Racial Bias in Bail Decisions
David Arnold, Will Dobbie, & Crystal S. Yang

Racial disparities exist at every stage of the U.S. criminal justice system. Compared to observably similar whites, blacks are more likely to be searched for contraband (Antonovics and Knight 2009), more likely to experience police force (Fryer 2016), more likely to be charged with a serious offense (Rehavi and Starr 2014), more likely to be convicted (Anwar, Bayer, and Hjalmarrson 2012), and more likely to be incarcerated (Abrams, Bertrand, and Mullainathan 2012). Racial disparities are particularly prominent in the setting of bail: in our data, black defendants are 3.6 percentage points more likely to be assigned monetary bail than white defendants and, conditional on being assigned monetary bail, receive bail amounts that are $9,923 greater. However, deter- mining whether these racial disparities are due to racial bias or statistical discrimination remains an empirical challenge.

To test for racial bias, Becker (1957, 1993) proposed an "outcome test" that compares the success or failure of decisions across groups at the margin. In our setting, the outcome test is based on the idea that rates of pre-trial misconduct will be identical for marginal white and marginal black defendants if bail judges are racially unbiased and the disparities in bail setting are solely due to (accurate) statistical discrimination (e.g., Phelps 1972, Arrow 1973). In contrast, marginal white defendants will have higher rates of pre-trial misconduct than marginal black defendants if these bail judges are racially biased against blacks, whether that racial bias is driven by racial animus, inaccurate racial stereotypes, or any other form of racial bias. The outcome test has been difficult to implement in practice, however, as comparisons based on average defendant outcomes are biased when whites and blacks have different risk distributions – the well-known infra-marginality problem (e.g., Ayres 2002).

In recent years, two seminal papers have developed outcome tests of racial bias that partially circumvent this infra-marginality problem. In the first paper, Knowles, Persico, and Todd (2001) show that if motorists respond to the race-specific probability of being searched, then all motorists of a given race will carry contraband with equal probability. As a result, the marginal and average success rates of police searches will be identical and there is not an infra-marginality problem. Knowles et al. (2001) find no difference in the average success rate of police searches for white and black drivers, leading them to conclude that there is no racial bias in police searches. In a second important paper, Anwar and Fang (2006) develop a test of relative racial bias based on the idea that the ranking of search and success rates by white and black police officers should be unaffected by the race of the motorist even when there are infra-marginality problems. Consistent with Knowles et al. (2001), Anwar and Fang (2006) find no evidence of relative racial bias in police searches, but note that their approach cannot be used to detect absolute racial bias. However, the prior literature has been critiqued for its reliance on restrictive assumptions about unobserved characteristics of blacks and whites (e.g., Brock et al. 2012).

In this paper, we propose a new outcome test for identifying racial bias in the context of bail decisions. Bail is an ideal setting to test for racial bias for a number of reasons. First, the legal objective of bail judges is narrow, straightforward, and measurable: to set bail conditions that allow most defendants to be released while minimizing the risk of pre-trial misconduct. In contrast, the objectives of judges at other stages of the criminal justice process, such as sentencing, are complicated by multiple hard-to-measure objectives, such as the balance between retribution and mercy. Second, mostly untrained bail judges must make on-the-spot judgments with limited information and little to no interaction with defendants. These institutional features make bail decisions particularly prone to the kind of inaccurate stereotypes or categorical heuristics that exacerbate racial bias (e.g., Fryer and Jackson 2008, Bordalo et al. 2016). Finally, bail decisions are extremely consequential for both white and black defendants, with prior work suggesting that detained defendants suffer about $30,000 in lost earnings and government benefits alone (Dobbie, Goldin, and Yang forthcoming).

To implement the Becker outcome test in our setting, we develop an instrumental variables (IV) estimator for racial bias that identifies the difference in pre-trial misconduct rates for white and black defendants at the margin of release. Though IV estimates are often criticized for the local nature of the estimates, we exploit the fact that the Becker test relies on (the difference between) exactly these kinds of local treatment effects for white and black defendants at the margin of release to test for racial bias. Specifically, we use the release tendencies of quasi-randomly assigned judges to identify local average treatment effects (LATEs) for white and black defendants near the margin of release. We then use the difference between these race-specific LATEs to estimate a weighted average of the racial bias among bail judges in our data.

In the first part of the paper, we formally establish the conditions under which our IV-based estimate of racial bias converges to the true level of racial bias. We show that two conditions must hold for our empirical strategy to yield consistent estimates of racial bias. The first is that our instrument for judge leniency becomes continuously distributed so that each race-specific IV estimate approaches a weighted average of treatment effects for defendants at the margin of release. The estimation bias from using a discrete instrument decreases with the number of judges and, in our data, is less than 1.1 percentage points. The second condition is that the judge IV weights are identical for white and black defendants near the margin of release so that we can interpret the difference in the race-specific LATEs as racial bias and not differences in how treatment effects from different parts of the distribution are weighted. This second condition is satisfied if, as is suggested by our data, there is a linear first-stage relationship between pre-trial release and our judge instrument.

The second part of the paper tests for racial bias in bail setting using administrative court data from Miami and Philadelphia. We find evidence of significant racial bias in our data, ruling out statistical discrimination as the sole explanation for the racial disparities in bail. Marginally released white defendants are 19.8 percentage points more likely to be rearrested prior to disposition than marginally released black defendants, with significantly more racial bias among observably high-risk defendants. Our IV-based estimates of racial bias are nearly identical if we account for other observable crime and defendant differences by race, suggesting that our results cannot be explained by black-white differences in certain types of crimes (e.g., the proportion of felonies versus misdemeanors) or black-white differences in defendant characteristics (e.g., the proportion with prior offenses versus no prior offenses). In sharp contrast to these IV results, naïve

OLS estimates indicate, if anything, racial bias against white defendants, highlighting the importance of accounting for both infra-marginality and omitted variables when estimating bias in the criminal justice system.

In the final part of the paper, we explore which form of racial bias is driving our findings. The first possibility is that, as originally modeled by Becker (1957, 1993), racial animus leads judges to discriminate against black defendants at the margin of release. This type of taste-based racial bias may be a particular concern in our setting due to the relatively low number of minority bail judges, the rapid-fire determination of bail decisions, and the lack of face-to-face contact between defendants and judges. A second possibility is that bail judges rely on incorrect inferences of risk based on defendant race due to anti-black stereotypes, leading to the relative over-detention of black defendants at the margin. These inaccurate anti-black stereotypes can arise if black defendants are over-represented in the right tail of the risk distribution, even when the difference in the riskiness of the average black defendant and the average white defendant is very small (Bordalo et al. 2016). As with racial animus, these racially biased prediction errors in risk may be exacerbated by the fact that bail judges must make quick judgments on the basis of limited information, with virtually no training and, in many jurisdictions, little experience working in the bail system.

We find three sets of facts suggesting that our results are driven by bail judges relying on inaccurate stereotypes that exaggerate the relative danger of releasing black defendants versus white defendants at the margin. First, we find that both white and black bail judges exhibit racial bias against black defendants and that racial bias varies across subsamples where there are no a priori reasons to believe that racial animus should vary, results that are inconsistent with most models of racial animus. Second, we find that our data are strikingly consistent with the theory of stereotyping developed by Bordalo et al. (2016). For example, we find that black defendants are sufficiently over- represented in the right tail of the predicted risk distribution, particularly for violent crimes, to rationalize observed racial disparities in release rates under a stereotyping model. We also find that there is no racial bias against Hispanics, who, unlike blacks, are not significantly over-represented in the right tail of the predicted risk distribution. Finally, we find substantially more racial bias when prediction errors (of any kind) are more likely to occur. For example, we find substantially less racial bias among both the full-time and more experienced part-time judges who are least likely to rely on simple race-based heuristics, and substantially more racial bias among the least experienced part-time judges who are most likely to rely on these heuristics.

Our findings are broadly consistent with parallel work by Kleinberg et al. (forthcoming), who use machine learning techniques to show that bail judges make significant prediction errors for defendants of all races. Using a machine learning algorithm to predict risk using a variety of inputs such as prior and current criminal charges, but *excluding* defendant race, they find that the algorithm could reduce crime and jail populations while simultaneously reducing racial disparities. Their results also suggest that variables that are unobserved in the data, such as a judge's mood or a defendant's demeanor at the bail hearing, are the source of prediction errors, not private information that leads to more accurate risk predictions. Our results complement Kleinberg et al. (forthcoming) by documenting one specific source of these prediction errors – racial bias among bail judges.

Our results also contribute to an important literature testing for racial bias in the criminal justice system. As discussed above, Knowles et al. (2001) and Anwar and Fang (2006) are seminal works in this area. Subsequent work by Antonovics and Knight (2009) finds that police officers in Boston are more likely to conduct a search if the race of the officer differs from the race of the driver, consistent with racial bias among police officers, and Alesina and La Ferrara (2014) find that death sentences of minority defendants convicted of killing white victims are more likely to be reversed on appeal, consistent with racial bias among juries. Conversely, Anwar and Fang (2015) find no racial bias against blacks in parole board release decisions, observing that among prisoners released by the parole board between their minimum and maximum sentence, the marginal prisoner is the same as the infra-marginal prisoner. Mechoulan and Sahuguet (2015) also find no racial bias against blacks in parole board release decisions, arguing that for a given sentence, the marginal prisoner is the same as the infra-marginal prisoner. Finally, Ayres and Waldfogel (1994) show that bail bond dealers in New Haven charge lower prices to minority defendants, suggesting that minorities, at least on average, have a lower probability of pre-trial misconduct than whites, and Bushway and Gelbach (2011) find evidence of racial bias in bail setting using a parametric framework that accounts for unobserved heterogeneity across defendants.

Our paper is also related to work using LATEs provided by IV estimators to obtain effects at the margin of the instrument (e.g., Card 1999, Gruber, Levine, and Staiger 1999) or to extrapolate to other estimands of interest (e.g., Heckman and Vyltacil 2005, Heckman, Urzua, and Vyltacil 2006). In recent work, Brinch, Mogstad, and Wiswall (2017) show that a discrete instrument can be used to identify marginal treatment effects using functional form assumptions. Kowalski (2016) similarly shows that it is possible to bound and estimate average treatment effects for always takers and never takers using functional form assumptions. Most recently, Mogstad, Santos, and Torgovitsky (2017) show that because a LATE generally places some restrictions on unknown marginal treatment effects, if it possible to recover information about other estimands of interest.

The remainder of the paper is structured as follows. Section I provides an overview of the bail system, describes the theoretical model underlying our analysis, and develops our empirical test for racial bias. Section II describes our data and empirical methodology. Section III presents the main results. Section IV explores potential mechanisms, and Section V concludes. An online appendix provides additional results, theoretical proofs, and detailed information on our institutional setting . . .

III. Results

In this section, we present our main results applying our empirical test for racial bias. We then compare the results from our empirical test with the alternative outcome-based tests developed by Knowles et al. (2001) and Anwar and Fang (2006).

A. Empirical Tests for Racial Bias

. . . In total, 17.8 percent of defendants are rearrested for a new crime prior to disposition, with 7.9 percent of defendants being rearrested for drug offenses, 6.7 percent of defendants being rearrested for property offenses, and 6.1 percent of defendants being rearrested for violent offenses. We find convincing evidence of racial bias against black defendants. In Panel A, we find that marginally released white defendants are 24.6 percentage points more likely to be rearrested for any crime compared to marginally detained white defendants (column 1). In contrast, the effect of pre-trial release on rearrest rates for the marginally released black defendants is a statistically in significant 4.9 percentage points (column 2). Taken together, these estimates imply that marginally released white defendants are 19.8 percentage points more likely to be rearrested prior to disposition than marginally released black defendants (column 3), consistent with racial bias against blacks. Importantly, we can reject the null hypothesis of no racial bias even assuming the maximum potential bias in our IV estimator of 1.1 percentage points (see Appendix B). Our results therefore rule out statistical discrimination as the sole determinant of racial disparities in bail.

In Panel B, we find suggestive evidence of racial bias against black defendants across all crime types, although the point estimates are too imprecise to make definitive conclusions. Most strikingly, we find that marginally released whites are about 9.3 percentage points more likely to be rearrested for a violent crime prior to disposition than marginally released blacks (p-value = 0.079). Marginally released white defendants are also 7.6 percentage points more likely to be rearrested for a drug crime prior to case disposition than marginally released black defendants (p-value = 0.171), and 12.9 percentage points more likely to be rearrested for a property crime (p-value = 0.054). These results suggest that judges are racially biased against black defendants even if they are most concerned about minimizing specific types of new crime, such as violent crimes . . .

IV. Potential Mechanisms

In this section, we attempt to differentiate between two alternative forms of racial bias that could explain our findings: (1) racial animus (e.g., Becker 1957, 1993) and (2) racially biased prediction errors in risk (e.g., Bordalo et al. 2016).

A. Racial Animus

The first potential explanation for our results is that judges either knowingly or unknowingly discriminate against black defendants at the margin of release as originally modeled by Becker (1957, 1993). Bail judges could, for example, harbor explicit animus against black defendants that leads them to value the freedom of black defendants less than the freedom of observably similar white defendants. Bail judges could also harbor implicit biases against black defendants – similar to those documented among both employers (Rooth 2010) and doctors (Penner et al. 2010) – leading to the relative over-detention of blacks despite the lack of any explicit animus. Racial animus may be a particular concern in bail setting due to the relatively low number of minority bail judges, the rapid-fire determination of bail decisions, and the lack of face-to-face contact between defendants and judges. Prior work has shown that it is exactly these types of settings where racial prejudice is most likely to translate into the disparate treatment of minorities (e.g., Greenwald et al. 2009).

One piece of evidence against this hypothesis is provided by the Anwar and Fang (2006) test discussed above, which indicates that bail judges are monolithic in their treatment of white and black defendants. Consistent with these results, we also find that IV estimates of racial bias are similar among white and black judges, although the confidence intervals for these estimates are extremely large. These estimates suggest that either racial animus is not driving our results, or that black and white bail judges harbor equal levels of racial animus towards black defendants. A second piece of evidence against racial animus comes from the subsample results discussed above, where we find that racial bias varies across groups where there are no a priori reasons to believe that racial animus should vary. Taken together, these results suggest that racial animus is unlikely to be the main driver of our results.

B. Racially Biased Prediction Errors in Risk

A second explanation for our results is that judges are making racially biased prediction errors in risk, potentially due to inaccurate anti-black stereotypes. Bordalo et al. (2016) show, for example, that representativeness heuristics – that is, probability judgments based on the most distinctive differences between groups – can exaggerate perceived differences between groups. In our setting, these kinds of race-based heuristics or anti-black stereotypes could lead bail judges to exaggerate the relative danger of releasing black defendants versus white defendants at the margin. These race-based prediction errors could also be exacerbated by the fact that bail judges must make quick judgments on the basis of limited information and with virtually no training . . .

Taken together, our results suggest that bail judges make racially biased prediction errors in risk. In contrast, we find limited evidence in support of the hypothesis that bail judges harbor racial animus towards black defendants. These results are broadly consistent with recent work by Kleinberg et al. (forthcoming) showing that bail judges make significant prediction errors in risk for all defendants, perhaps due to over-weighting the most salient case and defendant characteristics such as race and the nature of the charged offense. Our results also provide additional support for the stereotyping model developed by Bordalo et al. (2016), which suggests that probability judgments based on the most distinctive differences between groups – such as the significant over- representation of blacks relative to whites in the right tail of the risk distribution – can lead to anti-black stereotypes and, as a result, racial bias against black defendants.

V. Conclusion

In this paper, we test for racial bias in bail setting using the quasi-random assignment of bail judges to identify pre-trial misconduct rates for marginal white and marginal black defendants. We find evidence that there is substantial bias against black defendants, with the largest bias against black defendants with the highest predicted risk of rearrest. Our estimates are nearly identical if we account for observable crime and defendant differences by race, indicating that our results cannot be explained by black-white differences in the probability of being arrested for certain types of crimes (e.g., the proportion of felonies versus misdemeanors) or black-white differences in defendant characteristics (e.g., the proportion of defendants with prior offenses versus no prior offenses).

We find several pieces of evidence consistent with our results being driven by racially biased prediction errors in risk, as opposed to racial animus among bail judges. First, we find that both white and black bail judges are racially biased against black defendants, a finding that is inconsistent with most models of racial animus. Second, we find that black defendants are sufficiently over-represented in the right tail of the predicted risk distribution to rationalize observed racial disparities in release rates under a theory of stereotyping. Finally, racial bias is significantly higher among both part-time and inexperienced judges, and descriptive evidence suggests that experienced judges can better predict misconduct risk for all defendants. Taken together, these results are most consistent with bail judges relying on inaccurate stereotypes that exaggerate the relative danger of releasing black defendants versus white defendants at the margin.

The findings from this paper have a number of important implications. If racially biased prediction errors among inexperienced judges are an important driver of black-white disparities in pre-trial detention, our results suggest that providing judges with increased opportunities for training or on- the-job feedback could play an important role in decreasing racial disparities in the criminal justice system. Consistent with recent work by Kleinberg et al. (forthcoming), our findings also suggest that providing judges with data-based risk assessments may help decrease unwarranted racial disparities.

The empirical test developed in this paper can also be used to test for bias in other settings. Our test for bias is appropriate whenever there is the quasi-random assignment of decision makers and the objective of these decision makers is both known and well-measured. Our test can therefore be used to explore bias in settings as varied as parole board decisions, Disability Insurance applications, bankruptcy filings, and hospital care decision.