NBER WORKING PAPER SERIES

RACIAL BIAS IN BAIL DECISIONS

David Arnold
Will Dobbie
Crystal S. Yang

Working Paper 23421
http://www.nber.org/papers/w23421

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2017

We thank Pedro Bordalo, Leah Platt Boustan, David Deming, Hanming Fang, Hank Farber, Roland Fryer, Jonah Gelbach, Nicola Gennaioli, Edward Glaeser, Paul Goldsmith-Pinkham, Christine Jolls, Louis Kaplow, Larry Katz, Michal Kolesár, Ilyana Kuziemko, Magne Mogstad, Nicola Persico, Steven Shavell, Andrei Shleifer, David Silver, Alex Torgovitsky, and numerous seminar participants for helpful comments and suggestions. Molly Bunke, Kevin DeLuca, and Amy Wickett provided excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Racial Bias in Bail Decisions
David Arnold, Will Dobbie, and Crystal S. Yang
NBER Working Paper No. 23421
May 2017
JEL No. J15,J71,K14

ABSTRACT

This paper develops a new test for identifying racial bias in the context of bail decisions – a high-stakes setting with large disparities between white and black defendants. We motivate our analysis using Becker's (1957) model of racial bias, which predicts that rates of pre-trial misconduct will be identical for marginal white and marginal black defendants if bail judges are racially unbiased. In contrast, marginal white defendants will have a higher probability of misconduct than marginal black defendants if bail judges are racially biased against blacks. To test the model, we develop a new estimator that uses the release tendencies of quasi-randomly assigned bail judges to identify the relevant race-specific misconduct rates. Estimates from Miami and Philadelphia show that bail judges are racially biased against black defendants, with substantially more racial bias among both inexperienced and part-time judges. We also find that both black and white judges are biased against black defendants. We argue that these results are consistent with bail judges making racially biased prediction errors, rather than being racially prejudiced per se.

David Arnold
Industrial Relations Section
Louis A. Simpson International Bldg.
Princeton University
Princeton, NJ 08544-2098
dharnold@princeton.edu

Will Dobbie
Industrial Relations Section
Louis A. Simpson International Bldg.
Princeton University
Princeton, NJ 08544-2098
and NBER
wdobbie@princeton.edu

Crystal S. Yang
Harvard Law School
Griswold 301
Cambridge, MA 02138
and NBER
cyang@law.harvard.edu

Racial disparities exist at every stage of the criminal justice process. 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 11.2 percentage points more likely to be assigned monetary bail than white defendants and, conditional on being assigned monetary bail, have bail amounts that are $14,376 greater.[1] However, determining whether these racial disparities are due to racial bias or statistical discrimination remains an empirical challenge.

To distinguish between racial bias and statistical discrimination, Becker (1957) proposed an "outcome" test that uses the success or failure rates of decisions across groups at the margin. In our setting, Becker's 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 observed racial disparities in bail setting are solely due to statistical discrimination (e.g., Phelps 1972, Arrow 1973). In contrast, marginal white defendants will have a higher probability of pre-trial misconduct than marginal black defendants if bail judges are racially biased against blacks and the observed racial disparities in bail setting are driven at least in part by this racial bias. Thus, the key implication of the Becker test is that racial bias among bail judges can be estimated using the difference in pre-trial misconduct rates for white and black defendants at the margin of release. Importantly, however, researchers usually cannot observe which defendants are and are not on the margin of release and, thus, comparisons based on average defendant outcomes are biased if defendants have different risk distributions (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.[2]

[1] Authors' calculation for Miami-Dade and Philadelphia using the data described in Section II. Racial disparities in bail setting are also observed in other jurisdictions. For example, black felony defendants in state courts are nine percentage points more likely to be detained pre-trial compared to otherwise similar white defendants (McIntyre and Baradaran 2013).

[2] We replicate the Knowles et al. (2001) and Anwar and Fang (2006) tests in our data, finding no evidence of racial bias in either case. The differences between our test and the Knowles et al. (2001) and Anwar and Fang (2006) tests is that (1) we identify treatment effects for marginal defendants rather than the average defendant, and (2) we

In this paper, we propose a new outcome test for identifying absolute 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 may make bail decisions particularly prone to the kind of 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 $40,000 in lost earnings and government benefits alone (Dobbie, Goldin, and Yang 2016).[3]

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 distinguish between racial bias and statistical discrimination. 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 condition is that our instrument for judge leniency is continuous so that each race-specific IV estimate approaches a weighted average of treatment effects for defendants at the margin of release. With 177 bail judges in our sample, we argue that this condition is approximately true in our data.[4] 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 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

identify absolute rather than relative bias. See Section III.D for additional details on why the Knowles et al. (2001) and Anwar and Fang (2006) tests yield different results than our test.

[3]See also Gupta, Hansman, and Frenchman (2016), Leslie and Pope (2016), and Stevenson (2016).

[4]In the online appendix, we show that an additional functional form assumption on the distribution of marginal treatment effects allows us to consistently estimate racial bias with a discrete instrument. We also characterize the estimation bias from a discrete instrument when no additional functional form assumptions are made. We show that, under reasonable assumptions, our interpretation of the IV estimates remains valid and that we can calculate bounds on the estimation bias from using a discrete instrument. In practice, we find that the maximum estimation bias from using a discrete instrument in our setting is less than 0.5 percentage points, as the distance between any two judge leniency measures in our data is relatively small.

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 18.0 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 and among drug offenders, prior offenders, and defendants charged with felonies. 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 a prior offense versus no prior offense). In sharp contrast to these IV results, however, naïve OLS estimates indicate no racial bias against black 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 the potential mechanisms driving our results. One possibility is that, as originally modeled by Becker (1957), racially prejudiced judges discriminate against black defendants at the margin of release due to either explicit or implicit bias against blacks. This type of taste-based racial prejudice 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. Prior work suggests that it is exactly these types of settings where racial prejudice is most likely to translate into adverse outcomes for minorities (e.g., Greenwald et al. 2009). 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 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 average white defendant is very small (Bordalo et al. 2016). As with racial prejudice, these racially biased prediction errors 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 and predicting defendant risk.

We find three sets of facts suggesting that bail judges make racially biased prediction errors, but are not racially prejudiced per se. First, we find that both white and black bail judges exhibit racial bias against black defendants, a finding that is inconsistent with most models of racial prejudice. Second, we find that our data are strikingly consistent with the theory of stereotyping developed by Bordalo et al. (2016). 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 model of representativeness-based discounting. We also find that there is no racial bias against Hispanics, who, unlike blacks, are not over-represented in the right tail of the predicted risk distribution. Third, we find substantially more racial bias against blacks in situations where prediction errors (of any kind) are more likely to occur. For example, we find that racial bias is substantially lower among the types of bail judges that are least likely to rely on simple race-based heuristics: full-time judges in Philadelphia, who hear an average of 6,239 cases per year, and the most experienced part-time judges in Miami, who hear at least a few thousand cases during their career. Conversely, we find much larger racial bias among the least experienced part-time judges in Miami who hear just a few hundred bail cases in their career and who may be more likely to rely on race-based heuristics. We argue that these results are most consistent with bail judges, particularly inexperienced bail judges, relying on race-based heuristics that exaggerate the relative danger of releasing black defendants versus white defendants at the margin.

These findings are broadly consistent with parallel work by Kleinberg et al. (2017), who use machine learning techniques to show that bail judges make significant prediction errors for defendants of all races. Using a machine 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 compliment Kleinberg et al. (2017) by documenting one specific source of these prediction errors – racial bias among bail judges.

Our results 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.[5]

[5]There is also a large literature examining racial bias in other settings. The outcome test has been used to test for discrimination in the labor market (Charles and Guryan 2008) and the provision of healthcare (Chandra and Staiger 2010, Anwar and Fang 2012), while non-outcome based tests have been used to test for discrimination in the criminal justice system (Pager 2003, Anwar, Bayer, and Hjalmarsson 2012, Rehavi and Starr 2014, Agan and Starr 2016), the labor market (Goldin and Rouse 2000, Bertrand and Mullainathan 2004, Glover, Pallais, and Pariente forthcoming), the credit market (Ayres and Siegelman 1995, Bayer, Ferreira, and Ross 2016), the housing market (Edelman, Luca, and Svirsky 2017), and in sports (Price and Wolfers 2010, Parsons et al. 2011), among a variety of other settings.

Our paper is also related to an emerging literature extrapolating from the LATEs provided by IV estimators (e.g., Heckman and Vyltacil 2005, Heckman, Urzua, and Vyltacil 2006). Brinch, Mogstad, and Wiswall (forthcoming) 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, it is possible to recover information about other estimands of interest. In the online appendix, we show that we can consistently estimate racial bias when there are a small number of judges using similar functional form assumptions on the distribution of marginal treatment effects.

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.

I. An Empirical Test of Racial Bias

In this section, we motivate and develop our empirical test for racial bias in bail setting. Our theoretical framework closely follows the previous literature on the outcome test in the criminal justice system (e.g., Becker 1957, Knowles et al. 2001, Anwar and Fang 2006, Antonovics and Knight 2009). Consistent with the prior literature, we show that we can test for racial bias by comparing treatment effects for the marginal black and marginal white defendants. We then develop an estimator that identifies these race-specific treatment effects using an IV approach that exploits the quasi-random assignment of cases to judges.

A. Overview of the Bail System

In the United States, bail judges are granted considerable discretion to determine which defendants should be released before trial. Bail judges are meant to balance two competing objectives when deciding whether to detain or release a defendant before trial. First, bail judges are directed to release all but the most dangerous defendants before trial to reduce jail expenses and increase defendant well-being. Second, bail judges are instructed to minimize the risk of pre-trial misconduct by setting the appropriate conditions for release. Importantly, bail judges are not supposed to assess guilt or punishment at the bail hearing.

The conditions of release are set at a bail hearing typically held within 24 to 48 hours of a defendant's arrest. In most jurisdictions, bail hearings last only a few minutes and are held through a video-conference to the detention center such that judges can observe each defendant's demeanor. During the bail hearing, the assigned bail judge considers factors such as the nature of the alleged

See Fryer (2011) and Bertrand and Duflo (2016) for partial reviews of the literature.

offense, the weight of the evidence against the defendant, the nature and probability of danger that the defendant's release poses to the community, the likelihood of flight based on factors such as the defendant's employment status and living situation, and any record of prior flight or bail violations, among other factors (Foote 1954). Because bail judges are granted considerable discretion in setting the appropriate bail conditions, there are substantial differences across judges in the same jurisdiction (e.g., Dobbie et al. 2016, Gupta et al. 2016, Leslie and Pope 2016, Stevenson 2016).

The assigned bail judge has a number of potential options when setting a defendant's bail conditions. For example, the bail judge can release low-risk defendants on a promise to return for all court appearances, known as release on recognizance (ROR). For defendants who pose a higher risk of flight or new crime, the bail judge can allow release but impose non-monetary conditions such as electronic monitoring or periodic reporting to pre-trial services. The judge can also require defendants to post a monetary amount to secure release, typically 10 percent of the total bail amount. If the defendant fails to appear at the required court appearances or commits a new crime while out on bail, either he or the bail surety forfeits the 10 percent payment and is liable for the remaining 90 percent of the total bail amount. In practice, the median bail amount is $5,000 in our sample, and only 31 percent of defendants are able to meet the required monetary conditions to secure release. Bail may also be denied altogether for defendants who commit the most serious crimes such as first- or second-degree murder.

One important difference between jurisdictions is the degree to which bail judges specialize in conducting bail hearings. For example, in our setting, Philadelphia bail judges are full-time specialists who are tasked with setting bail seven days a week throughout the entire year. In contrast, the bail judges we study in Miami are part-time nonspecialists who assist the bail court by serving weekend shifts once or twice per year. These weekend bail judges spend their weekdays as trial court judges. We discuss the potential importance of these institutional features in Section IV.

B. Model of Judge Behavior

This section develops a theoretical framework that allows us to define an outcome-based test of racial bias in bail setting. We begin with a model of taste-based racial bias that closely follows Becker (1957). We then present an alternative model of racially biased prediction errors, which generates the same empirical predictions as the taste-based model.

Taste-Based Discrimination: Let i denote defendants and $\mathbf{V}_i$ denote all case and defendant characteristics considered by the bail judge, excluding defendant race r_i. The expected cost of release for defendant i conditional on observable characteristics $\mathbf{V}_i$ and race r_i is equal to the expected probability of pre-trial misconduct $\mathbb{E}[\alpha_i|\mathbf{V}_i, r_i]$ times the cost of misconduct C. For simplicity, we normalize $C = 1$, so that the expected cost of release conditional on observable characteristics is equal to $\mathbb{E}[\alpha_i|\mathbf{V}_i, r_i]$. Moving forward, we also simplify our notation by letting the expected cost of release conditional on observables be denoted by $\mathbb{E}[\alpha_i|r_i]$.

The benefit of releasing defendant i assigned to judge j is denoted by $t_r^j(\mathbf{V}_i)$, where we explicitly allow for the benefits to be a function of the observable case and defendant characteristics $\mathbf{V}_i$. The benefit of release $t_r^j(\mathbf{V}_i)$ includes cost savings from reduced jail time and private gains to defendants, such as an improved bargaining position with the prosecutor and increased labor force participation. Importantly, we allow the benefit of release $t_r^j(\mathbf{V}_i)$ to vary by race $r \in W, B$ to allow for judge preferences to differ for white and black defendants.

Definition 1. Following Becker (1957), we define judge j as racially biased against black defendants if $t_W^j(\mathbf{V}_i) > t_B^j(\mathbf{V}_i)$. Thus, for racially biased judges, there is a higher benefit of releasing white defendants than releasing observably identical black defendants.

Finally, we assume that bail judges are risk neutral and maximize the net benefit of pre-trial release. Thus, bail judge j will release defendant i if and only if the cost of pre-trial release is less than the expected benefit of release:

$$\mathbb{E}[\alpha_i | r_i = r] \leq t_r^j(\mathbf{V}_i) \tag{1}$$

Given this decision rule, the marginal defendant for judge j and race r is the defendant i for whom the expected cost of release is exactly equal to the benefit of release, i.e. $\mathbb{E}[\alpha_i^j | r_i = r] = t_r^j(\mathbf{V}_i)$. We simplify our notation moving forward by letting this expected cost of release for the marginal defendant for judge j and race r be denoted by α_r^j.

Based on the above framework and Definition 1, the model yields the familiar outcome-based test for racial bias from Becker (1957):

Proposition 1. If judge j is racially biased against black defendants, then $\alpha_W^j > \alpha_B^j$. Thus, for racially biased judges, the expected cost of release for the marginal white defendant is higher than the expected cost of release for the marginal black defendant.

Proposition 1 predicts that the marginal white and marginal black defendant should have the same probability of pre-trial misconduct if judge j is racially unbiased, but that the marginal white defendant should have a higher probability of misconduct than the marginal black defendant if judge j is racially biased against black defendants.

Racially Biased Prediction Errors: In the taste-based model of discrimination outlined above, we assume that judges agree on the expected cost of release, $\mathbb{E}[\alpha_i | r_i]$, but not the benefit of release, $t_r^j(\mathbf{V}_i)$. An alternative approach is to assume that judges vary in their predictions of the expected cost of release, as would be the case if there were race-specific prediction errors (e.g., if judges systematically overestimate the cost of release for black defendants relative to white defendants). We show that a model motivated by racially biased prediction errors can generate the same predictions as a model of taste-based discrimination.

Let i again denote defendants and $\mathbf{V}_i$ denote all case and defendant characteristics considered by the bail judge, excluding defendant race r_i. The benefit of releasing defendant i assigned to judge j is now defined as $t(\mathbf{V}_i)$, which does not vary by judge.

The expected cost of release for defendant i conditional on observable characteristics $\mathbf{V}_i$ is equal to the expected probability of pre-trial misconduct, $\mathbb{E}^j[\alpha_i|\mathbf{V}_i, r_i]$, which varies across judge. We can write the expected cost of release as:

$$\mathbb{E}^j[\alpha_i|\mathbf{V}_i] = \mathbb{E}[\alpha_i|\mathbf{V}_i, r_i = r] + \tau_r^j(\mathbf{V}_i) \tag{2}$$

where $\tau_r^j(\mathbf{V}_i)$ is a prediction error that is allowed to vary by judge and defendant race. To simplify our notation, we let the true probability of pre-trial misconduct conditional on all variables observed by the judge be denoted by $\mathbb{E}[\alpha_i|r_i]$.

Definition 2. We define judge j as making racially biased prediction errors against black defendants if $\tau_B^j(\mathbf{V}_i) > \tau_W^j(\mathbf{V}_i)$. Thus, judges making racially biased prediction errors systematically overestimate the cost of release for black defendants relative to white defendants.

Following the taste-based model, bail judge j will release defendant i if and only if the benefit of pre-trial release is greater than the expected cost of release:

$$\mathbb{E}^j[\alpha_i|\mathbf{V}_i, r_i] = \mathbb{E}[\alpha_i|r_i] + \tau_r^j(\mathbf{V}_i) \leq t(\mathbf{V}_i) \tag{3}$$

Given the above setup, it is straightforward to show that the prediction error model can be reduced to the taste-based model of discrimination outlined above if we relabel $t(\mathbf{V}_i) - \tau_r^j(\mathbf{V}_i) = t_r^j(\mathbf{V}_i)$. As a result, we can generate identical empirical predictions using the prediction error and taste-based models.

Following this logic, our model of racially biased prediction errors yields a similar outcome-based test for racial bias:

Proposition 2. If judge j systematically overestimates the expected cost of release of black defendants relative to white defendants, then $\alpha_W^j > \alpha_B^j$. Thus, for judges who make racially biased prediction errors, the expected cost of release for the marginal white defendant is higher than the expected cost of release for the marginal black defendant.

Proposition 2 predicts that the marginal white and marginal black defendant should have the same probability of pre-trial misconduct if judge j does not systematically make prediction errors that vary with race, but that the marginal white defendant should have a higher probability of misconduct than the marginal black defendant if judge j systematically overestimates the expected cost of release of black defendants relative to white defendants.

Regardless of the underlying behavioral model that drives the differences in judge behavior, the empirical predictions generated by these outcome-based tests are identical: if there is racial bias against black defendants, then marginal white defendants will have a higher probability of misconduct than marginal black defendants. In contrast, if observed racial disparities in bail setting are solely due to statistical discrimination, then marginal white defendants will not have a higher probability of misconduct than marginal black defendants.

However, the interpretation of racial bias does depend on the underlying behavioral model. In a taste-based model, a higher misconduct rate for marginal white versus marginal black defendants implies that judges are racially prejudiced against black defendants. In a prediction error model, the same empirical finding implies that judges systematically overestimate the relative risk of black defendants relative to white defendants. We will return to this issue in Section IV when we discuss more speculative evidence that allows us to differentiate between racial bias due to taste and racial bias due to prediction errors.

C. Empirical Test of Racial Bias in Bail Setting

The goal of our analysis is to empirically test for racial bias in bail setting using the rate of pre-trial misconduct for white defendants and black defendants at the margin of release. Following the theory model, let the true weighted average across all bail judges, $j = 1...J$, of treatment effects at the margin of release for defendants of race r be given by:

$$\alpha_r^* = \sum_{j=1}^{J} \lambda^j \cdot \alpha_r^j \tag{4}$$

where λ^j are non-negative weights which sum to one, which will be described in further detail below, and α_r^j is the treatment effect for a defendant of race r at the margin of release for judge j. Intuitively, α_r^* represents a weighted average across all judges of the treatment effects for defendants of race r at the margin of release.

Following this notation, the true weighted average of racial bias among bail judges D^* is given by:

$$\begin{aligned} D^* &= \sum_{j=1}^{J} \lambda^j \left(\alpha_W^j - \alpha_B^j \right) \\ &= \sum_{j=1}^{J} \lambda^j \alpha_W^j - \sum_{j=1}^{J} \lambda^j \alpha_B^j \\ &= \alpha_W^* - \alpha_B^* \end{aligned} \tag{5}$$

where λ^j are again non-negative weights which sum to one, such that D^* represents a weighted average across all judges of the difference in treatment effects for white defendants at the margin of release and black defendants at the margin of release. In theory, there are many sensible weighting schemes, λ^j, for racial bias. In practice, we let λ^j be defined as the standard IV weights (Imbens and Angrist 1994), i.e. weights that depend on the size of the subpopulation whose pre-trial release decision is changed if they are assigned to a more or less lenient judge. Thus, we give more weight to judges whose release preferences impact the pre-trial release status of a greater number of defendants.

In the following section, we formally establish the conditions under which we can consistently estimate D^* using the random assignment of cases to bail judges. We begin by assessing the bias that arises from simple OLS estimates. We then turn to our IV estimator for racial bias and show that our estimator yields a consistent estimate of D^* under two conditions: (1) that the instrument for pre-trial release Z_i is a continuous measure of judge leniency and (2) that the IV weights are constant by race, a condition that is satisfied if the first-stage relationship between pre-trial release and our preferred measure of Z_i is linear.

Bias with OLS Estimates: Let defendant i's probability of pre-trial misconduct, Y_i, be given by the following relationship:

$$Y_i = \alpha_W Released_i \cdot White_i + \alpha_B Released_i \cdot Black_i + \beta \mathbf{X}_i + \mathbf{U}_i + \varepsilon_i \tag{6}$$

where $Released_i$ is an indicator for being released before trial, $White_i$ and $Black_i$ are race indicators, $\mathbf{X}_i$ denotes characteristics of the defendant observed by both the econometrician and bail judge, and $\mathbf{U}_i$ denotes characteristics observed by the bail judge but not the econometrician. In practice, $\mathbf{X}_i$ includes variables such age, gender, type of crime, and prior offenses, while $\mathbf{U}_i$ include characteristics such as the defendant's physical appearance and any information conveyed during the bail hearing. ε_i is the idiosyncratic defendant-level variation that is unobserved by both the econometrician and the judge.

OLS estimates of α_W and α_B from Equation (6) will typically not recover unbiased estimates of the true rate of pre-trial misconduct for white and black defendants at the margin of release for two reasons. First, characteristics observable to the judge but not the econometrician, $\mathbf{U}_i$, may be correlated with $Released_i$, resulting in omitted variable bias. For example, bail judges may be more likely to release defendants who both appear to be less dangerous during the bail hearing and who are, in fact, less likely to have an incident of pre-trial misconduct. In this scenario, OLS estimates of Equation (6) will be biased downwards from the true average treatment effect.

The second, and more important, reason OLS estimates will not recover unbiased estimates of treatment effects for white and black defendants at the margin of release is that the treatment effect of pre-trial release may be correlated with judges' decision rules, meaning that the average treatment effect identified by OLS will not be equal to the marginal treatment effect required by our test (e.g., Ayres 2002). Thus, even if the econometrician observes the full set of observables known to the bail judge, $\mathbf{X}_i$ and $\mathbf{U}_i$, OLS estimates are still not sufficient to test for racial bias unless one is willing to assume constant treatment effects across the entire distribution of defendants (i.e. assuming that the average treatment effect is equal to the marginal treatment effect). In our model, we explicitly rule out constant treatment effects by allowing judges' race-specific decision rules to be correlated with the expected treatment effect, $\mathbb{E}[\alpha_i | r_i = r]$ (see Equation 1). In this scenario, the average treatment effect will be an underestimate of the marginal treatment effect required by our outcome test.

In this paper, we identify racial bias in the presence of both omitted variables and infra-marginality issues using the local nature of instrumental variables estimators to estimate causal treatment effects for individuals at the margin of release. We now formally establish the conditions under which our judge IV strategy yields consistent estimates of racial bias in bail setting.

Defining our IV Estimator: Before defining our estimator, we briefly review the econometric properties of a race-specific IV estimator that uses judge leniency as an instrumental variable for pre-trial release. Let Z_i be a scalar measure of the assigned judge's propensity for pre-trial release that takes on values ordered $\{z_0, ..., z_J\}$, where $J+1$ is the number of total judges in the bail system. For example, a value of $z_j = 0.5$ indicates that judge j releases 50 percent of all defendants. In practice, we construct Z_i using a standard leave-out procedure that captures the pre-trial release tendency of judges across both white and black defendants. As will be described in further detail in Section II.B, we make a standard monotonicity assumption that the judge ordering produced by the scalar Z_i is the same for both white and black defendants in our main results. We relax this monotonicity assumption in Section III.C by separately calculating our leave-out judge leniency measure by defendant race.

Following Imbens and Angrist (1994), a race-specific IV estimator using Z_i as an instrumental variable for pre-trial release is valid and well-defined under the following three assumptions:

Assumption 1. [Existence]. Pre-trial release is a nontrivial function of Z_i such that a first stage exists:

$$Cov(Released_i, Z_i) \neq 0$$

Assumption 1 ensures that there is a first-stage relationship between our instrument Z_i and the probability of pre-trial release.

Assumption 2. [Exclusion Restriction]. Z_i is uncorrelated with unobserved determinants of Y_i:

$$Cov(Z_i, \mathbf{v}_i) = 0$$

where $\mathbf{v}_i = \mathbf{U}_i + \varepsilon_i$. Assumption 2 ensures that our instrument Z_i is orthogonal to characteristics unobserved by the econometrician, $\mathbf{v}_i$. In other words, Assumption 2 assumes that the assigned judge only affects pre-trial misconduct through the channel of pre-trial release.

Assumption 3. [Monotonicity]. The impact of judge assignment on the probability of pre-trial release is monotonic if for each z_{j-1}, z_j pair:

$$R_i(z_j) - R_i(z_{j-1}) \geq 0$$

where $R_i(z_j)$ equals 1 if defendant i is released if assigned to judge j. Assumption 3 implies that any defendant released by a strict judge would also be released by a more lenient judge, and any defendant detained by a lenient judge would also be detained by a more strict judge.

Under these assumptions, the race-specific IV estimator that uses judge leniency as an instrumental variable for pre-trial release can be expressed as a weighted average of pairwise treatment effects:

$$\alpha_r^{IV} = \sum_{j=1}^{J} \lambda_r^j \cdot \alpha_r^{j,j-1} \tag{7}$$

where λ_r^j are the standard non-negative IV weights which sum to one (Imbens and Angrist 1994), which are previously described in Equation (5). The weights λ_r^j depend on the size of the subpopulation whose treatment status is altered by changing the value of the instrument from z_j to z_{j-1}, as well as the probability of being assigned a particular judge. Each pairwise treatment effect $\alpha_r^{j,j-1}$ captures the treatment effects of compliers within each j, $j-1$ pair. In the potential outcomes framework, $\alpha_r^{j,j-1} = \mathbb{E}[Y_i(1) - Y_i(0)|R_i(z_j) - R_i(z_{j-1}) = 1, r_i = r]$, with $Y_i(1)$ being an indicator for pre-trial misconduct for defendant i if released before trial, $Y_i(0)$ being an indicator for pre-trial misconduct for defendant i if detained before trial, and $R_i(z_j)$ being equal to 1 if defendant i is released if assigned to judge j.

And using the definition of α_r^{IV} from Equation (7), our IV estimator for racial bias can be expressed as:

$$D^{IV} = \alpha_W^{IV} - \alpha_B^{IV} = \sum_{j=1}^{J} \lambda_W^j \alpha_W^{j,j-1} - \sum_{j=1}^{J} \lambda_B^j \alpha_B^{j,j-1} \tag{8}$$

where each pairwise LATE, $\alpha_r^{j,j-1}$, is again the average treatment effect of compliers between judges $j-1$ and j and the weights, λ_r^j, depend on the proportion of compliers between judges j and $j-1$.

Consistency of our IV Estimator: Building on the standard IV framework, we can now establish the two conditions under which our IV estimator for racial bias D^{IV} provides a consistent estimate of D^*. The first condition for our IV estimator D^{IV} to provide a consistent estimate is that our judge leniency measure Z_i is continuously distributed over some interval $[\underline{z}, \bar{z}]$. Formally, as our instrument becomes continuous, for any judge j and any $\epsilon > 0$, there exists a judge k such that $|z_j - z_k| < \epsilon$. Following Angrist, Graddy, and Imbens (2000), as our instrument becomes continuously distributed, each pairwise treatment effect converges to the treatment effect for a defendant at the margin of release at z_j:

$$\alpha_r^j = \alpha_r(z = z_j) = \lim_{dz \to 0} \mathbb{E}[Y_i(1) - Y_i(0)|R_i(z_j) - R_i(z_j - dz) = 1, r_i = r] \tag{9}$$

Proposition 3. As Z_i becomes continuously distributed, each race-specific IV estimate, α_r^{IV}, converges to a weighted average of treatment effects for defendants at the margin of release.

Proof. See Appendix B.

Intuitively, each defendant becomes marginal to a judge as the distance between any two judge leniency measures converges to zero, i.e. the instrument becomes more continuous. Therefore, under this first condition, each race-specific IV estimate approaches a weighted average of treatment effects for defendants at the margin of release. In the limit, the weights of our race-specific IV estimates depend on both the derivative of the probability of release with respect to leniency and the probability density function of our judge leniency measure, i.e. the continuous analog to λ_r^j in Equation (7).

The second condition for our IV estimator D^{IV} to provide a consistent estimate of racial bias D^* is that the weights on the pairwise LATEs must be equal across race. Equal weights ensure that the race-specific IV estimates from Equation (7), α_W^{IV} and α_B^{IV}, provide the same weighted averages of $\alpha_W^{j,j-1}$ and $\alpha_B^{j,j-1}$. If the weights $\lambda_W^j = \lambda_B^j = \lambda^j$, our IV estimator can then be rewritten as a simple weighted average of the difference in pairwise LATEs for white and black defendants:

$$D^{IV} = \sum_{j=1}^{J} \lambda^j (\alpha_W^{j,j-1} - \alpha_B^{j,j-1}) \tag{10}$$

Proposition 4. Our IV estimator D^{IV} provides a consistent estimate of racial bias D^* if (1) λ_r^j is constant by race and (2) Z_i is continuous. The requirement that λ_r^j is constant by race holds if and only if the proportion of compliers shifted by moving across judges is constant by race for each z_{j-1}, z_j pair:

$$\frac{Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)}{Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B)} = c \tag{11}$$

where c is some constant.

Proof. See Appendix B.

In practice, a linear first-stage relationship between pre-trial release and our judge leniency measure by race is a sufficient condition for ensuring that the proportion of compliers shifted by moving from judge $j - 1$ to j is constant by race (see Appendix B). We show below that a linear first stage for each race is consistent with our data (see Figure 1), indicating that the equal weights assumption is unlikely to be violated in our setting.

Under these two conditions, our estimator D^{IV} provides a consistent estimate of the complier-weighted average of racial bias across all judges within a court. Importantly, our estimator allows for any relationship between the leniency of each judge j and judge j's racial bias. For example, our interpretation of D^{IV} remains valid even if lenient judges are biased against black defendants while stricter judges are biased against white defendants. In this scenario, the magnitude of and direction of D^{IV} depend on the distribution of compliers across the lenient and strict judges.

Potential Bias with a Discrete Instrument: The consistency of our judge IV estimator discussed above relies on the condition that our judge instrument is continuous. With a discrete rather than continuous instrument, each defendant is no longer marginal to a particular judge. Because of this infra-marginality concern in the context of a discrete instrument, D^{IV} may no longer provide a consistent estimate of D^*.

There are two approaches to addressing this infra-marginality bias with a discrete instrument. The first is to place additional functional form assumptions on the distribution of the underlying marginal treatment effects to allow for the consistent estimation of racial bias (e.g., Brinch et al. forthcoming). In Appendix B, we show that a sufficient condition for D^{IV} to provide a consistent estimate of true racial bias D^* is that the marginal treatment effects can be well approximated by linear splines with knots at points in the support of leniency. Thus, it remains possible to consistently estimate racial bias when there are a small number of judges if one is willing to make functional form assumptions on the distribution of marginal treatment effects.

A second approach is to characterize the maximum potential bias of our IV estimator D^{IV} relative to the true level of racial bias D^* when there are no additional functional form assumptions on the distribution of marginal treatment effects.

Proposition 5. If Assumptions 1-3 are satisfied and the first-stage relationship is linear, the maximum bias of our IV estimator D^{IV} from the true level of racial bias D^* is given by $\max_j(\lambda^j)(\alpha^{max} - \alpha^{min})$, where α^{max} is the largest treatment effect among compliers, α^{min} is the smallest treatment effect among compliers, and λ^j is given by:

$$\lambda^j = \frac{(z_j - z_{j-1}) \cdot \sum_{l=j}^{J} \pi^l (z_l - \mathbb{E}[Z])}{\sum_{m=1}^{J} (z_j - z_{j-1}) \cdot \sum_{l=m}^{J} \pi^l (z_l - \mathbb{E}[Z])} \tag{12}$$

where π^j is the probability of being assigned to judge j.

Proof. See Appendix B.

The maximum bias of D^{IV} relative to D^* decreases as (1) the distance in leniency between any two judges decreases and (2) the heterogeneity in treatment effects among compliers decreases. Intuitively, if the distance between adjacent judges is large, then the IV estimator incorporates information from infra-marginal defendants in estimating treatment effects. In the limit, as the distance between judges shrinks, all compliers are at the margin of release, and so the potential bias from infra-marginal defendants goes to zero. Similarly, holding fixed the distance between the judge leniency measures, the bias in our estimator decreases as the heterogeneity in treatment effects among compliers decreases. For example, in the extreme, if treatment effects are homogeneous among compliers such that $\alpha^{max} = \alpha^{min}$, our IV estimator D^{IV} continues to provide a consistent estimate of D^*.

In Appendix B, we calculate the maximum bias of D^{IV} relative to D^* when our instrument is discrete. This maximum bias can be estimated using the empirical distribution of judge leniency in our data, the closed form solution for the weights λ^j when the first stage is linear, and worst case assumptions regarding treatment effect heterogeneity between white and black compliers. This calculation indicates that in our setting, the true level of racial bias D^* is within 0.5 percentage points of D^{IV}. We find similar results when we place fewer parametric restrictions on the first stage (e.g., estimate the first stage using 100 separate bins).

D. Discussion and Extensions

In this section, we discuss some important assumptions underlying our test for racial bias, possible extensions to our test, and how they affect the interpretation of our results.

Racial Differences in Arrest Probability: Our test for racial bias assumes that any measurement error in the outcome is uncorrelated with race. This assumption would be violated if, for example, the police are more likely to rearrest black defendants conditional on having committed a new crime and judges minimize new crime, not just new arrests. In this scenario, we will overestimate the probability of pre-trial misconduct for black versus white defendants at the margin and, as a result, underestimate the true amount of racial bias in bail setting. It is therefore possible that our estimates reflect the lower bound on the true amount of racial bias among bail judges.

Omitted Objectives for Release: We also assume that judges do not consider other objectives or outcomes, or what Kleinberg et al. (2017) refer to as the "omitted payoff bias." We will have this kind of omitted payoff bias if, for example, bail judges consider how pre-trial detention impacts a defendant's employment status. This kind of omitted payoff bias will bias our estimates to the extent that these other outcomes or objectives are correlated with race. For example, if judges also minimize employment disruptions when setting bail, and white defendants at the margin of release are less likely to be employed compared to black defendants at the margin, we will again underestimate the true level of racial bias.

We explore the empirical relevance of an omitted payoff bias in several ways. First, we find in unreported results that our estimates are nearly identical if we measure pre-trial misconduct using both any rearrest and any failure to appear (although we can only conduct this test in Philadelphia where we observe missed court appearances). These results are also consistent with Kleinberg et al. (2017), who find similar evidence of prediction errors using rearrests or failures to appear. Second, as will be discussed below, we find similar estimates when we measure pre-trial misconduct using crime-specific rearrest rates to address the concern that judges may be most concerned about reducing violent crimes. Third, we note that Dobbie et al. (2016) find that white defendants at the margin of release are no more likely to be employed in the formal labor market up to four years after the bail hearing compared to black defendants at the margin of release. This goes against the idea that judges may be trading off minimizing pre-trial misconduct with maximizing employment. Finally, as will be discussed below, we find that racial bias against black defendants is larger for part-time and inexperienced judges compared to full-time and experienced judges. There are few conceivable stories where omitted payoffs differ by judge experience.

Taken together, we therefore believe that any omitted payoff bias is likely to be small in practice. This conclusion is also supported by the fact that bail judges are required by law to make release decisions with the narrow objective of minimizing the risk of pre-trial misconduct. Bail judges are also explicitly told not to consider other objectives in deciding who to release or detain. Moreover, bail judges feel enormous political pressure to solely minimize pre-trial misconduct. For example, one bail judge told NPR that elected bail judges feel enormous pressure to detain defendants, and end up setting high bail amounts rather than releasing defendants because "they will have less criticism from the public for letting someone out if that person gets out and commits another crime."[6]

[6] See http://www.npr.org/2016/12/17/505852280/states-and-cities-take-steps-to-reform-dishonest-bail-system

Judge Preferences for Non-Race Characteristics: Bail judges may also be biased across non-race characteristics such as crime type or crime severity. For example, judges may be biased against defendants charged with violent offenses for reasons having nothing to do with race. If black defendants are more likely to be charged with violent offenses, however, then our estimates will reflect both the direct effects of racial bias and the indirect effects of this "offense type" bias.

This possibility suggests two conceptually distinct tests for racial bias. Our preferred test includes both the direct and indirect effects of racial bias as any bias on non-race factors may, in fact, be motivated by race. For example, bail judges could be biased against offenses involving drugs compared to alcohol because blacks are more likely to be arrested for these drug crimes. However, it is also possible to test for the direct effects of racial bias, holding fixed all non-race characteristics such as crime severity and crime type (e.g., Barsky et al. 2002, Chandra and Staiger 2010). In Appendix B, we show that the direct effects of racial bias can be estimated using a re-weighting procedure under the assumption that judge preferences vary only by observable characteristics, i.e. $t_r^j(\mathbf{V}_i) = t_r^j(\mathbf{X}_i)$. In practice, however, this re-weighting procedure yields nearly identical estimates as our preferred non-weighted specifications.

II. Data and Instrument Construction

This section summarizes the most relevant information regarding our administrative court data from Philadelphia and Miami-Dade and the construction of our judge leniency measure. Further details on the cleaning and coding of variables are contained in Appendix C.

A. Data Sources and Descriptive Statistics

Philadelphia court records are available for all defendants arrested and charged between 2010-2014 and Miami-Dade court records are available for all defendants arrested and charged between 2006-2014. For both jurisdictions, the court data contain information on defendant's name, gender, race, date of birth, and zip code of residence. Because our ethnicity identifier does not distinguish between non-Hispanic white and Hispanic white, we match the surnames in our dataset to census genealogical records of surnames. If the probability a given surname is Hispanic is greater than 80 percent, we label this individual as Hispanic. In our main analysis, we include all defendants and compare outcomes for marginal black and marginal white (Hispanic and non-Hispanic) defendants. In robustness checks, we present results comparing marginal black and marginal non-Hispanic white defendants.[7]

The court data also include information on the original arrest charge, the filing charge, and the final disposition charge. We also have information on the severity of each charge based on state-specific offense grades, the outcome for each charge, and the punishment for each guilty disposition. Finally, the case-level data include information on attorney type, arrest date, and the date of and

[7] Appendix Table A1 presents results for marginal Hispanic defendants compared to non-Hispanic white defendants. Perhaps in some part because of measurement error in our coding of Hispanic ethnicity, we find no evidence of racial bias against Hispanics.

judge presiding over each court appearance from arraignment to sentencing. Importantly, the case-level data also include information on bail type, bail amount when monetary bail is set, and whether bail was met. Because the data contain defendant identifiers, we can measure whether a defendant committed pre-trial misconduct by whether the defendant was subsequently arrested for a new crime before the case was resolved.

We make three restrictions to the court data to isolate cases that are quasi-randomly assigned to judges. First, we drop a small set of cases with missing bail judge information. Second, we drop the 30 percent of defendants in Miami-Dade who never have a bail hearing because they post bail immediately following the arrest; below we show that the characteristics of defendants who have a bail hearing are uncorrelated with our judge leniency measure. Third, we drop all weekday cases in Miami-Dade because, as explained in Appendix D, bail judges in Miami-Dade are assigned on a quasi-random basis only on the weekends. The final sample contains 193,431 cases from 116,583 unique defendants in Philadelphia and 93,572 cases from 66,003 unique defendants in Miami-Dade.

Table 1 reports summary statistics for our estimation sample separately by race and pre-trial release status measured at three days within the bail hearing, as recent policy initiatives focus on this time period. In addition, three days is the time period over which the initial bail judge is most likely to affect pre-trial detention. Following the initial bail hearing, defendants have the opportunity to petition for a bail modification that could result in a different bail judge making a different detention decision. On average, black defendants are more 11.2 percentage points more likely to be assigned monetary bail compared to white defendants and receive bail amounts that are $14,376 greater than white defendants. Compared to white defendants, released black defendants are also 6.4 percentage points more likely to be rearrested for a new crime before case disposition. Released black defendants are also 4.1 percentage points, 1.0 percentage points, and 0.8 percentage points more likely to be rearrested for a drug, property, and violent crime, respectively.

B. Construction of the Instrumental Variable

We estimate the causal impact of pre-trial release for the marginal defendant using a measure of the tendency of a quasi-randomly-assigned bail judge to release a defendant pre-trial as an instrument for release. In both Philadelphia and Miami-Dade, there are multiple bail judges serving at each point in time in both jurisdictions, allowing us to utilize variation in bail setting across judges. Both jurisdictions also assign cases to bail judges in a quasi-random fashion in order to balance caseloads: Philadelphia utilizes a rotation system where three judges work together in five-day shifts, with one judge working an eight-hour morning shift (7:30AM-3:30PM), another judge working the eight-hour afternoon shift (3:30PM-11:30PM), and the final judge working the eight-hour evening shift (11:30PM-7:30AM). Similarly, bail judges in Miami-Dade rotate through the weekend felony and misdemeanor bail hearings. Additional details on the setting can be found in Appendix D.

We construct our instrument using a residualized, leave-out judge leniency measure that accounts for case selection following Dahl et al. (2014) and Dobbie et al. (2016). Because the judge assignment procedures in Philadelphia and Miami-Dade are not truly random as in other settings, selection may impact our estimates if we used a simple leave-out mean to measure judge leniency following the previous literature (e.g., Kling 2006, Aizer and Doyle 2015). For example, bail hearings following DUI arrests disproportionately occur in the evenings and on particular days of the week, leading to case selection. If certain bail judges are more likely to work evening or weekend shifts due to shift substitutions, the simple leave-out mean will be biased.

Given the rotation systems in both counties, we account for court-by-bail year-by-bail day of week fixed effects and court-by-bail month-by-bail day of week fixed effects. In Philadelphia, we add additional bail-day of week-by-bail shift fixed effects. Including these exhaustive court-by-time effects effectively limits the comparison to defendants at risk of being assigned to the same set of judges. With the inclusion of these controls, we can interpret the within-cell variation in the instrument as variation in the propensity of a quasi-randomly assigned bail judge to release a defendant relative to the other cases seen in the same shift and/or same day of the week.

Let the residual pre-trial release decision after removing the effect of these court-by-time fixed effects be denoted by:

$$Released^*_{ict} = Released_{ic} - \gamma \mathbf{X}_{ict} = Z_{ctj} + v_{ict} \tag{13}$$

where $\mathbf{X}_{ict}$ includes the respective court-by-time fixed effects. The residual release decision, $Released^*_{ict}$, includes our measure of judge leniency Z_{ctj}, as well as unobserved defendant level variation v_{ict}.

For each case, we then use these residual bail release decisions to construct the leave-out mean decision of the assigned judge within a bail year:

$$Z_{ctj} = \left(\frac{1}{n_{tj} - n_{itj}}\right)\left(\sum_{k=0}^{n_{tj}}(Released^*_{ikt}) - \sum_{c=0}^{n_{itj}} Released^*_{ict}\right) \tag{14}$$

where n_{tj} is the number of cases seen by judge j in year t and n_{itj} is the number of cases of defendant i seen by judge j in year t. We calculate the instrument across all case types (i.e. both felonies and misdemeanors), but allow the instrument to vary across years. In robustness checks, we allow judge tendencies to vary by defendant race.

The leave-out judge measure given by Equation (14) is the release rate for the first assigned judge after accounting for the court-by-time fixed effects. This leave-out measure is important for our analysis because regressing outcomes for defendant i on our judge leniency measure without leaving out the data from defendant i would introduce the same estimation errors on both the left- and right-hand side of the regression and produce biased estimates of the causal impact of being released pre-trial. In our two-stage least squares results, we use our predicted judge leniency measure, Z_{ctj}, as an instrumental variable for whether the defendant is released pre-trial.

Figure 1 presents the distribution of our residualized judge leniency measure for pre-trial release at the judge-by-year level for all defendants, white defendants, and black defendants. Our sample includes seven total bail judges in Philadelphia and 170 total bail judges in Miami-Dade. In Philadelphia, the average number of cases per judge is 27,633 during the sample period of 2010-2014, with the typical judge-by-year cell including 6,239 cases. In Miami-Dade, the average number of cases per judge is 550 during the sample period of 2006-2014, with the typical judge-by-year cell including 187 cases. Controlling for our vector of court-by-time effects, the judge release measure ranges from -0.164 to 0.205 with a standard deviation of 0.036. In other words, moving from the least to most lenient judge increases the probability of pre-trial release by 37.1 percentage points, a 72.3 percent change from the mean three-day release rate of 50.6 percentage points.

One question might be why judges differ in their bail decisions. Dobbie et al. (2016) show that defendants on the margin of pre-trial release are those for whom judges disagree about the appropriateness of non-monetary versus monetary bail, not those for whom judges disagree about the appropriateness of ROR versus other bail decisions. While interesting for thinking about the design of the bail determination process, however, it is not critical to our analysis to know precisely why some judges are more lenient than others. What is critical is that some judges are systematically more lenient than others, that judge assignment only impacts defendants through the pre-trial detention decision, and that defendants released by a strict judge would also be released by a lenient one. We consider below whether each of these conditions holds in our data.

Another question is how many and what types of defendants are compliers in our setting. In Appendix Table A2, we describe the characteristics of compliers in our sample following the approach developed by Abadie (2003) and extended by Dahl et al. (2014). Compliers in our sample are 12 percentage points more likely to be charged with a misdemeanor and 17 percentage points more likely to be charged with non-violent offenses compared to the average defendant. Compliers are not systematically different from the average defendant by race or prior criminal history, however. We also find that 13 percent of defendants in our sample are "compliers," meaning that they would have received a different bail outcome had their case been assigned to the most lenient judge instead of the most strict judge. In comparison, 53 percent of our sample are "never takers," meaning that they would be detained by all judges, and 34 percent are "always takers," meaning that they would be released pre-trial regardless of the judge assigned to the case.

C. Instrument Validity

Existence and Linearity of First Stage: To examine the first-stage relationship between bail judge leniency and whether a defendant is released pre-trial (*Released*), we estimate the following equation for individual i and case c, assigned to judge j at time t using a linear probability model:

$$Released_{ictj} = \gamma_0 + \gamma_1 Z_{ctj} \cdot White_i + \gamma_2 Z_{ctj} \cdot Black_i + \pi \mathbf{X}_{ict} + v_{ict} \tag{15}$$

where the vector $\mathbf{X}_{ict}$ includes court-by-time fixed effects. The error term v_{ict} is composed of characteristics unobserved by the econometrician but observed by the judge, as well as idiosyncratic variation unobserved to both the judge and econometrician. As described previously, Z_{ctj} are leave-out (jackknife) measures of judge leniency that are allowed to vary across years. Robust standard errors are two-way clustered at the individual and judge-by-shift level.

Figure 1 provides graphical representations of the first stage relationship, pooled and separately by race, between our residualized measure of judge leniency and the probability of pre-trial release controlling for our exhaustive set of court-by-time fixed effects, overlaid over the distribution of judge leniency. The graphs are a flexible analog to Equation (15), where we plot a local linear regression of actual individual pre-trial release against judge leniency. The individual rate of pre-trial release is monotonically increasing for both races, and approximately linearly increasing in our leniency measure. These results suggest that a linear first stage for defendants of both races, and thus the assumption of constant IV weights by race (Proposition 4), is likely valid in our setting.

Table 2 presents formal first stage results from Equation (15) for all defendants, white defendants, and black defendants. Columns 1, 3, and 5 begin by reporting results with only court-by-time fixed effects. Columns 2, 4, and 6 add our baseline crime and defendant controls: race, gender, age, whether the defendant had a prior offense in the past year, the number of charged offenses, indicators for crime type (drug, DUI, property, violent, other) and crime severity (felony or misdemeanor), and indicators for missing characteristics.

We find that our residualized judge instrument is highly predictive of whether a defendant is released pre-trial, with an F-statistic for the instrument of 501.8. Our results show that a defendant assigned to a bail judge that is 10 percentage points more likely to release a defendant pre-trial is 5.9 percentage points more likely to be released pre-trial. Judge leniency is also highly predictive of pre-trial release for both white and black defendants. A white defendant assigned to a bail judge that is 10 percentage points more likely to release a defendant pre-trial is 5.4 percentage points more likely to be released pre-trial and a black defendant assigned to a bail judge that is 10 percentage points more likely to release a defendant pre-trial is 6.4 percentage points more likely to be released pre-trial.

Exclusion Restriction: Table 3 verifies that assignment of cases to bail judges is random after we condition on our court-by-time fixed effects. Columns 1, 3, and 5 of Table 3 uses a linear probability model to test whether case and defendant characteristics are predictive of pre-trial release. These estimates capture both differences in the bail conditions set by the bail judges and differences in these defendants' ability to meet the bail conditions. We control for court-by-time fixed effects and two-way cluster standard errors at the individual and judge-by-shift level. For example, we find that black male defendants are 12.6 percentage points less likely to be released pre-trial compared to similar female defendants, while white male defendants are 11.5 percentage points less likely to be released pre-trial compared to similar female defendants. White defendants with a prior offense in the past year are 20.1 percentage points less likely to be released compared to defendants with no prior offense, while black defendants with a prior offense in the past year are 14.5 percentage points less likely to be released compared to defendants with no prior offense. Columns 2, 4, and 6 assess whether these same case and defendant characteristics are predictive of our judge leniency measure using an identical specification. We find that judges with differing leniencies are assigned cases with very similar defendants.

Even with random assignment, the exclusion restriction could be violated if bail judge assignment impacts the probability of pre-trial misconduct through channels other than pre-trial release. The assumption that judges only systematically affect defendant outcomes through pre-trial release is fundamentally untestable, and our estimates should be interpreted with this potential caveat in mind. However, we argue that the exclusion restriction assumption is reasonable in our setting. Bail judges exclusively handle one decision, limiting the potential channels through which they could affect defendants. In addition, we are specifically interested in short-term outcomes (pre-trial misconduct) which occur prior to disposition, further limiting the role of alternative channels that could affect longer-term outcomes. Finally, Dobbie et al. (2016) find that there are no independent effects of the money bail amount or the non-monetary bail conditions, and that bail judge assignment is uncorrelated with the assignment of public defenders and subsequent trial judges.

Monotonicity: The final condition needed to interpret our estimates as the LATE of pre-trial release is that the impact of judge assignment on the probability of pre-trial release is monotonic across defendants. In our setting, the monotonicity assumption requires that individuals released by a strict judge would also be released by a more lenient judge and that individuals detained by a lenient judge would also be detained by a stricter judge. If the monotonicity assumption is violated, our two-stage least squares estimates would still be a weighted average of pairwise local average treatment effects, but the weights would not sum to one (Angrist et al. 1996, Heckman and Vytlacil 2005). The monotonicity assumption is therefore necessary to interpret our estimates as a well-defined LATE.

An implication of the monotonicity assumption is that the first stage estimates should be non-negative for all subsamples. Appendix Table A3 present these first stage results using the full sample of cases to calculate our measure of judge leniency. We find that our residualized measure of judge leniency is consistently non-negative and sizable in all subsamples, in line with the monotonicity assumption. Appendix Figure A1 further explores how judges treat cases of observably different defendants by plotting our residualized judge leniency measures calculated separately by offense type, offense severity, and prior criminal history. Each plot reports the coefficient and standard error from an OLS regression relating each measure of judge leniency. Consistent with our monotonicity assumption, we find that the slopes relating the relationship between judge leniency in one group and judge leniency in another group are non-negative, suggesting that judge tendencies are similar across observably different defendants and cases.

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

We apply our proposed method to estimate the probability of pre-trial misconduct for white and black defendants on the margin of release. Specifically, we estimate the following two-stage least squares specification for individual i and case c, assigned to judge j at time t:

$$Y_{ict} = \beta_0 + \alpha_W^{IV} Released_{ic} \cdot White_i + \alpha_B^{IV} Released_{ic} \cdot Black_i + \beta_1 \mathbf{X}_{ict} + \mathbf{v}_{ict} \tag{16}$$

where the vector $\mathbf{X}_{ict}$ includes court-by-time fixed effects and defendant gender, age, whether the defendant had a prior offense in the past year, the number of charged offenses, indicators for crime type (drug, DUI, property, violent, or other), crime severity (felony or misdemeanor), and indicators for any missing characteristics. As described previously, the error term $\mathbf{v}_{ict} = \mathbf{U}_i + \varepsilon_{ict}$ consists of characteristics unobserved by the econometrician but observed by the judge, $\mathbf{U}_i$, and idiosyncratic variation unobserved by both the econometrician and judge, ε_{ict}. We instrument for pre-trial release with the interaction of defendant race and our measure of judge leniency, Z_{ctj}. Robust standard errors are two-way clustered at the individual and judge-by-shift level.

Table 4 presents estimates of Equation (16). Columns 1-2 reports two-stage least squares estimates of the causal effect of pre-trial release on the probability of rearrest prior to case disposition for marginal white defendants, α_W^{IV}, and marginal black defendants, α_B^{IV}, respectively. Column 3 reports our estimate of racial bias $D^{IV} = \alpha_W^{IV} - \alpha_B^{IV}$. Panel A presents results for the probability of rearrest for any crime prior to case disposition, while Panel B presents results for rearrest rates for drug, property, and violent offenses separately. In total, 20.8 percent of defendants are rearrested for a new crime prior to disposition, with 9.1 percent of defendants being rearrested for drug offenses and 5.9 percent of defendants being rearrested for property offenses.

We find convincing evidence of racial bias against black defendants. In Panel A, we find that marginally released white defendants are 18.5 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 insignificant 0.5 percentage points (column 2). Taken together, these estimates imply that marginally released white defendants are 18.0 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 0.5 percentage points (see Appendix B).

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 white defendants are 9.7 percentage points more likely to be rearrested for a drug crime prior to case disposition than marginally released black defendants (p-value = 0.024). Marginally released white defendants are also 3.0 percentage points more likely to be rearrested for a property crime compared to marginally released black defendants (p-value = 0.579), and marginally released whites are about 8.2 percentage points more likely to be rearrested for a violent crime prior to disposition than marginally released blacks (p-value = 0.036). 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.

In Appendix Table A4, we present results comparing outcomes for marginal non-Hispanic white defendants and marginal black defendants. We find very similar results consistent with racial bias against black defendants. Overall, these findings indicate significant racial bias against black defendants, driven largely by differences in the probability of committing a new drug crime for marginal white and marginal black defendants.[8] Our results therefore rule out statistical discrimination as the sole determinant of racial disparities in bail.

Our IV estimates for racial bias capture the difference in the weighted average treatment effects for white defendants and black defendants at the margin of release. To better understand the parts of the judge leniency distribution that drive these results, we estimate treatment effects for defendants at different margins of release by calculating marginal treatment effects (MTEs) over our judge leniency range. In practice, the MTE is estimated by taking the derivative of our outcome measure with respect to the predicted probability of being released (i.e. the propensity score). We estimates these MTEs in two steps. In the first step, we use our judge leniency measure to estimate the propensity score, capturing the variation in treatment status due solely to the instrument (Doyle 2007). In the second step, we compute the numerical derivative of a smoothed function relating rearrest prior to disposition to the propensity score following Heckman and Vytlacil (2006). Specifically, we residualize the rearrest prior to case disposition using court-by-time fixed effects and then estimate the relationship between the residualized variable and the propensity score using a local quadratic estimator. To obtain the MTE, we compute the numerical derivative of the local quadratic estimator. Figure 2 presents the MTEs, by defendant race, as a function of our judge leniency measure. Low propensity scores correspond to strict judges while high propensity scores correspond to lenient judges. Figure 2 reveals that the MTEs for white defendants lie strictly above the MTEs for black defendants, implying that marginally released white defendants are riskier than marginally released black defendants at all points in the distribution. These results, while less precise than our IV estimates, indicate that racial bias against black defendants arises at every part of the judge leniency distribution. These MTE results also suggest that we would find racial bias in bail setting regardless of the weighting scheme, and that our main results are not driven by the decision to use the standard IV weights, λ^j.

B. Subsample Results

To explore heterogeneous treatment effects, we combine all observable demographic and crime characteristics into a single risk index. In Table 5, we divide defendants into above and below median predicted risk, with those in the below median group having a 12.5 percent probability of rearrest prior to case disposition compared to 31.2 percent among defendants in the above median group.[9] We find that racial bias against black defendants is almost exclusively driven by those with the high-

[8]For completeness, Figure 1 provides a graphical representations of our reduced form results separately by race. Following the first stage results, we plot the reduced form relationship between our judge leniency measure and the residualized rate of rearrest prior to case disposition, estimated using local linear regression.

[9]In small samples, endogenous stratification may lead to biased estimates (e.g., Abadie, Chingos, and West 2014). We find identical results if we use a split-sample estimator to predict risk in a 5 percent random sample and estimate our two-stage least squares results in the remaining 95 percent of the sample.

est predicted risk of rearrest. Among high-risk defendants, marginally released white defendants are 36.5 percentage points more likely to be rearrested prior to case disposition than marginally released black defendants (p-value = 0.013). In contrast, we find no evidence of racial bias against black defendants among low-risk defendants (p-value = 0.752).

In Appendix Tables A5-A8, we explore additional subsample results. In Appendix Table A5, we analyze whether racial bias against black defendants is larger among those charged with drug offenses versus non-drug offenses. This subsample split is of particular interest because black defendants in our sample are more likely to be charged with drug offenses compared to white defendants, and conditional on being charged with a drug offense, are less likely to be released before trial. We find that our main results are largely driven by the differential treatment of white and black defendants charged with drug offenses. Among drug offenders, marginally released white defendants are 36.0 percentage points more likely to be rearrested prior to case disposition than marginally released black defendants (p-value = 0.024). In contrast, we find limited evidence of racial bias among defendants arrested for all other non-drug crimes (p-value = 0.313).

Another important dimension on which white and black defendants differ, and which affects the likelihood of pre-trial release, is the likelihood of having a prior offense from the last year. In Appendix Table A6, we find evidence that racial bias against black defendants is also driven by defendants with a prior in the past year. Among prior offenders, marginally released white defendants are 31.1 percentage points more likely to be rearrested prior to case disposition than marginally released black defendants (p-value = 0.014), whereas we find limited evidence of racial bias among defendants with no recent priors (p-value = 0.434). In Appendix Tables A7-A8, we also find that racial bias against black defendants is larger among defendants charged with felonies (p-value = 0.011) and defendants from below median income zip codes (p-value = 0.058).

C. Robustness

Our main results are robust to a number of alternative specifications. In Appendix Table A9, we present analogous re-weighted two-stage least squares with the weights chosen to match the distribution of observable characteristics by race. After re-weighting on observables, we find that marginally released white defendants are 15.9 percentage points more likely to be rearrested prior to case disposition than marginally released black defendants (p-value = 0.061), driven largely by differences in rearrest rates for drug crimes among marginal white and marginal black defendants (p-value = 0.025). These results indicate that even after accounting for differences in other observable characteristics by defendant race, bail judges appear to be directly racially biased against black defendants.

In Appendix Table A10, we present our main results clustering more conservatively at the individual and judge level. In Appendix Table A11, we reestimate the main results using a version of our instrument constructed separately for white and black defendants. By calculating the instrument separately by defendant race, we relax the monotonicity assumption and specifically allow for judge tendencies to vary across white and black defendants. In Appendix Table A12, we present our main results with bootstrap-clustered standard errors, which correct for estimation error in the construction of our judge leniency measure.[10] Under these alternative specifications, we continue to find that marginally released white defendants are significantly more likely to be rearrested prior to disposition than marginally released black defendants, evidence of racial bias against black defendants.

D. Comparison to Other Outcome Tests

In this section, we replicate the outcome tests from Knowles et al. (2001) and Anwar and Fang (2006) in our sample. In the context of bail setting, the Knowles et al. (2001) test relies on the prediction that, under the null hypothesis of no racial bias, the average pre-trial misconduct rate will not vary by defendant race. The Anwar and Fang (2006) test instead relies on the prediction that, under the null hypothesis of no relative racial bias, the relative treatment of white defendants compared to black defendants does not depend on judge race.

Appendix Table A13 presents results for the Knowles et al. (2001) test for absolute racial bias. We estimate an OLS regression of pre-trial release on an indicator for rearrest before case disposition for both white and black defendants. This OLS specification compares the average rearrest rates for white and black defendants conditional on observables. In contrast to our preferred IV test, the OLS results indicate that judges are not racially biased against black defendants (p-value = 0.424), indicating that there are omitted variables biasing the OLS estimates, that the marginal effect of pre-trial release is not equal to the average effect of pre-trial release, or both. While it is not possible to distinguish between these various explanations using our data, these results suggest that the Knowles et al. (2001) test is invalid in our setting.

Appendix Tables A14-A15 present results for the Anwar and Fang (2006) test for relative racial bias. Information on the race of each bail judge in our sample comes from official court directories and internet searches. In Miami, there are 91 white judges, 61 Hispanic judges, and 15 black judges in our sample. In Philadelphia, however, all seven bail judges in our sample are white, making it impossible to implement any tests of relative racial bias. We therefore restrict the sample for these tests to cases in Miami. See Appendix C for additional details on the coding of judge race.

Appendix Table A14 presents average release rates and average rearrest rates conditional on release by both judge and defendant race. Unlike Anwar and Fang (2006), we find that judges do not differ substantially in their treatment of black versus white defendants. For example, Panel A of Appendix Table A14 indicates that 34.5 percent of white defendants are released by white judges and 33.9 percent of white defendants are released by black judges. Similarly, black defendants are generally less likely to be released by both white judges (31.1 percent) and black judges (31.8 percent). These results suggest that judges are monolithic in their treatment of both white and

[10] We calculate the bootstrap-clustered standard errors using the procedure outlined in Cameron, Gelbach, and Miller (2008). First, we draw 500 bootstrap samples at the judge-by-shift level with replacement, re-constructing our measure of leniency within each bootstrap sample. Second, we run our two-stage least squares specification to estimate α_W^{IV}, α_B^{IV}, and D^{IV} within each of the 500 bootstrap samples. Finally, we use the standard deviations of these 500 estimates to calculate the bootstrap-clustered standard errors.

black defendants.

Appendix Table A15 presents bootstrapped p-values from a test of relative racial bias, i.e. whether white judges are more lenient for white defendants than black defendants and whether black judges are more lenient for black defendants than white defendants. Following Anwar and Fang (2006), the null hypothesis is that there is no reversal in the relative treatment by judge race. Consistent with our estimates from Appendix Table A14, we find no evidence of relative racial bias using the Anwar and Fang (2006) test for either pre-trial release rates (p-value = 0.364) or rearrest rates conditional on release (p-value = 0.412). These results suggest that both white and black judges are racially biased against black defendants. In results available upon request, we also find that the IV estimate of racial bias is similar among white and black judges in Miami, although the confidence intervals for these estimates are extremely large, making definitive conclusions impossible.

These results highlight the importance of accounting for both infra-marginality and omitted variables when estimating racial bias in the criminal justice system. The (false) finding of no racial bias using standard OLS specifications suggests that recent attempts to measure judge decisions using machine learning algorithms could be biased by these issues, as is extensively discussed by Kleinberg et al. (2017). Moreover, our finding that bail judges are monolithic in their treatment of white and black defendants and, as a result, that there is no relative racial bias in bail setting, highlights the importance of developing empirical tests that can detect absolute racial bias.

IV. Potential Mechanisms

In this section, we attempt to differentiate between two alternative theories for the racial bias observed in our setting: (1) racial prejudice (e.g., Becker 1957) and (2) racially biased prediction errors (e.g., Bordalo et al. 2016).

A. Racial Prejudice

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). Bail judges could, for example, harbor explicit prejudices against black defendants that lead 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 prejudice.[11] Racial prejudice 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

[11] Implicit bias is correlated with the probability of making negative judgments about the ambiguous actions by blacks (Rudman and Lee 2002), of exhibiting a variety of micro-behaviors indicating discomfort with minorities (McConnell and Leibold 2001), and of showing greater activation of the area of the brain associated with fear-driven responses to the presentation of unfamiliar black versus white faces (Phelps et al. 2000).

and judges. Prior work has shown that it is exactly these types of settings where racial prejudice is most likely to translate into adverse outcomes for minorities (e.g., Greenwald et al. 2009).

A partial test of this hypothesis is provided by the Anwar and Fang (2006) test discussed above. These results suggest that judges are monolithic in their treatment of both white and black defendants and, as a result, that there is no relative racial bias in bail setting. We also find that our IV estimate of racial bias is similar among white and black judges, although the confidence intervals for these estimates are extremely large. Taken together, these results suggest that racial prejudice is unlikely to be the main driver of our results.

B. Racially Biased Prediction Errors

A second explanation for our results is that judges are making racially biased prediction errors. Bordalo et al. (2016) show, for example, that representativeness heuristics – that is, probability judgments based on the most distinctive differences between groups – can lead to anti-black stereotypes that exaggerate the perceived differences between blacks and whites. In our setting, these kinds of race-based heuristics 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.[12]

Representativeness of Black and White Defendants: We first explore whether our data are consistent with the formation of the type of anti-black stereotypes described by Bordalo et al. (2016). These anti-black stereotypes should only be present if blacks are over-represented among the right tail of the predicted risk distribution. To test this, we calculate the likelihood ratios for black defendants relative to white defendants in our sample, $\mathbb{E}(x|Black)/\mathbb{E}(x|White)$, for all observed characteristics and outcomes. Figure 3 presents the distribution of the predicted risk of rearrest prior to case disposition using the full set of crime and defendant characteristics, as well as the likelihood ratios throughout the risk distribution.[13] Results for individual characteristics are presented in Appendix Table A16. Consistent with anti-black stereotypes, we find that black defendants are significantly over-represented in the right tail of the predicted risk distribution. Black defendants are 1.65 times

[12] For example, some jurisdictions do not require bail judges to have any legal education or certification other than a one-day training session, while in other jurisdictions bail hearings are conducted by "generalist" judges that have no specific training in bail setting and who only assist with bail hearings a few days a year. See https://bangordailynews.com/2011/03/22/business/maine's-bail-system-a-19th-century-holdoverpart\-1-of-4people-who-set-bail-in-maine-have-almost-no-legal-training. Recent reforms include increased training for bail judges and mandatory review of all bail determinations by a second judge. See http://www.nytimes.com/2015/10/02/nyregion/jonathan-lippman-bail-incarceration-new-york-state-chief-judge.html. Other jurisdictions encourage new bail judges to shadow experienced ones. See http://pinetreewatchdog.org/maines-bail-system-best-state-can-afford-or-a-threat-to-due-process/.

[13] Our measures of representativeness and predicted risk may be biased if judges base their decisions on variables that are not observed by the econometrician (e.g., demeanor at the bail hearing). Following Kleinberg et al. (2017), we can test for the importance of unobservables in bail decisions by splitting our sample into a training set to generate the risk predictions and a test set to test those predictions. We find that our measure of predicted risk from the training set is a strong predictor of true risk in the test set, indicating that our measure of predicted risk is not systematically biased by unobservables (see Appendix Figure A2).

more likely to be represented than whites among the top 25 percent of the predicted risk distribution, and 2.14 times more likely to be represented among the top five percent of the predicted risk distribution.

In Appendix E, we show that these black-white differences in the predicted risk distribution are large enough to rationalize the black-white differences in pre-trial release rates. To show this, we follow Bordalo et al. (2016) and assume that judges form beliefs about the distribution of risk through a representativeness-based discounting model. Under this model, the weight attached to a given risk type t is increasing in the representativeness of t. Formally, let $\pi_{t,r}$ be the probability that a defendant of race r is in risk category t. Let $\pi^{st}_{t,r}$ be the stereotyped belief that a defendant of race r is in risk category t. The stereotyped beliefs for black defendants, $\pi^{st}_{t,B}$, is given by:

$$\pi^{st}_{t,B} = \pi_{t,B} \frac{\left(\frac{\pi_{t,B}}{\pi_{t,W}}\right)^{\theta}}{\sum_{s \in T} \pi_{s,B} \left(\frac{\pi_{s,B}}{\pi_{s,W}}\right)^{\theta}} \tag{17}$$

where θ captures the extent to which representativeness distorts beliefs and the representativeness ratio, $\frac{\pi_{t,B}}{\pi_{t,W}}$, is equal to the probability a defendant is black given risk category t divided by the probability a defendant is white given risk category t. Using this approach, we find that a $\theta = 2.5$ can rationalize the average release rate for blacks. To understand how far these beliefs are from the true distribution of risk, we plot the stereotyped distribution for blacks with $\theta = 2.5$ alongside the true distribution of risk for blacks in Appendix Figure E1. These results indicate that a relatively modest shift in the true risk distribution for black defendants is sufficient to explain the large racial disparities we observe in our setting.

Further evidence on anti-black stereotypes comes from a comparison of the crime-specific distributions of risk. Black defendants are most over-represented in the right tail of the predicted risk distribution for new drug and new violent crimes, but not over-represented at all in the right tail of the risk distribution for new property crimes (see Appendix Figure A3). Consistent with anti-black stereotypes, we find strong evidence of racial bias for rearrests of new drug and new violent crimes, but weak evidence of racial bias for rearrests of new property crimes (see Table 4).

A final piece of evidence comes from an analysis of Hispanic defendants. Consistent with stereotyping, we find that the risk distributions of Hispanic and white defendants overlap considerably (see Appendix Figure A4) and that there is no bias against Hispanic defendants (see Appendix Table A1). Thus, all of our results are broadly consistent with bail judges making race-based prediction errors due to anti-black stereotypes and representativeness-based thinking, which in turn leads to the over-detention of black defendants at the margin of release.

Racial Bias and Prediction Errors: Another testable implication of race-based prediction errors is that racial bias should be larger in situations where prediction errors (of any kind) are more likely to occur. For example, Kleinberg et al. (2017) show that bail judges struggle to form accurate risk predictions for the most observably high-risk defendants. It is plausible that judges rely on stereotypical thinking and heuristics in exactly these types of situations. Consistent with this theory, we find significantly more racial bias among observably high-risk defendants (see Table 5). In contrast, there is no reason to believe that racial prejudice should be different for low- and high-risk defendants.

An alternative test uses a comparison of experienced and inexperienced judges. When a defendant violates the conditions of release (such as by committing a new crime), he or she is taken into custody and brought to court for a hearing during which the bail judge decides whether to revoke bail. As a result, judges may obtain more information on the actual risk of pre-trial misconduct for white and black defendants as they acquire greater on-the-job experience. Consistent with this idea, we find that more experienced bail judges are more likely to release defendants, but no more likely to make mistakes (see Appendix Figure A5).[14] Thus, while it appears plausible that prediction errors will decrease with experience, there is no reason to believe that racial prejudice will change with experience.[15]

Table 6 presents a series of estimates for judges with different levels of experience. We first exploit the fact that in Philadelphia, bail judges are full-time judges who specialize in setting bail 24 hours a day, seven days a week, hearing an average of 6,239 cases each year. Conversely, the Miami bail judges in our sample are part-time generalists who work as trial court judges on weekdays and assist the bail court on weekend, hearing an average of only 187 bail cases each year. If racially biased prediction errors decrease with on-the-job experience, the degree of racial bias as estimated under our test should be different across the two jurisdictions.

Columns 1-3 of Table 6 presents our estimates of racial bias, D^{IV}, separately by court. Column 1 reports the difference in pre-trial misconduct rates for marginal white and marginal black defendants in Miami, column 2 reports the difference in pre-trial misconduct rates for marginal white and marginal black defendants in Philadelphia, and column 3 reports the difference between the two jurisdictions. Consistent with racially biased prediction errors being more common among inexperienced judges, we find that racial bias is higher in Miami than Philadelphia (p-value = 0.094). In Miami, marginally released white defendants are 29.1 percentage points more likely to be rearrested compared to marginally released black defendants (p-value = 0.071). In Philadelphia, we find no statistically significant evidence of racial bias, suggesting the possible importance of experience in alleviating any prediction errors.

We can also exploit the substantial variation in the experience profiles of the Miami bail judges

[14]Ideally, we would estimate a series of IV specifications separately by both race and other relevant observable characteristics for each year of experience. In practice, however, these subsample estimates are too imprecise to be informative. Instead, we plot the relationship between judicial experience and both the residualized rate of pre-trial release and the residualized rate of rearrest prior to case disposition conditional on release (i.e. the mistake rate). Pre-trial release and rearrest prior to case disposition are both residualized using the full set of court-by-time fixed effects to control for any systematic differences in the types of defendants seen by judges.

[15]One potential concern is that intergroup contact can increase tolerance towards minority groups. For example, Van Laar et al. (2005) and Boisjoly et al. (2006) show that living with a minority group increases tolerance among white college students, Dobbie and Fryer (2013) show that teaching in a school with mostly minority children increases racial tolerance, and Clingingsmith et al. (2009) show that winning a lottery to participate in the Hajj pilgrimage to Mecca increases belief in equality and harmony of ethnic groups. However, it is not clear how these findings should be extrapolated to our setting, where judges primarily interact with blacks who are criminal defendants.

in our sample. For example, splitting by the median number of years hearing bail cases, the average experienced Miami judge has 9.5 years of experience working in the bail system, while the average inexperienced Miami judge has only 2.5 years of experience working in the bail system. Columns 4 through 6 of Table 6 present results separately for Miami judges with above- and below-median levels of experience. Consistent with our across-court findings, we find suggestive evidence that inexperienced judges are more racially biased than experienced judges (p-value = 0.290). Among inexperienced judges, marginally released white defendants are 48.0 percentage points more likely to be rearrested compared to marginally released black defendants (p-value = 0.071). Among experienced judges, marginally released white defendants are 19.3 percentage points more likely to be rearrested compared to marginally released black defendants (p-value = 0.228).

Taken together, our results suggest that bail judges make racially biased prediction errors, but are not racially prejudiced per se. These results are broadly consistent with recent work by Kleinberg et al. (2017) showing that bail judges make significant prediction errors 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 a prior offense versus no prior offense).

We find several pieces of evidence consistent with our results being driven by racially biased prediction errors, as opposed to racial prejudice 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 prejudice. 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 representativeness-based discounting. 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 race-based heuristics 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. (2017), 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 decisions.

References

[1] Abadie, Alberto. 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." Journal of Econometrics, 113(2): 231-263.

[2] Abadie, Alberto, Matthew M. Chingos, and Martin R. West. 2014. "Endogenous Stratification in Randomized Experiments." NBER Working Paper No. 19742.

[3] Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" Journal of Legal Studies, 41(2): 347-383.

[4] Agan, Amanda Y., and Sonja B. Starr. 2016. "Ban the Box, Criminal Records, and Statistical Discrimination: A Field Experiment." University of Michigan Law & Econ Research Paper No. 16-012.

[5] Aizer, Anna, and Joseph Doyle, Jr. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly-Assigned Judges." Quarterly Journal of Economics, 130(2): 759-803.

[6] Alesina, Alberto, and Eliana La Ferrara. 2014. "A Test of Racial Bias in Capital Sentencing." American Economic Review, 104(11): 3397-3433.

[7] Angrist, Joshua, Kathryn Graddy, and Guido W. Imbens. 2000. "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish." Review of Economic Studies, 67(3): 499-527.

[8] Angrist, Joshua, Guido W. Imbens, and Donald Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." Journal of the American Statistical Association, 91(434): 444-455.

[9] Antonovics, Kate, and Brian Knight. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." Review of Economics and Statistics, 91(1): 163-177.

[10] Anwar, Shamena, and Hanming Fang. 2006. "An Alternative Test of Racial bias in Motor Vehicle Searches: Theory and Evidence." American Economic Review, 96(1): 127-151.

[11] Anwar, Shamena, and Hanming Fang. 2012. "Testing for the Role of Prejudice in Emergency Departments Using Bounceback Rates." The B.E. Journal of Economic Analysis and Policy (Advances), 12(3): 1-47.

[12] Anwar, Shamena, and Hanming Fang. 2015. "Testing for Racial Prejudice in the Parole Board Release Process: Theory and Evidence." Journal of Legal Studies, 44(1): 1-37.

[13] Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." Quarterly Journal of Economics, 127(2): 1017-1055.

[14] Arrow, Kenneth J. 1973. The Theory of Discrimination. In *Discrimination in Labor Markets*, Ashenfelter, O. and A. Rees (Eds) Princeton University Press, 3-33.

[15] Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car." American Economic Review, 85(3): 304-321.

[16] Ayres, Ian, and Joel Waldfogel. 1994. "A Market Test for Race Discrimination in Bail Setting." Stanford Law Review, 46(5): 987-1047.

[17] Ayres, Ian. 2002. "Outcome Tests of Racial Disparities in Police Practices." Justice Research and Policy, 4(Special Issue): 131-142.

[18] Barsky, Robert, John Bound, Kerwin Kofi Charles, and Joseph P. Lupton. 2002. "Accounting for the Black-White Wealth Gap: A Nonparametric Approach." Journal of the American Statistical Association, 97(459): 663-673.

[19] Bayer, Patrick, Fernando Ferreira, and Stephen L. Ross. 2016. "The Vulnerability of Minority Homeowners in the Housing Boom and Bust." American Economic Journal: Economic Policy, 8(1): 1-27.

[20] Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.

[21] Bertrand, Marianne, and Esther Duflo. 2016. "Field Experiments on Discrimination." NBER Working Paper No. 22014.

[22] Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Loken, and Magne Mogstad. 2016. "Incarceration, Recidivism and Employment." Unpublished Working Paper.

[23] Bushway, Shawn D., and Jonah B. Gelbach. 2011. "Testing for Racial Discrimination in Bail Setting Using Nonparametric Estimation of a Parametric Model." Unpublished Working Paper.

[24] Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." American Economic Review, 96(5): 1890-1905.

[25] Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." Quarterly Journal of Economics, 131(4): 1753-1794.

[26] Brinch, Christian N., Magne Mogstad, and Matthew Wiswall. Forthcoming. "Beyond LATE with a Discrete Instrument." Journal of Political Economy.

[27] Chandra, Amitabh, and Douglas O. Staiger. 2010. "Identifying Provider Bias in Healthcare." NBER Working Paper No. 16382.

[28] Charles, Kerwin Kofi, and Jonathan Guryan. 2008. "Prejudice and Wages: an Empirical Assessment of Becker's The Economics of Discrimination." Journal of Political Economy, 116(5): 773-809.

[29] Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer. 2009. "Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering." Quarterly Journal of Economics, 124(3): 1133-1170.

[30] Demuth, Stephen. 2003. "Racial and Ethnic Differences in Pretrial Release Decisions and Outcomes: A Comparison of Hispanic, Black, and White Felony Arrestees." Criminology, 41(3): 873-908.

[31] Dahl, Gordon B., Andreas Ravndal Kostol, and Magne Mogstad. 2014. "Family Welfare Cultures." Quarterly Journal of Economics, 129(4): 1711-1752.

[32] Dobbie, Will, and Roland G. Fryer. 2013. "The Impact of Voluntary Youth Service on Future Outcomes: Evidence from Teach For America." The B.E. Journal of Economic Analysis and Policy, 15(3): 1031-1065.

[33] Dobbie, Will, Jacob Goldin, and Crystal Yang. 2016. "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." NBER Working Paper No. 22511.

[34] Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment." American Economic Journal: Applied Economics, 9(2): 1-22.

[35] Foote, Caleb. 1954. "Compelling Appearance in Court: Administration of Bail in Philadelphia." University of Pennsylvania Law Review, 102: 1031-1079.

[36] Frölich, Markus. 2007. "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates." Journal of Econometrics, 139(1): 35-75.

[37] Fryer, Roland G., and Matthew Jackson. 2008. "A Categorical Model of Cognition and Biased Decision-Making." The B.E. Journal of Theoretical Economics, 8(1): 1-42.

[38] Fryer, Roland G. 2011. "Racial Inequality in the 21st Century: The Declining Significance of Discrimination." Handbook of Labor Economics, 4(B): 855-971.

[39] Fryer, Roland G. 2016. "An Empirical Analysis of Racial Differences in Police Use of Force." NBER Working Paper No. 22399.

[40] Gelman, Andrew, Jeffrey Fagan, and Alex Kiss. 2012. "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.

[41] Glover, Dylan, Amanda Pallais, and William Pariente. Forthcoming. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." Quarterly Journal of Economics.

[42] Goldkamp, John S., and Michael R. Gottfredson. 1988. "Development of Bail/Pretrial Release Guidelines in Maricopa County Superior Court, Dade County Circuit Court and Boston Municipal Court. Washington, D.C." National Institute of Justice.

[43] Greenwald, Anthony G., T. Andrew Poehlman, Eric L. Uhlmann, and Mahzarin R. Banaji. 2009. "Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity." Journal of Personality and Social Psychology, 97(1): 17-41.

[44] Gupta, Arpit, Christopher Hansman, and Ethan Frenchman. 2016. "The Heavy Costs of High Bail: Evidence from Judge Randomization." Journal of Legal Studies, 45(2): 471-505.

[45] Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." Review of Economics and Statistics, 88(3): 389-432.

[46] Heckman, James J., and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." Econometrica, 73(3): 669-738.

[47] Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." Econometrica, 62(2): 467-475.

[48] Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human Decisions and Machine Predictions." NBER Working Paper No. 23180.

[49] Kling, Jeffrey. 2006. "Incarceration Length, Employment, and Earnings." American Economic Review, 96(3): 863-876.

[50] Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." Journal of Political Economy, 109(1): 203-232.

[51] Kowalski, Amanda E. 2016. "Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments." NBER Working Paper No. 22363.

[52] Leslie, Emily, and Nolan G. Pope. 2016. "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from NYC Arraignments." Unpublished Working Paper.

[53] McConnell, Allen R., and Jill M. Leibold. 2001. "Relations Among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes." Journal of Experimental Social Psychology, 37(5): 435-442.

[54] McIntyre, Frank, and Shima Baradaran. 2013. "Race, Prediction, and Pretrial Detention." Journal of Empirical Legal Studies, 10(4): 741-770.

[55] Mechoulan, Stéphane, and Nicolas Sahuguet. 2015. "Assessing Racial Disparities in Parole Release." Journal of Legal Studies, 44(1): 39-74.

[56] Mogstad, Magne, Andres Santos, and Alexander Torgovitsky. 2016. "Using Instrumental Variables for Inference about Policy-Relevant Treatment Effects." Unpublished Working Paper.

[57] Pager, Devah. 2003. "The Mark of a Criminal Record." American Journal of Sociology, 108(5): 937-975.

[58] Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. 2011. "Strike Three: Discrimination, Incentives, and Evaluation." American Economic Review 101(4): 1410-1435.

[59] Penner, Louis A., John F. Dovidio, Tessa V. West, Samuel L. Gaertner, Terrance L. Albrecht, Rhonda K. Dailey, and Tsveti Markova. 2010. "Aversive Racism and Medical Interactions

with Black Patients: A Field Study." Journal of Experimental Social Psychology, 46(2): 436-440.

[60] Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." American Economic Review, 62(4): 659-661.

[61] Phelps, Elizabeth A., Kevin J. O'Connor, William A. Cunningham, E. Sumie Funayama, J. Christopher Gatenby, John C. Gore, and Mahzarin R. Banaji. 2000. "Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation." Journal of Cognitive Neuroscience, 12(5): 729-738.

[62] Price, Joseph, and Justin Wolfers. 2010. "Racial Discrimination Among NBA Referees." Quarterly Journal of Economics, 125(4): 1859-1887.

[63] Rehavi, M. Marit, and Sonja B. Starr. 2014. "Racial Disparity in Federal Criminal Sentences." Journal of Political Economy, 122(6): 1320-1354.

[64] Rooth, Dan-Olof. 2010. "Automatic Associations and Discrimination in Hiring: Real World Evidence." Labour Economics, 17(3): 523-534.

[65] Rudman, Laurie A., and Matthew R. Lee. 2002. "Implicit and Explicit Consequences of Exposure to Violent and Misogynous Rap Music." Group Processes and Intergroup Relations, 5(2): 133-150.

[66] Shubik-Richards, Claire, and Don Stemen. 2010. "Philadelphia's Crowded, Costly Jails: The Search for Safe Solutions." Technical Report, Pew Charitable Trusts Philadelphia Research Initiative.

[67] Stevenson, Megan. 2016. "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes." Unpublished Working Paper.

[68] Van Laar, Colette, Shana Levin, Stacey Sinclair, and Jim Sidanius. 2005. "The Effect of University Roommate Contact on Ethnic Attitudes and Behavior." Journal of Experimental Social Psychology, 41(4): 329-345.

Table 1: Descriptive Statistics

	All Defendants		White		Black	
	Detained	Released	Detained	Released	Detained	Released
Panel A: Bail Type	(1)	(2)	(3)	(4)	(5)	(6)
Release on Recognizance	0.024	0.369	0.028	0.384	0.021	0.353
Non-Monetary Bail	0.052	0.227	0.056	0.209	0.049	0.247
Monetary Bail	0.925	0.404	0.917	0.407	0.930	0.400
Bail Amount (in thousands)	53.262	15.162	43.980	17.948	60.066	12.150
Panel B: Subsequent Bail Outcomes						
Bail Modification Petition	0.463	0.056	0.458	0.050	0.466	0.064
Released in 14 days	0.079	1.000	0.087	1.000	0.074	1.000
Released before Trial	0.376	1.000	0.376	1.000	0.376	1.000
Panel C: Defendant Characteristics						
Male	0.875	0.775	0.869	0.752	0.880	0.801
Age at Bail Decision	34.357	33.987	34.990	33.959	33.893	34.020
Prior Offense in Past Year	0.385	0.220	0.376	0.193	0.392	0.251
Panel D: Charge Characteristics						
Number of Offenses	3.349	2.402	2.956	2.420	3.638	2.381
Felony Offense	0.627	0.344	0.586	0.315	0.656	0.376
Misdemeanor Only	0.373	0.656	0.414	0.685	0.344	0.624
Any Drug Offense	0.287	0.407	0.278	0.377	0.295	0.443
Any DUI Offense	0.023	0.112	0.026	0.123	0.021	0.100
Any Violent Offense	0.264	0.200	0.223	0.214	0.294	0.184
Any Property Offense	0.351	0.193	0.358	0.189	0.346	0.197
Panel E: Outcomes						
Rearrest Prior to Disposition	0.190	0.202	0.177	0.173	0.200	0.236
Drug	0.068	0.102	0.060	0.083	0.073	0.124
Property	0.068	0.042	0.069	0.037	0.068	0.047
Violent	0.047	0.022	0.037	0.018	0.054	0.026
Observations	141,689	145,314	59,917	77,678	81,772	67,636

Note: This table reports descriptive statistics for the sample of defendants from Philadelphia and Miami-Dade counties. The sample consists of bail hearings that were quasi-randomly assigned from Philadelphia between 2010-2014 and from Miami-Dade between 2006-2014. We define pre-trial release based on whether a defendant was released within the first three days after the bail hearing. Information on race, gender, age, and criminal outcomes is derived from court records. See Appendix C for additional details on the sample and variable construction.

Table 2: Judge Leniency and Pre-Trial Release

	All Defendants		White		Black	
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-trial Release	0.572***	0.587***	0.533***	0.537***	0.612***	0.640***
	(0.035)	(0.033)	(0.046)	(0.043)	(0.044)	(0.041)
	[0.506]	[0.506]	[0.565]	[0.565]	[0.453]	[0.453]
Court x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Crime Controls	No	Yes	No	Yes	No	Yes
Observations	287,003	287,003	137,595	137,595	149,408	149,408

Note: This table reports first stage results. The regressions are estimated on the sample as described in the notes to Table 1. Judge leniency is estimated using data from other cases assigned to a bail judge in the same year following the procedure described in Section II.B. Columns 1, 3, and 5 begin by reporting results with only court-by-time fixed effects. Columns 2, 4, and 6 add the demographic and crime controls discussed in Section II.C. The sample mean of the dependent variable is reported in brackets. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 3: Test of Randomization

	All		White		Black	
	Pre-Trial Release	Judge Leniency	Pre-Trial Release	Judge Leniency	Pre-Trial Release	Judge Leniency
	(1)	(2)	(3)	(4)	(5)	(6)
Male	−0.12394***	0.00008	−0.11527***	0.00013	−0.12574***	0.00002
	(0.00255)	(0.00019)	(0.00322)	(0.00024)	(0.00378)	(0.00027)
Age at Bail Decision	−0.01361***	−0.00002	−0.01621***	−0.00007	−0.01234***	0.00004
	(0.00081)	(0.00007)	(0.00114)	(0.00010)	(0.00109)	(0.00009)
Prior Offense in Past Year	−0.17756***	0.00026*	−0.20655***	0.00001	−0.14502***	0.00042**
	(0.00200)	(0.00014)	(0.00292)	(0.00021)	(0.00261)	(0.00018)
Number of Offenses	−0.02685***	0.00004	−0.02320***	0.00005	−0.02712***	0.00003
	(0.00046)	(0.00003)	(0.00072)	(0.00004)	(0.00054)	(0.00003)
Felony Offense	−0.32579***	0.00002	−0.30763***	−0.00012	−0.33212***	0.00017
	(0.00277)	(0.00013)	(0.00380)	(0.00018)	(0.00365)	(0.00017)
Any Drug Offense	0.08375***	−0.00001	0.06379***	0.00004	0.10462***	−0.00004
	(0.00243)	(0.00022)	(0.00330)	(0.00027)	(0.00333)	(0.00029)
Any Violent Offense	−0.00052	−0.00013	0.04381***	−0.00048	−0.03541***	0.00023
	(0.00381)	(0.00022)	(0.00486)	(0.00031)	(0.00434)	(0.00026)
Any Property Offense	−0.02291***	−0.00026	−0.03273***	0.00002	−0.00801**	−0.00055**
	(0.00274)	(0.00022)	(0.00364)	(0.00030)	(0.00352)	(0.00028)
Joint F-Test	[0.00000]	[0.38774]	[0.00000]	[0.78955]	[0.00000]	[0.10715]
Observations	287,003	287,003	137,595	137,595	149,408	149,408

Note: This table reports reduced form results testing the random assignment of cases to bail judges. The regressions are estimated on the sample as described in the notes to Table 1. Judge leniency is estimated using data from other cases assigned to a bail judge in the same year following the procedure described in Section II.B. Columns 1, 3, and 5 report estimates from an OLS regression of pre-trial release on the variables listed and court-by-time fixed effects. Columns 2, 4, and 6 report estimates from an OLS regression of judge leniency on the variables listed and court-by-time fixed effects. The p-value reported at the bottom of the columns is for a F-test of the joint significance of the variables listed in the rows. Robust standard errors two-way clustered at the individual and the judge-by-shift level are reported in parentheses. ***=significant at 1 percent level, **=significant at 5 percent level, *=significant at 10 percent level.

Table 4: Pre-trial Release and Criminal Outcomes

	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.185***	0.005	0.180**
	(0.067)	(0.057)	(0.087)
	[0.174]	[0.216]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.077**	−0.020	0.097**
	(0.034)	(0.037)	(0.049)
	[0.073]	[0.096]	–
Property Crime	0.029	−0.001	0.030
	(0.045)	(0.033)	(0.054)
	[0.051]	[0.059]	–
Violent Crime	0.044	−0.038	0.082**
	(0.028)	(0.027)	(0.039)
	[0.026]	[0.042]	–
Observations	137,595	149,408	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 5: Results for High Risk and Low Risk Offenders

	High Risk			Low Risk		
	White	Black	Difference	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)
Rearrest Prior to Disposition	0.372***	0.006	0.365**	0.034	0.004	0.030
	(0.126)	(0.081)	(0.147)	(0.062)	(0.068)	(0.095)
	[0.271]	[0.282]	–	[0.106]	[0.126]	–
Panel B: Rearrest by Crime Type						
Drug Crime	0.165**	−0.017	0.182**	−0.006	−0.015	0.009
	(0.065)	(0.057)	(0.084)	(0.026)	(0.027)	(0.038)
	[0.130]	[0.149]	–	[0.032]	[0.024]	–
Property Crime	0.105	0.003	0.102	−0.031	−0.002	−0.029
	(0.091)	(0.050)	(0.103)	(0.027)	(0.031)	(0.042)
	[0.095]	[0.082]	–	[0.020]	[0.026]	–
Violent Crime	0.032	−0.028	0.060	0.056	−0.056	0.112*
	(0.040)	(0.034)	(0.052)	(0.038)	(0.043)	(0.059)
	[0.021]	[0.033]	–	[0.030]	[0.053]	–
Observations	56,879	86,622	–	80,716	62,786	–

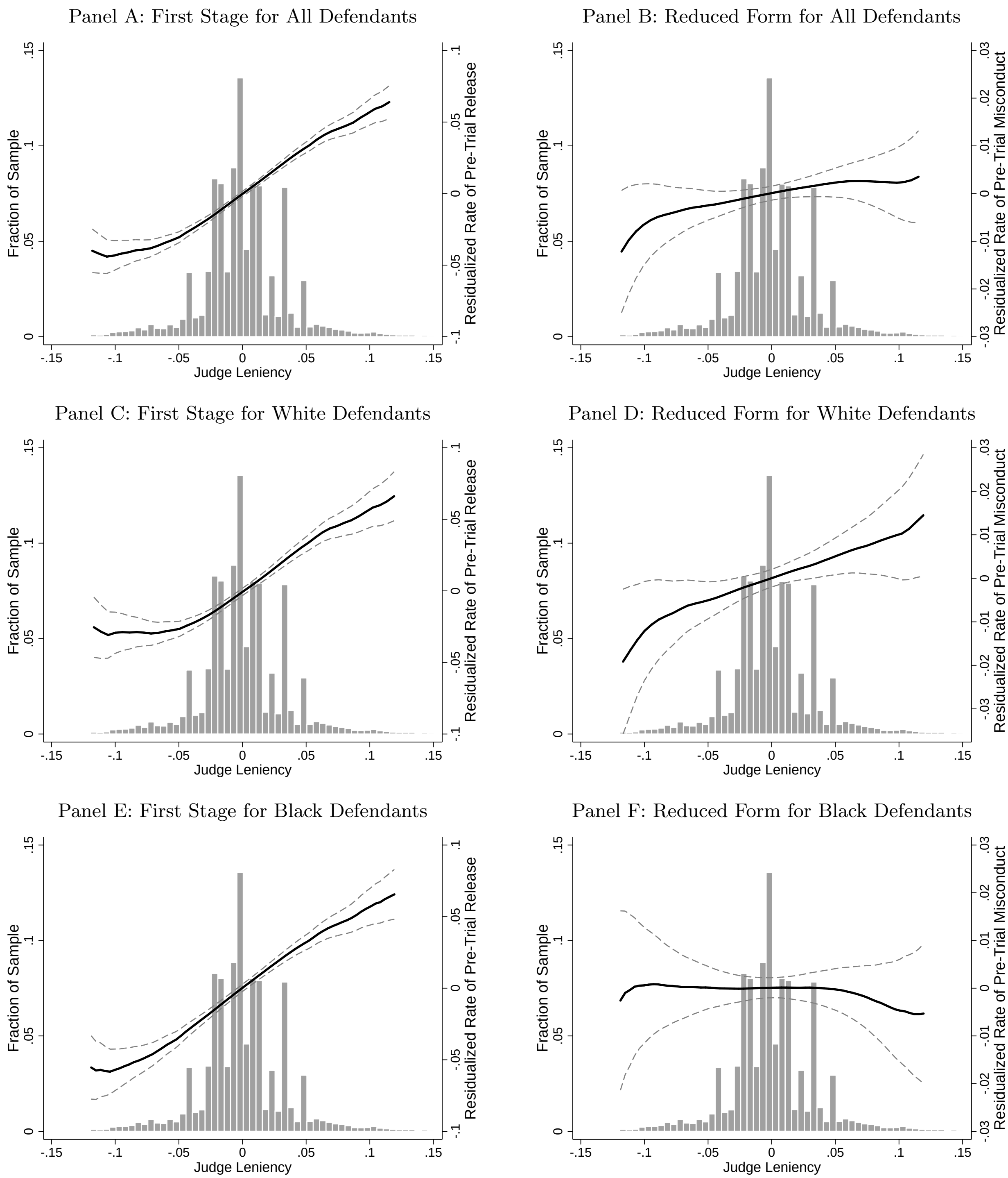
Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately for high-risk and low-risk defendants. Risk is computed by estimating a logistic regression of rearrest prior to case disposition conditional on release on the crime and demographic controls discussed in Section II. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 6: Pre-trial Release and Criminal Outcomes: The Role of Experience

	Judge Specialization			Judge Experience		
	Miami Non-Spec.	Philadelphia Specialist	Difference	Miami Low Exp.	Miami High Exp.	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)
Rearrest Prior to Disposition	0.291*	−0.014	0.305*	0.480**	0.193	0.288
	(0.161)	(0.094)	(0.182)	(0.231)	(0.160)	(0.272)
	[0.226]	[0.182]	–	[0.224]	[0.228]	–
Panel B: Rearrest by Crime Type						
Drug Crime	0.105	0.090	0.015	0.187*	0.070	0.117
	(0.065)	(0.060)	(0.085)	(0.108)	(0.080)	(0.130)
	[0.079]	[0.088]	–	[0.076]	[0.082]	–
Property Crime	0.070	−0.051	0.121	0.171	−0.028	0.198
	(0.087)	(0.045)	(0.097)	(0.142)	(0.102)	(0.169)
	[0.080]	[0.043]	–	[0.080]	[0.080]	–
Violent Crime	0.110*	0.041	0.069	0.102	0.124	−0.023
	(0.062)	(0.038)	(0.071)	(0.084)	(0.080)	(0.113)
	[0.049]	[0.027]	–	[0.050]	[0.047]	–
Observations	93,572	193,431	–	47,772	45,800	–

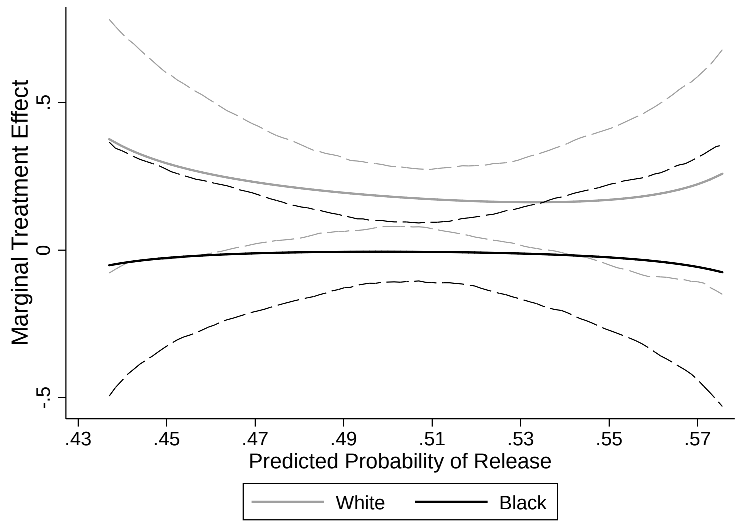
Note: This table reports two-stage least squares results of the estimate of racial bias separately by judge specialization and judge experience. Columns 1-3 report estimates for non-specialist bail judges in Miami-Dade and specialist bail judges in Philadelphia. Columns 4-6 report estimates for non-specialist bail judges in Miami with below and above median years of experience. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Figure 1: First Stage and Reduced Form



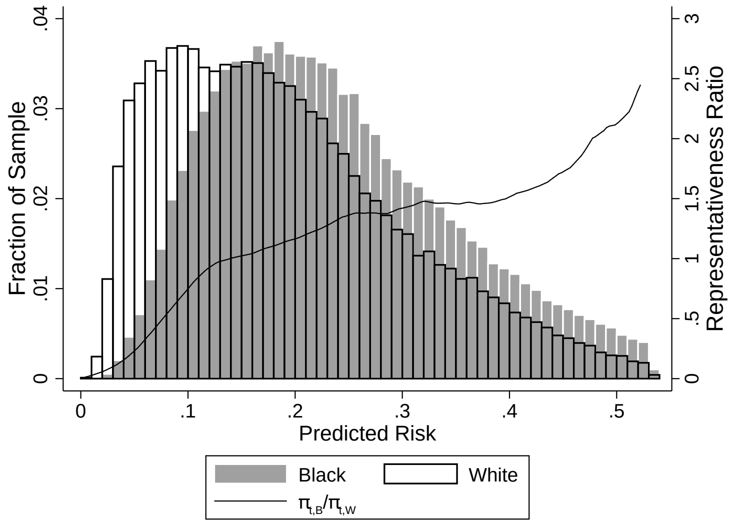
Note: These figures report the distribution of the judge leniency measure that is estimated using data from other cases assigned to a bail judge in the same year following the procedure described in Section II.B. Panels A-B pools all defendants. Panels C-D restricts the sample to white defendants. Panels E-F restricts the sample to black defendants. In the first figure in each Panel, the solid line is a local linear regression of pre-trial on judge leniency. In the second figure in each Panel, the solid line is a local linear regression of pre-trial misconduct on judge leniency. All regressions include the full set of court-by-time fixed effects.

Figure 2: Marginal Treatment Effects by Defendant Race



Note: This figure displays the estimated marginal treatment effects of release on rearrest separately for white and black defendants. The MTE is estimated in two steps. In the first step, we estimate probability of release (i.e. the propensity score) using only judge leniency in order to capture variation in treatment status due solely to the instrument (Doyle 2007). In the second step, we estimate the relationship between the propensity score and the residualized outcome (rearrest prior to disposition) using a local quadratic estimator (bandwidth = 0.065). The MTE is equal to the numerical derivative of the local quadratic estimator. Standard errors are computed using 500 bootstrap replications.

Figure 3: Predicted Risk Distribution by Race



Note: This figure reports the distribution of the risk of pre-trial misconduct separately by defendant race. Risk is computed by estimating a logit regression of the probability of rearrest prior to case disposition conditional on release on the crime and demographic controls discussed in Section II.C. The sample is described in the notes to Table 1.

Appendix A: Additional Results

Appendix Table A1: White-Hispanic Results

	White	Hispanic	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.150**	0.250**	−0.099
	(0.075)	(0.119)	(0.138)
	[0.196]	[0.191]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.103**	0.054	0.049
	(0.046)	(0.055)	(0.072)
	[0.073]	[0.083]	–
Property Crime	0.080*	0.001	0.079
	(0.045)	(0.080)	(0.091)
	[0.061]	[0.056]	–
Violent Crime	0.000	0.102**	−0.101*
	(0.033)	(0.049)	(0.058)
	[0.027]	[0.030]	–
Observations	35,468	78,554	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A2: Characteristics of Compliers by Race

	White			Black		
	P[X=x]	P[X=x|complier]	$\frac{P[X=x\|complier]}{P[X=x]}$	P[X=x]	P[X=x|complier]	$\frac{P[X=x\|complier]}{P[X=x]}$
Drug	0.313	0.388	1.241	0.348	0.335	0.963
	(0.001)	(0.026)	(0.082)	(0.001)	(0.022)	(0.064)
NonDrug	0.687	0.612	0.890	0.652	0.665	1.019
	(0.001)	(0.026)	(0.037)	(0.001)	(0.022)	(0.034)
Violent	0.192	0.006	0.029	0.204	0.016	0.080
	(0.001)	(0.019)	(0.101)	(0.001)	(0.019)	(0.095)
NonViolent	0.808	0.994	1.230	0.796	0.984	1.235
	(0.001)	(0.019)	(0.024)	(0.001)	(0.019)	(0.024)
Felony	0.433	0.249	0.574	0.529	0.375	0.707
	(0.001)	(0.024)	(0.056)	(0.001)	(0.023)	(0.042)
NonFelony	0.567	0.751	1.326	0.471	0.625	1.329
	(0.001)	(0.024)	(0.043)	(0.001)	(0.023)	(0.048)
Prior	0.273	0.333	1.223	0.328	0.384	1.168
	(0.001)	(0.021)	(0.076)	(0.001)	(0.019)	(0.058)
NonPrior	0.727	0.667	0.916	0.672	0.616	0.918
	(0.001)	(0.021)	(0.029)	(0.001)	(0.019)	(0.028)

Note: This table presents the sample distribution, complier distribution, and relative likelihood for different subgroups by race. Bootstrapped standard errors in parentheses are obtained using 500 replications.

Appendix Table A3: First Stage Results by Case Characteristics

	Crime Severity		Crime Type			Defendant Type	
	Misd.	Felony	Property	Drug	Violent	Prior	No Prior
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pre-trial Release	0.793***	0.383***	0.744***	0.615***	0.068	0.703***	0.533***
	(0.046)	(0.043)	(0.055)	(0.052)	(0.057)	(0.049)	(0.039)
	[0.643]	[0.360]	[0.371]	[0.598]	[0.475]	[0.369]	[0.369]
Court x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crime Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	148,269	138,734	95,012	71,113	56,791	86,552	200,451

Note: This table reports first stage subsample results. The regressions are estimated on the sample as described in the notes to Table 1. Judge leniency is estimated using data from other cases assigned to a bail judge in the same year following the procedure described in Section II.B. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A4: Non-Hispanic White-Black Results

	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.147*	0.006	0.141
	(0.076)	(0.057)	(0.091)
	[0.196]	[0.216]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.101**	−0.020	0.121**
	(0.047)	(0.037)	(0.057)
	[0.073]	[0.096]	–
Property Crime	0.080*	−0.001	0.080
	(0.045)	(0.033)	(0.053)
	[0.061]	[0.059]	–
Violent Crime	−0.003	−0.038	0.035
	(0.033)	(0.027)	(0.042)
	[0.027]	[0.042]	–
Observations	35,468	149,408	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately for blacks and non-Hispanic whites. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A5: Results for Drug Offenders vs. Other Crimes

	Drug			Other Crime		
	White	Black	Difference	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)
Rearrest Prior to Disposition	0.264**	−0.096	0.360**	0.156*	0.053	0.103
	(0.113)	(0.120)	(0.160)	(0.080)	(0.062)	(0.102)
	[0.221]	[0.268]	–	[0.153]	[0.189]	–
Panel B: Rearrest by Crime Type						
Drug Crime	0.258**	−0.093	0.351**	0.009	0.003	0.007
	(0.112)	(0.120)	(0.160)	(0.011)	(0.009)	(0.015)
	[0.220]	[0.267]	–	[0.006]	[0.005]	–
Property Crime	0.002	−0.002	0.004	0.038	0.008	0.031
	(0.012)	(0.010)	(0.016)	(0.061)	(0.046)	(0.074)
	[0.001]	[0.001]	–	[0.074]	[0.089]	–
Violent Crime	−0.003	−0.001	−0.001	0.064	−0.052	0.116**
	(0.007)	(0.006)	(0.009)	(0.040)	(0.038)	(0.056)
	[0.001]	[0.001]	–	[0.038]	[0.063]	–
Observations	43,057	51,955	–	94,538	97,453	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by crime type. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. Subgroup-specific means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A6: Results for Prior Offenders vs. No Prior Offenders

	Prior Offender			No Prior Offender		
	White	Black	Difference	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)
Rearrest Prior to Disposition	0.285***	−0.027	0.311**	0.118	0.032	0.086
	(0.097)	(0.081)	(0.126)	(0.083)	(0.073)	(0.110)
	[0.256]	[0.285]	–	[0.144]	[0.183]	–
Panel B: Rearrest by Crime Type						
Drug Crime	0.095*	−0.040	0.135*	0.073*	0.007	0.066
	(0.058)	(0.052)	(0.075)	(0.040)	(0.043)	(0.058)
	[0.109]	[0.130]	–	[0.059]	[0.080]	–
Property Crime	0.057	−0.013	0.070	0.005	0.009	−0.004
	(0.055)	(0.046)	(0.067)	(0.059)	(0.042)	(0.071)
	[0.079]	[0.080]	–	[0.041]	[0.048]	–
Violent Crime	0.034	−0.040	0.074	0.042	−0.043	0.084*
	(0.042)	(0.038)	(0.056)	(0.035)	(0.035)	(0.050)
	[0.032]	[0.049]	–	[0.024]	[0.038]	–
Observations	37,506	49,046	–	100,089	100,362	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by whether the defendant is a prior offender. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A7: Results for Felonies vs. Misdemeanors

	Felony			Misdemeanor		
	White	Black	Difference	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)
Rearrest Prior to Disposition	0.594**	−0.095	0.689**	0.034	0.074	−0.040
	(0.246)	(0.129)	(0.272)	(0.045)	(0.046)	(0.065)
	[0.217]	[0.249]	–	[0.141]	[0.180]	–
Panel B: Rearrest by Crime Type						
Drug Crime	0.275**	−0.009	0.284**	0.002	−0.025	0.027
	(0.113)	(0.088)	(0.136)	(0.023)	(0.025)	(0.034)
	[0.095]	[0.112]	–	[0.056]	[0.078]	–
Property Crime	0.134	−0.017	0.151	−0.008	0.016	−0.024
	(0.157)	(0.078)	(0.170)	(0.022)	(0.023)	(0.032)
	[0.086]	[0.085]	–	[0.025]	[0.030]	–
Violent Crime	0.118	−0.078	0.196*	0.017	−0.012	0.029
	(0.088)	(0.065)	(0.108)	(0.019)	(0.018)	(0.026)
	[0.040]	[0.058]	–	[0.016]	[0.023]	–
Observations	59,628	79,106	–	77,967	70,302	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by crime severity. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A8: Results for High-Income vs. Low-Income Offenders

	Above Median Income			Below Median Income		
	White	Black	Difference	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)
Rearrest Prior to Disposition	−0.070	0.017	−0.086	0.257**	0.011	0.246*
	(0.091)	(0.115)	(0.145)	(0.103)	(0.074)	(0.130)
	[0.154]	[0.211]	–	[0.181]	[0.213]	–
Panel B: Rearrest by Crime Type						
Drug Crime	0.055	−0.062	0.116	0.082	−0.002	0.084
	(0.048)	(0.064)	(0.079)	(0.050)	(0.045)	(0.068)
	[0.058]	[0.085]	–	[0.082]	[0.097]	–
Property Crime	−0.112*	0.043	−0.155*	0.093	−0.032	0.125
	(0.064)	(0.068)	(0.091)	(0.068)	(0.043)	(0.080)
	[0.047]	[0.066]	–	[0.049]	[0.056]	–
Violent Crime	−0.014	−0.047	0.033	0.009	−0.026	0.036
	(0.040)	(0.049)	(0.060)	(0.043)	(0.037)	(0.057)
	[0.023]	[0.040]	–	[0.027]	[0.041]	–
Observations	33,766	17,526	–	85,960	110,992	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately for high-income and low-income defendants. A defendant is classified as high-income if the zip code of residence has an average level of income greater than the median level of income in the city, while low-income is defined as coming from a zip code with an average level of income less than the median level of income in the city. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A9: Results Weighting by Case and Defendant Characteristics

	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.191***	0.032	0.159*
	(0.066)	(0.054)	(0.085)
	[0.174]	[0.216]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.082**	−0.019	0.101**
	(0.033)	(0.032)	(0.045)
	[0.073]	[0.096]	–
Property Crime	0.028	0.008	0.020
	(0.044)	(0.033)	(0.054)
	[0.051]	[0.059]	–
Violent Crime	0.045*	−0.021	0.066*
	(0.027)	(0.024)	(0.036)
	[0.026]	[0.042]	–
Observations	137,595	149,408	–

Note: This table reports weighted two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race. Results are re-weighted with the weights chosen to match the distribution of observable characteristics by race. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A10: Robustness to Clustering at the Judge Level

	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.185**	0.005	0.180*
	(0.082)	(0.059)	(0.098)
	[0.174]	[0.216]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.077*	−0.020	0.097*
	(0.045)	(0.036)	(0.051)
	[0.073]	[0.096]	–
Property Crime	0.029	−0.001	0.030
	(0.043)	(0.031)	(0.055)
	[0.051]	[0.059]	–
Violent Crime	0.044	−0.038	0.082**
	(0.031)	(0.026)	(0.038)
	[0.026]	[0.042]	–
Observations	137,595	149,408	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A11: Robustness to Race-Specific Leniency Measures

	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.181*	−0.032	0.213*
	(0.105)	(0.068)	(0.123)
	[0.196]	[0.216]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.129*	−0.050	0.179**
	(0.067)	(0.044)	(0.079)
	[0.073]	[0.096]	–
Property Crime	0.097	−0.006	0.103
	(0.062)	(0.037)	(0.070)
	[0.061]	[0.059]	–
Violent Crime	−0.038	−0.039	0.001
	(0.046)	(0.031)	(0.057)
	[0.027]	[0.042]	–
Observations	35,468	149,408	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race with judge leniency computed separately by race. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A12: Robustness to Bootstrap-Clustered Standard Errors

	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.185**	0.005	0.180*
	(0.089)	(0.063)	(0.108)
	[0.174]	[0.216]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.077*	−0.020	0.097*
	(0.039)	(0.045)	(0.057)
	[0.073]	[0.096]	–
Property Crime	0.029	−0.001	0.030
	(0.056)	(0.035)	(0.064)
	[0.051]	[0.059]	–
Violent Crime	0.044	−0.038	0.082*
	(0.035)	(0.031)	(0.047)
	[0.026]	[0.042]	–
Observations	137,595	149,408	–

Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Bootstrap standard errors clustered at the judge-by-shift level based on 500 simulations are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A13: OLS Results

	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.042***	0.037***	0.005
	(0.003)	(0.002)	(0.003)
	[0.174]	[0.216]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.023***	0.029***	−0.006***
	(0.001)	(0.002)	(0.002)
	[0.073]	[0.096]	–
Property Crime	0.003*	0.002	0.001
	(0.001)	(0.001)	(0.002)
	[0.051]	[0.059]	–
Violent Crime	−0.008***	−0.009***	0.002
	(0.001)	(0.001)	(0.001)
	[0.026]	[0.042]	–
Observations	137,595	149,408	–

Note: This table replicates the Knowles et al. (2001) test. The table reports OLS results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. The sample means of the dependent variables are reported in brackets. All specifications control for court-by-time fixed effects as well as the demographic and crime controls discussed in Section II.C. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A14: Pre-Trial Release and Pre-Trial Misconduct by Judge and Defendant Race

	Race of Judge	
	White	Black
Panel A: Release Rates	(1)	(2)
White	0.345	0.339
	(0.475)	(0.474)
Black	0.311	0.318
	(0.463)	(0.466)
Panel B: Pre-Trial Rearrest Rates		
White	0.175	0.174
	(0.380)	(0.379)
Black	0.253	0.273
	(0.435)	(0.446)

Note: This table presents average rates of pre-trial release and pre-trial misconduct conditional on release by defendant and judge race in Miami. The means are calculated using the Miami sample reported in Table 1. See text for additional details.

Appendix Table A15: p-values from Tests of Relative Racial Prejudice

	p-value
	(1)
Pre-Trial Release	0.364
Pre-Trial Rearrest	0.412

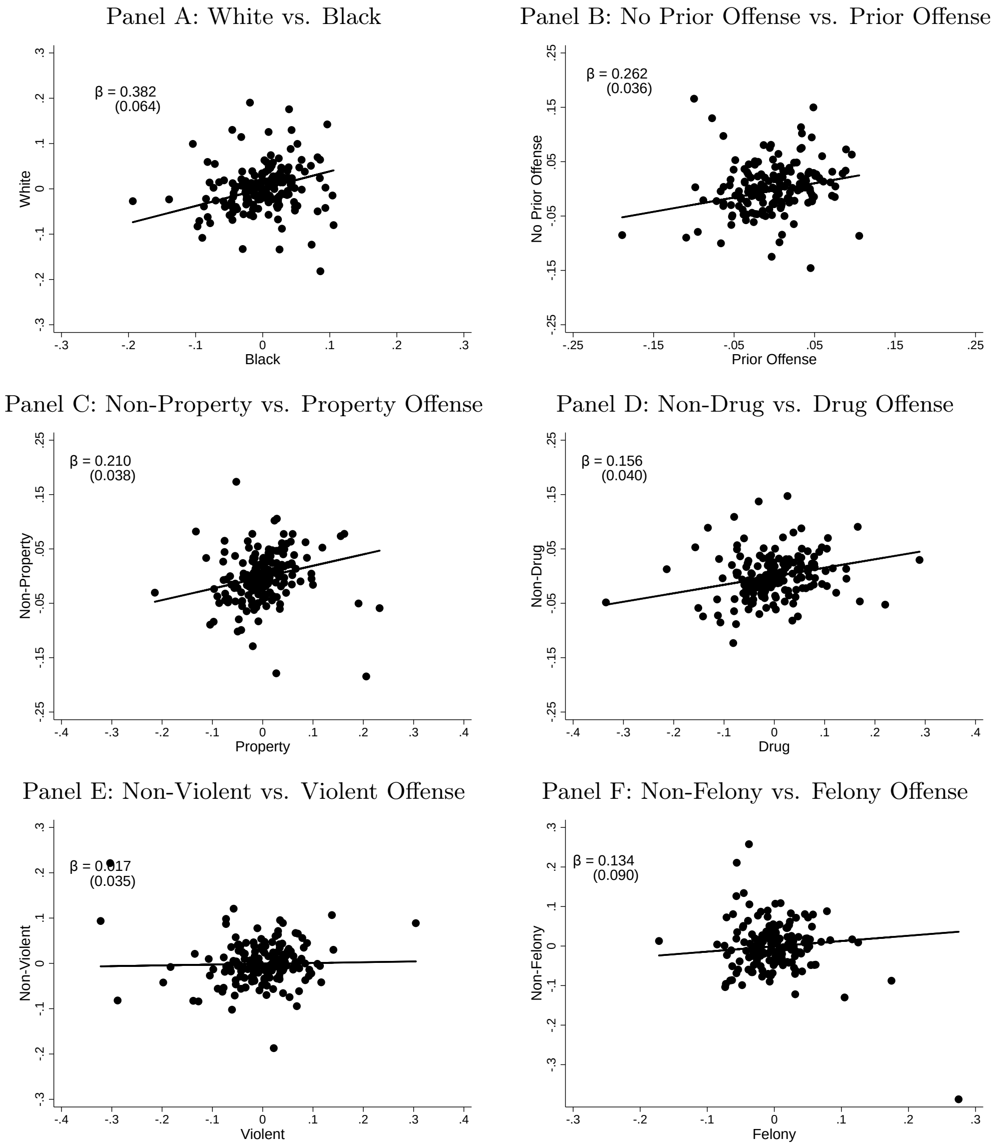
Note: This table replicates the Anwar and Fang (2006) test for pre-trial release rates and pre-trial misconduct rates. This table presents bootstrapped p-values testing for relative racial bias. The null hypothesis is rejected if white judges are more lenient on white defendants, and black judges are more lenient on black defendants.

Appendix Table A16: Representativeness Statistics

	$\mathbb{E}(x\|Black)/\mathbb{E}(x\|White)$
Panel A: Defendant Characteristics	(1)
Male	1.053
Age at Bail Decision	0.987
Prior Offense in Past Year	1.203
Panel B: Charge Characteristics	
Number of Offenses	1.157
Felony Offense	1.221
Misdemeanor Only	0.831
Any Drug Offense	1.085
Any DUI Offense	0.691
Any Violent Offense	1.118
Any Property Offense	1.059
Panel C: Outcomes	
Rearrest Prior to Disposition	1.239
Drug Crime	1.318
Property Crime	1.157
Violent Crime	1.588
Observations	287,003

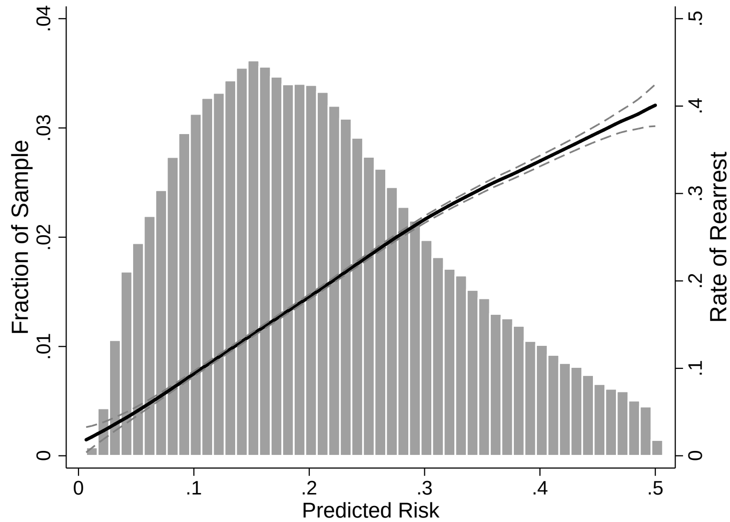
Note: This table reports the mean of the variable listed in the row given the defendant is black, divided by the mean of the variable listed in the row given the defendant is white. The sample is described in the notes to Table 1.

Appendix Figure A1: Judge Leniency by Defendant and Case Characteristics



Note: These figures show the correlation between our residualized measure of judge leniency for different groups of defendants over all available years of data. We also plot the linear best fit line estimated using OLS.

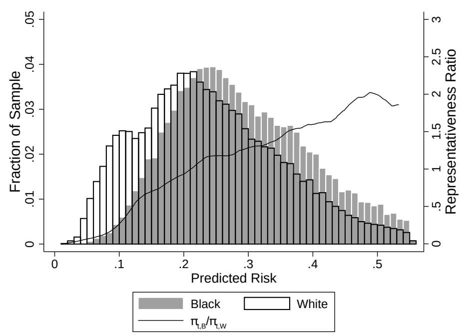
Appendix Figure A2: Relationship between Predicted Risk and True Risk



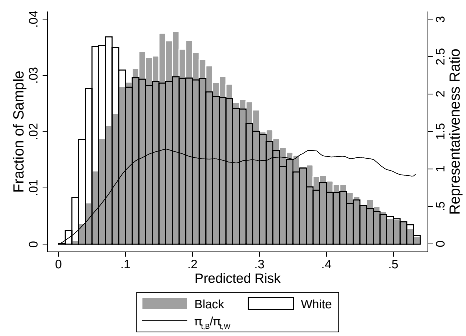
Note: This figure reports the distribution of the risk of pre-trial misconduct. Risk is computed by estimating a logistic regression of rearrest prior to case disposition conditional on release on the crime and demographic controls discussed in Section II.C. The solid line is a local linear regression of true risk on predicted risk. The sample is described in the notes to Table 1.

Appendix Figure A3: Predicted Crime-Specific Risk Distributions by Race

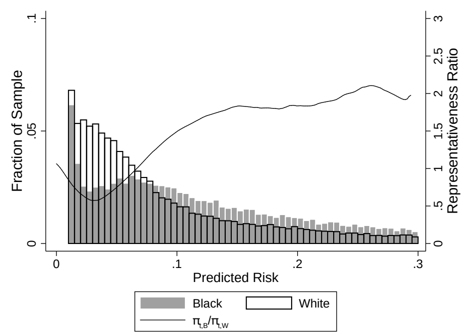
Panel A: Rearrest for Drug Crime



Panel B: Rearrest for Property Crime

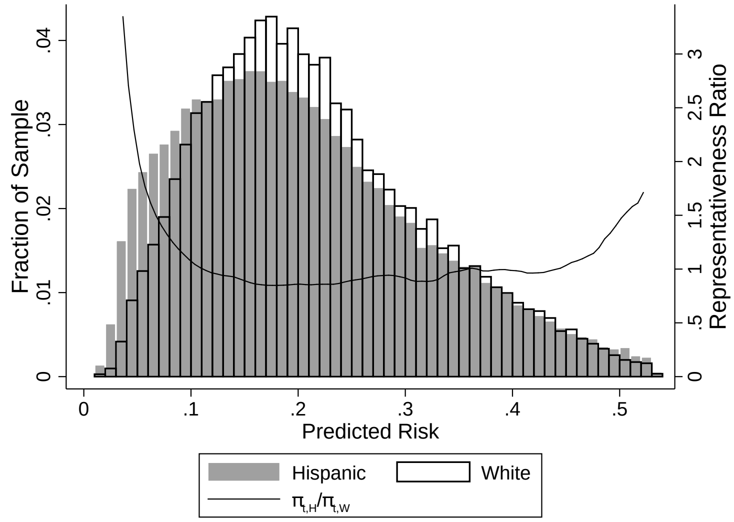


Panel C: Rearrest for Violent Crime



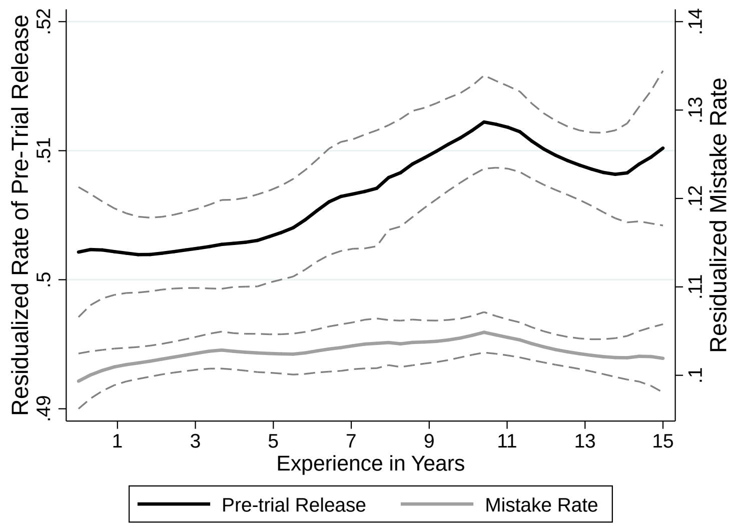
Note: This figure reports the distribution of the risk of crime-specific pre-trial misconduct separately by defendant race. Risk is computed by estimating a logit regression of the probability of rearrest prior to case disposition conditional on release on the crime and demographic controls discussed in Section II.C. Panel C omits defendants with a less than 1 percentage probability of committing a violent crime (73 percent of black defendants and 76 percent of white defendants). The sample is described in the notes to Table 1.

Appendix Figure A4: Predicted Risk Distribution by Hispanic versus White



Note: This figure reports the distribution of the risk of pre-trial misconduct separately by Hispanic and white defendants. Risk is computed by estimating a logit regression of the probability of rearrest prior to case disposition conditional on release on the crime and demographic controls discussed in Section II.C. The sample is described in the notes to Table 1.

Appendix Figure A5: Probability of Release and Pre-trial Misconduct with Experience



Note: This figure plots the relationship between judicial experience and both the residualized rate of pre-trial release and the residualized rate of rearrest prior to case disposition conditional on release (i.e. the mistake rate). Pre-trial release and rearrest prior to case disposition are both residualized using the full set of court-by-time fixed effects to control for any systematic differences in the types of defendants seen by judges.

Appendix B: Proofs of Propositions

A. Proof of Proposition 3

Proposition 3 states that as our judge leniency measure Z_i becomes continuously distributed, each race-specific IV estimate, α_r^{IV}, converges to a weighted average of treatment effects for defendants at the margin of release.

To see why this proposition holds, first define the treatment effect for a defendant at the margin of release at z_j as:

$$\alpha_r^j = \alpha_r(z = z_j) = \lim_{dz \to 0} \mathbb{E}[Y_i(1) - Y_i(0) | R_i(z) - R_i(z - dz) = 1] \tag{B.1}$$

With a continuous instrument Z_i, Angrist, Graddy, and Imbens (2000) show that the IV estimate, α_r^{IV}, converges to:

$$\alpha_r = \int \lambda_r(z)\alpha_r(z)dz \tag{B.2}$$

where the weights, $\lambda_r(z)$ are given by:

$$\lambda_r(z) = \frac{\frac{\partial R_r}{\partial z}(z) \cdot \int_z^{\bar{z}} (y - \mathbb{E}[z]) \cdot f_z^r(y)dy}{\int_{\underline{z}}^{\bar{z}} \frac{\partial R_r}{\partial z}(v) \cdot \int_v^{\bar{z}} (y - \mathbb{E}[z]) \cdot f_z^r(y)dydv} \tag{B.3}$$

where $\frac{\partial R_r}{\partial z}$ is the derivative of the probability of release with respect to leniency and f_z^r is the probability density function of leniency. If $\frac{\partial R_r}{\partial z} \geq 0$ for all z, then the weights are nonnegative. Therefore, as Z_i becomes continuously distributed, our race-specific IV estimate will return a weighted average of treatment effects of defendants on the margin of release. $\square$

B. Proof of Proposition 4

Proposition 4 states that our IV estimator D^{IV} provides a consistent estimate of racial bias D^* if (1) λ_r^j is constant by race and (2) Z_i is continuous. The first condition – that λ_r^j is constant by race – holds if and only if the proportion of compliers shifted by moving across judges is constant by race for each z_{j-1}, z_j pair:

$$\frac{Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)}{Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B)} = c \tag{B.4}$$

where c is some constant.

To show why this proposition holds, we proceed in two steps. First, we show that our IV estimator is consistent if the IV weights by race are constant and Z_i is continuous. Second, we show that the assumption of constant weights by race holds if and only if Equation (B.4) is true.

We begin by showing that if λ_r^j is constant by race, then as Z_i becomes continuously distributed,

D^{IV} provides a consistent estimate of D^*. D^{IV} is given by:

$$D^{IV} = \alpha_W^{IV} - \alpha_B^{IV} = \sum_{j=1}^{J} \lambda_W^j \alpha_W^{j,j-1} - \sum_{j=1}^{J} \lambda_B^j \alpha_B^{j,j-1} \tag{B.5}$$

If $\lambda_r^j = \lambda^j$, then:

$$D^{IV} = \sum_{j=1}^{J} \lambda^j \left(\alpha_W^{j,j-1} - \alpha_B^{j,j-1} \right) \tag{B.6}$$

Following Proposition 3, as Z_i becomes continuously distributed, we can rewrite D^{IV} as:

$$D^{IV} = \int \lambda(z) \left(\alpha_W(z) - \alpha_B(z) \right) dz = D^* \tag{B.7}$$

Therefore, in the limit, D^{IV} estimates a weighted average of differences in treatment effects for defendants at the margin of release, and therefore provides a consistent estimate of true racial bias.

Next, we show that the weights λ_r^j are constant by race if and only if:

$$\frac{Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)}{Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B)} = c \tag{B.8}$$

where c is some constant.

To begin, Imbens and Angrist (1994) show that the weights in the IV estimator with a multi-valued instrument are given by the formula:

$$\lambda_r^j = \frac{(P(Released|z_j, r) - P(Released|z_{j-1}, r)) \cdot \sum_{l=j}^{J} \pi_r^l (g(z_l) - \mathbb{E}[g(Z)])}{\sum_{m=1}^{J} (P(Released|z_m, r) - P(Released|z_{m-1}, r)) \cdot \sum_{l=m}^{J} \pi_r^l (g(z_l) - \mathbb{E}[g(Z)])} \tag{B.9}$$

where $g(Z)$ is the instrumental variable and π_r^j is the probability of being assigned judge j for defendant race r. In our setting, we use judge leniency as our instrument, and so $g(Z) = Z$.

To simplify notation, let $\phi_r^j = \sum_{l=j}^{J} \pi_r^l (z_l - \mathbb{E}[Z])$. Under the exclusion restriction (Assumption 2), the probability of being assigned to any particular judge should not differ by defendant race. Therefore, π_r^l and $\mathbb{E}[Z]$ are independent of race. Going forward, we we drop the r subscript on ϕ_r^j as this term does not depend on race.

First, we prove that if Equation (B.4) holds, then the IV weights are the same by race:

$$\begin{aligned}
\lambda_W^j =& \frac{(Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W))\phi^j}{\sum_{m=1}^{J} Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)\phi^m} \\
=& \frac{c(Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B))\phi^j)}{\sum_{m=1}^{J} c(Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B))\phi^m} \\
=& \frac{(Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B))\phi^j}{\sum_{m=1}^{J} (Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B))\phi^m} = \lambda_B^j
\end{aligned}$$

where the first equality follows from Imbens and Angrist (1994) and the second equality follows by substituting in Equation (B.4).

Next, we prove that if the IV weights are constant by race, then Equation (B.4) holds. To do so, we prove the contrapositive statement. Suppose Equation (B.4) does not hold, so that there exists z_j and z_k such that:

$$\frac{Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)}{Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B)} = c_1 \tag{B.10}$$

$$\frac{Pr(Released|z_k, r = W) - Pr(Released|z_{k-1}, r = W)}{Pr(Released|z_k, r = B) - Pr(Released|z_{k-1}, r = B)} = c_2 \tag{B.11}$$

where $c_1 \neq c_2$. To simplify notation, denote the denominator of λ_W^j as D_W and the denominator for λ_B^j as D_B, which is constant for all j. Then:

$$\frac{\lambda_W^j}{\lambda_B^j} = \frac{1}{c_1}\frac{D_B}{D_W} \tag{B.12}$$

while

$$\frac{\lambda_W^k}{\lambda_B^k} = \frac{1}{c_2}\frac{D_B}{D_W} \tag{B.13}$$

where Equation (B.12) and Equation (B.13) follow by substituting Equation (B.10) and Equation (B.11) into the formula for the IV weights. If $c_1 \neq c_2$ then $\frac{\lambda_W^j}{\lambda_B^j} \neq \frac{\lambda_W^k}{\lambda_B^k}$. Therefore, either $\lambda_W^j \neq \lambda_B^j$ or $\lambda_W^k \neq \lambda_B^k$, implying the weights cannot be equal by race. □

Sufficiency of Linear First Stage: We now show that a linear first stage is sufficient for the weights in our IV estimator to be the same by race. Let the first stage relationship between pre-trial release and Z_i is given by a linear probability model of the form:

$$Released_i = \gamma_0 + \gamma_W Z_i \cdot White_i + \gamma_B Z_i \cdot Black_i + \pi \mathbf{X}_i + \mathbf{U}_i + \varepsilon_i \tag{B.14}$$

such that the proportion of compliers shifted by moving from judge $j-1$ to j is constant by race.

If the first stage is linear for each race, then:

$$Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r) = \gamma_r(z_j - z_{j-1}) \tag{B.15}$$

Then, it is straightforward to show:

$$\frac{Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)}{Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B)} = \frac{\gamma_W(z_j - z_{j-1})}{\gamma_B(z_j - z_{j-1})} = \frac{\gamma_W}{\gamma_B} \tag{B.16}$$

where $\frac{\gamma_W}{\gamma_B}$ is constant for all j. □

C. Additional Functional Form Assumptions for Consistency

Our empirical strategy relies on the assumption that Z_i is continuous for our IV estimator, D^{IV}, to provide a consistent estimate of racial bias. However, D^{IV} will not in general be a consistent estimate of racial bias with a discrete instrument. In this section, we discuss additional assumptions necessary for D^{IV} to be consistent in this case.

To do so, we link our theoretical model to the marginal treatment effect (MTE) literature, which will allow us to illustrate the shape restrictions on the MTE necessary to interpret our IV estimator as an estimate of racial bias.

To begin, note that we characterize the pre-trial release decision as:

$$R_i(z_j) = \mathbb{1}\{\mathbb{E}[\alpha_i|r_i] \leq t_r^j\} \tag{B.17}$$

Let $F_{\alpha,r}$ be the cumulative density function of $\mathbb{E}[\alpha_i|r_i]$, which we assume is continuous on the interval $[0, 1]$. The model of release above has the same empirical content as:

$$R_i(z_j) = \mathbb{1}\{F_{\alpha,r}(\mathbb{E}[\alpha_i|r_i]) \leq F_{\alpha,r}(t_r^j)\} = \mathbb{1}\{U_{i,r} \leq Pr(Released|z_j, r)\} \tag{B.18}$$

where $U_{i,r} \in [0, 1]$ by construction. The second equality follows because $F_{\alpha,r}(t_r^j)$ is the probability of release given a judge with preferences t_r^j, which is simply the probability of release given assignment to judge j. By writing the release decision as a latent-index model of the form in Equation (B.18), we can map our model to the framework of Heckman and Vytlacil (2005). Following Heckman and Vytlacil (2005), the race-specific MTE is defined as:

$$MTE_r(u) = \mathbb{E}[Y_i(1) - Y_i(0)|r_i = r, U_{i,r} = u] \tag{B.19}$$

Recall that under our model, we want to estimate the treatment effect of defendants at the margin of release:

$$\alpha_r^j = \mathbb{E}[Y_i(1) - Y_i(0)|r_i = r, \mathbb{E}[\alpha_i|r_i] = t_r^j] \tag{B.20}$$

Given Equation (B.18) and Equation (B.19), the treatment effect for defendants at the margin of release is equivalent to:

$$\mathbb{E}[Y_i(1) - Y_i(0)|r_i = r, U_{i,r} = Pr(Released|z_j, r)] = MTE_r(Pr(Released|z_j, r)) \tag{B.21}$$

We can now discuss the structural assumptions we must place on the MTE function so that our IV estimator is a consistent estimate of racial bias. To simplify notation, let $Pr(Released|z_j, r) = p_r^j$, which we refer to as the propensity score. Below, we show that if the MTE can be well approximated by linear splines, with knots at points in the support of the propensity score, then the IV estimator is a consistent estimate of a weighted average of true racial bias. Specifically, if:

$$MTE_r(u) \approx \sum_{j=1}^{J} \mathbb{1}\{u \in [p_r^{j-1}, p_r^j]\}[\theta_r^{1,j} + \theta_r^{2,j} u] \tag{B.22}$$

then we may interpret D^{IV} as a consistent estimate of racial bias.

The strength of this assumption depends on the distribution of leniency. If Z_i becomes continuous, then the propensity score also becomes continuously distributed, implying this formulation imposes no structure on the MTE, consistent with Proposition 3. With only two judges, it imposes that the MTE is linear for compliers. The more points in the distribution of leniency, the more we can accommodate non-linearities into the MTE.

Our restrictions on the MTE are similar to Brinch et al. (forthcoming) who estimate the MTE in settings with a discrete instrument. With a binary instrument, they impose that the MTE is linear (exactly the same as our restriction). With k points in the distribution of leniency, they impose that the MTE is a polynomial of order no higher than $k-1$. We do not utilize our restrictions to estimate the MTE itself, but rather to interpret the our IV estimator for racial bias.

To show why Equation (B.22) implies that D^{IV} is a consistent estimate of racial bias, first consider a case with two judges with $p_r^0 < p_r^1$. As shown in Heckman and Vytlacil (2005), the LATE is related to the MTE by the following formula:

$$LATE_r^{0,1} = \frac{1}{p_r^1 - p_r^0} \int_{p_r^0}^{p_r^1} MTE_r(u) du \tag{B.23}$$

The relevant estimates for racial discrimination are $MTE_r(p_r^0)$ and $MTE_r(p_r^1)$ (i.e. the treatment effects for defendants at the margin of release). In contrast, the LATE is an average of MTE_r between these points. To relate the average to the endpoints, we assume $MTE_r(u)$ is linear over the interval $[p_r^0, p_r^1]$, so that:

$$\begin{aligned}
\frac{1}{p_r^1 - p_r^0} \int_{p_r^0}^{p_r^1} MTE_r(u) du &= \frac{1}{p_r^1 - p_r^0} \int_{p_r^0}^{p_r^1} [\theta_r^{1,1} + \theta_r^{2,1} u] du \\
&= \frac{1}{p_r^1 - p_r^0} \left[\theta_r^{1,1} u + \theta_r^{2,1} \frac{u^2}{2} \right]_{p_r^0}^{p_r^1} \\
&= \frac{\theta_r^{1,1} + \theta_r^{2,1} p_r^0}{2} + \frac{\theta_r^{1,1} + \theta_r^{2,1} p_r^1}{2} \\
&= \frac{MTE_r(p_r^0) + MTE_r(p_r^1)}{2} \\
&= \frac{t_r^0 + t_r^1}{2}
\end{aligned}$$

where the first line follows from substituting in Equation (B.22) and the last line follows from the fact $MTE_r(p_r^j) = \alpha_r^j = t_r^j$. By assuming the MTE is linear between these points, we may write the

LATE as a simple average of t_r^0 and t_r^1. This is true for all $j-1$, j pairs. Therefore:

$$\begin{aligned} D^{IV} &= \sum_{j=1}^{J} \lambda^j (\alpha_W^{j,j-1} - \alpha_B^{j,j-1}) \\ &= \frac{\lambda^1}{2}(t_W^0 - t_B^0) + \sum_{j=1}^{J-1} \frac{\lambda^j + \lambda^{j+1}}{2}(t_W^j - t_B^j) + \frac{\lambda^J}{2}(t_W^J - t_B^J) \\ &= \sum_{j=0}^{J} \tilde{\lambda}^j (t_W^j - t_B^j) = D^{**} \end{aligned}$$

where

$$\tilde{\lambda}^j = \begin{cases} \frac{\lambda^1}{2} & j = 0 \\ \frac{\lambda^j + \lambda^{j+1}}{2} & j \in [1, J-1] \\ \frac{\lambda^J}{2} & j = J \end{cases} \tag{B.24}$$

and $\sum \tilde{\lambda}^j = 1$. Note that the weights $\tilde{\lambda}^j$ differ slightly from the IV weights λ^j. We define this new weighted average of racial bias as D^{**}. Therefore, we have shown that under the functional form assumption in Equation (B.22), D^{IV} is a consistent estimate of D^{**}, which is a weighted average of true racial bias.

D. Proof of Proposition 5

Proposition 5 states that if Assumptions 1-3 are satisfied and the first stage relationship is linear, the maximum bias of our IV estimator D^{IV} from the true level of racial bias D^* is given by $\max_j(\lambda^j)(\alpha^{max} - \alpha^{min})$, where α^{max} is the largest treatment effect among compliers, α^{min} is the smallest treatment effect among compliers, and λ^j is given by:

$$\lambda^j = \frac{(z_j - z_{j-1}) \cdot \sum_{l=j}^{J} \pi^l (z_l - \mathbb{E}[Z])}{\sum_{m=1}^{J} (z_j - z_{j-1}) \cdot \sum_{l=m}^{J} \pi^l (z_l - \mathbb{E}[Z])} \tag{B.25}$$

where π^j is the probability of being assigned to judge j.

To prove that this proposition holds, we proceed in five steps. First, we derive an upper bound of D^{IV} by replacing $\alpha_W^{j,j-1}$ with its maximum possible value for every j and replacing $\alpha_B^{j,j-1}$ with its minimum possible value for every j. Second, we derive a lower bound of D^{IV} by replacing $\alpha_W^{j,j-1}$ with its minimum possible value for every j and replacing $\alpha_B^{j,j-1}$ with its maximum value for every j. Third, we show that the upper bound and lower bound of D^{IV} both converge to D^* as Z_i becomes continuously distributed. Fourth, we develop a formula for the maximum potential bias with a discrete instrument using the derived upper and lower bounds, and provide intuition for how we derive this estimation bias. Fifth, we show how to empirically estimate the maximum potential bias in the case of a discrete instrument.

To begin, note that under the assumption of equal weights (Proposition 4), our IV estimator for racial bias is given by:

$$D^{IV} = \sum_{j=1}^{J} \lambda^j \left(\alpha_W^{j,j-1} - \alpha_B^{j,j-1} \right)$$

Recall that under our theory model, compliers for judge j and $j-1$ are individuals such that $t_r^{j-1}(\mathbf{V}_i) < \mathbb{E}[\alpha_i | r_i] \leq t_r^j(\mathbf{V}_i)$. For illustrative purposes, we drop conditioning on $\mathbf{V}_i$. Under this definition of compliers, we know that:

$$\alpha_r^{j,j-1} \in (t_r^{j-1}, t_r^j] \tag{B.26}$$

Given Equation (B.26), we can derive an upper bound of D^{IV}:

$$\begin{aligned} D^{IV} &= \sum_{j=1}^{J} \lambda^j \left(\alpha_W^{j,j-1} - \alpha_B^{j,j-1} \right) \\ &< \sum_{j=1}^{J} \lambda^j \left(t_W^j - t_B^{j-1} \right) \\ &= \sum_{j=1}^{J} \lambda^j \left(t_W^j - (t_B^j - \Delta_B^j) \right) \\ &= \underbrace{\sum_{j=1}^{J} \lambda^j \left(t_W^j - t_B^j \right)}_{D^* = \text{true racial bias}} + \underbrace{\sum_{j=1}^{J} \lambda^j \Delta_B^j}_{\text{infra-marginality bias}} \end{aligned}$$

where $\Delta_B^j = t_B^j - t_B^{j-1}$. This bound follows from replacing $\alpha_W^{j,j-1} = t_W^j$ and replacing $\alpha_B^{j,j-1} = t_B^{j-1}$. Specifically, we replace $\alpha_W^{j,j-1}$ with its maximum possible value for every j and replace $\alpha_B^{j,j-1}$ with its minimum possible value for every j. As can be seen, this upper bound of D^{IV} is comprised of two components: (1) the true level of racial bias, D^* and (2) an "infra-marginality" bias.

Similarly, given Equation (B.26), we can also provide a lower bound of D^{IV}:

$$
\begin{aligned}
D^{IV} &= \sum_{j=1}^{J} \lambda^j \left(\alpha_W^{j,j-1} - \alpha_B^{j,j-1} \right) \\
&> \sum_{j=1}^{J} \lambda^j \left(t_W^{j-1} - t_B^{j} \right) \\
&= \sum_{j=1}^{J} \lambda^j \left(t_W^{j} - \Delta_W^{j} - t_B^{j} \right) \\
&= \underbrace{\sum_{j=1}^{J} \lambda^j \left(t_W^{j} - t_B^{j} \right)}_{D^* = \text{true racial bias}} \quad - \quad \underbrace{\sum_{j=1}^{J} \lambda^j \Delta_W^{j}}_{\text{infra-marginality bias}}
\end{aligned}
$$

where this lower bound comes from substituting $\alpha_W^{j,j-1} = t_W^{j-1}$ and $\alpha_B^{j,j-1} = t_B^{j}$. Specifically, we replace $\alpha_W^{j,j-1}$ with its minimum possible value for every j and replace $\alpha_B^{j,j-1}$ with its maximum possible value for every j. As can be seen, this lower bound of D^{IV} is again comprised of two components: (1) the true level of racial bias, D^* and (2) an "infra-marginality" bias.

Therefore, D^{IV} is bounded above and below by:

$$
\sum_{j=1}^{J} \lambda^j \left(t_W^{j} - t_B^{j} \right) - \sum_{j=1}^{J} \lambda^j \Delta_W^{j} < D^{IV} < \sum_{j=1}^{J} \lambda^j \left(t_W^{j} - t_B^{j} \right) + \sum_{j=1}^{J} \lambda^j \Delta_B^{j} \tag{B.27}
$$

Alternatively, expressed with respect to the true level of racial bias D^* , the difference between D^{IV} and D^* is bounded by:

$$
-\sum_{j=1}^{J} \lambda^j \Delta_W^{j} < D^{IV} - D^* < \sum_{j=1}^{J} \lambda^j \Delta_B^{j} \tag{B.28}
$$

Given that λ^j are non-negative weights which sum to one, $\sum_{j=1}^{J} \lambda^j \Delta_r^j \leq \max_j \Delta_r^j$ (i.e. the average is less than the maximum). If Z_i becomes continuous, then $\Delta_r^j \to 0$ for all j, and so infra-marginality bias shrinks to zero. Intuitively, at the limit, every complier is at the margin, and so there is no infra-marginality bias. As a result, D^{IV} converges to D^* as Z_i becomes continuous.

We now bound the infra-marginality bias when leniency is not continuous, i.e. when our instrument is discrete. Note that $\Delta_r^j = t_r^j - t_r^{j-1} = \alpha_r^j - \alpha_r^{j-1}$, where $\alpha_r^j - \alpha_r^{j-1} > 0$. Without loss of generality, assume $\lambda^1 \geq \lambda^2$. Then:

$$
\lambda^1 \left(\alpha_r^1 - \alpha_r^0 \right) + \lambda^2 \left(\alpha_r^2 - \alpha_r^1 \right) \leq \lambda^1 \left(\alpha_r^2 - \alpha_r^0 \right) \tag{B.29}
$$

We can continue in this manner to bound the infra-marginality bias. Without loss of generality, assume $\lambda^1 \geq \lambda^j$ for all j. Then

$$\begin{aligned}
\sum_{j=1}^{J} \lambda^j \left(\alpha_r^j - \alpha_r^{j-1}\right) &\leq \lambda^1(\alpha_r^2 - \alpha_r^0) + \sum_{j=3}^{J} \lambda^j(\alpha_r^j - \alpha_r^{j-1}) \\
&\leq \lambda^1(\alpha_r^3 - \alpha_r^0) + \sum_{j=4}^{J} \lambda^j(\alpha_r^j - \alpha_r^{j-1}) \\
&\ldots \\
&\leq \lambda^1(\alpha_r^J - \alpha_r^0)
\end{aligned}$$

Note that $\alpha_r^0 = \alpha_r^{min}$ (the smallest treatment effect is associated with the most strict judge) and $\alpha_r^J = \alpha_r^{max}$ (the largest treatment effect is associated with the most lenient judge), then the infra-marginality bias is bounded by:

$$\max_j(\lambda^j)(\alpha_r^{max} - \alpha_r^{min}) \tag{B.30}$$

In practice, we do not observe $\alpha_r^{max} - \alpha_r^{min}$. While we find evidence for limited heterogeneity in treatment effects in our setting, we take a conservative approach and assume the worst-case scenario. In other words, we assume that $\alpha_r^{max} - \alpha_r^{min} = 1$. We assume that there are defendants who are rearrested with probability 1 if released but never rearrested if detained such that $\alpha_r^{max} = 1$, and also that there are defendants whose rearrest probability is unaffected by release status such that $\alpha_r^{min} = 0$. Because the weights λ^j are identified in our data, under the worst-case scenario, the maximum bias due to infra-marginality concerns can be conservatively estimated to be equal to $\max_j(\lambda^j)$.

From Equation (B.28), we thus know that the maximum potential bias between D^{IV} and D^* is bounded by:

$$-\max_j(\lambda^j) < D^{IV} - D^* < \max_j(\lambda^j) \tag{B.31}$$

Intuition of Maximum Bias Formula: To illustrate intuitively how we bound the maximum estimation bias between D^{IV} and D^*, it is helpful to consider a simple two judge case. Assume that there is no racial bias such that both judges use the same release thresholds for both white and black defendants, $t_W^j = t_B^j$ for both $j = 1$ and $j = 2$. Under this scenario, $D^* = 0$.

Suppose that the lenient judge releases defendants with an expected pre-trial misconduct rate of less than 20 percent, while the strict judge releases defendants with an expected pre-trial misconduct rate of less than 10 percent. Then, the race-specific LATEs estimated using our judge IV strategy is the average treatment effect of all defendants with expected misconduct rates between 10 and 20 percent (i.e. compliers).

Within this range of compliers, suppose that all black defendants have expected rates of pre-trial misconduct of 10 percent, while all white defendants have expected rates of pre-trial misconduct of 20 percent (i.e. the distribution of compliers differs by race). If so, our IV estimator will yield a

LATE for whites ($\alpha_W^{IV} = 0.2$) that is larger in magnitude than the LATE for blacks ($\alpha_B^{IV} = 0.1$), causing us to estimate $D^{IV} > 0$. Our IV estimator would thus lead us to conclude that there was racial bias despite the fact that there is no true racial bias ($D^* = 0$). A similar exercise can be used to show that we may find $D^{IV} = 0$ even if $D^* > 0$.

Intuitively, this "infra-marginality bias" arises because not all compliers are marginal in the case of a discrete instrument. Because the distribution of treatment effects for both white and black defendants may be different among compliers, our IV estimator can lead us to erroneously find racial bias where none exists. Conversely, this infra-marginality problem could also lead us to conclude there is no racial bias when in fact both judges are racially biased against blacks.

Specifically, in the case of only two judges, the maximum estimation bias due to infra-marginality concerns is $\max_j(\lambda^j) = 1$ because 100 percent of compliers fall within the two judges. In this case, without further assumptions on the distribution of treatment effects by race, any value of D^{IV} is consistent with no true racial bias ($D^* = 0$), racial bias blacks ($D^* > 0$), or racial bias against whites ($D^* < 0$). As a result, in the two judge case, one would need to make additional assumptions to ensure the consistency of our estimator. For example, in this two judge case, one could assume that the distribution of treatment effects is uniform across the two thresholds for defendants of both races. With uniform distribution of treatment effects by race, our IV estimator would yield $\alpha_W^{IV} = 0.15$ and $\alpha_B^{IV} = 0.15$. Thus, we would find $D^{IV} = 0$, yielding a consistent estimate of true racial bias.

In order to bound the extent of the infra-marginality problem in our setting where there are many judges, we assume the worst-case scenario in an analogous way to the two judge example. With multiple judges, we also assume that $\alpha_r^{max} - \alpha_r^{min} = 1$ (the maximum possible heterogeneity among compliers). Because the weights λ^j are identified in our data, the maximum bias due to infra-marginality concerns can be conservatively estimated to be equal to $\max_j(\lambda^j)$.

Estimating Maximum Bias in our Setting: We now illustrate how we empirically estimate the maximum potential bias of our IV estimator from the true level of racial bias by using the formula in Proposition 5. Again, because we cannot observe $\alpha^{max} - \alpha^{min}$, we take the most conservative approach and assume that this value is equal to 1.

Recall from before that the weights λ_r^j are given by the following formula:

$$\lambda_r^j = \frac{(Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r)) \cdot \sum_{l=j}^{J} \pi_r^l (g(z_l) - \mathbb{E}[g(Z)])}{\sum_{m=1}^{J} (Pr(Released|z_m, r) - Pr(Released|z_{m-1}, r)) \cdot \sum_{l=m}^{J} \pi_r^l (g(z_l) - \mathbb{E}[g(Z)])} \tag{B.32}$$

As discussed in 4, under the exclusion restriction (Assumption 2), the probability of being assigned to any particular judge should not differ by defendant race. Therefore, π_r^l and $\mathbb{E}[Z]$ are independent of race. Also, we use judge leniency as our instrument so $g(Z) = Z$. Given a linear first stage, $Pr(Released|z_j, r) - Pr(Released|z_{j-1}) = \gamma_r(z_j - z_{j-1})$. Substituting this expression into (B.32) and simplifying yields:

$$\lambda^j = \frac{(z_j - z_{j-1}) \cdot \sum_{l=j}^{J} \pi^l (z_l - \mathbb{E}[Z])}{\sum_{m=1}^{J} (z_j - z_{j-1}) \cdot \sum_{l=m}^{J} \pi^l (z_l - \mathbb{E}[Z])} \tag{B.33}$$

We use Equation (B.33) to estimate the maximum bias of our estimator by replacing π^j and $\mathbb{E}[Z]$ with their empirical counterparts:

$$\hat{\pi}^j = \sum_{i=1}^{N} \frac{\mathbb{1}\{Z_i = z_j\}}{N} \tag{B.34}$$

$$\mathbb{E}[Z] = \frac{1}{N} \sum_{i=1}^{N} Z_i \tag{B.35}$$

Plugging these quantities into the formula for the weights yields an estimate of the weight attached to each pairwise LATE. We then take the maximum of our weights and interpret this estimate as the maximum potential bias between our IV estimator and the true level racial bias. This procedure yields a maximum bias of 0.005 or 0.5 percentage points.

From Equation (B.28), we know:

$$D^* < D^{IV} + \max_j(\lambda^j) = D^{IV} + .005$$
$$D^* > D^{IV} - \max_j(\lambda^j) = D^{IV} - .005$$

Therefore, in our setting, the true level of racial bias is bounded within 0.5 percentage points of our IV estimate for racial bias. □

E. Re-weighting Procedure to Allow Judge Preferences for Non-Race Characteristics

In this section, we show that a re-weighting procedure can be used to estimate direct racial bias (i.e. racial bias which cannot be explained by the composition of crimes). To begin, let the weights for all white defendants be equal to 1. We construct the weights for a black defendant with observables equal to $\mathbf{X}_i = x$ as:

$$\Psi(x) = \frac{Pr(W|x)Pr(B)}{Pr(B|x)Pr(W)} \tag{B.36}$$

where $Pr(W|x)$ is the probability of being white given observables $\mathbf{X}_i = x$, $Pr(B|x)$ is the probability of being black given observables $\mathbf{X}_i = x$, $Pr(B)$ is the unconditional probability of being black, and $Pr(W)$ is the unconditional probability of being white.

Define the covariate-specific LATE as:

$$\alpha_r^{j,j-1}(x) = \mathbb{E}[Y_i(1) - Y_i(0) | R_i(z_j) - R_i(z_{j-1}) = 1 | r_i = r, \mathbf{X}_i = x] \tag{B.37}$$

As noted by Fröhlich (2007), the unconditional LATE can be expressed as:

$$\alpha_r^{j,j-1} = \sum_{x \in X} \alpha_r^{j,j-1}(x) \frac{Pr(Released|z_j, x, r) - Pr(Released|z_{j-1}, x, r)}{Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r)} P(x|r) \tag{B.38}$$

Given a linear first stage:

$$\frac{Pr(Released|z_j, x, r) - Pr(Released|z_{j-1}, x, r)}{Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r)} = 1 \tag{B.39}$$

Therefore, in the re-weighted sample, $\alpha_B^{j,j-1}$ is given by:

$$\begin{aligned}
\alpha_B^{j,j-1} &= \sum_{x \in X} \alpha_B^{j,j-1}(x) Pr(x|B) \Psi(x) \\
&= \sum_{x \in X} \alpha_B^{j,j-1}(x) Pr(x|B) \frac{Pr(W|x)Pr(B)}{Pr(B|x)Pr(W)} \\
&= \sum_{x \in X} \alpha_B^{j,j-1}(x) \frac{Pr(B|x)Pr(x)}{Pr(B)} \frac{Pr(W|x)Pr(B)}{Pr(B|x)Pr(W)} \\
&= \sum_{x \in X} \alpha_B^{j,j-1}(x) \frac{Pr(W|x)Pr(x)}{Pr(W)} \\
&= \sum_{x \in X} \alpha_B^{j,j-1}(x) Pr(x|W)
\end{aligned}$$

Where line 2 follows by plugging in the formula for $\Psi(x)$ and lines 3 and 5 follow from Bayes' rule. Given that the weights for all white defendants are equal to 1, D^{IV} is given by:

$$D^{IV} = \sum_{j=1}^{J} \lambda^j \left(\sum_{x \in X} Pr(x|W) \left(\alpha_W^{j,j-1}(x) - \alpha_B^{j,j-1}(x) \right) \right) \tag{B.40}$$

$\square$

Appendix C: Data Appendix

Judge Leniency: We calculate judge leniency as the leave-one-out mean residualized pre-trial release decisions of the assigned judge within a bail year. We use the residual pre-trial release decision after removing court-by-time fixed effects. In our main results, we define pre-trial release based on whether a defendant was released within the first three days after the bail hearing.

Release on Recognizance: An indicator for whether the defendant was released on recognizance (ROR), where the defendant secures release on the promise to return to court for his next scheduled hearing. ROR is used for defendants who show minimal risk of flight, no history of failure to appear for court proceedings, and pose no apparent threat of harm to the public.

Non-Monetary Bail: An indicator for whether the defendant was released on non-monetary bail, also known as conditional release. Non-monetary conditions include monitoring, supervision, halfway houses, and treatments of various sorts, among other options.

Monetary Bail: An indicator for whether the defendant was assigned monetary bail. Under monetary bail, a defendant is generally required to post a bail payment to secure release, typically 10 percent of the bail amount, which can be posted directly by the defendant or by sureties such as bail bondsmen.

Bail Amount: Assigned monetary bail amount in thousands, set equal to missing for defendants who receive non-monetary bail or ROR.

Race: Information on defendant race is missing for the Philadelphia data prior to 2010.

Hispanic: We match the surnames in our data to census genealogical records of surnames. If the probability a given surname is Hispanic is greater than 80 percent, we label the defendant as Hispanic.

Prior Offense in Past Year: An indicator for whether the defendant had been charged for a prior offense in the past year of the bail hearing within the same county, set to missing for defendants who we cannot observe for a full year prior to their bail hearing.

Number of Offenses: Total number of charged offenses.

Felony Offense: An indicator for whether the defendant is charged with a felony offense.

Misdemeanor Offense: An indicator for whether the defendant is charged with only misdemeanor offenses.

Rearrest: An indicator for whether the defendant was rearrested for a new crime prior to case disposition.

Race: We collect information on judge race from court directories and conversations with court officials. All judges in Philadelphia are white. Information on judge race in Miami is missing for two of the 170 judges in our sample.

Experience: We use historical court records back to 1999 to compute experience, which we define as the difference between bail year and start year (earliest 1999). In our sample, years of experience range from zero to 15 years.

Appendix D: Institutional Details

Philadelphia County: Immediately following arrest in Philadelphia County, defendants are brought to one of six police stations around the city where they are interviewed by the city's Pre-Trial Services Bail Unit. The Bail Unit operates 24 hours a day, seven days a week, and interviews all adults charged with offenses in Philadelphia through videoconference, collecting information on the arrested individual's charge severity, personal and financial history, family or community ties, and criminal history. The Bail Unit then uses this information to calculate a release recommendation based on a four-by-ten grid of bail guidelines that is presented to the bail judge. However, these bail guidelines are only followed by the bail judge about half the time, with judges often imposing monetary bail instead of the recommended non-monetary options (Shubik-Richards and Stemen 2010).

After the Pre-Trial Services interview is completed and the charges are approved by the Philadelphia District Attorney's Office, the defendant is brought in for a bail hearing. Since the mid-1990s, bail hearings have been conducted through videoconference by the bail judge on duty, with representatives from the district attorney and local public defender's offices (or private defense counsel) also present. However, while a defense lawyer is present at the bail hearing, there is no real opportunity for defendants to speak with the attorney prior to the hearing. At the hearing itself, the bail judge reads the charges against the defendant, informs the defendant of his right to counsel, sets bail after hearing from representatives from the prosecutor's office and the defendant's counsel, and schedules the next court date. After the bail hearing, the defendant has an opportunity to post bail, secure counsel, and notify others of the arrest. If the defendant is unable to post bail, he is detained but has the opportunity to petition for bail modification in subsequent court proceedings.

Miami-Dade County: The Miami-Dade bail system follows a similar procedure, with one important exception. As opposed to Philadelphia where all defendants are required to have a bail hearing, most defendants in Miami-Dade can avoid a bail hearing and be immediately released following arrest and booking by posting an amount designated by a standard bail schedule. The bail schedule ranks offenses according to their seriousness and assigns an amount of bond that must be posted to permit a defendant's release. Critics have argued that this kind of standardized bail schedule discriminates against poor defendants by setting a fixed price for release according to the charged offense rather than taking into account a defendant's ability to pay, or propensity to flee or commit a new crime. Approximately 30 percent of all defendants in Miami-Dade are released prior to a bail hearing, with the other 70 percent attending a bail hearing (Goldkamp and Gottfredson 1988).

If a defendant is unable to post bail immediately in Miami-Dade, there is a bail hearing within 24 hours of arrest where defendants can argue for a reduced bail amount. Miami-Dade conducts separate daily hearings for felony and misdemeanor cases through videoconference by the bail judge on duty. At the bail hearing, the court will determine whether or not there is sufficient probable cause to detain the arrestee and if so, the appropriate bail conditions. The bail amount may be lowered, raised, or remain the same as the scheduled bail amount depending on the case situation and the arguments made by defense counsel and the prosecutor. While monetary bail amounts at this stage often follow the standard bail schedule, the choice between monetary versus non-monetary bail conditions varies widely across judges in Miami-Dade (Goldkamp and Gottfredson 1988).

Institutional Features Relevant to the Empirical Design: Our empirical strategy exploits variation in the pre-trial release tendencies of the assigned bail judge. There are three features of the Philadelphia and Miami-Dade bail systems that make them an appropriate setting for our research design. First, there are multiple bail judges serving simultaneously, allowing us to measure variation in bail decisions across judges. At any point in time, Philadelphia has six bail judges that only make bail decisions. In Miami-Dade, weekday cases are handled by a single bail judge, but weekend cases are handled by approximately 60 different judges on a rotating basis. These weekend bail judges are trial court judges from the misdemeanor and felony courts in Miami-Dade that assist the bail court with weekend cases.

Second, the assignment of judges is based on rotation systems, providing quasi-random variation in which bail judge a defendant is assigned to. In Philadelphia, the six bail judges serve rotating eight-hour shifts in order to balance caseloads. Three judges serve together every five days, with one bail judge serving the morning shift (7:30AM-3:30PM), another serving the afternoon shift (3:30PM-11:30PM), and the final judge serving the night shift (11:30PM-7:30AM). While it may be endogenous whether a defendant is arrested in the morning or at night or on a specific day of the week, the fact that these six bail judges rotate through all shifts and all days of the week allows us to isolate the independent effect of the judge from day-of-week and time-of-day effects. In Miami-Dade, the weekend bail judges rotate through the felony and misdemeanor bail hearings each weekend to ensure balanced caseloads during the year. Every Saturday and Sunday beginning at 9:00AM, one judge works the misdemeanor shift and another judge works the felony shift. Because of the large number of judges in Miami-Dade, any given judge works a bail shift approximately once or twice a year.

Third, there is very limited scope for influencing which bail judge will hear the case, as most individuals are brought for a bail hearing shortly following the arrest. In Philadelphia, all adults arrested and charged with a felony or misdemeanor appear before a bail judge for a formal bail hearing, which is usually scheduled within 24 hours of arrest. A defendant is automatically assigned to the bail judge on duty. There is also limited room for influencing which bail judge will hear the case in Miami-Dade, as arrested felony and misdemeanor defendants are brought in for their hearing within 24 hours following arrest to the bail judge on duty. However, given that defendants can post bail immediately following arrest in Miami-Dade without having a bail hearing, there is the possibility that defendants may selectively post bail depending on the identity of the assigned bail judge. It is also theoretically possible that a defendant may self-surrender to the police in order to strategically time their bail hearing to a particular bail judge. As a partial check on this important assumption of random assignment, we test the relationship between observable characteristics and bail judge assignment.

Appendix E: Model of Stereotypes

In this section, we consider whether a model of stereotypes can generate the pre-trial release rates we observe in our data. To do so, we assume a functional form for how judges form perceptions of risk and ask if this model can match the patterns we observe in the data.

Calculating Predicted Risk: We begin by estimating predicted risk by regressing the probability a defendant is rearrested prior to disposition on observables for released defendants. We then split the predicted risk measure into 100 equal sized bins. One potential concern with this procedure is that observably high-risk defendants may actually be low-risk based on variables observed by the judges, but not by the econometrician. To better understand the importance of this issue, we follow Kleinberg et al. (2017) and split our sample into a training and test set. We predict risk in the training set and then project the predictions onto the test set. We find that predicted risk is a strong predictor of true risk, indicating that the defendants released by judges do not have unusual unobservables which make their outcomes systematically diverge from what is expected (see Appendix Figure A2). This is true for both white and black defendants. Therefore, in this section, we interpret the predicted distributions of risk based on observables as the true distributions of risk.

No Stereotypes Benchmark: Following the construction of our predicted risk measure, we compute the fraction of black defendants that would be released if they were treated the same as white defendants. This calculation will serve as a benchmark for the stereotype model discussed below. To make this benchmark calculation, we assume judges accurately predict the risk of white defendants so that we can generate a relationship between release and risk, which we can then apply to black defendants. Under this assumption, we find that the implied release rate for black defendants is 52 percent if they were treated the same as white defendants. This implied release rate is lower than the true release rate of white defendants (55 percent), but higher than the true release rate for black defendants (45 percent), consistent with our main finding that judges over-detain black defendants.

Model with Stereotypes: We can now consider whether a simple model of stereotypes can rationalize the difference in true release rates. Following Bordalo et al. (2016), we assume judges form beliefs about the distribution of risk through a representativeness-based discounting model. Basically, the weight attached to a given risk type t is increasing in the representativeness of t. Formally, let $\pi_{t,r}$ be the probability that a defendant of race r is in risk category $t \in \{1, ..., 100\}$. In our data, a defendant with $t = 1$ has a zero expected expected probability of being rearrested before trial while a defendant with $t = 100$ has an expected 50 percent probability of being rearrested before trial.

Let $\pi^{st}_{t,r}$ be the stereotyped belief that a defendant of race r is in risk category t. The stereotyped beliefs for black defendants, $\pi^{st}_{t,B}$, is given by:

$$\pi^{st}_{t,B} = \pi_{t,B} \frac{\left(\frac{\pi_{t,B}}{\pi_{t,W}}\right)^{\theta}}{\sum_{s \in T} \pi_{s,B} \left(\frac{\pi_{s,B}}{\pi_{s,W}}\right)^{\theta}} \tag{E.1}$$

where θ captures the extent to which representativeness distorts beliefs and the representativeness ratio, $\frac{\pi_{t,B}}{\pi_{t,W}}$, is equal to the probability a defendant is black given risk category t divided by the probability a defendant is white given risk category t. Recall from Figure 3 that representativeness of blacks is strictly increasing in risk. Therefore, a representativeness-based discounting model will over-weight the right tail of risk for black defendants.

To compute the stereotyped distribution, we first assume a value of θ, and then compute $\pi_{t,r}$ for every risk category t and race r. We can then compute $\pi^{st}_{t,B}$ by plugging in the values for $\pi_{t,r}$ and the assumed value of θ into Equation (E.1).

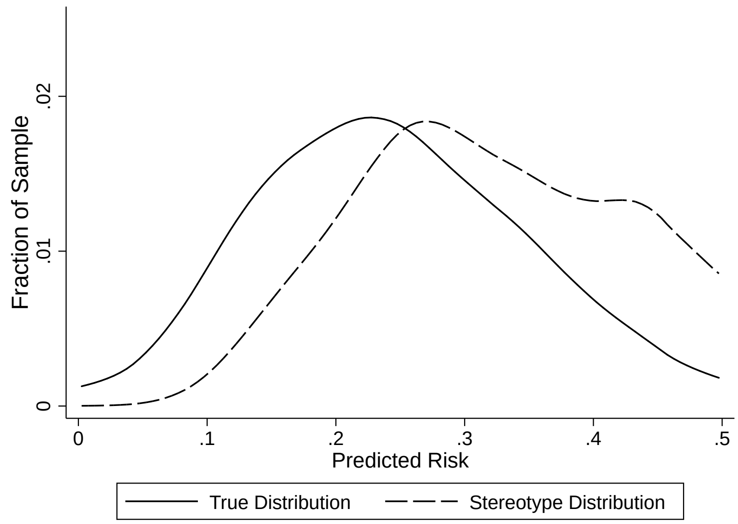
From the distribution of $\pi^{st}_{t,B}$, we compute the implied average release rate by multiplying the fraction of defendants believed to be at a given risk level by the probability of release for that risk level and summing up over all risk levels. Formally,

$$\mathbb{E}[Released_i = 1|r_i = B] = \sum_{s=1}^{100} \pi^{st}_{s,B} \mathbb{E}[Released_i = 1|t = s, r_i = B] \tag{E.2}$$

In the equation above, we cannot compute $\mathbb{E}[Released_i = 1|t = s, r_i = B]$ given that we explicitly assume judges make prediction errors for black defendants. That is, we do not know at what rate judges would release black defendants with risk equal to s, given that judges do not accurately predict risk for black defendants. However, in a stereotypes model, we can replace $\mathbb{E}[Released_i = 1|t = s, r_i = B] = \mathbb{E}[Released_i = 1|t = s, r_i = W]$ (i.e. given that if there is no taste-based discrimination, then conditional on perceived risk, the release rate will be equal between races). Under our additional assumption that judges accurately predict the risk of whites, we can estimate $\mathbb{E}[Released_i = 1|t = s, r_i = W]$ for all s. Therefore, we can compute every value on the right hand side of Equation (E.2), from which we can back out the average release rate for black defendants from the stereotyped distribution.

We find that $\theta = 2.5$ rationalizes the average release rate for blacks we observe in the data (45 percent). That is, if judges use a representativeness-based discounting model with $\theta = 2.5$ to form perceptions of the risk distribution, we would expect judges to release 45 percent of all black defendants. To understand how far these beliefs are from the true distribution of risk, we plot the stereotyped distribution for blacks with $\theta = 2.5$ alongside the true distribution of risk for blacks in Appendix Figure E1.

Appendix Figure E1: Stereotyped and True Distribution of Risk for Black Defendants



Note: This figure plots the true distribution of risk for black defendants alongside the perceived distribution of risk for black defendants. The stereotyped beliefs are generated by representativeness-based discounting model with $\theta = 2.5$. This value of θ rationalizes an average release rate of black defendants equal to 45 percent, the actual rate of release in the data.