations. The expectation in the integrand may be written as

$$
\begin{aligned}
&\mathbb{E}[Hit_i \mid R_i = r, Z_i = z, V_i = v] \\
&= \mathbb{E}[Search_i \times Guilty_i \mid R_i = r, Z_i = z, V_i = v] \\
&= \mathbb{E}[Guilty_i \mid Search_i = 1, R_i = r, Z_i = z, V_i = v]\ \mathbb{E}[Search_i \mid R_i = r, Z_i = z, V_i = v] \\
&= \mathbb{E}[Guilty_i \mid G(r, z, v) > T_i, R_i = r, Z_i = z, V_i = v]\ \mathbb{E}[Search_i \mid R_i = r, Z_i = z, V_i = v] \\
&= \mathbb{E}[Guilty_i \mid R_i = r, Z_i = z, V_i = v]\ \mathbb{E}[Search_i \mid R_i = r, Z_i = z, V_i = v] \\
&= G(r, z, v)\ F_{T|R}(G(r, z, v) \mid r),
\end{aligned}
$$

where the first equality follows by definition of Hit_i; the second equality follows by law of iterated expectations, and that $Search_i \times Guilty_i = 0$ when $Search_i = 0$; the third equality follows from the definition of $Search_i$; the fourth equality follows from $T_i \perp\!\!\!\perp Guilty_i \mid R_i, Z_i, V_i$ from Corollary 1; and the final equality follows by definition of $G(\cdot, \cdot, \cdot)$, as well as from (A.1)–(A.3). Substituting this expression for $\mathbb{E}[Hit_i \mid R_i = r, Z_i = z, V_i = v]$ into (A.4) completes the derivation of the hit rate.

A.1.4 Testing for bias using η instead of σ

I show below that η depends on the race of the driver if an only if σ depends on the race of the driver. For all $g > 0$,

$$
\begin{aligned}
& \eta(\sigma(\cdot; w); g) = \eta(\sigma(\cdot; m); g) \\
\iff \quad & g\sigma(g; w) = g\sigma(g; m) \\
\iff \quad & \sigma(g; w) = \sigma(g; m).
\end{aligned}
$$

For $g = 0$, consider the graph of η,

$$
\begin{aligned}
& (\sigma(0,w), \eta(\sigma(\cdot;w);0)) = (\sigma(0,m), \eta(\sigma(\cdot;m);0)) \\
\iff & (\sigma(0,w), 0 \times \sigma(0;w)) = (\sigma(0,m), 0 \times \sigma(0;m)) \\
\iff & (\sigma(0,w), 0) = (\sigma(0,m), 0).
\end{aligned}
$$

So there exists a $g \in [0,1]$ such that $\eta(\sigma(\cdot;w);g) \neq \eta(\sigma(\cdot;m);g)$ if and only if $\sigma(g;w) \neq \sigma(g;m)$. Then $\eta(\sigma(\cdot;r);\cdot)$ depends on $r \in \{w,m\}$ if and only if $\sigma(\cdot;r)$ depends on $r \in \{w,m\}$.

A.1.5 Proof of Proposition 1

Proof. The proof proceeds in three steps. First, I show that there is a linear relationship between $\mathbb{E}[\mathit{Search}_i \mid R_i, Z_i]$ and $\mathbb{E}[\mathit{Hit}_i \mid R_i, Z_i]$. Second, I show that this linear relationship may be recovered by a linear IV regression. Third, I show that the officer is biased if the IV estimands differ by race.

The assumption that $\mathit{Var}[\mathit{Search}_i \mid R_i = r, Z_i] > 0$ is to rule out the cases where $\sigma(g_1;r) = \sigma(g_2;r) > 0$, or $\sigma(g_2;r) = 0$. In the first case, the observed search rates indicate the proportion of drivers the officer searches, regardless of their risk. If these rates differ across race, then the officer is immediately revealed to be biased. In the second case where $\sigma(g_2;r) = 0$, it must be that $\sigma(g_1;r) = 0$, since $\sigma(\cdot;r)$ is a non-decreasing function. It follows that no searches are observed at all, so $\alpha_0(r) = 0$ and α_1 is not well defined. But this is a trivial case that corresponds to an officer who never searches any driver, and the absence of any searches fully reveals the officer's preferences. So for the remainder of the proof, I assume $\sigma(g_2;r) > 0$ for $r \in \{w,m\}$, but allow $\sigma(g_1;r)$ to be 0.

To show the linear relationship between $\mathbb{E}[\mathit{Hit}_i \mid R_i, Z_i]$ and $\mathbb{E}[\mathit{Search}_i \mid R_i, Z_i]$, write the search and hit rates as

$$\mathbb{E}[Search_i \mid R_i = r, Z_i = z] = \sigma(g_1; r) + p_{r,z}(g_2)(\sigma(g_2; r) - \sigma(g_1; r)), \tag{A.5}$$

$$\mathbb{E}[Hit_i \mid R_i = r, Z_i = z] = g_1 \ \sigma(g_1; r) + p_{r,z}(g_2)(g_2 \ \sigma(g_2; r) - g_1 \ \sigma(g_1; r)). \tag{A.6}$$

Solving for $p_{r,z}(g_2)$ in (A.5), I have

$$p_{r,z}(g_2) = \frac{\mathbb{E}[Search_i \mid R_i = r, Z_i = z] - \sigma(g_1; r)}{\sigma(g_2; r) - \sigma(g_1; r)}.$$

Substituting this expression for $p_{r,z}(g_2)$ into (A.6) and grouping terms, I have

$$\mathbb{E}[Hit_i \mid R_i = r, Z_i = z] = \alpha_0(r) + \alpha_1(r)\mathbb{E}[Search_i \mid R_i = r, Z_i = z], \tag{A.7}$$

where

$$\begin{aligned}
\alpha_0(r) &= -\frac{\sigma(g_1; r)\sigma(g_2; r)(g_2 - g_1)}{\sigma(g_2; r) - \sigma(g_1; r)} \leq 0, \\
\alpha_1(r) &= \frac{g_2 \ \sigma(g_2; r) - g_1 \ \sigma(g_1; r)}{\sigma(g_2; r) - \sigma(g_1; r)} \quad > 0.
\end{aligned}$$

This establishes the linear relationship between $\mathbb{E}[Search_i \mid R_i, Z_i]$ and $\mathbb{E}[Hit_i \mid R_i, Z_i]$, and that $\alpha_0(r) \leq 0$ and $\alpha_1(r) > 0$.

To show that α_0 and α_1 are identified by a linear IV regression, let

$$\begin{aligned}
X_i' &\equiv (1, Search_i) \\
W_i' &\equiv (1, Z_i) \\
\alpha(r)' &\equiv (\alpha_0(r), \alpha_1(r)).
\end{aligned}$$

To simplify the proof, suppose $\mathcal{Z} = \{0, 1\}$ so that $\alpha(r)$ is just identified. Then the IV estimand is

$$
\begin{aligned}
&\mathbb{E}[W_i X_i' \mid R_i = r]^{-1}\ \mathbb{E}[W_i\ Hit_i \mid R_i = r] \\
&= \mathbb{E}[\mathbb{E}[W_i X_i' \mid R_i = r, W_i] \mid R_i = r]^{-1} \mathbb{E}[\mathbb{E}[W_i\ Hit_i \mid R_i = r, W_i] \mid R_i = r] \\
&= \mathbb{E}[W_i\ \mathbb{E}[X_i' \mid R_i = r, W_i] \mid R_i = r]^{-1} \mathbb{E}[W_i\ \mathbb{E}[Hit_i \mid R_i = r, W_i] \mid R_i = r] \\
&= \mathbb{E}[W_i\ \mathbb{E}[X_i' \mid R_i = r, W_i] \mid R_i = r]^{-1} \mathbb{E}[W_i\ \mathbb{E}[X_i' \mid R_i = r, W_i] \mid R_i = r]\ \alpha(r) \\
&= \alpha(r),
\end{aligned}
$$

where the first equality if by law of iterated expectations; the second equality is by linearity of expecations; the third equality follows from (A.7) and linearity of expectations; and the final equality follows from matrix algebra.

From the definitions of $\alpha_0(r)$ and $\alpha_1(r)$, it follows that the officer is biased if $\alpha(w) \neq \alpha(m)$. To see why the converse does not hold, suppose $\sigma(g_1; w) = \sigma(g_1; m) = 0$, $\sigma(g_2; w) \neq \sigma(g_2; m)$, and $\sigma(g_2; w), \sigma(g_2; m) > 0$. Then $\alpha_0(w) = \alpha_0(m) = 0$ and $\alpha_1(w) = \alpha_1(m) = g_2$, even though the officer has different search preferences for white and minority drivers with risk g_2. ■

A.1.6 Proof of Corollary 3

Proof. To build off the proof of Proposition 1, suppose Z_i is unobserved by the researcher. Suppose also that the officer is unbiased so that $\alpha(0) = \alpha(1) = \alpha$. Then the hit rate observed by the researcher for race r is

$$
\begin{aligned}
\mathbb{E}[Hit_i \mid R_i = r] &= \mathbb{E}[\mathbb{E}[Hit_i \mid R_i = r = Z_i] \mid R_i = r] \\
&= \mathbb{E}[\alpha_0 + \alpha_1 \mathbb{E}[Search_i \mid R_i = r, Z_i] \mid R_i = r] \\
&= \alpha_0 + \alpha_1 \mathbb{E}[\mathbb{E}[Search_i \mid R_i = r, Z_i] \mid R_i = r] \\
&= \alpha_0 + \alpha_1 \mathbb{E}[Search_i \mid R_i = r],
\end{aligned}
$$

where the first equality is by law of iterated expectations; the second equality follows from (A.7); the third equality follows from linearity of expectations; and the final equality follows from the law of iterated expectations.

Proposition 1 implies that $\alpha_0 \leq 0$ and $\alpha_1 > 0$ if the officer is unbiased. Then by contraposition, if $\alpha_0 > 0$ or $\alpha_1 \leq 0$, then the officer must be biased. ■

A.2 Constraints in the bilinear programming problem

This section provides some examples of how to impose linear constraints in the bilinear program, as well as motivates monotonicity restriction (1.11) on the distributions of risk.

A.2.1 Imposing linear constraints

Consider the vector of variables $\mathbf{x}' = (x_0, \ldots, x_K)$. The monotonicity constraint

$$x_0 \leq x_1 \leq \cdots \leq x_K \tag{A.8}$$

may be written as

$$\begin{bmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix} \mathbf{x} \geq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

To reverse the direction of monotonicity, simply reverse the inequalities. Linear constraints of the form

$$\sum_{k=0}^{K} a_k x_k \lesseqqgtr b \tag{A.9}$$

may be written as

$$\mathbf{a}'\mathbf{x} \lesseqqgtr b,$$

where $\mathbf{a}' = (a_0, \ldots, a_K)$.

To ensure that the search probabilities $\varsigma_r = (\sigma(g_0; r), \ldots, \sigma(g_K; r))$ that are being optimized over are consistent with being a CDF of $T_i \mid R_i = r$ for $r \in \{w, m\}$, ς_r must be non-decreasing in index k, and each element must be in the unit interval. The non-decreasing property of ς_r takes the form of (A.8), and the bounds on each element of ς_r take the form of (A.9) (i.e., choose $\mathbf{a}$ to be a standard basis vector).

To ensure that the distribution of risk $\mathbf{p}_{r,z}$ is consistent with being a PMF, the elements of $\mathbf{p}_{r,z}$ must be in the unit interval and sum to 1. Both of these constraints take the form of (A.9). The researcher may also choose to impose monotonicity constraints on $\mathbf{p}_{r,z}$. These will take the form of (A.8).

The researcher may want to rank the average risk of drivers by race R_i, setting Z_i, or both. This constraint is straightforward to impose. To see how, write the average risk conditional on race and setting as

$$\begin{aligned}\mathbb{E}[Guilty_i \mid R_i = r, Z_i = z] &= \sum_{k=0}^{K} g_k \mathbf{p}_{r,z,k} \\ &= \mathbf{g}'\mathbf{p}_{r,z},\end{aligned}$$

where $\mathbf{g}' = (g_0, \ldots, g_K)$ is the vector of discretized risks. Then the ranking

$$\mathbb{E}[Guilty_i \mid R_i = r_1, Z_i = z_1] \leq \mathbb{E}[Guilty_i \mid R_i = r_2, Z_i = z_2]$$

takes the form

$$\begin{aligned} & \sum_{k=0}^{K} g_k \mathbf{p}_{r_1,z_2,k} \leq \sum_{k=0}^{K} g_k \mathbf{p}_{r_2,z_2,k} \\ \iff & \sum_{k=0}^{K} g_k \mathbf{p}_{r_1,z_1,k} - \sum_{k=0}^{K} g_k \mathbf{p}_{r_2,z_2,k} \leq 0 \\ \iff & \mathbf{g}'(\mathbf{p}_{r_1,z_1} - \mathbf{p}_{r_2,z_2}) \leq 0. \end{aligned}$$

This restriction has the same form as (A.9), with $\mathbf{a}' = (\mathbf{g}', -\mathbf{g}')$ and $\mathbf{x}' = (\mathbf{p}'_{r_1,z_1}, \mathbf{p}'_{r_2,z_2})$.

A.2.2 Motivating restrictions on the distribution of risk

To provide an example for how the PDF of risk for drivers stopped may be decreasing in risk, I consider the following model for traffic stops. Let $Stop_i \in \{0, 1\}$ denote the stop decision of an officer for driver i. Data is only available for drivers who are stopped, for whom $Stop_i = 1$. Let $\mathcal{U}^p_{P,i}(R_i)$ denote the random utility of stop decision p for driver i, and $\mathcal{U}^s_{S,i}(Guilty_i; R_i)$ denote the random utility of searching driver i. The search utilities $\{\mathcal{U}^s_{S,i}\}$ are as in the main paper, except I have included the additional 'S' subscript to distinguish it from the utilities from stopping a driver.

Before stopping the driver, the officer observes R_i, Z_i, and V_i^{pre}, where V_i^{pre} is a subvector of $V_i' \equiv (V_i^{\text{pre}'}, V_i^{\text{post}'})$. So V_i^{pre} contains variables that the officer observes without having to make a stop, such as the make of the vehicle and the speed it was traveling at; and V_i^{post} includes variables that the officer only observes after stopping and interacting with the driver, such as the demeanor of the driver and the smell of the vehicle interior. As in the main paper, the researcher observes no components of V_i. The officer also knows the stop utilities $\{\mathcal{U}^p_{P,i}\}$ before stopping the driver, similar to how he knows $\{\mathcal{U}^s_{S,i}\}$ before searching the driver.

To make his stop decision, the officer considers the expected utility from stopping a driver and not stopping a driver. The reason why he maximizes the expected utility is because he does not know whether he will search the driver afterwards, and if he does, whether the driver will be guilty. So the officer's stop decision may be expressed as

$$\begin{aligned}
Stop_i \equiv \underset{p \in \{0,1\}}{\arg\max}\, & \mathbb{1}\{p = 1\} \left(\mathcal{U}^1_{P,i}(R_i) + \mathbb{E}[\mathcal{U}^{Search_i}_{S,i}(Guilty_i; R_i) \mid R_i, Z_i, V_i^{\text{pre}}]\right) \\
& + \mathbb{1}\{p = 0\}\mathcal{U}^0_{P,i}(R_i) \\
= \mathbb{1} & \left\{\mathcal{U}^1_{P,i}(R_i) + \mathbb{E}[\mathcal{U}^{Search_i}_{S,i}(Guilty_i; R_i) \mid R_i, Z_i, V_i^{\text{pre}}] \geq \mathcal{U}^0_{P,i}(R_i)\right\} \\
= \mathbb{1} & \left\{\mathbb{E}[\mathcal{U}^{Search_i}_{S,i}(Guilty_i; R_i) \mid R_i, Z_i, V_i^{\text{pre}}] \geq T_i^{Stop}\right\},
\end{aligned}$$

where $T_i^{Stop} \equiv \mathcal{U}^0_{P,i}(R_i) - \mathcal{U}^1_{P,i}(R_i)$ is a random utility threshold. To distinguish between the thresholds for stop and search decisions, let T_i^{Search} denote the utility threshold for searches.

Assumption 5. $\{\mathcal{U}^p_{P,i}\} \perp\!\!\!\perp (\{\mathcal{U}^s_{S,i}\}, Z_i, V_i^{\text{pre}})$.

Corollary 6. $T_i^{Stop} \perp\!\!\!\perp (T_i^{Search}, Z_i, V_i^{\text{pre}})$.

The independence between $\{\mathcal{U}^p_{P,i}\}$ and $\{\mathcal{U}^s_{S,i}\}$ is imposed to ensure that Assumption 1(ii)–1(iii) in the main paper is satisfied. To see why, suppose the stop and search preferences are correlated and let V_i^{post} contain $\{\mathcal{U}^p_{P,i}\}$. Then Assumption 1(iii) is immediately violated. Assumption 1(ii) is also violated since the officer's draws of $\{\mathcal{U}^s_{S,i}\}$ may differ for drivers i and j of race r with $\mathcal{U}^1_{P,i}(r) \neq \mathcal{U}^1_{P,j}(r)$. The independence between $\{\mathcal{U}^p_{P,i}\}$ and (Z_i, V_i^{pre}) is not require and is imposed to simplify the model.

Note that Assumption 5 does not imply there is no relationship between officers' stop and search preferences. That is, it does not preclude officers who are eager to stop minority drivers to also be eager to search minority drivers. Instead, it imposes that the draws of the random utilities/thresholds in the stop and search decision are independent of each other. This is admittedly a strong assumption, but without it, other strong assumptions are required in order to detect bias in searches while explicitly modeling traffic stops.

The probability the officer stops a driver is then

$$\begin{aligned}
&\mathbb{P}\{Stop_i = 1 \mid R_i = r, Z_i = z, V_i^{pre} = v\} \\
&= \mathbb{P}\{\mathbb{E}[\mathcal{U}_{S,i}^{Search_i}(Guilty_i; R_i) \mid R_i, Z_i, V_i^{\text{pre}}] \geq T_i^{Stop} \mid R_i = r, Z_i = z, V_i^{pre} = v\} \\
&= F_{T^{Stop}|R}(\mathbb{E}[\mathcal{U}_{S,i}^{Search_i}(Guilty_i; R_i) \mid R_i = r, Z_i = z, V_i^{\text{pre}} = v] \mid r),
\end{aligned}$$

where the last equality follows from Corollary 6. To see that this probability depends on the risk of the driver, we can apply the law of iterated expectations to the expectation inside of the CDF,

$$
\begin{aligned}
&\mathbb{E}[\mathcal{U}_{S,i}^{Search_i}(Guilty_i; R_i) \mid R_i = r, Z_i = z, V_i^{pre} = v] \\
&= \sum_{s=0}^{1} \mathbb{E}[\mathcal{U}_{S,i}^{s}(Guilty_i; R_i) \mid Search_i = s, R_i = r, Z_i = z, V_i^{pre} = v] \times \\
&\qquad\qquad\qquad\qquad \mathbb{P}\{Search_i = s \mid R_i = r, Z_i = z, V_i^{pre} = v\}.
\end{aligned}
$$

Consider the terms in the summand when $s = 1$. Applying the law of iterated expectations again, I have

$$
\begin{aligned}
&\mathbb{E}[\mathcal{U}_{S,i}^{1}(Guilty_i; R_i) \mid S_i = 1, R_i = r, Z_i = z, V_i^{pre} = v] \\
&= \mathbb{E}[\mathbb{E}[\mathcal{U}_{S,i}^{1}(Guilty_i; R_i) \mid S_i = 1, R_i = r, Z_i = z, V_i] \mid S_i = 1, R_i = r, Z_i = z, V_i^{pre} = v] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathcal{U}_{S,i}^{1}(Guilty_i; R_i) \mid \underbrace{G(r, z, V_i)}_{\text{Risk}} \geq T_i^{Search}, R_i = r, Z_i = z, V_i\right] \middle| \begin{array}{l} S_i = 1, R_i = r, \\ Z_i = z, V_i^{pre} = v \end{array}\right]
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{P}\{Search_i = 1 \mid R_i = r, Z_i = z, V_i^{pre} = v\} \\
&= \mathbb{E}[\mathbb{P}\{Search_i = 1 \mid R_i = r, Z_i = z, V_i\} \mid R_i = r, Z_i = z, V_i^{pre} = v] \\
&= \mathbb{E}[F_{T^{Search}|R}(\underbrace{G(r, z, V_i)}_{\text{Risk}} \mid r) \mid R_i = r, Z_i = z, V_i^{pre} = v],
\end{aligned}
$$

where the last equality follows from the model in Section 2.2 of the main paper.

Suppose that the reduced form relationship between $\mathbb{P}\{Stop_i = 1 \mid R_i = r, Z_i = z\}$ and $G(R_i, Z_i, V_i)$ is as shown in the top panel of Figure A.1, where the officer has a 1% probability of stopping a driver with zero risk, and a 50% probability of stopping a driver with unit risk. Denote this relationship by by $\pi_{r,z}$, i.e.,

$$
\pi_{r,z}(g) \equiv \mathbb{P}\{Stop_i = 1 \mid R_i = r, Z_i = z, G_i = g\}.
$$

Suppose also that the population distribution of risk is as shown in the middle panel, and is equal to a beta distribution with shape parameters 1 and 9 and a mean of 0.1, i.e., 10% of drivers carry contraband. Denote the density by $f_{G|R,Z}(\cdot \mid r, z)$. Then conditional on being stopped, the distribution of risk is as shown in the bottom panel of Figure A.1 and may be written as

$$f_{G|Stop,R,Z}(g \mid 1, r, z) = \frac{\pi_{r,z}(g)\ f_{G|R,Z}(g \mid r, z)}{\int_0^1 \pi_{r,z}(g')\ f_{G|R,Z}(g' \mid r, z)\ dg'}.$$

So in spite of the officer's preference for stopping high-risk drivers, the proportion of low-risk drivers in population is sufficiently large so that the density of risk post-stop is strictly decreasing.

A.2.3 Recovering the identified sets $\mathcal{B}_k$ and $\mathcal{E}$

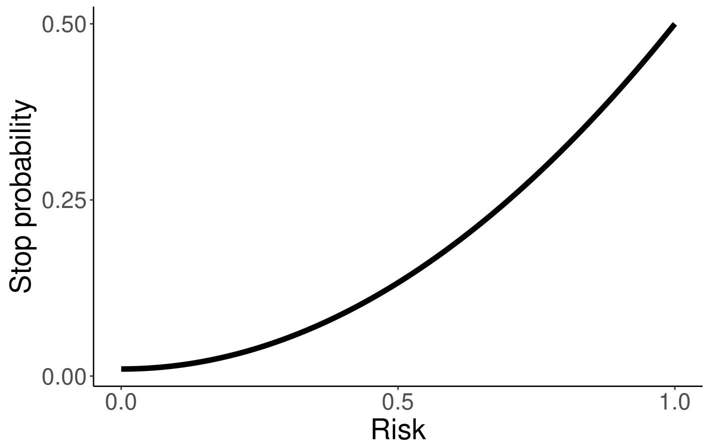
The identified set for $\beta(g_k)$ may be recovered by solving

$$\begin{aligned}
Q^{\star}_{\beta_k}(b) \equiv & \min_{\varsigma_w, \varsigma_m, \{\mathbf{p}_{w,z}\}, \{\mathbf{p}_{m,z}\}} \sum_{r,z} \left| \varsigma_r' \mathbf{p}_{r,z} - \mathbf{m}^S_{r,z} \right| + \sum_{r,z} \left| (\mathbf{g} \odot \varsigma_r)' \mathbf{p}_{r,z} - \mathbf{m}^H_{r,z} \right| \\
\text{s.t.} \quad & \varsigma_{m,k} - \varsigma_{w,k} = b \\
& \mathbf{A} \begin{bmatrix} \varsigma_w \\ \varsigma_m \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b}
\end{aligned}$$

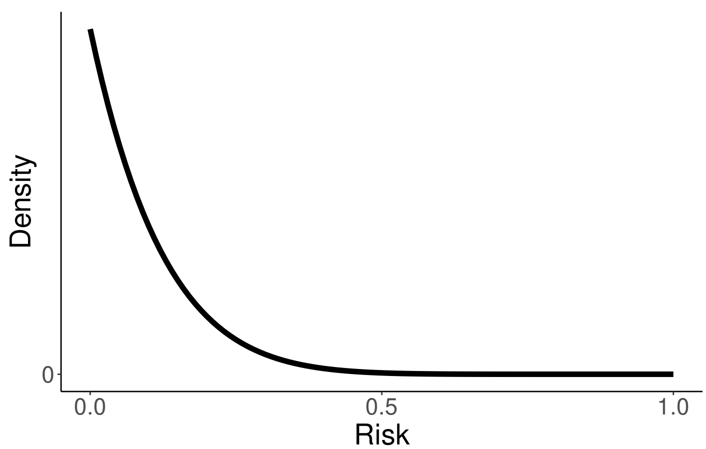
for all $b \in [-1, 1]$. Then $b \in \mathcal{B}_k$ if and only if $Q^{\star}_{\beta_k}(b) = 0$.

Figure A.1: Monotone-decreasing density for risk of drivers stopped

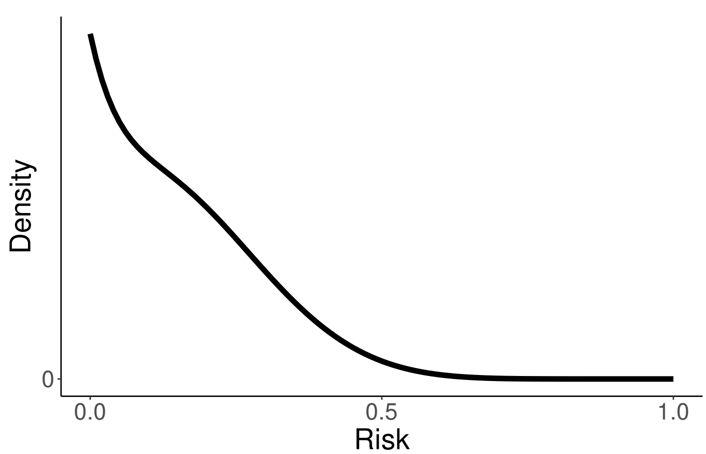
(a) Probability of stopping a driver



(b) Population distribution of risk



(c) Sample distribution of risk



Likewise, the identified set for $\mathbb{E}[\beta(G_i);\omega]$ may be recovered by solving

$$
\begin{aligned}
Q^{\star}_{\mathbb{E}[\beta(G);\omega]}(b) \equiv & \min_{\varsigma_w,\varsigma_m,\{\mathbf{p}_{w,z}\},\{\mathbf{p}_{m,z}\}} \sum_{r,z} \left|\varsigma_r' \mathbf{p}_{r,z} - \mathbf{m}^S_{r,z}\right| + \sum_{r,z} \left|(\mathbf{g}\odot\varsigma_r)'\mathbf{p}_{r,z} - \mathbf{m}^H_{r,z}\right| \\
\text{s.t.} \quad & \omega'(\varsigma_m - \varsigma_w) = b \\
& \mathbf{A}\begin{bmatrix} \varsigma_w \\ \varsigma_m \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b}
\end{aligned}
$$

for all $b \in [-1, 1]$. Then $b \in \mathcal{E}$ if and only if $Q^{\star}_{\mathbb{E}[\beta(G);\omega]}(b) = 0$.

A.2.4 Constructing confidence intervals for $\beta(g_k)$ and $\mathbb{E}[\beta(G_i);\omega]$

The confidence intervals for $\beta(g_k)$ for $k = 1, \ldots, K$ may be constructed by inverting the test for racial bias. To determine whether $b \in [-1, 1]$ is in the confidence interval, the researcher must solve

$$
\begin{aligned}
\widehat{Q}^{\star}_{\beta(g_k)}(b) \equiv & \min_{\varsigma_w,\varsigma_m,\{\mathbf{p}_{r,z}\}} \sum_{r,z} \widehat{\mathbf{w}}^S_{r,z} \left|\varsigma_r' \mathbf{p}_{r,z} - \widehat{\mathbf{m}}^S_{r,z}\right| + \sum_{r,z} \widehat{\mathbf{w}}^H_{r,z} \left|(\mathbf{g}\odot\varsigma_r)'\mathbf{p}_{r,z} - \widehat{\mathbf{m}}^H_{r,z}\right| \\
\text{s.t.} \quad & \varsigma_{m,k} - \varsigma_{w,k} = b \\
& \mathbf{A}\begin{bmatrix} \varsigma_w \\ \varsigma_m \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b},
\end{aligned}
$$

which is the BP problem in (1.16) with the additional constraint that the intensity of bias at g_k is equal to b. The researcher can then construct the test statistic

$$\widehat{\tau}_{\beta(g_k)}(b) = \frac{\widehat{Q}^{\star}_{\beta(g_k)}(b) - \widehat{Q}^{\star}_{B}}{\widehat{Q}^{\star}_{B}},$$

which compares the fit of the model when the officer is restricted to have $\beta(g_k) = b$ against the fit without the restriction. The distribution of $\widehat{\tau}_{\beta(g_k)}(b)$ may be estimated using the bootstrap, and the hypothesis that $\beta(g_k) = b$ is rejected if the α-quantile of the bootstrap distribution is sufficiently large, for some value of $\alpha \in [0,1]$. If the hypothesis is not rejected, then b enters into the $(1-\alpha)$-confidence interval of $\beta(g_k)$. Again, this heuristic approach is not guaranteed to generate confidence intervals with the correct coverage probabilities, but may still be informative and is a stand-in until a formal method for inference is developed.

The confidence intervals for $\mathbb{E}[\beta(G_i);\omega]$ maybe constructed in the same way. First solve

$$\begin{aligned}
\widehat{Q}^{\star}_{\beta(g_k)}(b) \equiv \min_{\varsigma_w, \varsigma_m, \{\mathbf{p}_{r,z}\}} & \sum_{r,z} \widehat{\mathbf{w}}^{S}_{r,z} \left| \varsigma_r' \mathbf{p}_{r,z} - \widehat{\mathbf{m}}^{S}_{r,z} \right| + \sum_{r,z} \widehat{\mathbf{w}}^{H}_{r,z} \left| (\mathbf{g} \odot \varsigma_r)' \mathbf{p}_{r,z} - \widehat{\mathbf{m}}^{H}_{r,z} \right| \\
\text{s.t.} \quad & \mathbf{q}_w' \mathbf{P}_w (\varsigma_m - \varsigma_w) = b \\
& \mathbf{A} \begin{bmatrix} \varsigma_w \\ \varsigma_m \\ \mathbf{p}_{w,1} \\ \vdots \\ \mathbf{p}_{m,|\mathcal{Z}|} \end{bmatrix} \leq \mathbf{b},
\end{aligned}$$

which is the BP problem in (1.16) with the additional constraint that the average intensity of bias is equal to b. The following test statistic may then be constructed,

$$\widehat{\tau}_{\mathbb{E}[\beta(G_i);\omega]}(b) = \frac{\widehat{Q}^{\star}_{\mathbb{E}[\beta(G_i);\omega]}(b) - \widehat{Q}^{\star}_{B}}{\widehat{Q}^{\star}_{B}},$$

which compares the fit of the model when the officer is restricted to have $\mathbb{E}[\beta(G_i);\omega] = b$

against the fit without the restriction. The same bootstrap procedure above may then be applied to determine whether b is in the confidence interval of $\mathbb{E}[\beta(G_i);\omega]$.

A.3 Bernstein polynomials

In this section, I briefly discuss some properties of Bernstein polynomials. See Farouki and Rajan (1988), Doha, Bhrawy, and Saker (2011), and Farouki (2012) for more details.

The Bernstein basis of degree L is defined by

$$\mathbf{b}_l^L(g) \equiv \binom{L}{l}(1-g)^{L-l}g^l$$

for $l = 0, \dots L$ and $g \in [0,1]$. A Bernstein polynomial of degree L has the form

$$f(g) = \sum_{l=0}^{L} \theta_l \mathbf{b}_l^L(g)$$

for some $\theta \equiv (\theta_0, \dots, \theta_L)$.

Suppose $\sigma^\star$ is modeled as a Bernstein polynomial, i.e., $\sigma^\star(g) = f(g)$. To see how this affects the bilinear program, consider the bilinear terms $\varsigma' \mathbf{p}_{r,z}$ from the objective function of the bilinear program in Proposition 2. These terms become

$$\varsigma' \mathbf{p}_{r,z} = \sum_{k=0}^{K} \sum_{l=0}^{L} \mathbf{b}_l^L(g_k) \underbrace{\theta_l \mathbf{p}_{r,z,k}}_{\text{Bilinear terms}},$$

where $\{\mathbf{b}_l^L(g_k)\}_{l=0,\dots,L;k=0\dots K}$ are known values. The BP program optimizes over θ and $\{\mathbf{p}_{r,z}\}$.

Imposing shape constraints on Bernstein polynomials is straightforward. The polynomial $f(g)$ satisfies

$$\min_l \theta_l \leq f(g) \leq \max_l \theta_l.$$

So a Bernstein polynomial may be bounded above or below simply by bounding its coefficients. For example, f may be constrained to be in the unit interval by imposing the restriction $0 \leq \theta_l \leq 1$ for $l = 1, \dots, L$.

To impose that f is monotonic increasing, add the restriction

$$\theta_0 \leq \theta_1 \leq \cdots \leq \theta_L,$$

which has the same form as (A.8). To impose that f is monotonic decreasing, simply reverse the inequalities.

The derivative of a Bernstein polynomial is also a Bernstein polynomial. So the shape constraints above may be used to constrain the derivatives of $f(g)$ as well. Doha et al. (2011) show that the q^{th} derivative of $f(g)$ is

$$f^{(q)}(g) = \sum_{l=0}^{L} \sum_{i=-q}^{q} \theta_{l-i}\ C_i(l, L, q)\ \mathbf{b}_l^L(g),$$

where

$$C_i(l, L, q) = q! \sum_{j=0}^{q} (-1)^{j+q} \binom{q}{j} \binom{l}{j+i} \binom{L-l}{q-j-i}.$$

So $f^{(q)}(g)$ is a Bernstein polynomial of degree L with coefficients $\left\{\sum_{i=-q}^{q} \theta_{l-i}\ C_i(l, L, q)\right\}_{l=0}^{L}$, where $C_i(l, L, q)$ are known constants. The derivatives may then also be restricted to fall within some interval and be monotonic.

A product of Bernstein polynomials is also a Bernstein polynomial. For instance, let

$$h(g) = \sum_{n=0}^{N} \pi_n \mathbf{b}_n^N(g)$$

for some $\pi \equiv (\pi_0, \ldots, \pi_N)$. Then

$$f(g)\ h(g) = \sum_{i=0}^{L+N} \left[\sum_{j=\max\{0,i-N\}}^{\min\{L,i\}} \underbrace{\frac{\binom{L}{j}\binom{N}{i-j}}{\binom{L+N}{i}}}_{\text{Known}} \underbrace{\theta_j\ \pi_{i-j}}_{\text{Bilinear terms}} \right] \mathbf{b}_i^{L+N}(g). \tag{A.10}$$

This means it is possible to model both σ and $\{\mathbf{p}_{r,z}\}$ as Bernstein polynomials. The bilinear program optimizes over θ and π.

Using the fact that the integral of any Bernstein basis polynomial of degree L over the unit interval is $(1+L)^{-1}$, the integral of (A.10) is

$$\begin{aligned}\int_0^1 f(g)\ h(g)\ dg &= \int_0^1 \sum_{i=0}^{L+N} \left[\sum_{j=\max\{0,i-N\}}^{\min\{L,i\}} \frac{\binom{L}{j}\binom{N}{i-j}}{\binom{L+N}{i}} \theta_j\ \pi_{i-j} \right] \mathbf{b}_i^{L+N}(g)\ dg \\ &= \sum_{i=0}^{L+N} \left[\sum_{j=\max\{0,i-N\}}^{\min\{L,i\}} \frac{\binom{L}{j}\binom{N}{i-j}}{\binom{L+N}{i}} \theta_j\ \pi_{i-j} \right] \int_0^1 \mathbf{b}_i^{L+N}(g)\ dg \\ &= \frac{1}{1+L+N} \sum_{i=0}^{L+N} \left[\sum_{j=\max\{0,i-N\}}^{\min\{L,i\}} \underbrace{\frac{\binom{L}{j}\binom{N}{i-j}}{\binom{L+N}{i}}}_{\text{Known}} \underbrace{\theta_j\ \pi_{i-j}}_{\text{Bilinear terms}} \right].\end{aligned} \tag{A.11}$$

Then by letting $f(g)$ in denote the search probability (i.e., $\sigma(g;r)$) and $h(g)$ denote the distribution of risk (i.e., $f_{G|R,Z}(g \mid r,z)$), equation (A.11) may be interpreted as the search rate of an officer, $\mathbb{P}\{Search_i \mid R_i = r, Z_i = z\}$. To obtain the expression of the hit rate, $\mathbb{P}\{Hit_i \mid R_i = r, Z_i = z\}$, I can substitute the expression for $f(g)$ in (A.11) for that of $g\ f(g)$. The expression for $g\ f(g)$ may be obtained using (A.10) since g and $f(g)$ are Bernstein polynomials,[1]

$$g\ f(g) = \sum_{l=0}^{L} \frac{\binom{L}{l}}{\binom{L+1}{l+1}} \theta_l \mathbf{b}_{l+1}^{L+1}(g). \tag{A.12}$$

So parameterizing σ and $f_{G|R,Z}$ allows for continuous risk while preserving the bilinear programming framework.

If $f(g)$ in (A.12) were a density function of a random variable, then the integral of (A.12) is the first moment of a random variable. Equation (A.12) can be generalized to express higher order moments as well.

[1]Specifically, g is a Bernstein polynomial of degree 1 with a coefficient $\theta = (0, 1)$.

Lemma 2. *Let random variable $G \in [0,1]$ have density h, where h is a Bernstein polynomial of degree N with coefficients $\pi_0, \ldots, \pi_N$. Then for $L \geq 1$,*

$$g^L h(g) = \sum_{i=0}^{L+N} \widetilde{\pi}_i \ \mathbf{b}_i^{L+N}(g).$$

where

$$\widetilde{\pi}_i = \begin{cases} 0 & \text{if } i < L, \\ \frac{\binom{N}{i-L}}{\binom{L+N}{i}} \pi_{i-L} & \text{if } i \geq L. \end{cases} \tag{A.13}$$

Proof. Equation (A.13) may be proven by recursively applying (A.10) via induction.

To show that (A.13) holds for $L = 1$, consider (A.10) when f is a Bernstein polynomial of degree 1 with coefficients $(\theta_0, \theta_1) = (0, 1)$ so that $f(g) = g$. Let $\widetilde{\pi}_i$ denote the i^{th} coefficient for the polynomial $\widetilde{h}(g) = f(g)\ h(g)$, i.e., $\widetilde{\pi}_i$ is the inner summation in (A.10),

$$\widetilde{\pi}_i = \sum_{j=\max\{0,i-N\}}^{\min\{L,i\}} \frac{\binom{L}{j}\binom{N}{i-j}}{\binom{L+N}{i}} \theta_j \ \pi_{i-j}. \tag{A.14}$$

The index j is restricted to be either 0 or 1, with the summand in (A.14) being 0 whenever $j = 0$ since $\pi_0 = 0$.

- If $i = 0$, then $j \in \{0\}$ and therefore $\widetilde{\pi}_0 = 0$.

- If $i \geq 1$, then $j \in \{0, 1\}$ and therefore

$$\begin{aligned} \widetilde{\pi}_i &= \frac{\binom{1}{1}\binom{N}{i-1}}{\binom{1+N}{i}} \pi_1 \ \pi_{i-1} \\ &= \frac{\binom{N}{i-1}}{\binom{1+N}{i}} \ \pi_{i-1}. \end{aligned}$$

This establishes that (A.13) holds for $L = 1$.

To complete the inductive step, suppose $\widetilde{h}(g) = g^L\ h(g)$ has coefficients $\widetilde{\pi}$ that satisfy (A.13) for arbitrary L so that $\widetilde{h}$ is a Bernstein polynomial of degree $L+N$. Then it suffices to show that (A.13) continues to hold for $g\ \widetilde{h}(g)$, whose coefficients are denoted by $\widehat{\pi}$. From (A.10) we have

$$\begin{aligned}
\widehat{\pi}_i &= \begin{cases} 0 & \text{if } i = 0 \\ \frac{\binom{L+N}{i-1}}{\binom{1+L+N}{i}}\ \widetilde{\pi}_{i-1} & \text{if } i \geq 1 \end{cases} \\
&= \begin{cases} 0 & \text{if } i < L+1 \\ \frac{\binom{L+N}{i-1}}{\binom{1+L+N}{i}}\ \frac{\binom{N}{(i-1)-L}}{\binom{L+N}{i-1}} \pi_{(i-1)-L} & \text{if } i \geq L+1 \end{cases} \\
&= \begin{cases} 0 & \text{if } i < L+1 \\ \frac{\binom{N}{i-(L+1)}}{\binom{(L+1)+N}{i}}\ \pi_{i-(L+1)} & \text{if } i \geq L+1 \end{cases}
\end{aligned}$$

where the second equality follows from substituting in the expression for $\widetilde{\pi}$, and the final equality follows from collecting terms. The final expression satisfies (A.13). ■

Corollary 7. *It follows from Lemma 2 that*

$$\begin{aligned}
\int_0^1 g^L\ h(g)\ dg &= \frac{1}{L+N+1} \sum_{i=0}^{L+N} \widetilde{\pi}_i \\
&= \frac{1}{L+N+1} \sum_{i=0}^{N} \frac{\binom{N}{i}}{\binom{L+N}{L+i}} \pi_i.
\end{aligned}$$

Proof. The first equality follows from

$$\int_0^1 \mathbf{b}_n^N(g)\ dg = \frac{1}{N+1}$$

and

$$g^L\ h(g) = \sum_{i=0}^{L+N} \widetilde{\pi}_i \mathbf{b}_n^{L+N}(g).$$

The second equality follows from omitting the first L terms in $g^L\ h(g)$, whose coefficients are equal to 0, and substituting in the expression for $\widetilde{\pi}$ stated in (A.13). ■

An application of Lemma 2 is to impose restrictions on the distribution of risk over time. For example, suppose the distribution of drivers stopped and search preferences are fixed across time, but the officer's ability to assess the risk of drivers changes across time. Let $G_{i,t}$ denote the distribution of (post-stop) risk in time period t. If officers have rational expectations, then

$$\mathbb{E}[G_{i,t_1} \mid R_i, Z_i] = \mathbb{E}[G_{i,t_2} \mid R_i, Z_i] \text{ for any } t_1, t_2. \tag{A.15}$$

However, an officer's ability to asses risk may improve over time so that the variance of risk grows over time,

$$Var[G_{i,t_1} \mid R_i, Z_i] \leq Var[G_{i,t_2} \mid R_i, Z_i] \text{ for any } t_1 < t_2. \tag{A.16}$$

This can happen if officers become more observant with experience such that the vector V_i grows over time.[2] Lemma 2 makes it straightforward to constrain the distributions of

[2]Consider the variance of risk when the officer observes V_{i1} versus (V_{i1}, V_{i2}), where I have suppressed the notation for race and setting for brevity,

$$\begin{aligned}
&Var[\mathbb{E}[Guilty_i \mid V_{i1}, V_{i2}]] - Var[\mathbb{E}[Guilty_i \mid V_{i1}]] \\
&= \mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}, V_{i2}]^2] - \mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}, V_{i2}]]^2 - \left(\mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}]^2] - \mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}]]^2\right) \\
&= \mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}, V_{i2}]^2] - \mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}]^2],
\end{aligned}$$

where the first equality follows by definition of the variance, and the second equality follows from applying the law of iterated expectations and collecting terms. The final line may be shown to be non-negative using risk $G_{i,t}$ to satisfy (A.15)–(A.16) over time.[3] Moreover, because the officer faces different distributions of risk over time, the search and hit rates from different periods may be treated as distinct moments, thereby increasing the number of moments for identifying the officer's search preferences.

$$\begin{aligned} Var[G_{i,t_1} \mid R_i, Z_i] &\geq Var[G_{i,t_2} \mid R_i, Z_i] \\ \Rightarrow \mathbb{E}[G^2_{i,t_1} \mid R_i, Z_i] - \mathbb{E}[G_{i,t_1} \mid R_i, Z_i]^2 &\geq \mathbb{E}[G^2_{i,t_2} \mid R_i, Z_i] - \mathbb{E}[G_{i,t_2} \mid R_i, Z_i]^2 \\ \mathbb{E}[G^2_{i,t_1} \mid R_i, Z_i] &\geq \mathbb{E}[G^2_{i,t_2} \mid R_i, Z_i] \end{aligned}$$

for $t_1 < t_2$, where the second line follows from the definition of $Var[G_{i,t} \mid R_i, Z_i]$, and the third line follows from $\mathbb{E}[G_{i,t_1} \mid R_i, Z_i] = \mathbb{E}[G_{i,t_2} \mid R_i, Z_i]$.

An implication of the officer becoming more observant over time is that his hit rate should increase over time, which is another constraint that may be added to the model.

Lemma 3. *The hit rate of the officer increases with the number of driver characteristics observed by the officer.*

Proof. Let the driver's risk be $G_i \equiv \mathbb{E}[Guilty_i \mid V_{i1}, V_{i2}]$, where I have suppressed the notation

Jensen's inequality,

$$\begin{aligned} \mathbb{E}[Guilty_i \mid V_{i1}]^2 &= \mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}, V_{i2}] \mid V_{i1}]^2 \\ &\leq \mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}, V_{i2}]^2 \mid V_{i1}] \\ \Rightarrow \mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}]^2] &\leq \mathbb{E}[\mathbb{E}[Guilty_i \mid V_{i1}, V_{i2}]^2]. \end{aligned}$$

[3] Lemma 2 provides expressions for the moments of risk. Equation (A.15) can be satisfied by restricting the first moments of risk to be constant over time conditional on race and setting. Equation (A.16) can be satisfied by restricting the second moments of risk to be non-increasing over time conditional on race and setting, since

$$\begin{aligned} Var[G_{i,t_1} \mid R_i, Z_i] - Var[G_{i,t_2} \mid R_i, Z_i] &= \mathbb{E}[G^2_{i,t_1} \mid R_i, Z_i] - \mathbb{E}[G_{i,t_1} \mid R_i, Z_i]^2 - \\ &\quad \left(\mathbb{E}[G^2_{i,t_2} \mid R_i, Z_i] - \mathbb{E}[G_{i,t_2} \mid R_i, Z_i]^2\right) \\ &= \mathbb{E}[G^2_{i,t_1} \mid R_i, Z_i] - \mathbb{E}[G^2_{i,t_2} \mid R_i, Z_i], \end{aligned}$$

where the second equality follows because the first moment of risk is constant across time conditional on race and setting.

for race and setting for brevity. The hit rate is the average risk of drivers searched. It suffices to show that the distribution of risk conditional on being searched when the officer observes (V_{i1}, V_{i2}) first-order stochastically dominates that of when he observes only V_{i1}.

When the officer is inexperienced and only observes V_{i1}, he bases his search decision on $\mathbb{E}[\mathit{Guilty}_i \mid V_{i1}] = \mathbb{E}[G_i \mid V_{i1}]$. Let f^{Inexp} denote the density of risk in this scenario. So the density of risk conditional on V_{i1} and being searched is

$$\begin{aligned} f^{\mathit{Inexp}}_{G|\mathit{Search},V_1}(g \mid 1, v_1) &= \frac{\int_{\mathcal{V}(v_1,g)} \sigma(\mathbb{E}[G_i \mid V_{i1} = v_1])\ f_{V_1,V_2}(v_1, v_2)\ dv_2}{\int_{\mathrm{supp}(V_2)} \sigma(\mathbb{E}[G_i \mid V_{i1} = v_1])\ f_{V_1,V_2}(v_1, v_2)\ dv_2} \\ &= \frac{\int_{\mathcal{V}(v_1,g)}\ f_{V_1,V_2}(v_1, v_2)\ dv_2}{\int_{\mathrm{supp}(V_2)}\ f_{V_1,V_2}(v_1, v_2)\ dv_2}, \end{aligned}$$

where

$$\mathcal{V}(v_1, g) \equiv \{v_2 : \mathbb{E}[\mathit{Guilty}_i \mid V_{i1} = v_1, V_{i2} = v_2] = g\}.$$

When the officer is experienced and observes both (V_{i1}, V_{i2}), he bases his search decision on risk G_i. Let f^{Exp} denote the density of risk in this scenario. So the density of risk conditional on V_{i1} and being searched is

$$\begin{aligned} f^{\mathit{Exp}}_{G|\mathit{Search},V_1}(g \mid 1, v_1) &= \frac{\int_{\mathcal{V}(v_1,g)} \sigma(\mathbb{E}[G_i \mid V_{i1} = v_1, V_{i2} = v_2])\ f_{V_1,V_2}(v_1, v_2)\ dv_2}{\int_{\mathrm{supp}(V_2)} \sigma(\mathbb{E}[G_i \mid V_{i1} = v_1, V_{i2} = v_2])\ f_{V_1,V_2}(v_1, v_2)\ dv_2} \\ &= \frac{\sigma(g)\ \int_{\mathcal{V}(v_1,g)} f_{V_1,V_2}(v_1, v_2)\ dv_2}{\int_{\mathrm{supp}(V_2)} \sigma(\mathbb{E}[G_i \mid V_{i1} = v_1, V_{i2} = v_2])\ f_{V_1,V_2}(v_1, v_2)\ dv_2}. \end{aligned}$$

Taking the ratio of the densities of risk for the experienced and inexperienced officers, we have

$$\frac{f^{\mathit{Exp}}_{G|\mathit{Search},V_1}(g \mid 1, v_1)}{f^{\mathit{Inexp}}_{G|\mathit{Search},V_1}(g \mid 1, v_1)} \propto \sigma(g),$$

which is non-decreasing in g. This holds for all values of v_1. Therefore, conditional on being searched, the distribution of risk when the officer observes (V_{i1}, V_{i2}) first-order stochastically dominates that of when he only observes V_{i1}. It follows that the average risk of drivers searched when the officer observes (V_{i1}, V_{i2}) must be at least as large as that of when he only observes V_{i1}. ■

A.4 Additional results

A.4.1 Alternative estimates of average bias

Figure A.2 presents the estimated bounds on the average bias for the 8 officers who fail the test. However, unlike the estimates in the main paper, the bias is restricted to be either non-negative (anti-minority) or non-positive (anti-white). The red bounds correspond to the bias being averaged over the distribution of risk for white (minority drivers and indicate how much more white drivers would be searched if they were treated as minorities. The blue bounds correspond to the bias being averaged over the distribution of risk of minority drivers and indicate how much more minority drivers are being searched compared to if they were treated as whites.

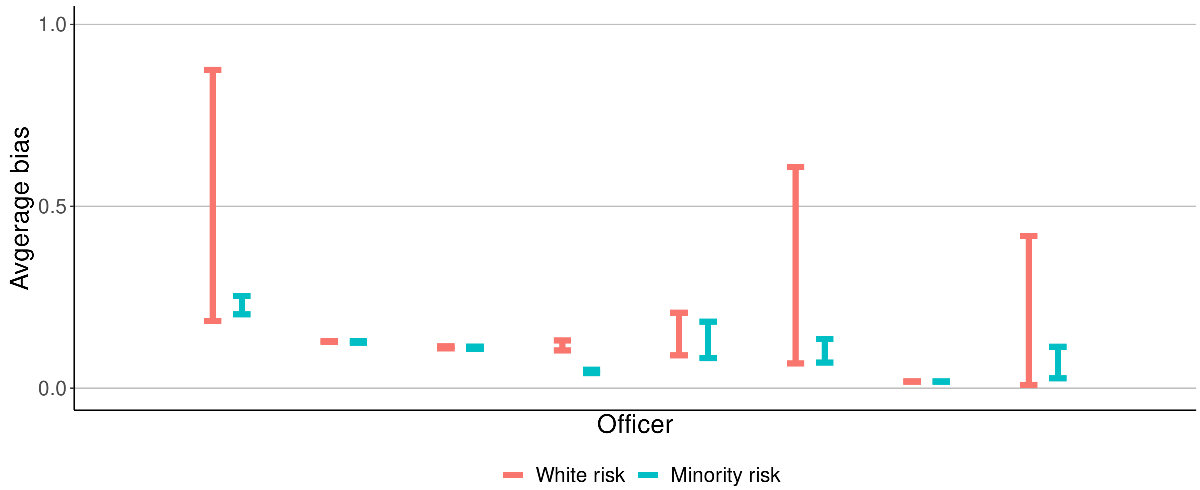
Figure A.2a presents the estimated bounds when the officer is restricted to be biased against minorities. Consider first the red bounds. Weighting the officers by the number of stops they have made for white drivers, the average of the red lower bounds is 6.5 percentage points, which is 0.3 percentage points greater than the estimates from the main paper. Now consider the blue bounds. Weighting the officers by the number of stops they have made for minority drivers, the average lower bound is 6.1 percentage points, which is 1.6 percentage points greater than the estimates in the main paper.

Figure A.2b presents the estimated bounds when the officer is restricted to be biased against whites. With the exception of the rightmost blue bounds, all the estimated bounds are very close to 0. This indicates that, given the constraint that the bias is non-positive, the data is best fit when the bias is close to 0, which provides evidence against such a constraint. The large negative lower bound for the rightmost bar in Figure A.2b is because the estimation procedure was unable to converge within a reasonable time and a conservative estimate is reported.

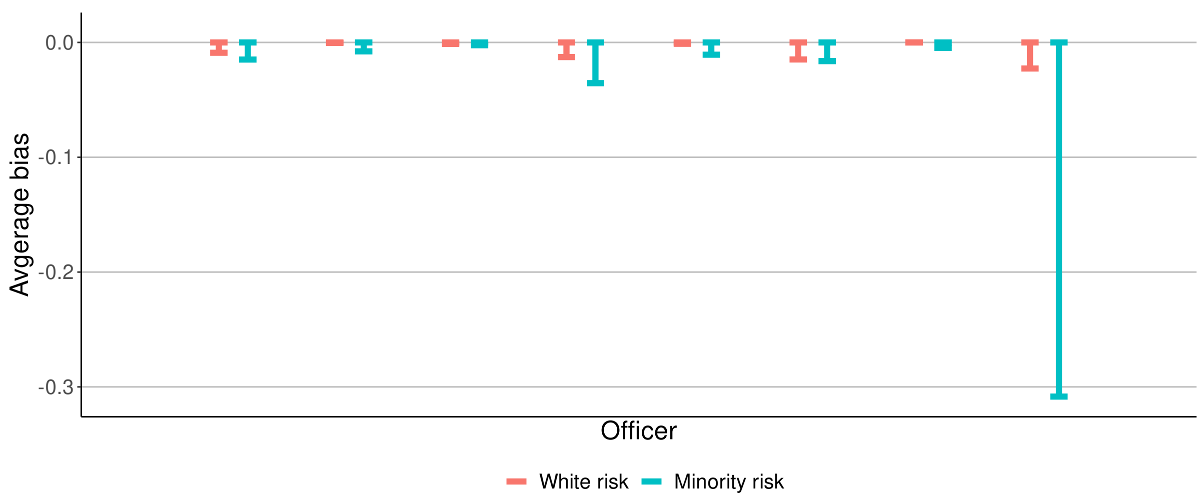
Figure A.3 presents the estimated bounds when the risk is continuous, without restrictions on the direction of bias. The bounds are estimated by modeling σ and $F_{G|R,Z}$ using Bernstein

Figure A.2: Bounds on average bias $\mathbb{E}[\beta(G_i);\omega]$ for biased officers

(a) With constraint $\beta(G_i) \geq 0$

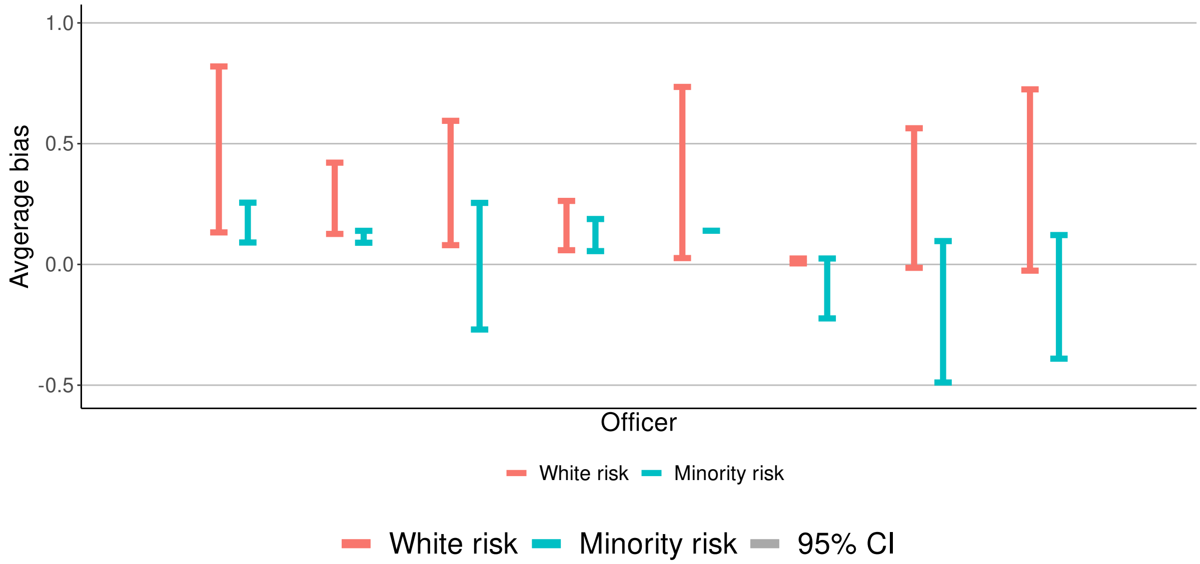


(b) With constraint $\beta(G_i) \leq 0$



Note: When the constraint $\beta(G_i) \geq 0$ is imposed, the bias is restricted to be non-negative so that officers search minority drivers at least as much as equally risky white drivers on average. When the constraint $\beta(G_i) \leq 0$ is imposed, the bias is restricted to be non-positive so that officers search white drivers at least as much as equally risky minority drivers on average. Red (blue) bounds indicate the average bias when ω is equal to the distribution of risk for white (minority) drivers.

Figure A.3: Bounds on average bias $\mathbb{E}[\beta(G_i); \omega]$ for biased officers, continuous risk



Note: σ and $F_{G|R,Z}$ are modeled using Bernstein polynomials of degree 20. Positive average bias indicates that the officer searches minority drivers more often than equally risky white drivers on average. Red (blue) bounds indicate the average bias when ω is equal to the distribution of risk for white (minority) drivers.

polynomials of degree 20. The average of the red lower bounds is 2.8 percentage points, which is 3.4 percentage points less than the estimates from the main paper, but is still 42% of the observed search rate of white drivers (6.7%). The average of the blue lower bounds is -25.3 percentage points, which is almost 30 percentage points below the average reported in the main paper and suggests that officers are in fact biased against white drivers. The large differences between the estimates in Figure 1.8 of the main paper and Figure A.3 are partly because the former estimates involve solving a more difficult computational problem with more unknown parameters and complex bilinear constraints. As a result, many of the bounds presented in Figure A.3 are suboptimal and very conservative.[4] Future versions of the paper will include improved estimates for the case where risk is continuously distributed, as well as when there are additional constraints on the direction of bias.

[4] The bounds shown are the solutions to the dual problem returned by Gurobi after running for 30 minutes on 25 processors. For many of these estimates, no solution to the primal problem is returned, so the mixed integer programming gap is unknown.

A.4.2 Estimates for all 50 officers

The weights ω are chosen so that $\mathbb{E}[\beta(G_i);\omega]$ measures the average difference in the probability that equally risky white and minority drivers are searched, under the counterfactual where the distribution of risk for minority drivers is equal to that of white drivers in the data.

Figure A.4: All officer level estimates

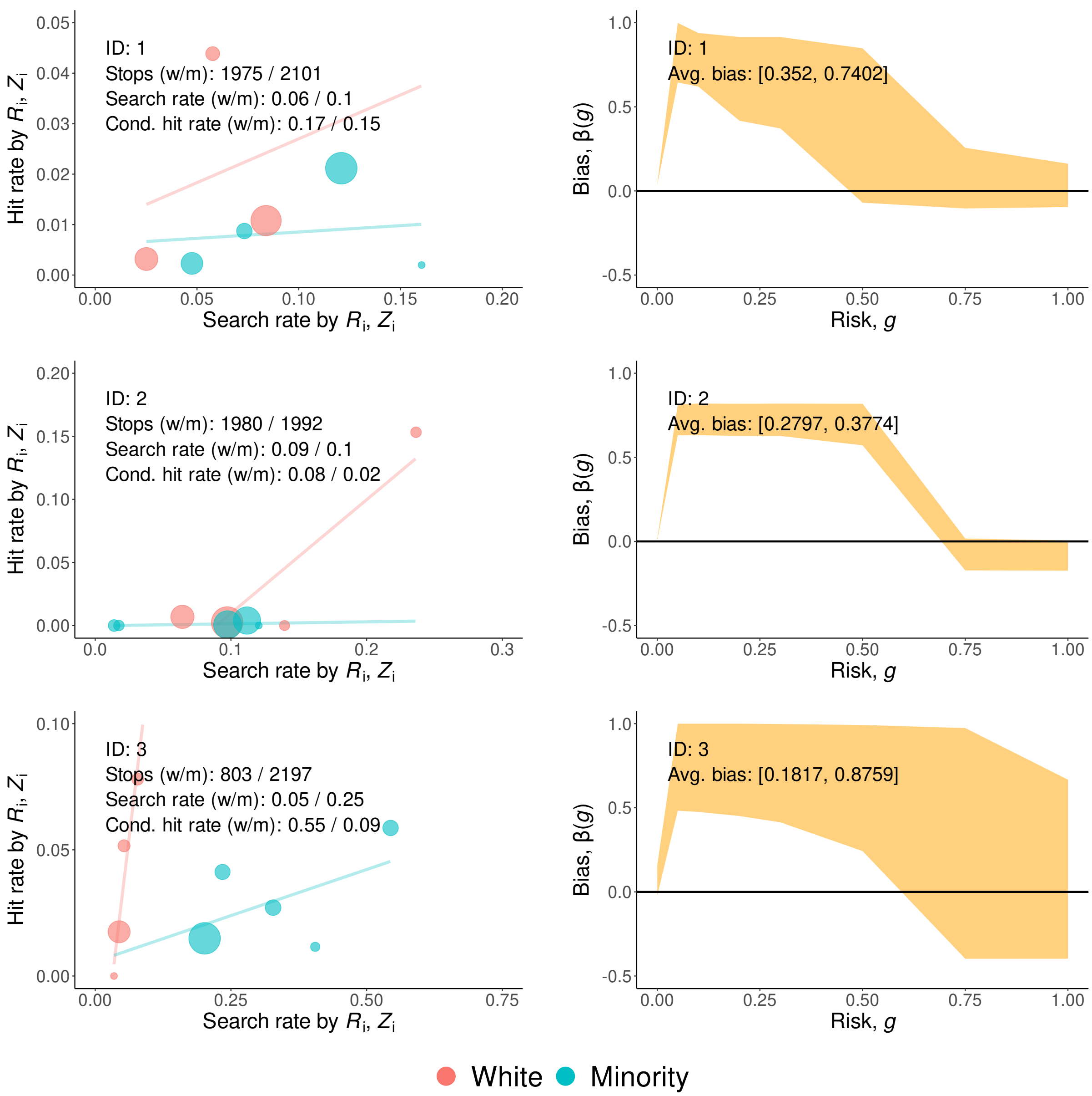


Figure A.4: All officer level estimates (continued)

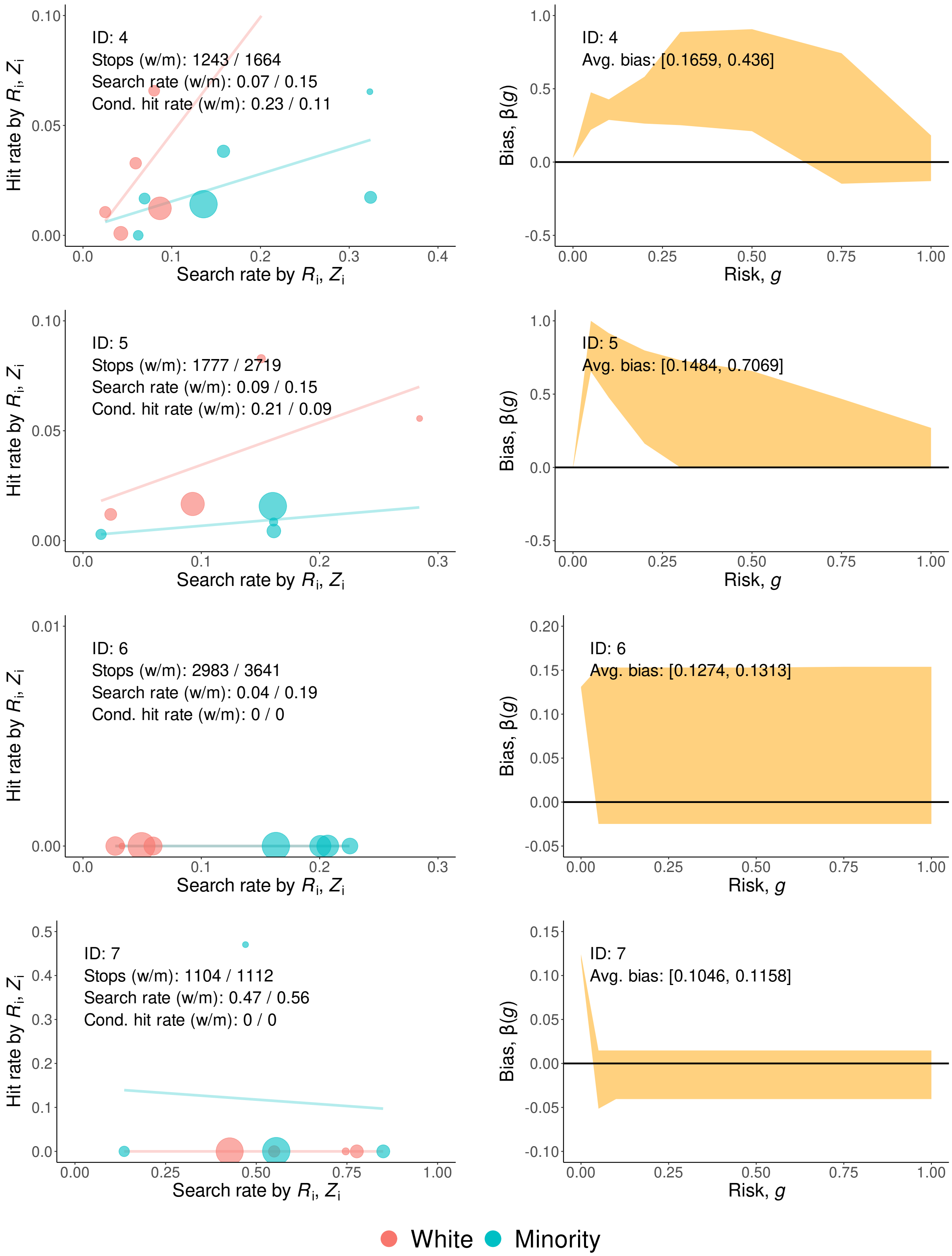


Figure A.4: All officer level estimates (continued)

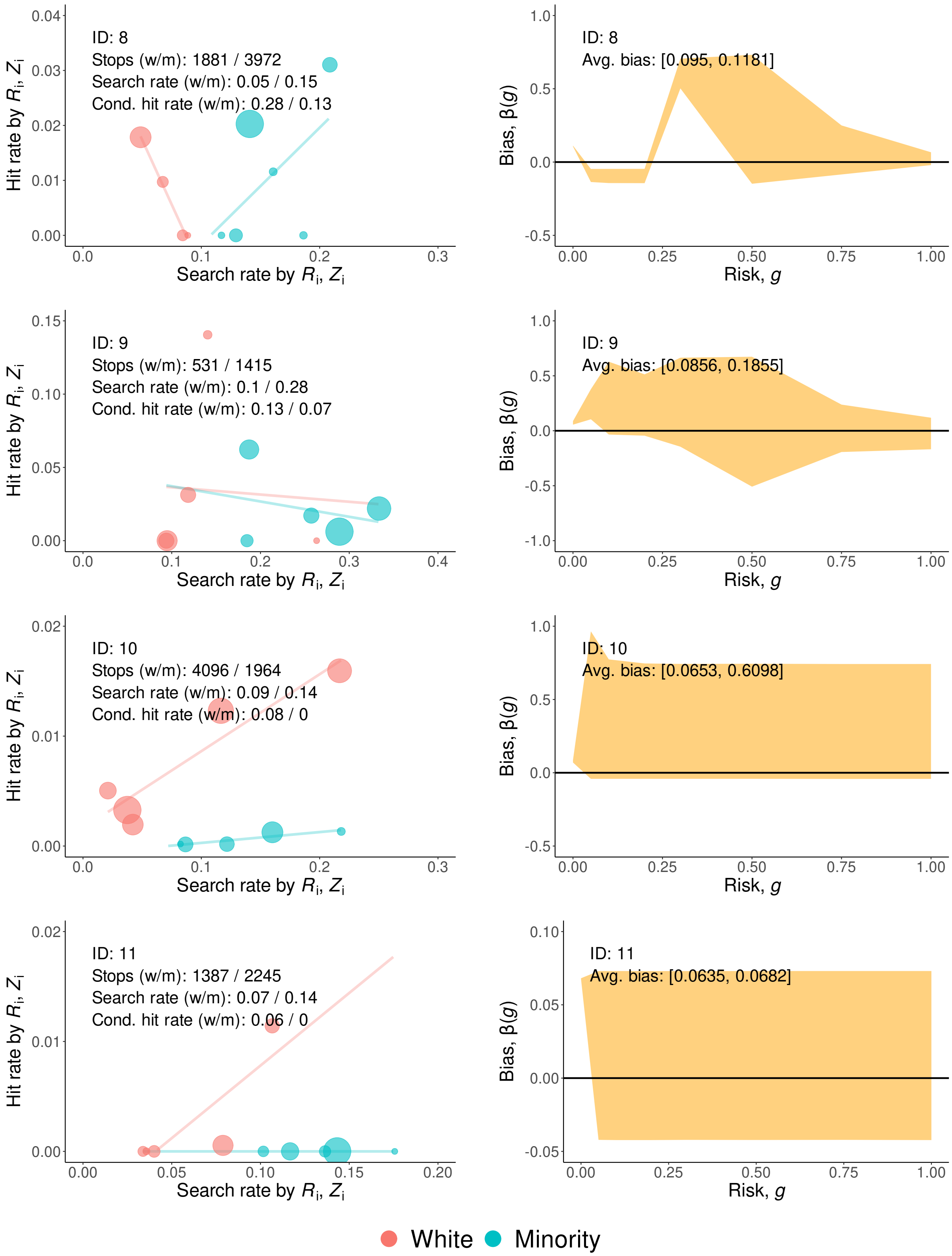


Figure A.4: All officer level estimates (continued)

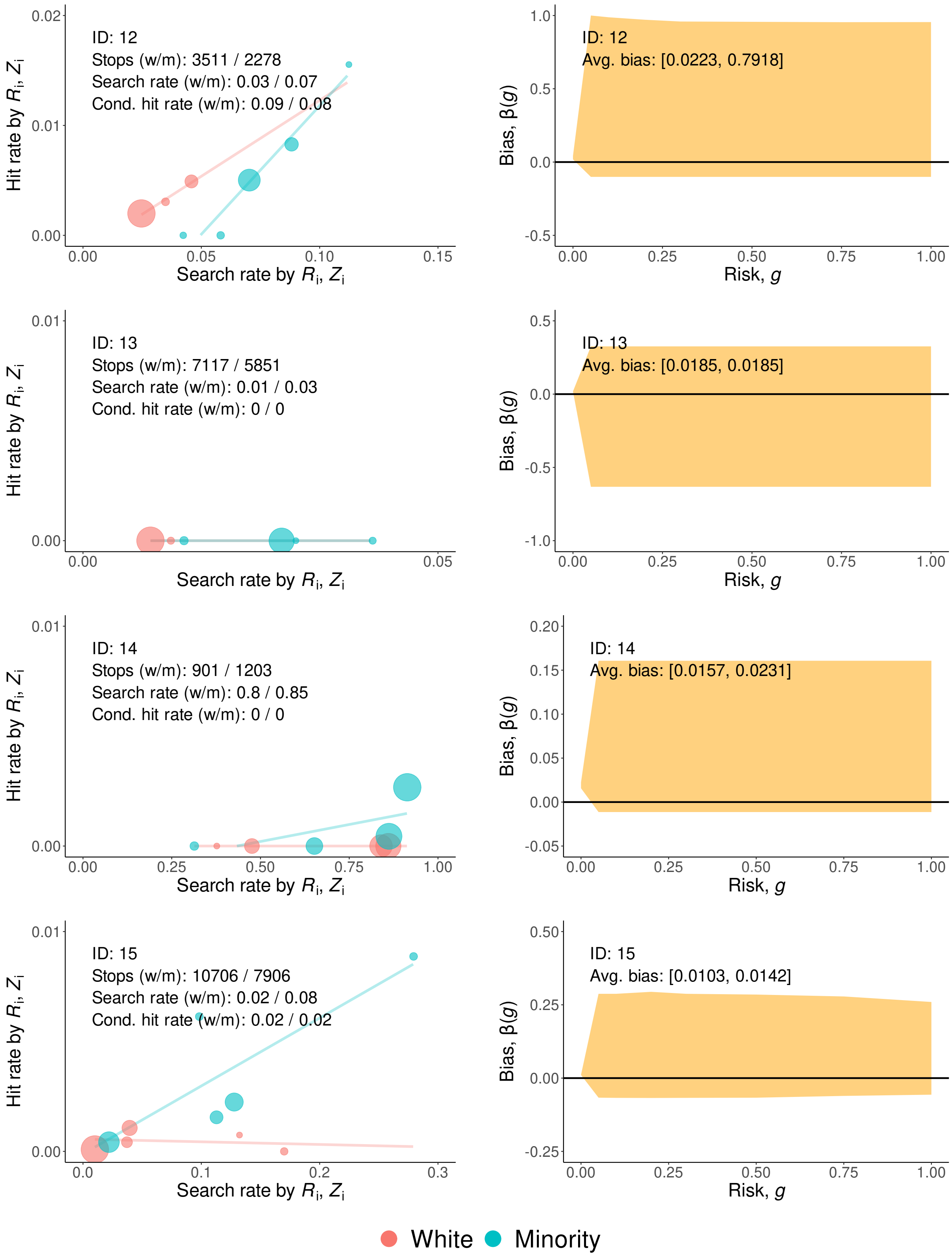


Figure A.4: All officer level estimates (continued)

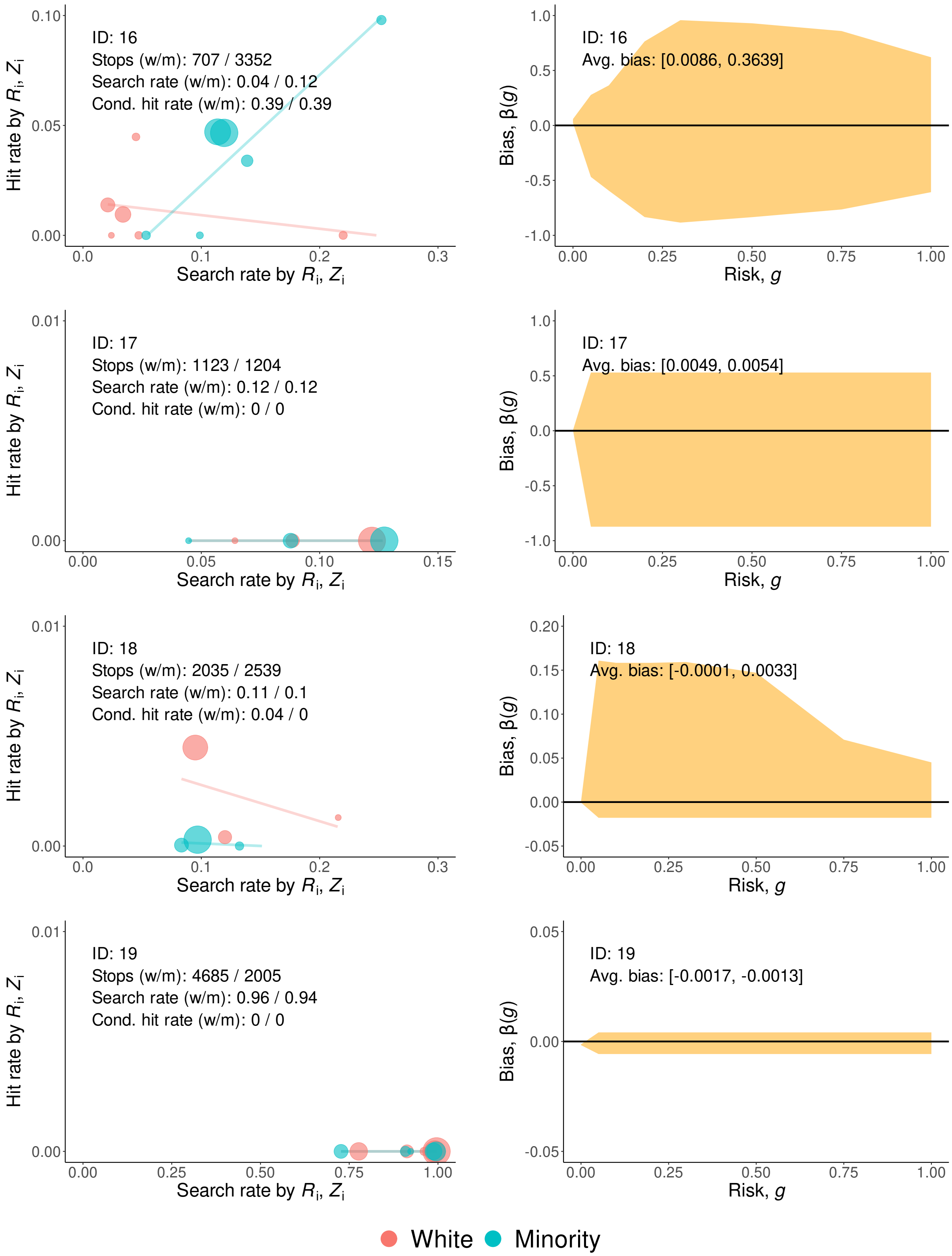


Figure A.4: All officer level estimates (continued)

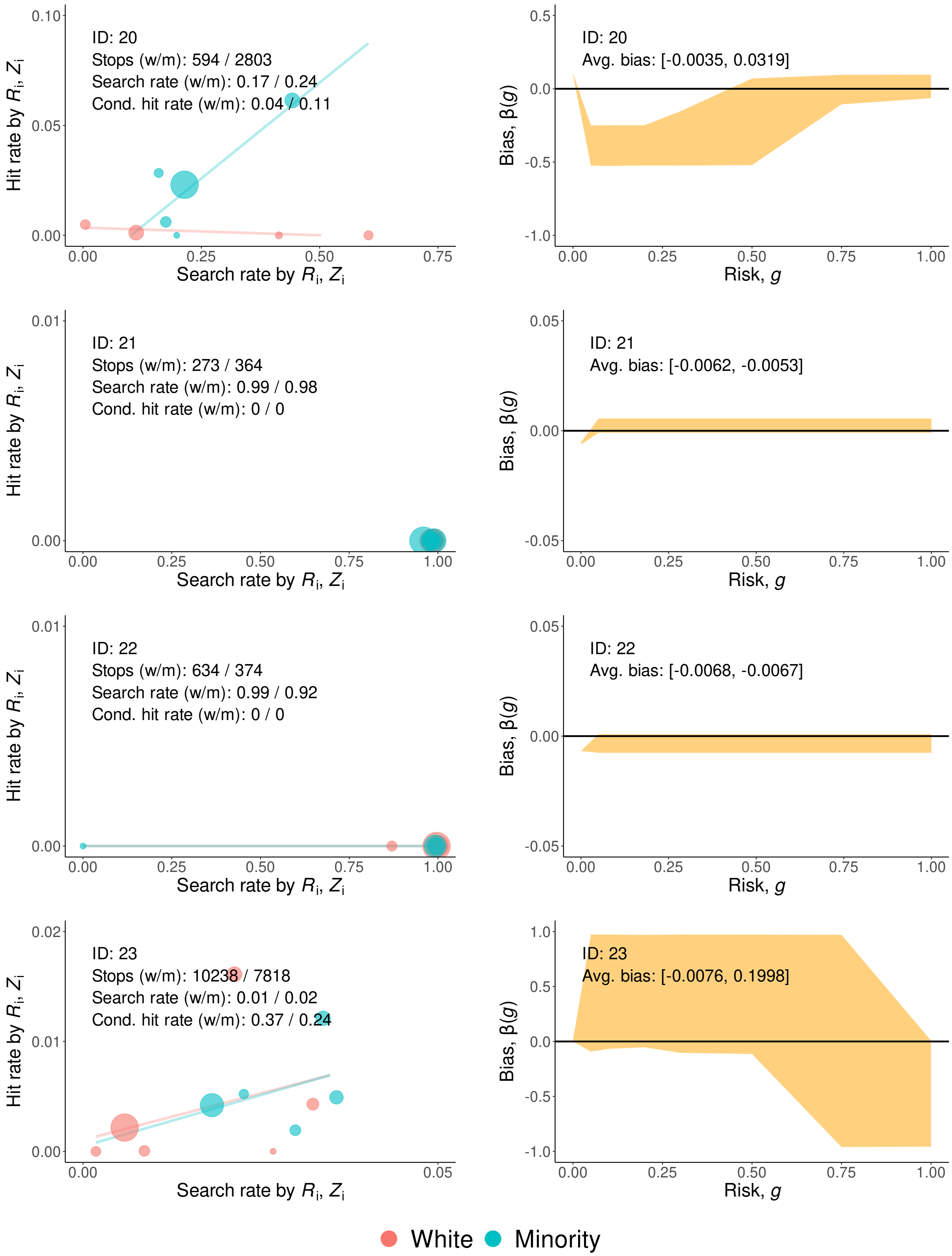


Figure A.4: All officer level estimates (continued)

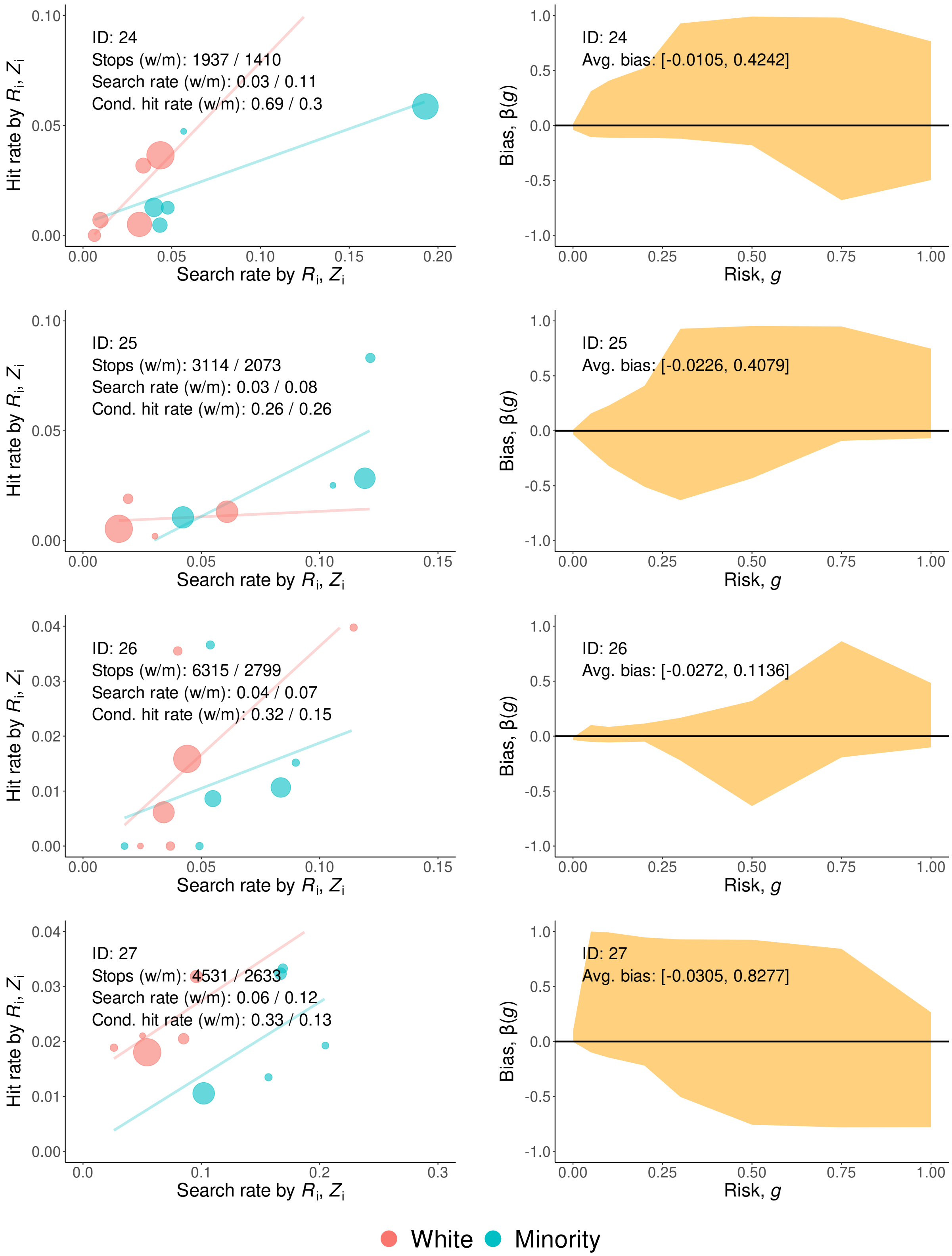


Figure A.4: All officer level estimates (continued)

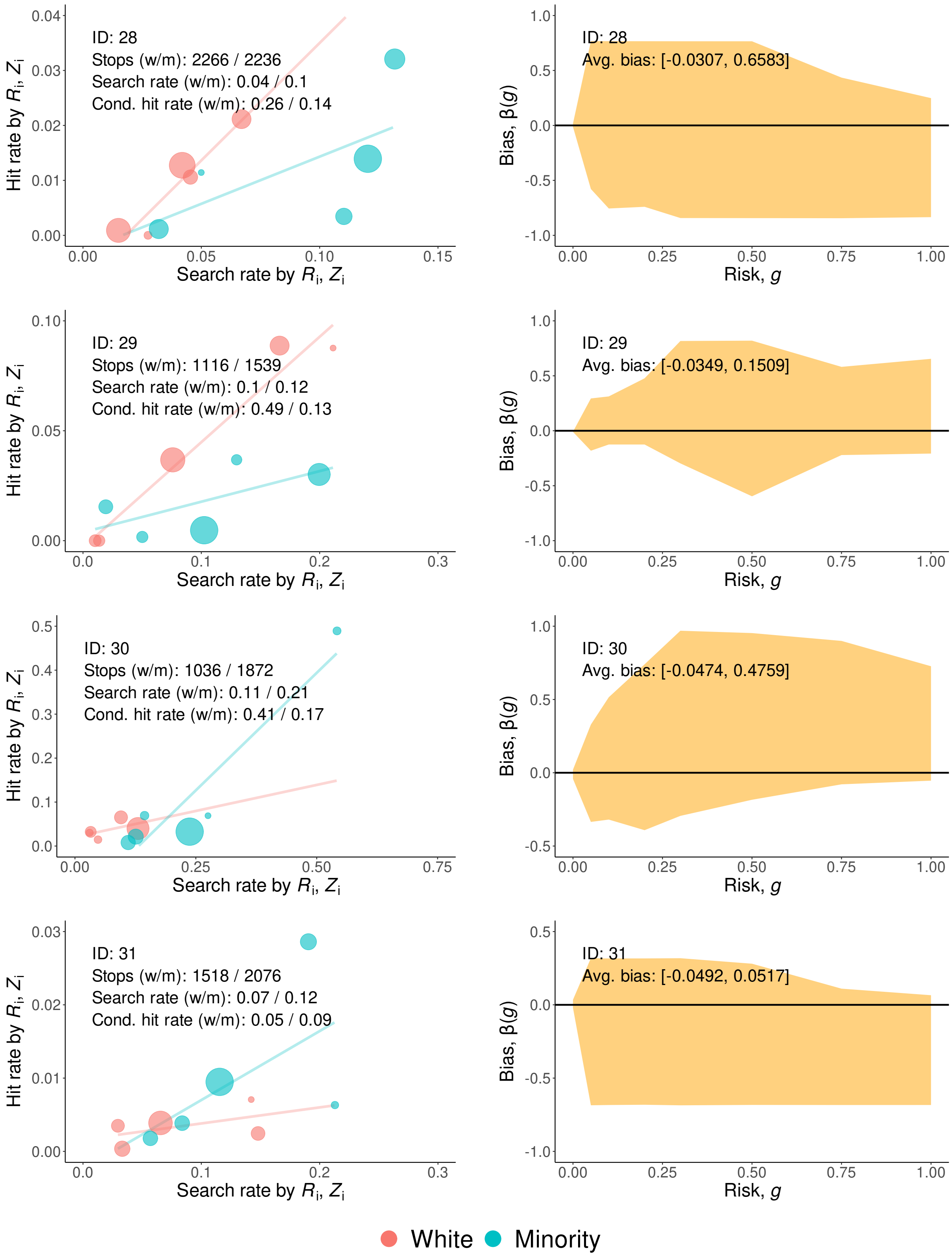


Figure A.4: All officer level estimates (continued)

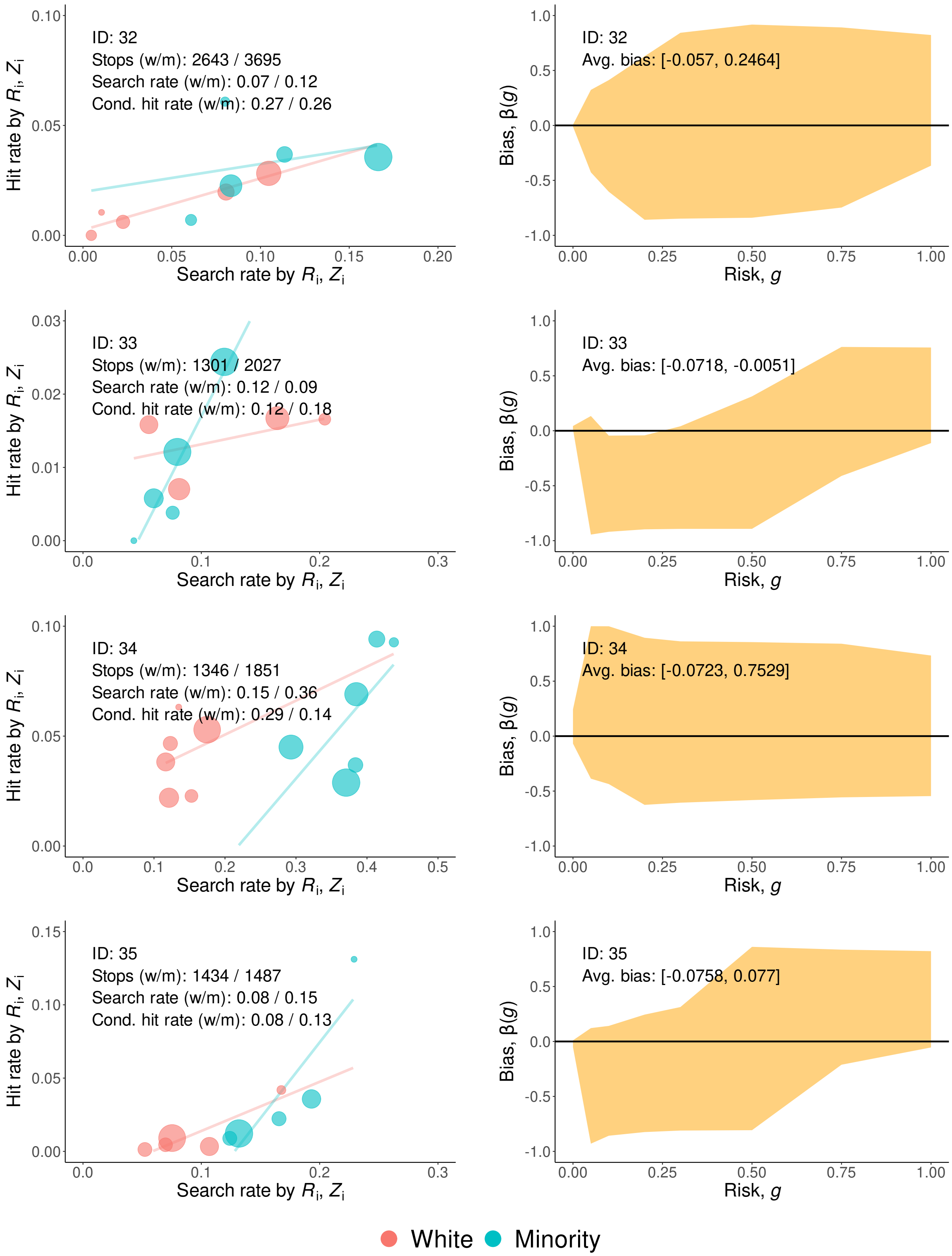


Figure A.4: All officer level estimates (continued)

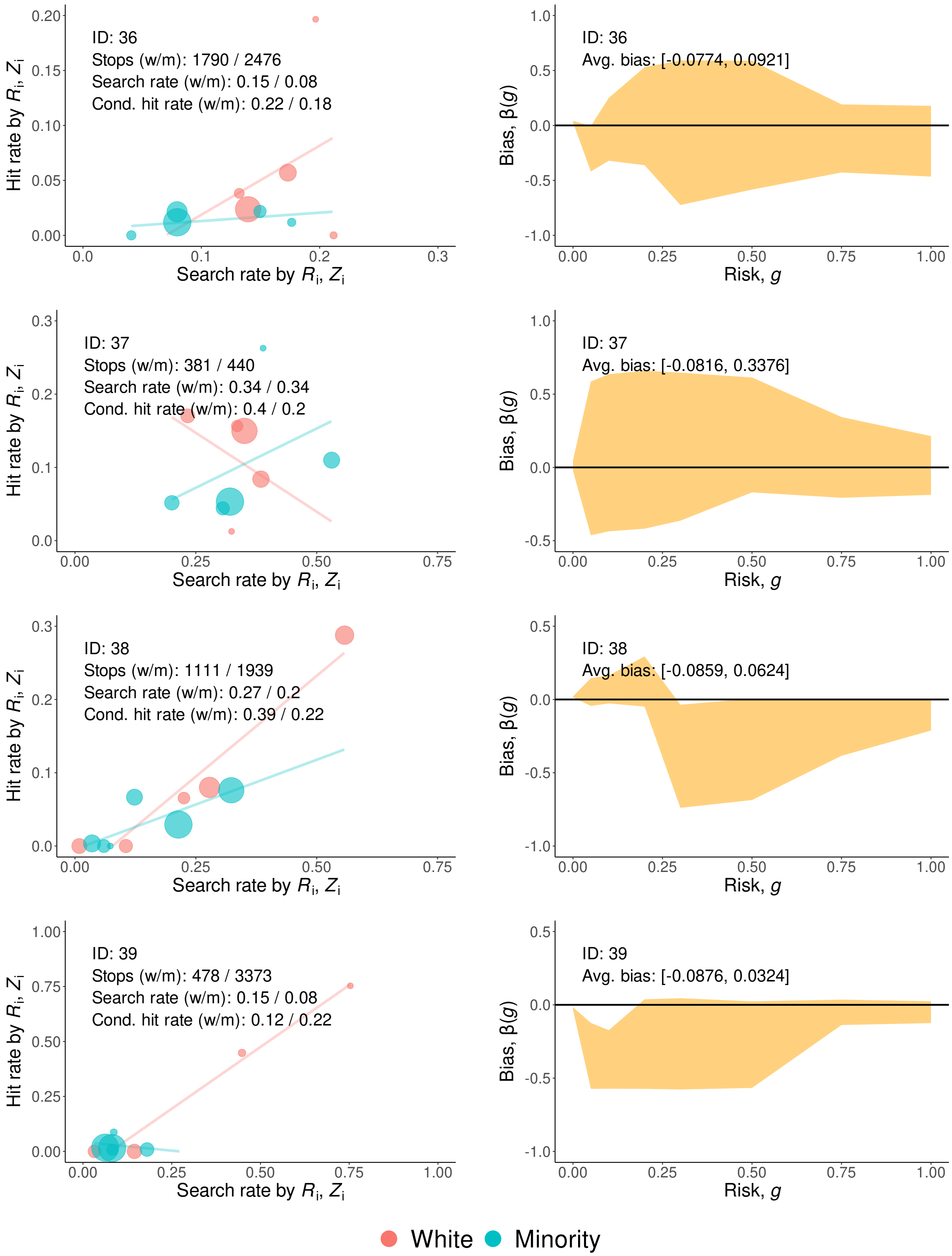


Figure A.4: All officer level estimates (continued)

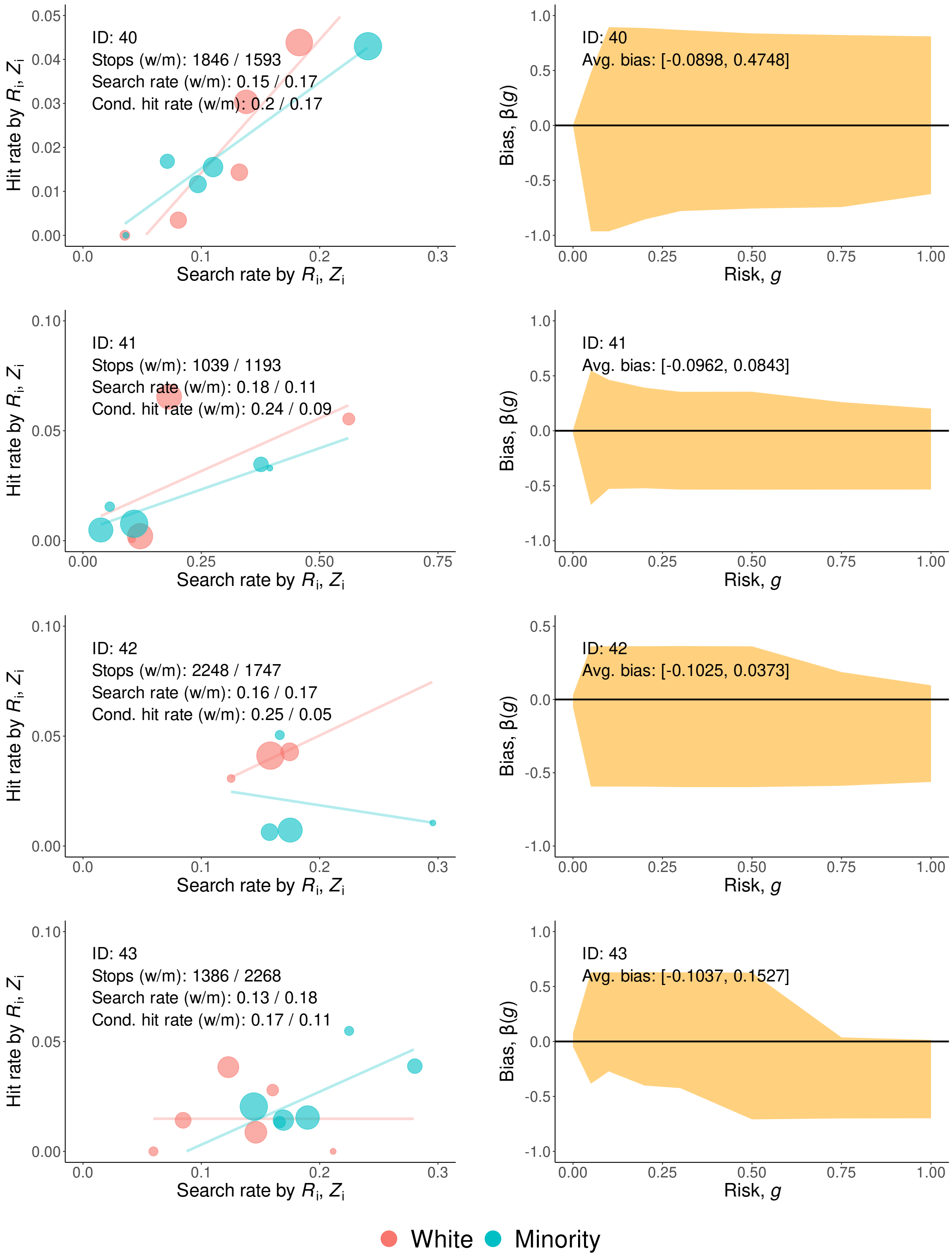


Figure A.4: All officer level estimates (continued)

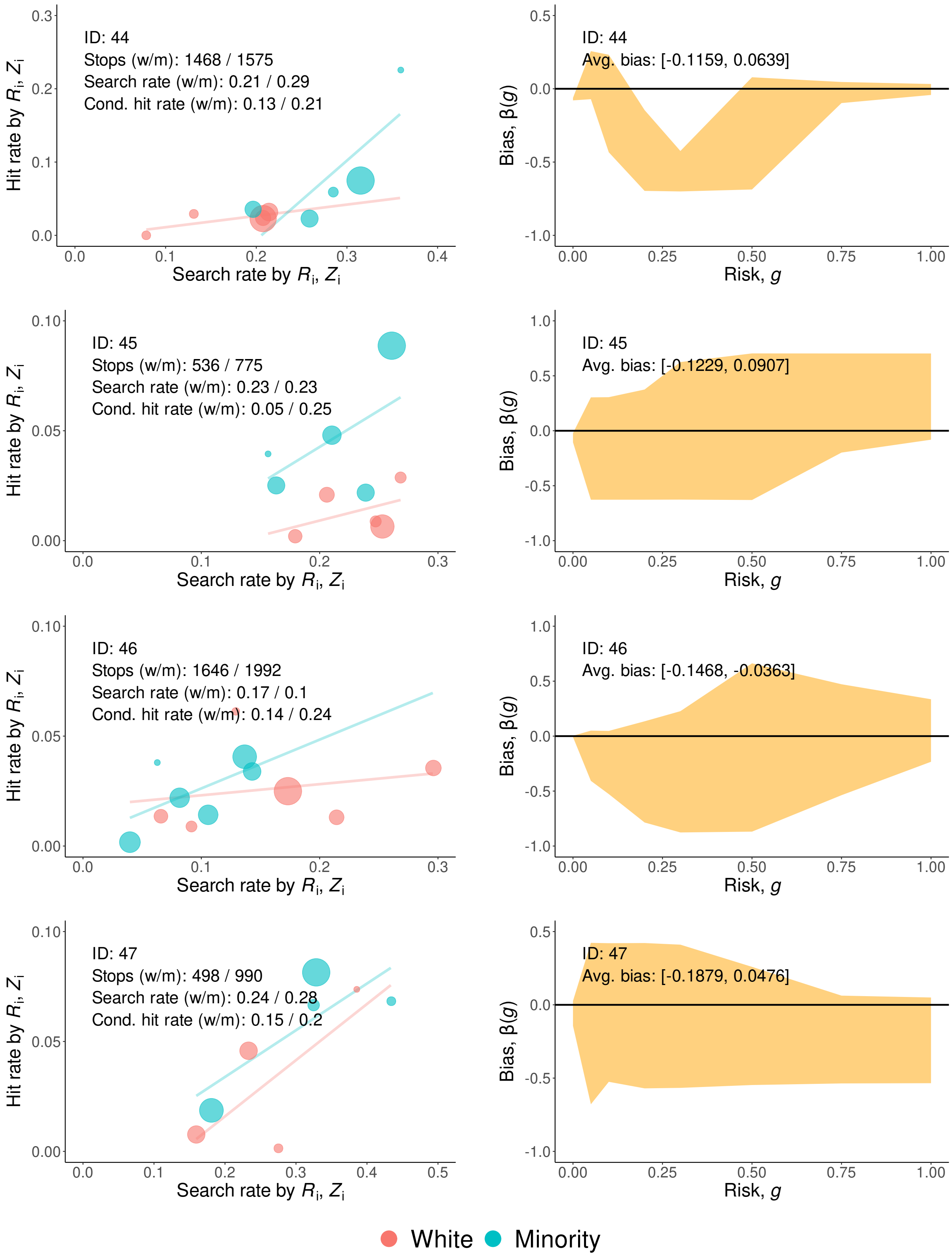
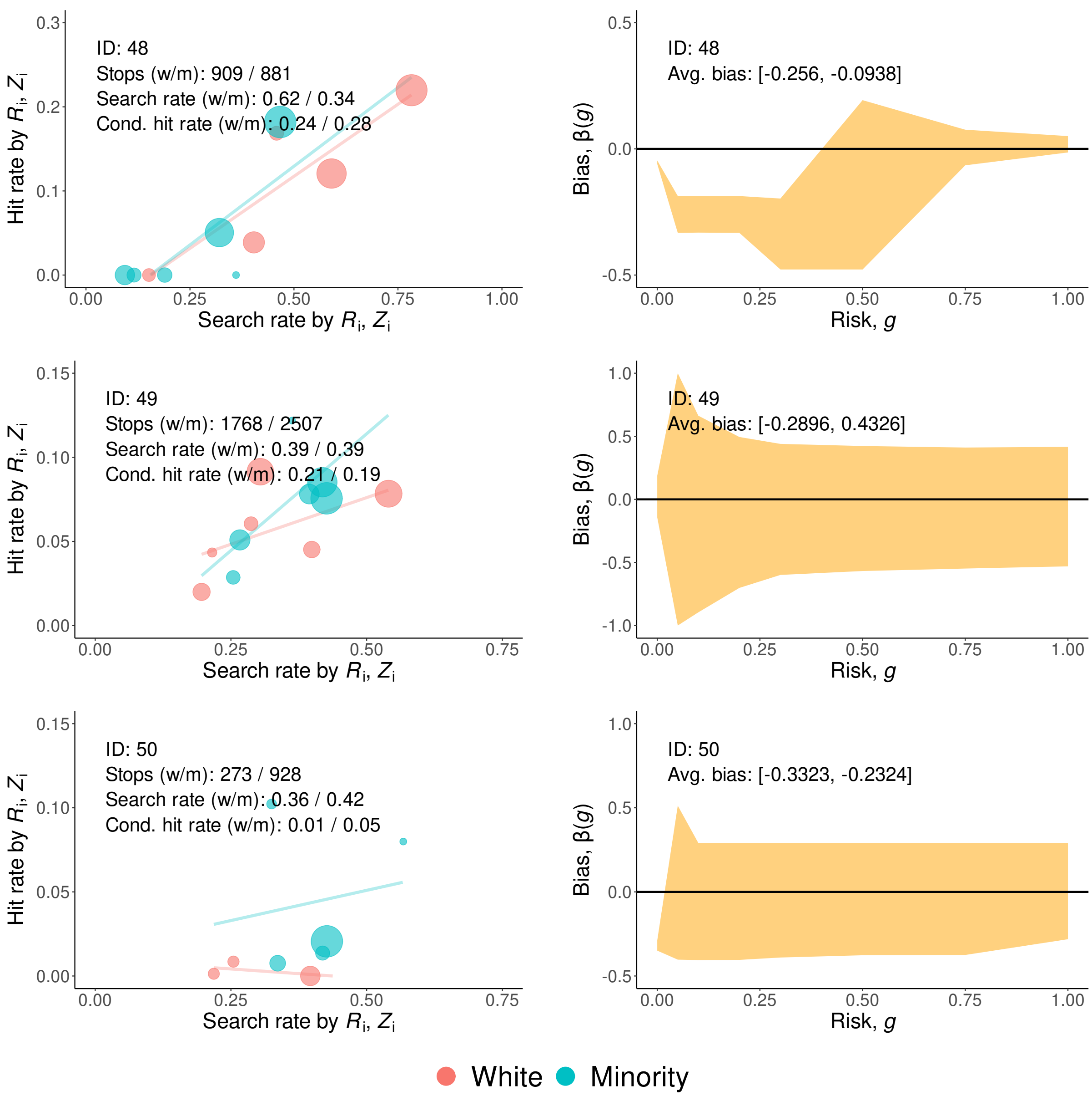


Figure A.4: All officer level estimates (continued)



APPENDIX B

APPENDIX TO CHAPTER 3

B.1 Proofs

Since all results are derived under the null that $\gamma(P) = \gamma_0$, we assume without loss of generality that $\gamma(P) = \gamma_0 = 0$. We use $a \lesssim b$ to denote that there exists $l > 0$ such that $a \leq lb$. We use $||$ to denote the Euclidean norm.

B.1.1 Derivation of Equation (3.3)

Equation (3.3) can be derived from equation (3.2) based on the following observation. For any $x \in \mathbb{R}$, we can decompose $x = x^+ - x^-$, where $x^+ = \max\{0, x\}$ and $x^- = \max\{0, -x\}$. It follows that $|x| = x^+ + x^-$. From this, we can write $b_j = b_j^+ - b_j^-$ and $r_j = r_j^+ - r_j^-$. Additionally, we introduce the following variables,

$$
\begin{aligned}
u_i &= \max\{0, Y_i - X_i'b - Z_i'r\} \\
v_i &= \max\{0, -Y_i + X_i'b + Z_i'r\} \\
\sigma_i &= \max\{0, |Y_i - X_i'b - Z_i'r| - \epsilon\} \\
s_i &= |Y_i - X_i'b - Z_i'r| - \epsilon - \sigma.
\end{aligned}
$$

The $n \times 1$ vectors u, v, σ, and s are obtained by stacking u_i, v_i, σ_i, and s_i, respectively, across all n observations. The first two constraints in (3.3) ensure that the decomposition of each term in (3.2) into the difference of its positive and negative components is consistent with the data. ■

B.1.2 Proof of Theorem 1

Follows immediately from Lemma 8, Lemma 9, and Lemma 10. ■

B.1.3 Derivation of Remark 3.3.3

As shown in Lemma 7,

$$
\begin{aligned}
n^{1/2}(\hat{\beta}_n - \beta) = \frac{1}{2} E[X_i X_i' f_Y(X_i'\beta - \epsilon | X_i, Z_i)]^{-1} n^{-1/2} \\
\times \sum_{1 \le i \le n} X_i (I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\}) + o_P(1)
\end{aligned}
$$

under Assumption 2(a)-(e). Symmetry of F under Assumption 3 implies that

$$
\begin{aligned}
n^{-1/2} \sum_{1 \le i \le n} X_i (I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\}) \\
\xrightarrow{d} N(0, E[X_i X_i' (I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\})^2]).
\end{aligned}
$$

By the law of iterated expectations, the variance term may be expressed as

$$
\begin{aligned}
& E[X_i X_i' (I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\})^2] \\
& = E[X_i X_i' \; E[(I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\})^2 \mid X_i, Z_i]].
\end{aligned}
$$

Note that

$$
\begin{aligned}
& E[(I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\})^2 | X_i, Z_i] \\
& = E[I\{Y_i \ge X_i'\beta + \epsilon\} + I\{Y_i \le X_i'\beta - \epsilon\} | X_i, Z_i] \\
& = 2E[F_Y(X_i'\beta - \epsilon | X_i, Z_i)] \; ,
\end{aligned}
$$

where the last equality follows from Assumption 2(c). Substituting this into the expression for the asymptotic variance completes the derivation. ■

B.1.4 Proof of Corollary 5

Follows immediately from Theorem 1 and the duality between hypotheses tests and confidence regions. ■

B.1.5 Auxiliary Lemmas

Lemma 4. *Suppose U_i, $1 \le i \le n$ are i.i.d. random variables where $E|U_i|^r < \infty$. Then*

$$n^{-1/r} \max_{1 \le i \le n} |U_i| \xrightarrow{P} 0 \ .$$

Proof of Lemma 4. Note that for all $\eta > 0$,

$$\begin{aligned} P\left\{n^{-1/r} \max_{1 \le i \le n} |U_i| > \eta\right\} &\le P\left\{\max_{1 \le i \le n} |U_i|^r > n\eta^r\right\} \\ &\le nP\{|U_i|^r > n\eta^r\} \le \frac{n}{n\eta^r} E[|U_i|^r I\{|U_i|^r > n\eta^r\}] = \frac{1}{\eta^r} E[|U_i|^r I\{|U_i|^r > n\eta^r\}] \to 0 \ , \end{aligned}$$

where the convergence follows from the dominated convergence theorem and $E|U_i|^r \to 0$. ■

Lemma 5. *Suppose P satisfies Assumption 2(b)-(d). Then*

$$S(b) = E[\max\{|Y - X'b - \epsilon|, 0\}]$$

is uniquely minimized at $b = \beta$.

Proof of Lemma 5. Follows immediately upon noting that Theorem 9.8 of Steinwart and Christmann (2008) holds under Assumption 2(b)-(d). ■

Lemma 6. *Suppose P satisfies Assumption 2(b)-(d) and λ_n satisfies Assumption 4. Then,*

$$\hat{\beta}_n \xrightarrow{P} \beta \ .$$

Proof of Lemma 6. Define

$$S_n(b) = n^{-1} \sum_{1 \le i \le n} \max\{0, |Y_i - X_i' b| - \epsilon\} + \lambda_n \|b\|_1 \ .$$

To begin with, note that $S_n(b)$ is convex in b. Without any loss of generality suppose $\beta = 0$. For any $\delta > 0$, let B_δ denote the closed δ-ball around 0. By definition,

$$S_n(\hat{\beta}_n) \le S_n(0) \ . \tag{B.1}$$

For all $b \in \mathbb{R}^{d_x} \backslash B_\delta$, we have by convexity that

$$S_n(b_\delta) \le \frac{\delta}{|b|} S_n(b) + \frac{|b| - \delta}{|b|} S_n(0) \ ,$$

where

$$b_\delta = \frac{\delta}{|b|} b \ .$$

Therefore

$$S_n(b) \ge \frac{|b|}{\delta} S_n(b_\delta) - \frac{|b| - \delta}{\delta} S_n(0) \ . \tag{B.2}$$

Since $\{b : |b| = 1\}$ is compact and $S(b)$ is continuous in b, we have by Lemma 5 that there exists $\eta > 0$ such that

$$\min_{b:|b|=1} S(b) \ge S(0) + \eta \ . \tag{B.3}$$

By Lemma 2.6.18 of Van Der Vaart and Wellner (1996), $\{b \to |y - x'b| - \epsilon + \lambda_n \|b\|_1 : |b| = 1\}$ is a VC class, thus Donsker, and thus Glivenko-Cantelli, i.e.,

$$\sup_{b:|b|=1} |S_n(b) - S(b)| = o_P(1) \ . \tag{B.4}$$

Combining (B.2), (B.3), and (B.4), and that $S_n(0) \xrightarrow{P} S(0)$, we have that

$$\{|\hat{\beta}_n - \beta| > \delta\} \Rightarrow \{S_n(\hat{\beta}_n) \geq S_n(0) + \eta + o_P(1)\} ,$$

which has probability approaching zero because of (B.1). ■

Lemma 7. *Suppose P satisfies Assumption 2(a)-(e). Then,*

$$n^{1/2}(\hat{\beta}_n - \beta) = \frac{1}{2} E[X_i X_i' f_Y(X_i'\beta - \epsilon | X_i, Z_i)]^{-1} n^{-1/2} \times \sum_{1 \leq i \leq n} X_i (I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\}) + o_P(1) .$$

Proof of Lemma 7. Define

$$\hat{L}_n = n^{-1/2} \sum_{1 \leq i \leq n} X_i (I\{Y_i - X_i'\hat{\beta}_n > \epsilon\} - I\{Y_i - X_i'\hat{\beta}_n < -\epsilon\}) \tag{B.5}$$

$$L_n = n^{-1/2} \sum_{1 \leq i \leq n} X_i (I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\}) . \tag{B.6}$$

To begin with, note that

$$\left| n^{-1/2} \sum_{1 \leq i \leq n} X_i \hat{a}_i^+ - \hat{L}_n \right| \lesssim n^{-1/2} \max_{1 \leq i \leq n} |X_i| \sum_{1 \leq i \leq n} (I\{Y_i = X_i'\hat{\beta}_n\} + I\{Y_i - X_i'\hat{\beta}_n = \epsilon\}) = o_P(1) , \tag{B.7}$$

because of by Lemma 1, Lemma 4, Assumption 2(a), and that the number of support vectors are bounded. By similar arguments,

$$\left| n^{-1/2} \sum_{1 \leq i \leq n} X_i \hat{a}_i^+ \right| \leq \left| n^{-1/2} \lambda_n \mathbf{1}_{d_x} \right| + o_P(1) = o_P(1) , \tag{B.8}$$

where the second equality follows from (3.7) and the last follows from Lemma 4. Next, we write

$$
\begin{aligned}
\hat{L}_n - L_n = n^{-1/2} \sum_{1 \le i \le n} X_i (I\{Y_i \le X_i' \beta + \epsilon\} - I\{Y_i \le X_i' \hat{\beta}_n + \epsilon\}) \\
+ n^{-1/2} \sum_{1 \le i \le n} X_i (I\{Y_i \le X_i' \beta - \epsilon\} - I\{Y_i \le X_i' \hat{\beta}_n - \epsilon\}) \qquad \text{(B.9)} \\
= R_{1,n} + R_{2,n} + R_{1,n}^- + R_{2,n}^- ,
\end{aligned}
$$

where

$$
\begin{aligned}
R_{1,n} &= n^{-1/2} \sum_{1 \le i \le n} X_i I\{Y_i \le X_i'\beta + \epsilon\} - E[X_i I\{Y_i \le X_i'\beta + \epsilon\}] \\
&\quad - (X_i I\{Y_i \le X_i'\hat{\beta}_n + X_i' t + \epsilon\} - E[X_i I\{Y_i \le X_i'\beta + X_i' t + \epsilon\}])|_{t = \hat{\beta}_n - \beta} \\
R_{2,n} &= n^{1/2} E[X_i (I\{Y_i \le X_i'\beta + \epsilon\} - I\{Y_i \le X_i'\beta + n^{-1/2} X_i' t + \epsilon\})]|_{t = n^{1/2}(\hat{\beta}_n - \beta)} ,
\end{aligned}
$$

and similarly for $R_{1,n}^-$ and $R_{2,n}^-$.

Since $\hat{\beta}_n - \beta = o_P(1)$ by Lemma 6, $R_{1,n} \xrightarrow{P} 0$ by the similar arguments as those used in the last part of the proof of Lemma A.1 of Bai et al. (2019). For $R_{2,n}$, note that

$$
\begin{aligned}
R_{2,n} &= n^{1/2} E[X_i (F_Y(X_i'\beta + \epsilon | X_i, Z_i) - F_Y(X_i'\beta + n^{-1/2} X_i' t + \epsilon | X_i, Z_i))]|_{t = n^{1/2}(\hat{\beta}_n - \beta)} \\
&= -n^{1/2} E[X_i n^{-1/2} X_i' t f_Y(X_i'\beta + \epsilon + s_i n^{-1/2} X_i' t | X_i, Z_i)]|_{t = n^{1/2}(\hat{\beta}_n - \beta)} \\
&= -n^{1/2} (\hat{\beta}_n - \beta) E[X_i X_i' f_Y(X_i'\beta + \epsilon + s_i n^{-1/2} X_i' t | X_i, Z_i)]|_{t = n^{1/2}(\hat{\beta}_n - \beta)} \qquad \text{(B.10)}
\end{aligned}
$$

where $s_i \in [0, 1]$ is a random variable. The first equality above holds by the law of iterated expectation and the second holds by the mean-value theorem. A similar decomposition holds for $R_{1,n}^-$ and $R_{2,n}^-$.

We then argue that $L_n = O_P(1)$. Indeed, by Assumption 2(c), the conditional distributions are symmetric so that the individual terms of L_n are i.i.d. mean zero and therefore $L_n = O_P(1)$ by the central limit theorem.

By applying similar arguments as those used to establish Lemma A.2 of Bai et al. (2019), where assumptions are satisfied under Assumption 2(a)-(e), and noting that $L_n = O_p(1)$, it follows from (B.5), (B.6), (B.7), (B.8), and (B.10) that

$$n^{1/2}(\hat{\beta}_n - \beta) = O_p(1) \ . \tag{B.11}$$

and

$$n^{1/2}(\hat{\beta}_n - \beta)(E[X_iX_i'\Big(f_Y(X_i'\beta+\epsilon|X_i,Z_i)+f_Y(X_i'\beta-\epsilon|X_i,Z_i)\Big)]+o_P(1)) = -L_n + o_P(1) \ . \tag{B.12}$$

The proof is finished by plugging (B.11) in (B.12), and noting that

$$f_Y(X_i'\beta+\epsilon|X_i,Z_i) = f_Y(X_i'\beta-\epsilon|X_i,Z_i)$$

by Assumption 2(c). ■

Lemma 8. *Suppose P satisfies Assumption 2(a)-(d) and ϵ and λ_n satisfies Assumption 4. Then,*

$$n^{-1/2}\mathbf{Z}_n'\hat{a}^+ \xrightarrow{d} N(0, 2E[\tilde{Z}_i\tilde{Z}_i'F_Y(X_i'\beta-\epsilon|X_i,Z_i)]) \ ,$$

where

$$\tilde{Z}_i = Z_i - E[Z_iX_i'f_Y(X_i'\beta-\epsilon|X_i,Z_i)]E[X_iX_i'f_Y(X_i'\beta-\epsilon|X_i,Z_i)]^{-1}X_i$$

Proof of Lemma 8. It follows from Lemma 6, Lemma 7, and similar arguments used to establish Lemma A.1 of Bai et al. (2019) that

$$n^{-1/2}\mathbf{Z}_n'\hat{a}^+ = n^{-1/2}\sum_{1\le i\le n}\tilde{Z}_i(I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\}) + o_P(1) \ .$$

Note that

$$
\begin{aligned}
&E[(I\{Y_i - X_i'\beta > \epsilon\} - I\{Y_i - X_i'\beta < -\epsilon\})^2 | X_i, Z_i] \\
&= E[I\{Y_i \geq X_i'\beta + \epsilon\} + I\{Y_i \leq X_i'\beta - \epsilon\} | X_i, Z_i] \\
&= 2E[F_Y(X_i'\beta - \epsilon | X_i, Z_i)] \ ,
\end{aligned}
$$

where the last equality follows from Assumption 2(c). The lemma now follows from the Central Limit Theorem and Assumption 2(a). ■

Lemma 9. *Suppose P satisfies Assumption 2(a) and Assumption 3(a). Then,*

$$
n^{-1}\mathbf{Z}_n'\mathbf{M}_n\mathbf{Z}_n \xrightarrow{P} E[Z_iZ_i'] - E[Z_iX_i']E[X_iX_i']^{-1}E[X_iZ_i'] \ .
$$

Proof of Lemma 9. Follows from Assumption 2(a), Assumption 3(a), and an application of the weak law of large numbers. ■

Lemma 10. *Suppose P satisfies Assumptions 2(a)-(d) and 3, and λ_n satisfies Assumption 4. Then,*

$$
\hat{p}_n \xrightarrow{P} 2p_\epsilon \ .
$$

Proof of Lemma 10. We consider

$$
\frac{1}{n}\sum_{1 \leq i \leq n} I\{Y_i \leq X_i'\beta + Z_i'\gamma_0 - \epsilon + X_i'(\hat{\beta}_n - \beta)\} \ ,
$$

and the other half follows similarly. By Lemma 6, since Assumptions 2(a)-(d) and 4 hold, $\hat{\beta}_n \xrightarrow{P} \beta$. Fix $\eta > 0$. For any $\delta > 0$, consider the empirical process indexed by the class of functions

$$
\{t \to I\{y \leq x'\beta + z'\gamma_0 - \epsilon + x't\} : \|t\| \in [0, \delta]\} \ .
$$

It is easy to see the class of functions is VC by Lemma 9.12 of Kosorok (2008), so that is Donsker hence Glivenko-Cantelli by Theorem 2.6.7 of Van Der Vaart and Wellner (1996), i.e.,

$$\sup_{t\in[0,\delta]} \left| \frac{1}{n} \sum_{1\le i\le n} I\{Y_i \le X_i'\beta + Z_i'\gamma_0 - \epsilon + X_i't\} - P\{Y_i \le X_i'\beta + Z_i'\gamma_0 - \epsilon + X_i't\} \right| \xrightarrow{P} 0 \ .$$

Next,

$$P\{Y_i \le X_i'\beta + Z_i'\gamma_0 - \epsilon + X_i't\}$$

is continuous at $t = 0$. Since $\hat{\beta}_n - \beta = o_P(1)$, with probability approaching 1, $\|\hat{\beta}_n - \beta\| \le \delta$, and therefore

$$\left| \frac{1}{n} \sum_{1\le i\le n} I\{Y_i \le X_i'\beta + Z_i'\gamma_0 - \epsilon + X_i't\} - P\{Y_i \le X_i'\beta + Z_i'\gamma_0 - \epsilon\} \right| \le \eta + \eta_\delta \ ,$$

where $\eta_\delta \to 0$ as $\delta \to 0$. Let $\delta \to 0$ to finish the proof. ■