Racial Bias in Bail Decisions*

David Arnold† Will Dobbie‡ Crystal S. Yang§

November 2017

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, 1993) 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 higher rates of misconduct than marginal black defendants if bail judges are racially biased, whether that bias is driven by racial animus, inaccurate racial stereotypes, or any other form of bias. 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 find suggestive evidence that this racial bias is driven by bail judges relying on inaccurate stereotypes that exaggerate the relative danger of releasing black defendants.

*We gratefully acknowledge the coeditors Lawrence Katz and Andrei Shleifer, and five anonymous referees for many valuable insights and suggestions. We also thank Josh Angrist, David Autor, 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, Michal Kolesár, Amanda Kowalski, Ilyana Kuziemko, Magne Mogstad, Nicola Persico, Steven Shavell, David Silver, Alex Torgovitsky, and numerous seminar participants for helpful comments and suggestions. Molly Bunke, Kevin DeLuca, Nicole Gandre, James Reeves, and Amy Wickett provided excellent research assistance.

†Princeton University. Email: dharnold@princeton.edu

‡Princeton University and NBER. Email: wdobbie@princeton.edu

§Harvard Law School and NBER. Email: cyang@law.harvard.edu

Racial disparities exist at every stage of the U.S. criminal justice system. Compared to observably similar whites, blacks are more likely to be searched for contraband (Antonovics and Knight 2009), more likely to experience police force (Fryer 2016), more likely to be charged with a serious offense (Rehavi and Starr 2014), more likely to be convicted (Anwar, Bayer, and Hjalmarrson 2012), and more likely to be incarcerated (Abrams, Bertrand, and Mullainathan 2012). Racial disparities are particularly prominent in the setting of bail: in our data, black defendants are 3.6 percentage points more likely to be assigned monetary bail than white defendants and, conditional on being assigned monetary bail, receive bail amounts that are $9,923 greater.[1] However, determining whether these racial disparities are due to racial bias or statistical discrimination remains an empirical challenge.

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

In recent years, two seminal papers have developed outcome tests of racial bias that partially circumvent this infra-marginality problem. In the first paper, Knowles, Persico, and Todd (2001) show that if motorists respond to the race-specific probability of being searched, then all motorists of a given race will carry contraband with equal probability. As a result, the marginal and average success rates of police searches will be identical and there is not an infra-marginality problem. Knowles et al. (2001) find no difference in the average success rate of police searches for white and black drivers, leading them to conclude that there is no racial bias in police searches. In a second important paper, Anwar and Fang (2006) develop a test of relative racial bias based on the idea that the ranking of search and success rates by white and black police officers should be unaffected by the race of the motorist even when there are infra-marginality problems. Consistent with Knowles et al. (2001), Anwar and Fang (2006) find no evidence of relative racial bias in police searches, but note that their approach cannot be used to detect absolute racial bias.[2] However, the prior literature has

[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 are that (1) we identify treatment effects for marginal defendants rather than the average defendant, and (2) we 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.

been critiqued for its reliance on restrictive assumptions about unobserved characteristics of blacks and whites (e.g., Brock et al. 2012).

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

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

The second part of the paper tests for racial bias in bail setting using administrative court

[3] See also Dobbie et al. (forthcoming), Gupta, Hansman, and Frenchman (2016), Leslie and Pope (2016), and Stevenson (2016) for evidence on the non-financial consequences of bail decisions.

data from Miami and Philadelphia. We find evidence of significant racial bias in our data, ruling out statistical discrimination as the sole explanation for the racial disparities in bail. Marginally released white defendants are 19.8 percentage points more likely to be rearrested prior to disposition than marginally released black defendants, with significantly more racial bias among observably high-risk defendants. Our IV-based estimates of racial bias are nearly identical if we account for other observable crime and defendant differences by race, suggesting that our results cannot be explained by black-white differences in certain types of crimes (e.g., the proportion of felonies versus misdemeanors) or black-white differences in defendant characteristics (e.g., the proportion with prior offenses versus no prior offenses). In sharp contrast to these IV results, naïve OLS estimates indicate, if anything, racial bias against <u>white</u> defendants, highlighting the importance of accounting for both infra-marginality and omitted variables when estimating bias in the criminal justice system.

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

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

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

Our results also contribute to an important literature testing for racial bias in the criminal justice system. As discussed above, Knowles et al. (2001) and Anwar and Fang (2006) are seminal works in this area. Subsequent work by Antonovics and Knight (2009) finds that police officers in Boston are more likely to conduct a search if the race of the officer differs from the race of the driver, consistent with racial bias among police officers, and Alesina and La Ferrara (2014) find that death sentences of minority defendants convicted of killing white victims are more likely to be reversed on appeal, consistent with racial bias among juries. Conversely, Anwar and Fang (2015) find no racial bias against blacks in parole board release decisions, observing that among prisoners released by the parole board between their minimum and maximum sentence, the marginal prisoner is the same as the infra-marginal prisoner. Mechoulan and Sahuguet (2015) also find no racial bias against blacks in parole board release decisions, arguing that for a given sentence, the marginal prisoner is the same as the infra-marginal prisoner. Finally, Ayres and Waldfogel (1994) show that bail bond dealers in New Haven charge lower prices to minority defendants, suggesting that minorities, at least on average, have a lower probability of pre-trial misconduct than whites, and Bushway and Gelbach (2011) find evidence of racial bias in bail setting using a parametric framework that accounts for unobserved heterogeneity across defendants.[4]

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

[4]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 forthcoming), the labor market (Goldin and Rouse 2000, Bertrand and Mullainathan 2004, Glover, Pallais, and Pariente 2017), 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. See Fryer (2011) and Bertrand and Duflo (2016) for partial reviews of the literature.

(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.

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, Becker 1993, 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 avoid undue punishment for defendants who have not yet been convicted of a crime. Second, bail judges are instructed to minimize the risk of pre-trial misconduct by setting the appropriate conditions for release. In our setting, pre-trial misconduct includes both the risk of new criminal activity and the risk of failure to appear for a required court appearance. 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 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. forthcoming, 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 $6,000 in our sample, and only 57 percent of defendants 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 explore the potential importance of these institutional features in Section IV.

B. Model of Judge Behavior

This section develops a stylized 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, 1993). We then present an alternative model of racially biased prediction errors, which generates similar 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 $\mathbb{E}[\alpha_i|\mathbf{V}_i, r_i]$, which includes the likelihood of both new crime and failure to appear, times the cost of misconduct C, which includes the social cost of any new crime or failures to appear. 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 perceived benefit of releasing defendant i assigned to judge j is denoted by $t_r^j(\mathbf{V}_i)$, which is a function of observable case and defendant characteristics $\mathbf{V}_i$. The perceived benefit of release $t_r^j(\mathbf{V}_i)$ includes social cost savings from reduced jail time, private gains to defendants from an improved bargaining position with the prosecutor or increased labor force participation, and personal benefits to judge j from any direct utility or disutility from being known as either a lenient or tough judge, respectively. Importantly, we allow the perceived 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, 1993), 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 perceived benefit of releasing white defendants than releasing observably identical black defendants.

For simplicity, we assume that bail judges are risk neutral and maximize the perceived net benefit of pre-trial release. We also assume that bail judges solely decide whether to release or detain a defendant, i.e., we abstract away from the fact that bail judges may set different levels of monetary bail that depend on a defendant's ability to pay. We discuss possible extensions to the model that account for these features below.

Under these assumptions, the model implies that bail judge j will release defendant i if and only if the expected cost of pre-trial release is less than the perceived 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 perceived 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, 1993):

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

Racially Biased Prediction Errors in Risk: In the taste-based model of discrimination outlined above, we assume that judges agree on the (true) expected cost of release, $\mathbb{E}[\alpha_i | r_i]$, but not the perceived benefit of release, $t_r^j(\mathbf{V}_i)$. An alternative approach is to assume that judges disagree on their (potentially inaccurate) predictions of the expected cost of release, as would be the case if judges systematically overestimate the cost of release for black defendants relative to white defendants. We show that a model motivated by these kinds of racially biased prediction errors in risk 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 perceived benefit of releasing defendant i assigned to judge j is now defined as $t(\mathbf{V}_i)$, which does not vary by judge.

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

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

where $\tau_r^j(\mathbf{V}_i)$ is a prediction error that is allowed to vary by judge j and defendant race r_i. To simplify our notation, we let the true expected 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 in risk 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 true 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 perceived cost of release:

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

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 in risk yields a similar outcome-based test for racial bias:

Proposition 2. If judge j systematically overestimates the true 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 in risk, the true expected cost of release for the marginal white defendant is higher than the true expected cost of release for the marginal black defendant.

Parallel to Proposition 1, Proposition 2 predicts that marginal white and marginal black defendants should have the same probability of pre-trial misconduct if judge j does not systematically make prediction errors in risk that vary with race, but marginal white defendants should have a higher probability of misconduct if judge j systematically overestimates the true 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, marginal white defendants will not have a higher probability of misconduct than marginal black defendants if observed racial disparities in bail setting are solely due to statistical discrimination.[5] Of course, finding higher misconduct rates for marginal white versus marginal black defendants does have a different interpretation depending on the underlying behavioral model. We will return to this issue in Section IV when we discuss more speculative evidence that allows us to differentiate between these two forms of racial bias.

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:

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

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. By definition, $\alpha_r^j = t_r^j$, where t_r^j represents judge j's threshold for release for defendants of race r. 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:

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

where λ^j are again non-negative weights which sum to one. From Equation (4), we can express D^* as 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 that depend in part on the size of the subpopulation whose pre-trial release decision is changed if they are assigned to a more or less lenient judge (Imbens and Angrist 1994). Section III discusses robustness checks that explore alternative weighting schemes.

[5] In our setting, models of statistical discrimination suggest that blacks may be treated worse than observably identical whites if either (1) blacks are, on average, riskier given an identical signal of risk (e.g., Phelps 1972, Arrow 1973) or (2) blacks have less precise signals of risk (e.g., Aigner and Cain 1977). In both types of statistical discrimination models, however, judges use race to form <u>accurate</u> predictions of risk, both on average and at the margin of release. As a result, neither form of statistical discrimination will lead to marginal white defendants having a higher probability of misconduct than marginal black defendants.

Bias with OLS Estimates: Let defendant i's probability of pre-trial misconduct, Y_i, be given by the following relationship, estimated separately for white and black defendants:

$$Y_i = \alpha_W Released_i + \beta_W \mathbf{X}_i + \mathbf{U}_i + \varepsilon_i \tag{6}$$

$$Y_i = \alpha_B Released_i + \beta_B \mathbf{X}_i + \mathbf{U}_i + \varepsilon_i \tag{7}$$

where $Released_i$ is an indicator for being released before trial, $\mathbf{X}_i$ 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 as age, gender, type of crime, prior pre-trial misconduct, 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 Equations (6) and (7) will typically not recover unbiased estimates of the true rate of pre-trial misconduct 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 commit pre-trial misconduct. In this scenario, OLS estimates of Equations (6) and (7) may be biased downwards from the true effect of pre-trial release, both on average and at the margin.

The second, and more important, reason OLS estimates will not recover unbiased estimates of treatment effects at the margin of release is that the average treatment effect identified by OLS may not be equal to the treatment effect at the margin required by the outcome 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. 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).

Defining our IV Estimator: We now formally establish the conditions under which our judge IV strategy yields consistent estimates of racial bias in bail setting. Before defining our IV 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 below, 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. This monotonicity assumption also implies that judges agree on how to rank-order defendants in terms of who should be released. 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:

$$\alpha_r^{IV} = \sum_{j=1}^{J} \lambda_r^j \cdot \alpha_r^{j,j-1} \tag{8}$$

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). 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.

Using the definition of α_r^{IV} from Equation (8), our IV estimator for racial bias can be expressed as:

$$D^{IV} = \alpha_W^{IV} - \alpha_B^{IV} = \sum_{j=1}^{J} \lambda_W^j \alpha_W^{j,j-1} - \sum_{j=1}^{J} \lambda_B^j \alpha_B^{j,j-1} \tag{9}$$

Consistency of our IV Estimator: Our IV estimator D^{IV} provides a consistent estimate of D^* under two conditions. The first is that our judge leniency measure Z_i is continuously distributed over some interval $[\underline{z}, \bar{z}]$. 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. Under this first condition, each race-specific IV estimate approaches a weighted average of treatment effects for defendants at the margin of release. 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, an assumption that is satisfied if the first-stage relationship between pre-trial release and our preferred measure of Z_i is linear for each race. This equal weights assumption ensures that the race-specific IV estimates from Equation (8), α_W^{IV} and α_B^{IV}, provide the same weighted averages of $\alpha_W^{j,j-1}$ and $\alpha_B^{j,j-1}$.

Proposition 3. Our IV estimator D^{IV} provides a consistent estimate of racial bias D^* if (1) Z_i is continuous and (2) λ_r^j is constant by race. 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:

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

where c is some constant. A linear first-stage relationship between pre-trial release and our judge leniency measure for each 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.

Proof. See Appendix B.

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 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 and D^{IV} may no longer provide a consistent estimate of D^*.[6] In Appendix B, we show that when Proposition 3 is satisfied

[6] One approach to estimating D^* with a discrete instrument 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.,

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 are the standard IV weights. The maximum bias of our estimator therefore decreases as (1) the heterogeneity in treatment effects among compliers decreases ($\alpha^{max} \to \alpha^{min}$) and (2) the maximum of the judge weights decreases ($\max_j(\lambda^j) \to 0$), as would occur when there are more judges distributed over the range of the instrument.[7]

To estimate the maximum bias in our setting, we use the empirical distribution of judge leniency in our data, the closed form solution for the IV weights λ^j when the first stage is linear, and worst case assumptions regarding treatment effect heterogeneity between white and black compliers (i.e. $\alpha^{max} - \alpha^{min} = 1$). This calculation indicates that, in our setting, the maximum bias of our IV estimator D^{IV} from the true level of racial bias D^* is 1.1 percentage points. See Appendix B for additional details.

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, judges minimize new crime, not just new arrests, and the police are more likely to rearrest black defendants conditional on having committed a new crime. In this scenario, we will overestimate the probability of pre-trial misconduct for black versus white defendants at the margin and, as a

Brinch et al. 2017). 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.

[7] To better understand why the number of judges may affect the maximum bias of our estimator, it is helpful to start with a simple two judge case where both judges use the same release thresholds for both white and black defendants, $t^j_W = t^j_B$, such that there is no racial bias, $D^* = 0$. Suppose that the more lenient judge releases defendants with an expected pre-trial misconduct rate of less than 20 percent, while the more strict judge releases defendants with an expected pre-trial misconduct rate of less than 10 percent. Then, the race-specific LATEs estimated using our IV strategy is the average treatment effect of all defendants with expected misconduct rates between 10 and 20 percent. 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. Then, our IV estimator will yield a LATE for whites ($\alpha^{IV}_W = 0.2$) that is larger in magnitude than the LATE for blacks ($\alpha^{IV}_B = 0.1$), causing us to estimate $D^{IV} = 0.1 > 0$. Our IV estimator would thus lead us to incorrectly conclude that there was racial bias. A similar exercise can be used to show that we may find $D^{IV} = 0$ even if $D^* > 0$. Under the worst-case scenario where we assume the maximum heterogeneity in treatment effects ($\alpha^{max} - \alpha^{min} = 1$), the maximum bias is $\max_j(\lambda^j) = 1$ because 100 percent of compliers fall within the two judges. In this case, this "infra-marginality bias" makes our IV estimator uninformative on the true level of racial bias. However, using the same logic, it is straightforward to show that the magnitude of this "infra-marginality bias" decreases when there are many judges because the share of compliers within any two judges decreases, thus decreasing $\max_j(\lambda^j)$.

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 to the extent that judges minimize new crime.

Omitted Objectives for Release: We also assume that judges do not consider other objectives or outcomes, or what Kleinberg et al. (forthcoming) refer to as "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 and this outcome is 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 omitted payoff bias in several ways. First, as will be discussed below, we find that our estimates are nearly identical if we measure pre-trial misconduct using only rearrests versus using rearrests or failures to appear. These results are also consistent with Kleinberg et al. (forthcoming), who find similar evidence of prediction errors using rearrests or failures to appear. Second, as will be discussed below, we also 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. (forthcoming) 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."[8]

Racially Differences in Ability to Pay Monetary Bail: In our model, we abstract away from the fact that bail judges may set different levels of monetary bail that, by law, should take into account a defendant's ability to pay. Extending our model to incorporate these institutional details means that racial bias could also be driven by judges systematically over-predicting the relative ability of black defendants to pay monetary bail at the margin. This type of racial bias could occur if, for

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

example, judges rely on fixed bail schedules that do not account for any racial differences in the ability to pay monetary bail.

We explore the empirical relevance of racial differences in ability to pay monetary bail in two ways. First, we test whether the assignment of non-monetary (versus monetary) bail has a larger impact on the probability of release for marginal black defendants. If judges systematically over-predict black defendants' ability to pay monetary bail at the margin, then the assignment of non-monetary bail will increase the probability of pre-trial release more for marginal black defendants compared to marginal white defendants. To test this idea, Panel A of Appendix Table A1 presents two-stage least squares estimates of the impact of non-monetary bail on pre-trial release using a leave-out measure based on non-monetary bail decisions as an instrumental variable. We find that the assignment of non-monetary bail has a nearly identical impact on the pre-trial release rates for marginal black defendants and marginal white defendants. These results run counter to the hypothesis that judges systematically over-predict the ability of black defendants to pay monetary bail.

We can also directly estimate racial bias in the setting of non-monetary (versus monetary) bail to incorporate any additional bias stemming from this margin. We can estimate these effects using a two-stage least squares regression of pre-trial misconduct on non-monetary bail, again using a leave-out measure based on non-monetary bail decisions as an instrumental variable. Panel B of Appendix Table A1 presents these estimates. We find nearly identical estimates of racial bias when focusing on the non-monetary versus monetary bail decision.[9] These results further suggest that judges are not systematically over-predicting the ability of black defendants to pay monetary bail.

Judge Preferences for Non-Race Characteristics: Another extension to our model concerns two distinct views about what constitutes racial bias. The first is that racial bias includes not only any bias due to phenotype, but also bias due to seemingly non-race factors that are correlated with, if not driven by, race. For example, according to this view, bail judges could be biased against defendants charged with drug offenses because blacks are more likely to be charged with these types of crimes. Our preferred estimates are consistent with this broader view of racial bias, measuring the disparate treatment of black and white defendants for all reasons unrelated to true risk of pre-trial misconduct, including reasons related to seemingly non-race characteristics such as crime type.

A second view is that racial bias is disparate treatment due to phenotype alone, not other correlated factors such as crime type. In Appendix B, we show that it is possible to test for this narrower form of racial bias using a re-weighting procedure that weights the distribution of observables of blacks to match observables of whites. This narrower test for racial bias relies on the assumption that judge preferences vary only by observable characteristics, i.e. $t_r^j(\mathbf{V}_i) = t_r^j(\mathbf{X}_i)$. We

[9] To compare these estimates to our preferred pre-trial release estimates, we scale the estimated treatment effects by the "first stage" effect of non-monetary bail on pre-trial release given by the Panel A estimates described above. For example, in our main results, we find that marginal white defendants are 19.8 percentage points more likely to be rearrested for any crime prior to disposition compared to marginal black defendants. If we scale the estimates in Panel B of Appendix Table A1 by those in Panel A of Appendix Table A1, we find that marginal white defendants are 18.7 percentage points more likely to be rearrested compared to marginal black defendants ($D^{IV} = \frac{0.102}{0.445} - \frac{0.020}{0.476} = 0.187$).

find nearly identical estimates of racial bias using this re-weighting procedure, suggesting that judge preferences over non-race characteristics are a relatively unimportant contributor to our findings.

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 70 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.[10]

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 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 was subsequently arrested for a new crime before case disposition. In Philadelphia, we also observe whether a defendant failed to appear for a required court appearance.

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 or missing race 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

[10] Appendix Table A2 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.

sample contains 162,836 cases from 93,914 unique defendants in Philadelphia and 93,417 cases from 65,944 unique defendants in Miami-Dade.

Table 1 reports summary statistics for our estimation sample separately by race and pre-trial release status. On average, black defendants are 3.6 percentage points more likely to be assigned monetary bail compared to white defendants and receive bail amounts that are $9,923 greater than white defendants. Conversely, black defendants are 2.0 percentage points and 1.6 percentage points less likely to be released on their own recognizance or to be assigned non-monetary bail with conditions compared to white defendants, respectively. As a result, black defendants are 2.4 percentage points more likely to be detained pre-trial compared to white defendants.

Compared to white defendants, released black defendants are also 1.9 percentage points more likely to be rearrested for a new crime before case disposition, our preferred measure of pre-trial misconduct. Released black defendants are also 0.9 percentage points, 0.7 percentage points, and 3.0 percentage points more likely to be rearrested for a drug, property, and violent crime, respectively. Finally, released black defendants are 1.4 percentage points more likely to fail to appear in court and 4.1 percentage points more likely to commit any form of pre-trial misconduct (either rearrest for a new crime or failure to appear) compared to white defendants.[11]

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. (forthcoming). 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 (e.g., Kling 2006, Aizer and Doyle 2015). For example, in our setting, bail hearings for DUI arrests disproportionately occur in the evenings and on particular days of the week. If certain bail judges are more likely to work evenings or weekends due to shift substitutions, the simple leave-out mean will be biased.

[11] We find that approximately four percent of detained defendants are rearrested for a new crime prior to case disposition – an outcome that should be impossible. In robustness checks, we show that our results are unaffected by dropping these cases.

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. With these court-by-time effects, 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 during 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:

$$Released^{*}_{ict} = Released_{ic} - \omega \mathbf{X}_{ict} = Z_{ctj} + v_{ict} \tag{11}$$

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 leniency measure of the assigned judge within a bail year:

$$Z_{ctj} = \left(\frac{1}{n_{tj} - n_{itj}}\right)\left(\sum_{k=0}^{n_{tj}}(Released^{*}_{ikt}) - \sum_{c=0}^{n_{itj}} Released^{*}_{ict}\right) \tag{12}$$

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. Leaving out a defendant's own observations is important 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. We calculate our 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 leniency measure given by Equation (12) represents the release rate for the assigned bail judge after accounting for the court-by-time fixed effects. 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.[12]

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.[13] 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 23,262 during the sample period of 2010-2014, with the typical judge-by-year cell including 5,253 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

[12] Our leave-out procedure is essentially a reduced-form version of jackknife IV, with the leave-out leniency measure Z_{ctj} for judge j being algebraically equivalent to judge j's fixed effect from a leave-out regression of residualized pre-trial release on the full set of judge fixed effects. Jackknife IV and LIML estimates using the full set of judge fixed effects as instruments are presented in robustness checks.

[13] Characteristics of compliers in our setting are presented in Appendix Table A3.

cell including 179 cases. Controlling for the exhaustive set of court-by-time fixed effects, the judge release measure ranges from -0.212 to 0.184 with a standard deviation of 0.033. In other words, moving from the least to most lenient judge increases the probability of pre-trial release by 39.6 percentage points, a 56.7 percent change from the mean release rate of 69.8 percentage points.

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, estimated separately for white and black defendants:

$$Released_{ictj} = \gamma_W Z_{ctj} + \pi_W \mathbf{X}_{ict} + v_{ict} \tag{13}$$

$$Released_{ictj} = \gamma_B Z_{ctj} + \pi_B \mathbf{X}_{ict} + v_{ict} \tag{14}$$

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, for all defendants 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 Equations (13) and (14), 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 the assumption of constant IV weights by race (Proposition 3) is likely valid in our setting.

Table 2 presents formal first stage results from Equations (13) and (14) 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, whether the defendant had a prior history of pre-trial crime in the past year, whether the defendant had a prior history of failure to appear 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.

We find that our residualized judge instrument is highly predictive of whether a defendant is released pre-trial. 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 5.3 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 5.0 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 5.5 percentage points more likely to be released pre-trial.[14]

The probability of being released pre-trial does not increase one-for-one with our measure of judge leniency, likely because of attenuation bias due to sampling variation in the construction of our instrument. For instance, our judge leniency measure is computed over finite judge caseloads and judge leniency may drift over the course of the year or fluctuate with case characteristics, reducing the accuracy of our leave-out measure. It is important to note that attenuation bias due to sampling variation in our leniency measure does not bias our IV estimates since it affects both the first stage and reduced form proportionally.

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 use 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 10.4 percentage points less likely to be released pre-trial compared to similar female defendants, while white male defendants are 8.6 percentage points less likely to be released pre-trial compared to similar female defendants. White defendants with at least one prior offense in the past year are 16.8 percentage points less likely to be released compared to defendants with no prior offenses, while black defendants with at least one prior offense in the past year are 13.4 percentage points less likely to be released compared to defendants with no prior offenses. 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. (forthcoming) 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.

[14] First stage results by subsamples are presented in Appendix Table A4.

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, i.e. that judges agree on how to rank-order defendants in terms of who should be released. 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).

An implication of the monotonicity assumption is that the ranking of each judge's leniency measure for whites and the ranking of each judge's leniency measure for blacks should fit a 45-degree line up to sampling error. Significant divergence between the white and black judge ranks would signal that the restrictions imposed by monotonicity are violated. Appendix Figure A1 plots our residualized judge leniency measures calculated separately by race. Consistent with our monotonicity assumption, we find that the slope relating the relationship between judge leniency for whites and judge leniency for blacks is strongly positive, suggesting that judge tendencies are broadly similar for black and white defendants. However, the raw data also suggest some non-monotonic behavior, with a handful of judges being relatively more lenient for one race than another. These potential violations of monotonicity appear to be driven by judges where we observe relatively few cases, making it difficult to determine whether these patterns are due to true non-monotonic behavior or sampling error.

To formally test for violations of monotonicity, we regress each judge's leniency rank in the white distribution on each judge's leniency rank in the black distribution. This regression yields a coefficient on the rank in the black distribution equal to 0.824 (se=0.010), suggesting that while judges' leniency ranks are highly correlated in our data, they are not strictly monotonic in race. We find similar results using an overidentification Wald test that asks whether the monotonicity assumption is valid for every judge, not just valid on average (e.g., Angrist et al. 2017). Our results suggest that the non-monotonic behavior we observe in our data is driven by approximately 12.5 percent of judges who hear less than 10 percent of all cases. Given this evidence, Section III.C discusses alternative tests of racial bias that relax the monotonicity assumption.[15]

[15] The monotonicity assumption would also be violated if lenient judges are better at using unobservable information to predict the risk of pre-trial misconduct, as this would result in some high-risk defendants being released by only strict judges. Following Kleinberg et al. (forthcoming), we test for this possibility by examining pre-trial misconduct rates among observably identical defendants released by either lenient or strict judges. If the most lenient judges are better at using unobservable information to predict risk, then defendants released by these most lenient judges will have lower misconduct rates than observably identical defendants released by the less lenient judges. To implement this test, we first split judges into quintiles of leniency. We then calculate predicted risk using the machine learning algorithm described in Appendix F, but only in the sample of defendants assigned to the most-lenient quintile. Finally, we apply the risk predictions to defendants in all leniency quintiles and plot predicted risk against actual risk for each leniency quintile (see Appendix Figure A2). Following the above logic, if lenient judges are better at using unobservable information, then predicted risk should be systematically below actual risk. We find that predicted risk largely tracks true risk in all judge leniency quintiles, suggesting that lenient judges are neither more nor less skilled in predicting defendant risk. These results are broadly consistent with Kleinberg et al. (forthcoming), who find that judges more or less agree on how to rank-order defendants based on their observable characteristics.

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. In the results that follow, we measure pre-trial misconduct as the probability of rearrest prior to case disposition. Specifically, we estimate the following two-stage least squares specifications for individual i and case c, assigned to judge j at time t, estimated separately for white and black defendants:

$$Y_{ict} = \alpha_W^{IV} Released_{ic} + \beta_W \mathbf{X}_{ict} + \mathbf{v}_{ict} \quad (15)$$

$$Y_{ict} = \alpha_B^{IV} Released_{ic} + \beta_B \mathbf{X}_{ict} + \mathbf{v}_{ict} \quad (16)$$

where the vector $\mathbf{X}_{ict}$ includes court-by-time fixed effects and baseline crime and defendant controls. 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 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 Equations (15) and (16). In our main results, we use cases from both Philadelphia and Miami in order to reduce the infra-marginality bias that arises from estimating racial bias using a discrete number of judges (see Appendix B). Columns 1-2 report 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, 17.8 percent of defendants are rearrested for a new crime prior to disposition, with 7.9 percent of defendants being rearrested for drug offenses, 6.7 percent of defendants being rearrested for property offenses, and 6.1 percent of defendants being rearrested for violent offenses.[16]

We find convincing evidence of racial bias against black defendants. In Panel A, we find that marginally released white defendants are 24.6 percentage points more likely to be rearrested for any crime compared to marginally detained white defendants (column 1). In contrast, the effect of pre-trial release on rearrest rates for the marginally released black defendants is a statistically in-

[16] For completeness, Figure 1 provides a graphical representation 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.

significant 4.9 percentage points (column 2). Taken together, these estimates imply that marginally released white defendants are 19.8 percentage points more likely to be rearrested prior to disposition than marginally released black defendants (column 3), consistent with racial bias against blacks. Importantly, we can reject the null hypothesis of no racial bias even assuming the maximum potential bias in our IV estimator of 1.1 percentage points (see Appendix B). Our results therefore rule out statistical discrimination as the sole determinant of racial disparities in bail.

In Panel B, we find suggestive evidence of racial bias against black defendants across all crime types, although the point estimates are too imprecise to make definitive conclusions. Most strikingly, we find that marginally released whites are about 9.3 percentage points more likely to be rearrested for a violent crime prior to disposition than marginally released blacks (p-value = 0.079). Marginally released white defendants are also 7.6 percentage points more likely to be rearrested for a drug crime prior to case disposition than marginally released black defendants (p-value = 0.171), and 12.9 percentage points more likely to be rearrested for a property crime (p-value = 0.054). These results suggest that judges are racially biased against black defendants even if they are most concerned about minimizing specific types of new crime, such as violent crimes.[17]

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 using a machine learning algorithm described in Appendix F, with those in the below median group having a 12.7 percent probability of rearrest prior to case disposition compared to 32.9 percent among defendants in the above median group. We find that racial bias against black defendants is largely driven by those with the highest predicted risk of rearrest. Among high-risk defendants, marginally released white defendants are 98.3 percentage points more likely to be rearrested prior to case disposition than marginally released black defendants (p-value = 0.040). In contrast, we find more limited evidence of racial bias against black defendants among low-risk defendants (p-value = 0.386). See Appendix Table A5 for subsample results for a subset of the individual characteristics in our risk index.

C. Robustness

Our estimates of racial bias are robust to a number of potential concerns. Appendix Table A6 explores whether our main findings are subject to omitted payoff bias, which can arise if judges consider other outcomes such as failures to appear. While we only observe failures to appear in

[17] 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, Appendix Figure A4 presents MTEs by defendant race. 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 judge leniency distribution. These results indicate that racial bias against black defendants arises at every part of the judge leniency distribution. See Appendix E for details on how we estimate the MTEs.

Philadelphia, we find that our estimates are nearly identical when we use a measure of pre-trial misconduct defined as any rearrest or any failure to appear. To further explore the possibility that judges may only care about minimizing specific types of new crime, Appendix Table A7 presents estimates for a subset of more serious crime types for which estimates of social costs are available, such as murder, assault, and robbery, and weights each individual estimate of D^{IV} by the corresponding social cost. While our estimates become less precise given the infrequency of certain types of new crime, a social cost-weighted estimate of racial bias yields a lower bound of \$26,163 and an upper bound of \$69,376 for these more serious crimes, suggesting that marginally released white defendants create larger social costs than marginally released black defendants.

Appendix Table A8 explores the sensitivity of our main results to a number of different specifications, where each column reports our estimate of racial bias D^{IV}. Column 1 drops a small number of defendants who the data indicate were rearrested prior to disposition despite never being released. Column 2 presents re-weighted estimates with the weights chosen to match the distribution of observable characteristics by race. Column 3 presents results comparing outcomes for marginal non-Hispanic white defendants and marginal black defendants. Column 4 presents results clustering more conservatively at the individual and judge level. Column 5 presents results with bootstrap-clustered standard errors, which correct for estimation error in the construction of our judge leniency measure.[18] Columns 6 and 7 present results that use judge fixed effects as instruments for pre-trial release (F-statistic = 18.18) estimated using jackknife IV and LIML, respectively. Column 8 assesses whether monetary bail amounts have an independent effect on the probability of pre-trial misconduct – a potential violation of the exclusion restriction – by controlling for monetary bail amount as an additional regressor in both our first- and second-stage regressions. 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.

In a final specification check, column 9 of Appendix Table A8 relaxes our monotonicity assumption, which implies that the judge ordering produced by our instrument is the same for both white and black defendants, by separately calculating our leave-out judge leniency measure by defendant race. Using this race-specific instrument, we continue to find evidence of economically and statistically significant racial bias against black defendants. In Appendix E, we explore whether unequal IV weights by race, λ_r^j, could be driving this robustness result in two ways. First, we estimate semi-parametric MTE estimates using race-specific judge leniency measures and then ex post impose equal weights across race for each judge in our sample. We find a very similar estimate of racial bias under a simple weighted average of the race-specific MTEs, with $D^{IV} = 0.325$ (p-value=0.004). Second, we estimate lower and upper bounds on the true level of racial bias using non-parametric

[18] 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.

pairwise LATEs calculated separately for blacks and whites where we again ex post impose equal weights across race for each judge in our sample.[19] We find economically significant racial bias against black defendants in all specifications, although our results are less precise compared to our main estimates. See Appendix E for additional details.

D. Comparison to Other Outcome Tests

Appendix Tables A9-A11 replicate the outcome tests from Knowles et al. (2001) and Anwar and Fang (2006). 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 given by standard OLS specifications will not vary by defendant race. In contrast to our preferred IV test, however, standard OLS specifications suggest racial bias against <u>white</u> defendants. The Anwar and Fang (2006) test instead relies on the prediction that, under the null hypothesis of no relative racial bias, the treatment of black and white defendants will not depend on judge race. However, this test fails to find racial bias in our setting because both white and black judges are racially biased against black defendants. We also find that the IV estimate of racial bias is similar among white and black judges, although the confidence intervals for these estimates are large. Taken together, these results highlight the importance of accounting for both infra-marginality and omitted variables, as well as the importance of developing empirical tests that can detect absolute racial bias in the criminal justice system. See Arnold, Dobbie, and Yang (2017) for additional details on these results.

IV. Potential Mechanisms

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

A. Racial Animus

The first potential explanation for our results is that judges either knowingly or unknowingly discriminate against black defendants at the margin of release as originally modeled by Becker (1957, 1993). Bail judges could, for example, harbor explicit animus against black defendants that leads them to value the freedom of black defendants less than the freedom of observably similar white defendants. Bail judges could also harbor implicit biases against black defendants – similar to those documented among both employers (Rooth 2010) and doctors (Penner et al. 2010) – leading to the relative over-detention of blacks despite the lack of any explicit animus.[20] Racial animus may be a particular concern in bail setting due to the relatively low number of minority bail judges, the

[19] We thank an anonymous referee for this suggestion.

[20] 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).

rapid-fire determination of bail decisions, and the lack of face-to-face contact between defendants and judges. Prior work has shown that it is exactly these types of settings where racial prejudice is most likely to translate into the disparate treatment of minorities (e.g., Greenwald et al. 2009).

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

B. Racially Biased Prediction Errors in Risk

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

Representativeness of Black and White Defendants: We first explore whether our data are consistent with the formation of anti-black stereotypes that could lead to racially biased prediction errors. Extending Bordalo et al. (2016) to our setting, 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, Figure 2 presents the distribution of the predicted risk of rearrest prior to case disposition calculated using the full set of crime and defendant characteristics, as well as the likelihood ratios, $\mathbb{E}(x|Black)/\mathbb{E}(x|White)$, throughout the risk distribution.[21] Results for each individual characteristic in our predicted risk measure are also presented in Appendix Table A12. Consistent with the potential formation of anti-black stereotypes, we find that black defendants are significantly under-represented in the left tail of the predicted risk distribution and over-represented in the right tail of the predicted risk distribution. For example, black defendants are 1.2 times less likely than whites

[21] 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. (forthcoming), 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 A3).

to be represented among the bottom 25 percent of the predicted risk distribution, but 1.1 times more likely to be represented among the top 25 percent and 1.2 times more likely to be represented among the top five percent of the predicted risk distribution.

In Appendix F, 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 in the Bordalo et al. (2016) stereotypes model. First, as a benchmark for the stereotypes model, we compute the fraction of black defendants that would be released if judges applied the same release threshold for whites to blacks. We rank-order both black and white defendants using our predicted risk measure, finding that 70.8 percent of black defendants would be released pre-trial if judges use the white release threshold for both black and white defendants. By comparison, only 68.8 percent of black defendants are actually released pre-trial. Thus, to rationalize the black-white difference in release rates, the stereotypes model will require that judges believe that black defendants are riskier than they actually are.

In the stereotypes model, judges form beliefs about the distribution of risk through a representativeness-based discounting model, where 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. 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 the definition of $\pi^{st}_{t,B}$ from Equation (17), we can calculate the full stereotyped risk distribution for black defendants under different values of θ. For each value of θ, we can then calculate the implied release rate for black defendants under the above assumption that judges use the white release threshold for both black and white defendants. By iterating over different values of θ, we can then find the level of θ that equates the implied and true release rates for black defendants. Using this approach, we find that $\theta = 1.9$ can rationalize the true 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 $\theta = 1.9$ alongside the true distribution of risk for blacks in Appendix Figure A5. The mean predicted risk is 0.235 under the true distribution of risk for blacks, compared to 0.288 under the stereotyped distribution for blacks with $\theta = 1.9$. 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. See Appendix F for additional details on the stereotypes model and these calculations.

Further evidence in support of anti-black stereotypes comes from a comparison of the crime- and subsample-specific distributions of risk. Black defendants are most over-represented in the right tail of the predicted risk distribution for new violent crimes (see Appendix Figure A6), where we also tend to find strong evidence of racial bias. Black defendants charged with felonies are also substantially over-represented in the right tail of the predicted risk distribution for any new crime, as are black defendants with either a prior criminal history or prior history of pre-trial misconduct (see Appendix Figure A7). These are also the subsamples where we tend to find stronger evidence of racial bias (see Appendix Table A5).

A final piece of evidence in support of stereotyping comes from a comparison of the Hispanic and black distributions of risk. Recall that we find no evidence of racial bias against Hispanic defendants (see Appendix Table A2). Consistent with the stereotyping model, we also find that the risk distributions of Hispanic and white defendants overlap considerably. In contrast, the risk distribution for blacks is shifted to the right relative to both the Hispanic and white distributions (see Appendix Figure A8). 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 in Risk: We can also test for race-based prediction errors by examining situations where prediction errors (of any kind) are more likely to occur. The first such test uses a comparison of low- and high-risk defendants. Kleinberg et al. (forthcoming) 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 idea, we find significantly more racial bias among observably high-risk defendants (see Table 5). In contrast, there is no reason to believe that racial animus should be different for low- and high-risk defendants.

A second test for race-based prediction errors 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 be less likely to rely on inaccurate racial stereotypes as they acquire greater on-the-job experience, at least in settings with limited information and contact. Consistent with this idea, we find that more experienced bail judges are more likely to release defendants, but not make misclassification errors (see Appendix Figure A9). In contrast, while it appears plausible that race-based prediction errors will decrease with experience, there is no reason to believe that racial animus will change with experience.[22]

Table 6 presents a series of estimates for judges with different levels of experience. Columns 1-3 of Table 6 presents our estimates of racial bias, D^{IV}, separately by court. In Philadelphia, bail judges are full-time judges who specialize in setting bail 24 hours a day, seven days a week, hearing

[22]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.

an average of 5,253 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 179 bail cases each year. 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 = 0.269). In Miami, marginally released white defendants are 23.2 percentage points more likely to be rearrested compared to marginally released black defendants (p-value = 0.026). In Philadelphia, we find no statistically significant evidence of racial bias (p-value = 0.948), suggesting the possible importance of experience in alleviating any prediction errors.[23]

Columns 4 through 6 of Table 6 exploit the substantial variation in the experience profiles of the Miami bail judges in our sample. 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. Consistent with our across-court findings, we find suggestive evidence that inexperienced judges are more racially biased than experienced judges (p-value = 0.189). Among inexperienced judges, marginally released white defendants are 43.8 percentage points more likely to be rearrested compared to marginally released black defendants (p-value = 0.051). Among experienced judges, marginally released white defendants are 12.8 percentage points more likely to be rearrested compared to marginally released black defendants (p-value = 0.412).

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

V. Conclusion

In this paper, we test for racial bias in bail setting using the quasi-random assignment of bail judges to identify pre-trial misconduct rates for marginal white and marginal black defendants. We find evidence that there is substantial bias against black defendants, with the largest bias against black defendants with the highest predicted risk of rearrest. Our estimates are nearly identical if we account for observable crime and defendant differences by race, indicating that our results cannot be explained by black-white differences in the probability of being arrested for certain types of

[23]Our estimates from Philadelphia should be interpreted with some caution given that we only observe seven judges in our data. As a result, infra-marginality bias may a larger concern in Philadelphia, where the maximum bias of our estimator is 0.163 compared to 0.011 in Miami-Dade. See Appendix B.

crimes (e.g., the proportion of felonies versus misdemeanors) or black-white differences in defendant characteristics (e.g., the proportion of defendants with prior offenses versus no prior offenses).

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

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

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

References

[1] 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.

[2] Agan, Amanda Y., and Sonja B. Starr. Forthcoming. "Ban the Box, Criminal Records, and Statistical Discrimination: A Field Experiment." Quarterly Journal of Economics.

[3] Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." ILR Review, 30(2): 175-187.

[4] 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.

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

[6] 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.

[7] Angrist, Joshua, Peter Hull, Parag A. Pathak, Christopher Walters. 2017. "Leveraging Lotteries for School Value-Added: Testing and Estimation." Quarterly Journal of Economics, 132(2): 871-919.

[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] Arnold, David, Will Dobbie, and Crystal S. Yang. 2017. "Racial Bias in Bail Decisions." NBER Working Paper No. 23421.

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

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

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

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

[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] Becker, Gary S. 1993. "Nobel Lecture: The Economic Way of Looking at Behavior." Journal of Political Economy, 101(3): 385-409.

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

[23] 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.

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

[25] Brinch, Christian N., Magne Mogstad, and Matthew Wiswall. 2017. "Beyond LATE with a Discrete Instrument." Journal of Political Economy, 125(4): 98-1039.

[26] Brock, William A., Jane Cooley, Steven N. Durlauf, and Salvador Navarro. 2012. "On the Observational Implications of Taste-Based Discrimination in Racial Profiling." Journal of Econometrics, 166: 66-78.

[27] 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.

[28] Cameron, Colin A., Jonah B. Gelbach, and Douglas A. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Standard Errors." The Review of Economics and Statistics, 90(3): 414-427.

[29] Card, David. 1999. "The Causal Effect of Schooling on Earnings," in Handbook of Labor Economics. Orley Ashenfelter and David Card, eds. Amsterdam: North Holland.

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

[31] 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.

[32] 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.

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

[34] 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.

[35] Dobbie, Will, Jacob Goldin, and Crystal Yang. Forthcoming. "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." American Economic Review.

[36] Doyle, Joseph. 2007. "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." American Economic Review, 97(5): 1583-1610.

[37] 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.

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

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

[40] 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.

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

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

[43] 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.

[44] Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." Quarterly Journal of Economics, 132(3): 1219-1260.

[45] Goldin, Claudia and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians." American Economic Review, 90(4): 715-741.

[46] 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.

[47] 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.

[48] Gruber, Jonathan, Phillip Levine, and Douglas Staiger. 1999. "Abortion Legalization and Child Living Circumstances: Who is the "Marginal Child?" Quarterly Journal of Economics, 114(1): 263-291.

[49] 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.

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

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

[52] Heckman, James J., and Edward Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. Proceedings of the National Academy of Sciences." 96(8): 4730-4734.

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

[54] Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. Forthcoming. "Human Decisions and Machine Predictions." Quarterly Journal of Economics.

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

[56] 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.

[57] 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.

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

[59] 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.

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

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

[62] Mogstad, Magne, Andres Santos, and Alexander Torgovitsky. 2017. "Using Instrumental Variables for Inference about Policy-Relevant Treatment Effects." NBER Working Paper No. 23568.

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

[64] 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.

[65] 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.

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

[67] 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.

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

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

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

[71] 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.

[72] 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.

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

[74] 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	
	Released	Detained	Released	Detained	Released	Detained
Panel A: Bail Type	(1)	(2)	(3)	(4)	(5)	(6)
Release on Recognizance	0.258	0.000	0.269	0.000	0.249	0.000
Non-Monetary Bail w/ Conditions	0.195	0.030	0.203	0.033	0.189	0.028
Monetary Bail	0.547	0.970	0.527	0.967	0.562	0.972
Bail Amount (in thousands)	13.235	35.286	11.957	24.782	14.180	42.227
Panel B: Defendant Characteristics						
Male	0.811	0.893	0.796	0.890	0.822	0.895
Age at Bail Decision	33.911	35.092	34.070	36.296	33.794	34.296
Prior Offense in Past Year	0.287	0.466	0.272	0.464	0.299	0.466
Arrested on Bail in Past Year	0.185	0.262	0.181	0.256	0.188	0.266
Failed to Appear in Court in Past Year	0.071	0.057	0.070	0.054	0.071	0.059
Panel C: Charge Characteristics						
Number of Offenses	2.722	3.162	2.544	2.587	2.854	3.541
Felony Offense	0.482	0.538	0.450	0.473	0.506	0.581
Misdemeanor Only	0.518	0.462	0.550	0.527	0.494	0.419
Any Drug Offense	0.390	0.260	0.373	0.244	0.403	0.271
Any DUI Offense	0.084	0.007	0.091	0.007	0.079	0.007
Any Violent Offense	0.310	0.331	0.288	0.241	0.326	0.390
Any Property Offense	0.238	0.387	0.237	0.406	0.239	0.376
Panel D: Outcomes						
Rearrest Prior to Disposition	0.237	0.042	0.226	0.037	0.245	0.045
Rearrest Drug Crime	0.111	0.006	0.106	0.005	0.115	0.006
Rearrest Property Crime	0.086	0.022	0.082	0.022	0.089	0.022
Rearrest Violent Crime	0.078	0.021	0.061	0.013	0.091	0.026
Failure to Appear in Court	0.258	0.006	0.250	0.006	0.264	0.007
Failure to Appear in Court or Rearrest	0.348	0.044	0.325	0.039	0.366	0.048
Observations	178,765	77,488	76,015	30,831	102,750	46,657

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. 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	0.527***	0.525***	0.499***	0.498***	0.552***	0.549***
	(0.031)	(0.029)	(0.043)	(0.040)	(0.044)	(0.040)
	[0.698]	[0.698]	[0.711]	[0.711]	[0.688]	[0.688]
Court x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Crime Controls	No	Yes	No	Yes	No	Yes
Observations	256,253	256,253	106,846	106,846	149,407	149,407

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	−0.09424***	0.00025	−0.08593***	0.00034	−0.10379***	0.00018
	(0.00235)	(0.00020)	(0.00325)	(0.00030)	(0.00323)	(0.00026)
Age at Bail Decision	−0.01725***	−0.00004	−0.02250***	−0.00008	−0.01512***	−0.00002
	(0.00086)	(0.00007)	(0.00127)	(0.00013)	(0.00104)	(0.00008)
Prior Offense in Past Year	−0.14922***	0.00029	−0.16817***	0.00044	−0.13411***	0.00021
	(0.00287)	(0.00024)	(0.00445)	(0.00040)	(0.00362)	(0.00029)
Arrested on Bail in Past Year	0.01066***	0.00002	0.01967***	−0.00049	0.00495	0.00035
	(0.00355)	(0.00028)	(0.00552)	(0.00049)	(0.00439)	(0.00035)
Failed to Appear in Court in Past Year	0.03318***	−0.00022	0.03253***	−0.00004	0.03245***	−0.00035
	(0.00413)	(0.00021)	(0.00631)	(0.00035)	(0.00529)	(0.00026)
Number of Offenses	−0.02090***	−0.00001	−0.01829***	−0.00004	−0.02131***	0.00001
	(0.00053)	(0.00003)	(0.00085)	(0.00005)	(0.00063)	(0.00003)
Felony Offense	−0.17618***	−0.00005	−0.18817***	−0.00016	−0.16948***	0.00003
	(0.00257)	(0.00010)	(0.00397)	(0.00017)	(0.00323)	(0.00012)
Any Drug Offense	0.03514***	−0.00035	0.02558***	−0.00024	0.04069***	−0.00042*
	(0.00258)	(0.00022)	(0.00357)	(0.00033)	(0.00332)	(0.00026)
Any Violent Offense	0.01640***	0.00024	0.07515***	0.00030	−0.02443***	0.00019
	(0.00389)	(0.00021)	(0.00497)	(0.00036)	(0.00429)	(0.00024)
Any Property Offense	−0.04272***	−0.00002	−0.05560***	−0.00002	−0.03188***	−0.00002
	(0.00285)	(0.00023)	(0.00388)	(0.00035)	(0.00354)	(0.00028)
Joint F-Test	[0.00000]	[0.40846]	[0.00000]	[0.66049]	[0.00000]	[0.19096]
Observations	256,253	256,253	106,846	106,846	149,407	149,407

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	0.246***	0.049	0.198**
	(0.064)	(0.063)	(0.089)
	[0.172]	[0.182]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.084**	0.008	0.076
	(0.038)	(0.040)	(0.055)
	[0.077]	[0.081]	–
Property Crime	0.160***	0.031	0.129*
	(0.054)	(0.043)	(0.067)
	[0.065]	[0.068]	–
Violent Crime	0.095***	0.002	0.093*
	(0.035)	(0.039)	(0.053)
	[0.047]	[0.071]	–
Observations	106,846	149,407	–

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	1.111**	0.128	0.983**	0.071*	0.027	0.044
	(0.450)	(0.153)	(0.478)	(0.037)	(0.036)	(0.051)
	[0.261]	[0.267]	–	[0.091]	[0.092]	–
Panel B: Rearrest by Crime Type						
Drug Crime	0.234	0.008	0.226	0.048***	0.021	0.027
	(0.200)	(0.104)	(0.221)	(0.018)	(0.021)	(0.027)
	[0.123]	[0.128]	–	[0.035]	[0.030]	–
Property Crime	0.713*	0.095	0.618	0.053**	0.011	0.042
	(0.365)	(0.111)	(0.377)	(0.023)	(0.023)	(0.033)
	[0.108]	[0.106]	–	[0.026]	[0.027]	–
Violent Crime	0.454**	−0.019	0.473**	0.021	0.015	0.007
	(0.212)	(0.100)	(0.237)	(0.021)	(0.021)	(0.031)
	[0.069]	[0.100]	–	[0.028]	[0.039]	–
Observations	50,784	77,342	–	56,062	72,065	–

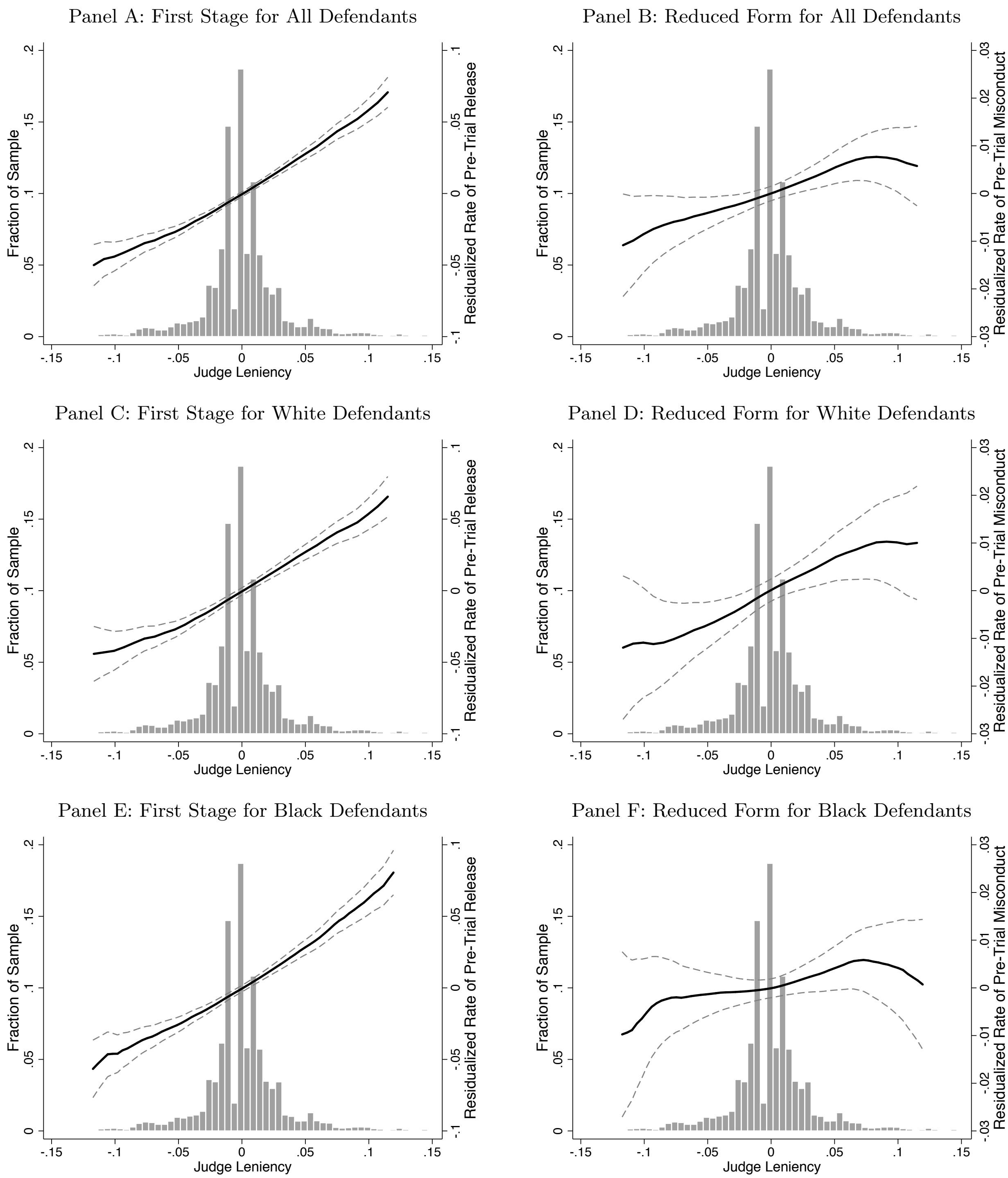
Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately for high-risk and low-risk defendants. Predicted risk is calculated using the machine learning algorithm described in Appendix F. The regressions are estimated 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
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Rearrest for All Crimes						
Rearrest Prior to Disposition	0.232**	0.011	0.221	0.438*	0.128	0.310
	(0.104)	(0.169)	(0.200)	(0.224)	(0.156)	(0.236)
	[0.149]	[0.194]	–	[0.148]	[0.152]	–
Panel B: Rearrest by Crime Type						
Drug Crime	0.080	0.046	0.034	0.136	0.048	0.087
	(0.063)	(0.126)	(0.142)	(0.102)	(0.078)	(0.110)
	[0.057]	[0.092]	–	[0.057]	[0.057]	–
Property Crime	0.168**	−0.070	0.238*	0.311*	0.097	0.215
	(0.078)	(0.101)	(0.129)	(0.163)	(0.113)	(0.169)
	[0.078]	[0.060]	–	[0.078]	[0.079]	–
Violent Crime	0.116*	−0.005	0.121	0.215*	0.065	0.150
	(0.061)	(0.106)	(0.124)	(0.119)	(0.085)	(0.128)
	[0.050]	[0.067]	–	[0.048]	[0.051]	–
Observations	93,417	162,836	–	47,692	45,725	–

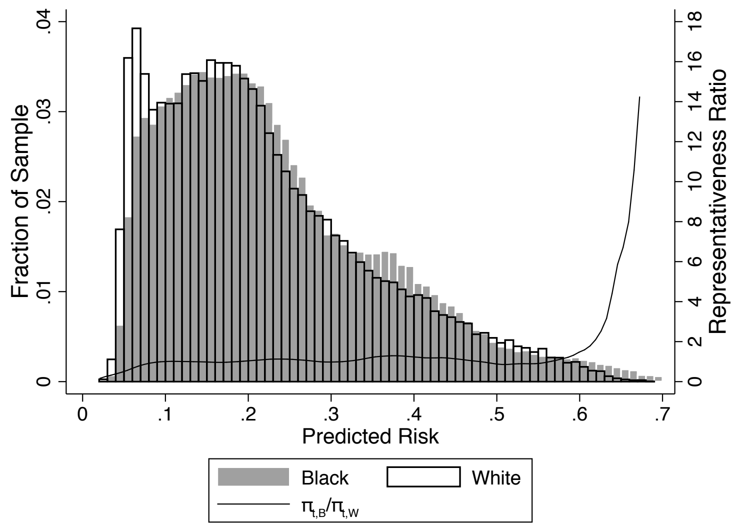
Note: This table reports two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct by judge specialization and judge experience. Columns 1-3 report estimates for non-specialist bail judges in Miami-Dade and specialist bail judges in Philadelphia. Columns 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. 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.

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 release 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: Predicted Risk Distribution by Defendant Race



Note: This figure reports the distribution of pre-trial misconduct risk separately for black and white defendants. Pre-trial misconduct risk is computed using a machine learning algorithm described in Appendix F. The solid line represents the representativeness ratio for black versus white defendants. See the text for additional details.

Appendix A: Additional Results

Appendix Table A1: Non-Monetary Bail Assigned and Criminal Outcomes

	White	Black	Difference
Panel A: Pre-Trial Release	(1)	(2)	(3)
Pre-trial Release	0.445***	0.476***	−0.031
	(0.037)	(0.037)	(0.048)
	[0.711]	[0.688]	–
Panel B: Rearrest			
Rearrest Prior to Disposition	0.102***	0.020	0.082*
	(0.034)	(0.033)	(0.048)
	[0.172]	[0.182]	–
Drug Crime	0.052**	−0.005	0.056*
	(0.022)	(0.022)	(0.031)
	[0.077]	[0.081]	–
Property Crime	0.067***	0.011	0.056
	(0.026)	(0.023)	(0.035)
	[0.065]	[0.068]	–
Violent Crime	0.060***	−0.002	0.062**
	(0.020)	(0.022)	(0.030)
	[0.047]	[0.071]	–
Observations	106,846	149,407	–

Note: This table reports two-stage least squares results of the impact of any form of non-monetary bail on the probability of release (Panel A) and probability of rearrest (Panel B) 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: White-Hispanic Results

	White	Hispanic	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.219***	0.340***	−0.121
	(0.077)	(0.121)	(0.141)
	[0.167]	[0.176]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.055	0.142**	−0.087
	(0.053)	(0.069)	(0.087)
	[0.066]	[0.087]	–
Property Crime	0.150**	0.226**	−0.076
	(0.059)	(0.107)	(0.121)
	[0.066]	[0.064]	–
Violent Crime	0.068	0.152**	−0.084
	(0.043)	(0.066)	(0.079)
	[0.045]	[0.052]	–
Observations	35,752	47,676	–

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. We further restrict to individuals who are matched to census genealogical records of surnames. 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 A3: 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	0.317	0.339	1.072	0.331	0.256	0.774
	(0.001)	(0.027)	(0.084)	(0.001)	(0.023)	(0.069)
NonDrug	0.683	0.661	0.967	0.669	0.744	1.112
	(0.001)	(0.027)	(0.039)	(0.001)	(0.023)	(0.034)
Violent	0.252	0.015	0.059	0.316	0.077	0.242
	(0.001)	(0.020)	(0.079)	(0.001)	(0.024)	(0.076)
NonViolent	0.748	0.985	1.318	0.684	0.923	1.350
	(0.001)	(0.020)	(0.027)	(0.001)	(0.024)	(0.035)
Felony	0.456	0.143	0.313	0.529	0.255	0.481
	(0.002)	(0.025)	(0.055)	(0.001)	(0.025)	(0.046)
NonFelony	0.544	0.857	1.576	0.471	0.745	1.584
	(0.002)	(0.025)	(0.046)	(0.001)	(0.025)	(0.052)
Prior	0.327	0.306	0.934	0.351	0.389	1.108
	(0.001)	(0.023)	(0.070)	(0.001)	(0.021)	(0.060)
NonPrior	0.673	0.694	1.032	0.649	0.611	0.942
	(0.001)	(0.023)	(0.034)	(0.001)	(0.021)	(0.032)
High Risk	0.475	0.206	0.434	0.518	0.321	0.621
	(0.002)	(0.024)	(0.051)	(0.001)	(0.023)	(0.044)
Low Risk	0.525	0.794	1.513	0.482	0.679	1.407
	(0.002)	(0.024)	(0.040)	(0.001)	(0.023)	(0.047)

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 A4: 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	0.837***	0.230***	0.699***	0.452***	0.137***	0.597***	0.476***
	(0.052)	(0.044)	(0.058)	(0.060)	(0.049)	(0.046)	(0.033)
	[0.721]	[0.674]	[0.607]	[0.785]	[0.685]	[0.587]	[0.587]
Court x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crime Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	128,409	127,844	55,432	83,277	74,193	87,424	168,829

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 A5: Subsample Results

	Felony	Misd.	Prior Offense	No Prior Offendse	Prior Rearrest	No Prior Rearrest	Low Income	High Income
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rearrest Prior to Disposition	1.011**	−0.012	0.258*	0.136	0.392*	0.172*	0.269*	−0.063
	(0.472)	(0.048)	(0.145)	(0.112)	(0.209)	(0.101)	(0.141)	(0.175)
Panel B: Rearrest by Crime Type								
Drug Crime	0.373	0.001	0.092	0.054	0.180	0.037	0.083	0.093
	(0.250)	(0.026)	(0.091)	(0.069)	(0.125)	(0.062)	(0.081)	(0.122)
Property Crime	0.438	0.047	0.138	0.089	0.181	0.119*	0.214**	−0.043
	(0.335)	(0.031)	(0.120)	(0.075)	(0.178)	(0.070)	(0.105)	(0.134)
Violent Crime	0.435*	0.002	0.178**	0.032	0.310**	0.037	0.155*	−0.005
	(0.239)	(0.028)	(0.087)	(0.068)	(0.125)	(0.063)	(0.083)	(0.118)
Observations	127,844	128,409	87,424	168,829	53,309	202,944	174,233	45,057

Note: This table reports the difference in two-stage least squares estimates of the impact of pre-trial release on the probability of pre-trial misconduct between white and black defendants by different subsamples. 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.

Appendix Table A6: Pre-trial Release and Criminal Outcomes

	White	Black	Difference
Panel A: FTA or Rearrest for All Crimes	(1)	(2)	(3)
Pre-trial Misconduct	0.277***	0.103	0.174*
	(0.066)	(0.064)	(0.091)
	[0.242]	[0.267]	–
Panel B: FTA or Rearrest by Crime Type			
FTA or Drug Crime	0.133***	0.073*	0.060
	(0.043)	(0.043)	(0.060)
	[0.170]	[0.191]	–
FTA or Property Crime	0.201***	0.092*	0.110
	(0.058)	(0.047)	(0.072)
	[0.171]	[0.187]	–
FTA or Violent Crime	0.134***	0.056	0.078
	(0.040)	(0.043)	(0.059)
	[0.158]	[0.189]	–
Observations	106,846	149,407	–

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, where pre-trial misconduct equals one if the defendant fails to appear in court or is rearrested prior to disposition. 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: Social Cost of Crime Results

	D^{IV} Estimate	Lower Bound	Upper Bound
	(1)	(2)	(3)
Rearrest for Murder	0.004	\$4,301,817	\$11,559,713
	(0.008)		
Rearrest for Rape	−0.024**	\$187,680	\$343,859
	(0.011)		
Rearrest for Robbery	0.049	\$73,196	\$333,701
	(0.031)		
Rearrest for Assault	0.086*	\$41,046	\$109,903
	(0.045)		
Rearrest for Burglary	0.042	\$50,291	\$50,291
	(0.048)		
Rearrest for Theft	0.101*	\$9,598	\$9,974
	(0.057)		
Rearrest for Drug	0.076	\$2,544	\$2,544
	(0.055)		
Rearrest for DUI	0.008	\$25,842	\$25,842
	(0.008)		

Note: This table reports the difference in two-stage least squares estimates of the impact of pre-trial release on the probability of pre-trial misconduct between white and black defendants for different crimes. 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.

Appendix Table A8: Robustness Results

	Drop Impossible	Re-Weight Char.	Drop Hispanics	Cluster Judge Level	Cluster Bootstrap	JIVE	LIML	Control Bail $	Race-Specific Leniency
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rearrest Prior to Disposition	0.163**	0.211**	0.171*	0.198**	0.198*	0.173*	0.156	0.212**	0.364**
	(0.082)	(0.091)	(0.096)	(0.092)	(0.111)	(0.100)	(0.114)	(0.095)	(0.159)
Panel B: Rearrest by Crime Type									
Drug Crime	0.088*	0.087*	0.047	0.076	0.076	0.031	0.009	0.076	0.039
	(0.053)	(0.052)	(0.066)	(0.052)	(0.069)	(0.067)	(0.060)	(0.058)	(0.067)
Property Crime	0.067	0.139**	0.119*	0.129*	0.129*	0.092	0.101	0.162**	0.315**
	(0.057)	(0.068)	(0.068)	(0.066)	(0.089)	(0.075)	(0.095)	(0.071)	(0.151)
Violent Crime	0.073	0.110**	0.067	0.093	0.093	0.116*	0.103*	0.091*	0.090
	(0.047)	(0.054)	(0.058)	(0.061)	(0.066)	(0.064)	(0.058)	(0.055)	(0.065)
Observations	252,992	256,253	184,876	256,253	256,253	256,253	256,253	256,253	256,253

Note: This table reports the difference in two-stage least squares estimates of the impact of pre-trial release on the probability of pre-trial misconduct between white and black defendants. Column 1 drops cases where defendants are reported as being detained but are rearrested prior to disposition. Column 2 re-weights cases so that the white and black samples have identical observable characteristics. Column 3 drops Hispanic whites from the sample. Column 4 reports standard errors clustered by defendant and judge. Column 5 reports bootstrap standard errors clustered at the judge shift level using 500 replications. Column 6 reports estimates using jackknife instrumental variables (JIVE), instrumenting for pre-trial release using a full set of judge indicators and judge indicators interacted with race. Column 7 reports estimates limited-information maximum likelihood (LIML), instrumenting for pre-trial release using a full set of judge indicators and judge indicators interacted with race. Column 8 instruments for monetary bail amount with a leave-out leniency measure constructed using bail amounts. Column 9 reports estimates using a race-specific leave-out leniency measure to instrument for pre-trial release. All regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. 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: OLS Results

	White	Black	Difference
Panel A: Rearrest for All Crimes	(1)	(2)	(3)
Rearrest Prior to Disposition	0.181***	0.188***	−0.007**
	(0.003)	(0.002)	(0.004)
	[0.172]	[0.182]	–
Panel B: Rearrest by Crime Type			
Drug Crime	0.097***	0.103***	−0.006**
	(0.002)	(0.002)	(0.002)
	[0.077]	[0.081]	–
Property Crime	0.067***	0.073***	−0.006*
	(0.002)	(0.002)	(0.003)
	[0.065]	[0.068]	–
Violent Crime	0.052***	0.063***	−0.010***
	(0.002)	(0.002)	(0.002)
	[0.047]	[0.071]	–
Observations	106,846	149,407	–

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 A10: Pre-Trial Release and Pre-Trial Misconduct by Judge and Defendant Race

	Race of Judge	
	White	Black
Panel A: Release Rates	(1)	(2)
White	0.557	0.552
	(0.497)	(0.497)
Black	0.535	0.530
	(0.499)	(0.499)
Panel B: Pre-Trial Rearrest Rates		
White	0.207	0.202
	(0.405)	(0.402)
Black	0.280	0.294
	(0.449)	(0.456)

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 A11: p-values from Tests of Relative Racial Prejudice

	p-Value
	(1)
Pre-Trial Release	0.782
Pre-Trial Rearrest	0.580

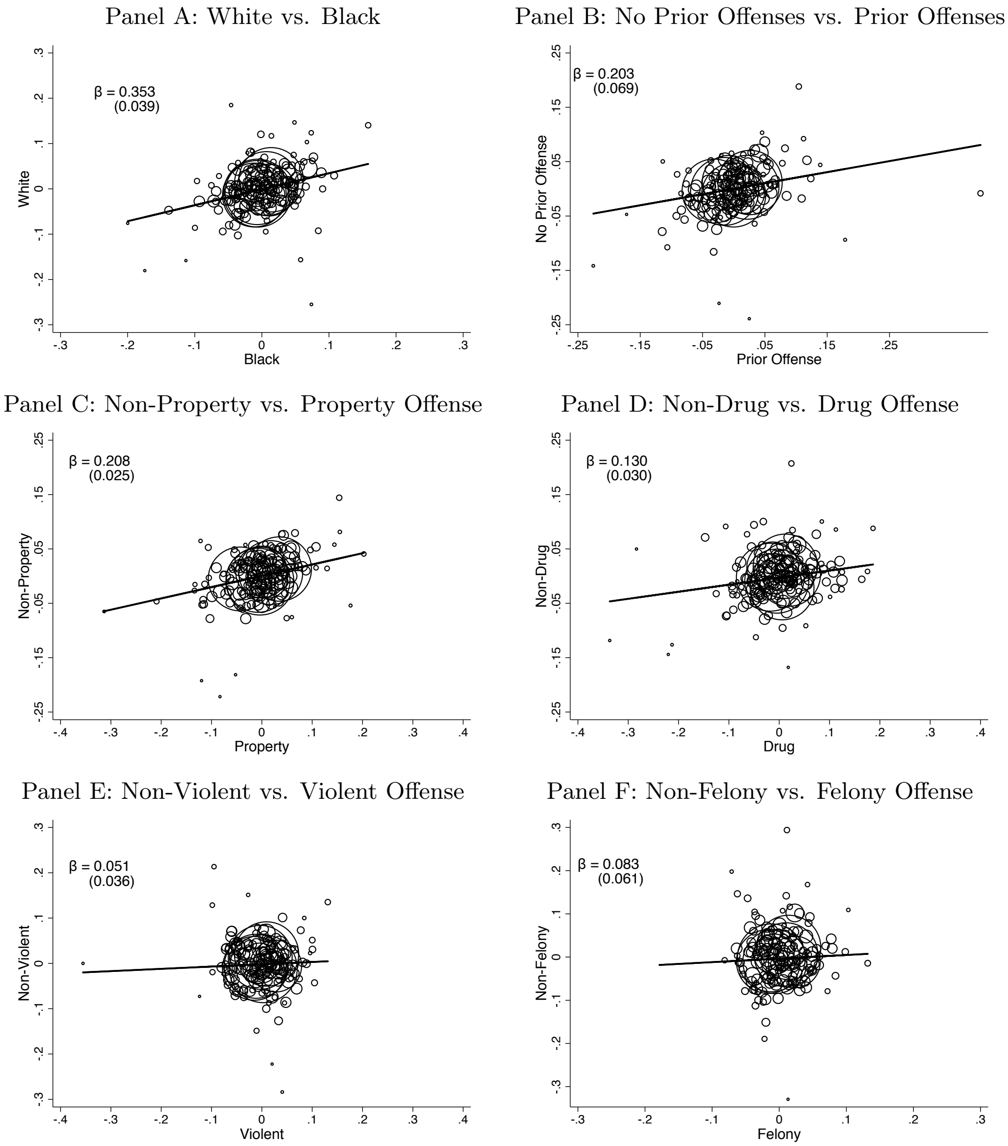
Note: This table replicates the Anwar and Fang (2006) test for pre-trial release rates and pre-trial misconduct rates. This table presents bootstrapped p-values testing for relative racial bias. The null hypothesis is rejected if white judges are more lenient on white defendants, and black judges are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

Appendix Table A12: Representativeness Statistics

	$\mathbb{E}(x\|Black)/\mathbb{E}(x\|White)$
Panel A: Defendant Characteristics	(1)
Male	1.026
Age at Bail Decision	0.978
Prior Offense in Past Year	1.072
Arrested on Bail in Past Year	1.048
Failed to Appear in Court in Past Year	1.028
Panel B: Charge Characteristics	
Number of Offenses	1.200
Felony Offense	1.160
Misdemeanor Only	0.866
Any Drug Offense	1.077
Any DUI Offense	0.839
Any Violent Offense	1.260
Any Property Offense	0.983
Panel C: Outcomes	
Rearrest Prior to Disposition	1.061
Drug Crime	1.059
Property Crime	1.044
Violent Crime	1.496
Failure to Appear in Court	0.983
Failure to Appear in Court or Rearrested	1.102
Observations	256,253

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.

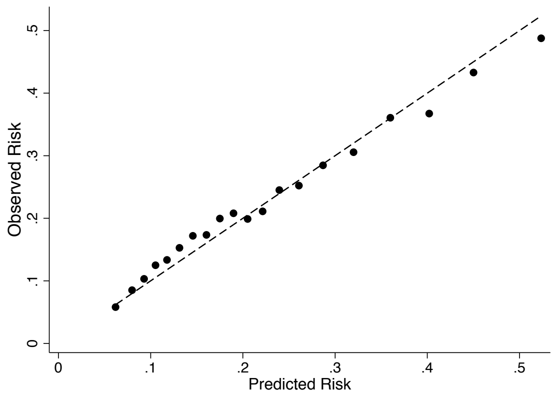
Appendix Figure A1: Judge Leniency by Defendant and Case Characteristics



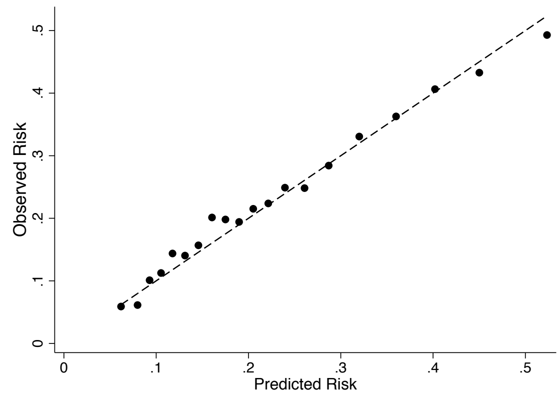
Note: These figures show the correlation between our residualized measure of judge leniency for different groups of defendants over all available years of data. We also plot the linear best fit line estimated using OLS.

Appendix Figure A2: Predicted and Actual Risk by Judge Leniency

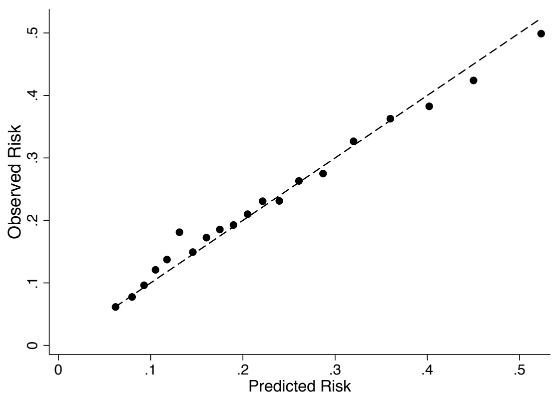
Panel A: First Quintile of Judge Leniency



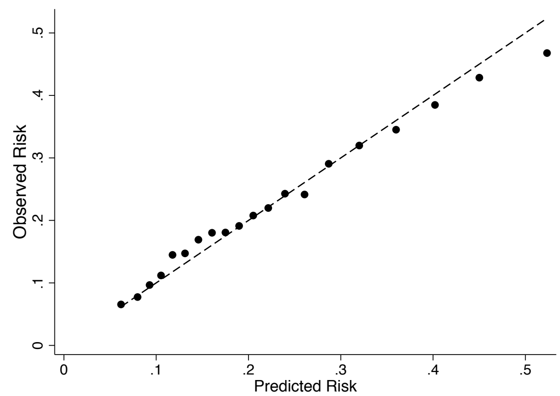
Panel B: Second Quintile of Judge Leniency



Panel C: Third Quintile of Judge Leniency

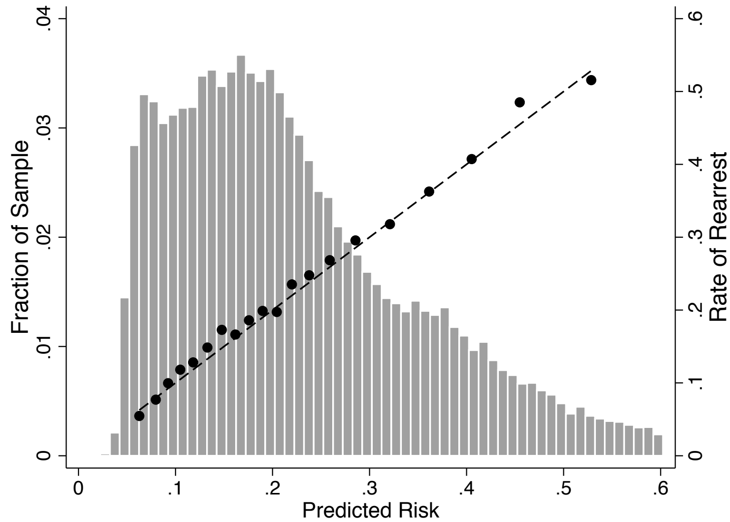


Panel D: Fourth Quintile of Judge Leniency



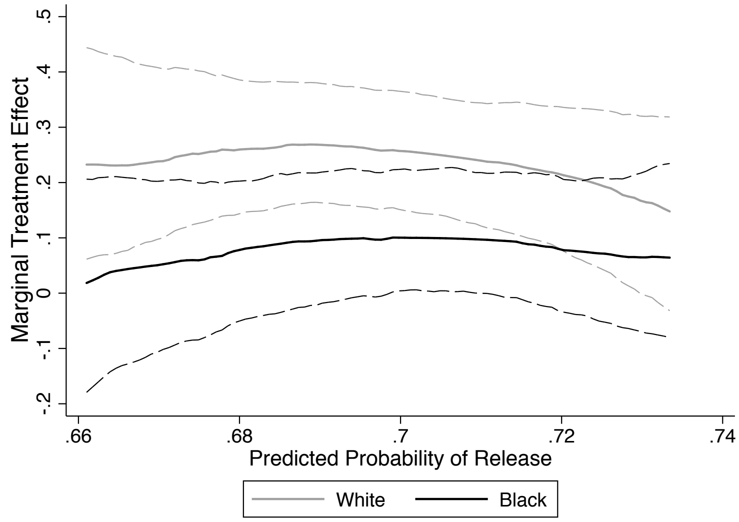
Note: These figures plot predicted pre-trial misconduct risk against actual pre-trial misconduct for different judge-leniency quintiles. Predicted risk is calculated using only cases from the most lenient quintile of judges and the machine learning algorithm described in Appendix F. See the text for additional details.

Appendix Figure A3: Relationship between Predicted Risk and True Risk



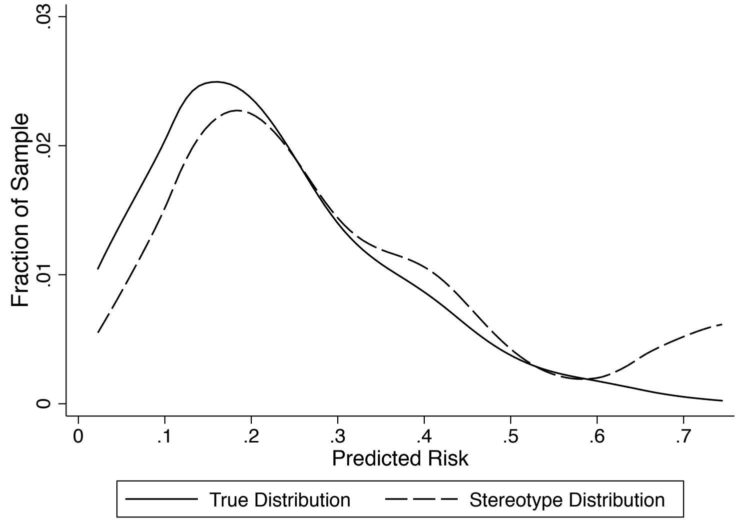
Note: This figure reports the distribution of the pre-trial misconduct risk and plots the predicted pre-trial misconduct risk against actual pre-trial misconduct for the test sample. Predicted risk is calculated using the machine learning algorithm described in Appendix F. The dashed line is the 45 degree line. See the text and for additional details.

Appendix Figure A4: Marginal Treatment Effects by Defendant Race



Note: This figure displays the estimated marginal treatment effects (MTEs) of release on rearrest separately for white and black defendants. To estimate each MTE, we first estimate the predicted probability of release using only judge leniency. We then estimate the relationship between the predicted probability of release and rearrest prior to disposition using a local quadratic estimator (bandwidth = 0.045). Finally, we use the numerical derivative of the local quadratic estimator to calculate the MTE at each point in the distribution. Standard errors are computed using 500 bootstrap replications. See the text for additional details.

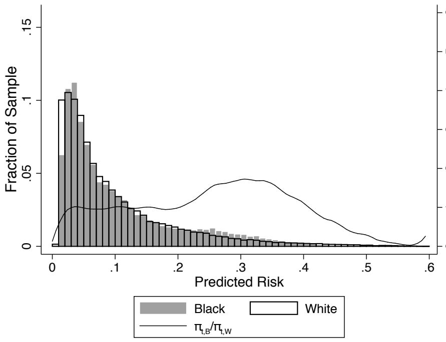
Appendix Figure A5: Stereotyped and True Distribution of Risk for Black Defendants



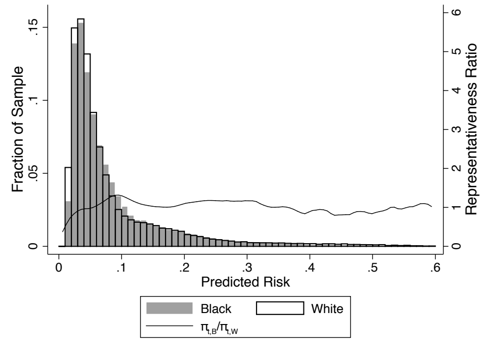
Note: This figure plots the true distribution of risk for black defendants alongside the perceived distribution of risk for black defendants. The stereotyped beliefs are generated by representativeness-based discounting model with $\theta = 1.9$. This value of θ rationalizes an average release rate of black defendants equal to 68.8 percent, the actual rate of release in the data. See the text and Appendix F for additional details.

Appendix Figure A6: Crime-Specific Predicted Risk Distributions by Race

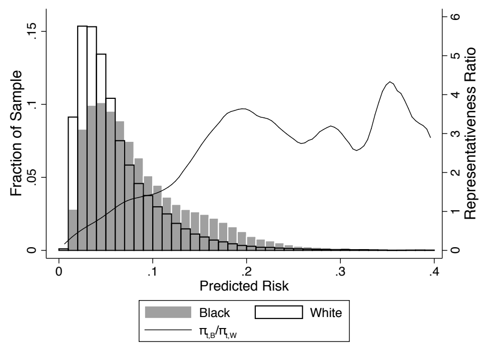
Panel A: Rearrest for Drug Crime



Panel B: Rearrest for Property Crime

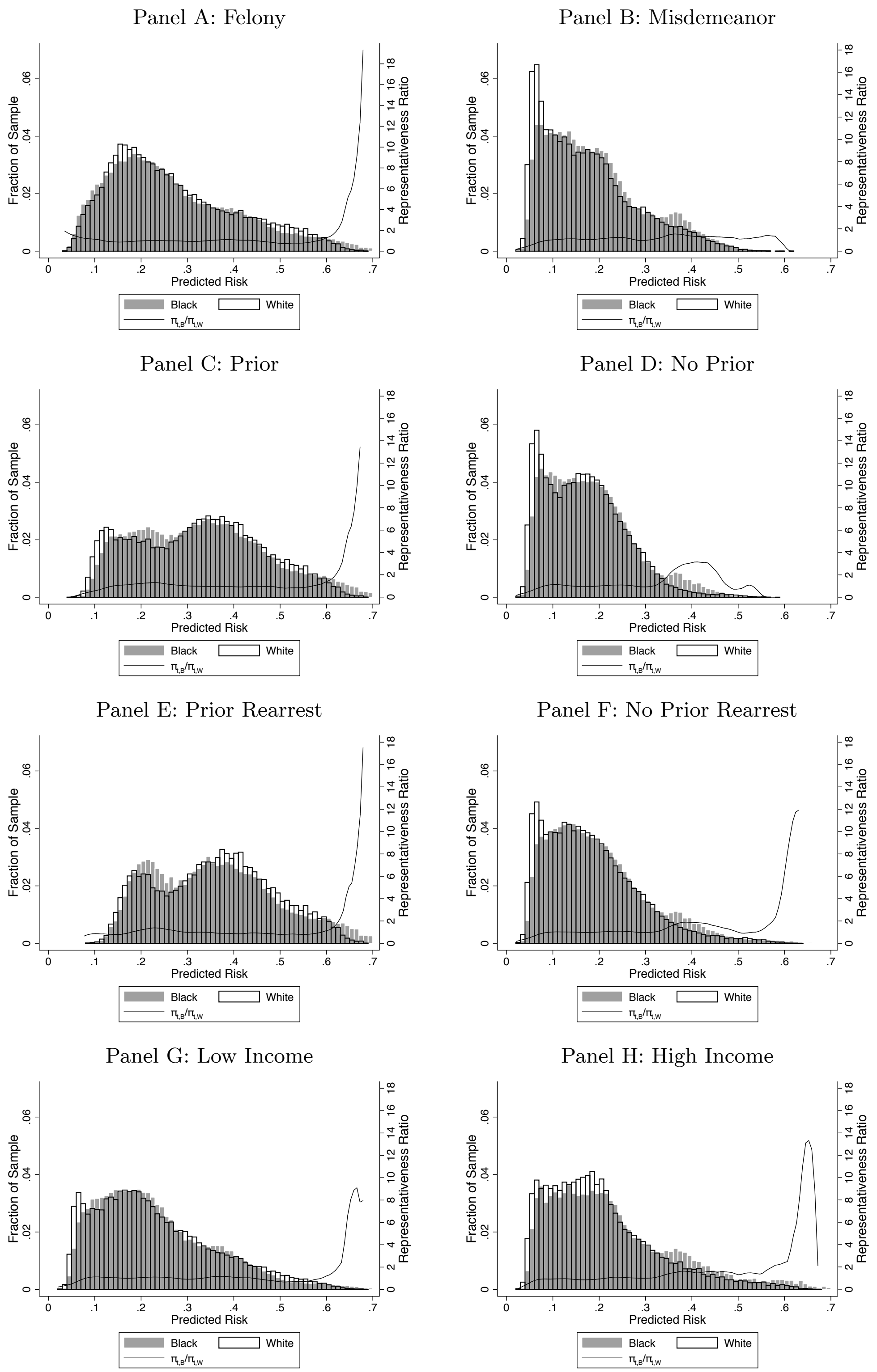


Panel C: Rearrest for Violent Crime



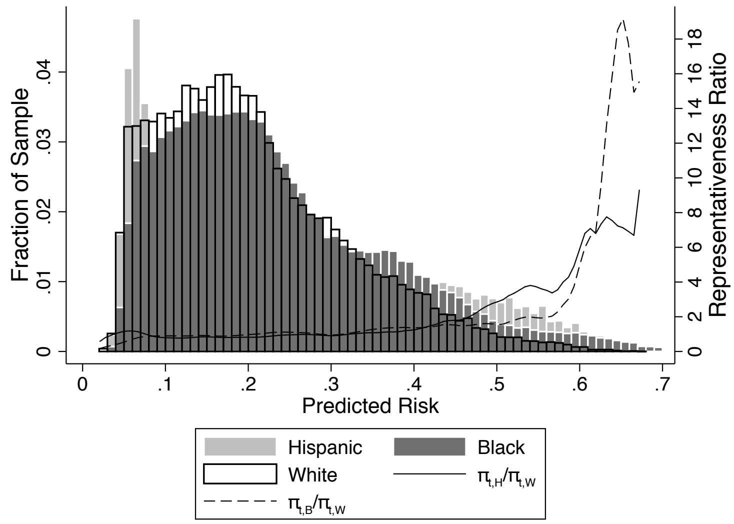
Note: These figures report the distribution of crime-specific risk separately by defendant race. Predicted risk is calculated using the machine learning algorithm described in Appendix F. The solid line in each figure represents the representativeness ratio for black versus white defendants. See the text for additional details.

Appendix Figure A7: Subgroup-Specific Predicted Risk Distributions by Race



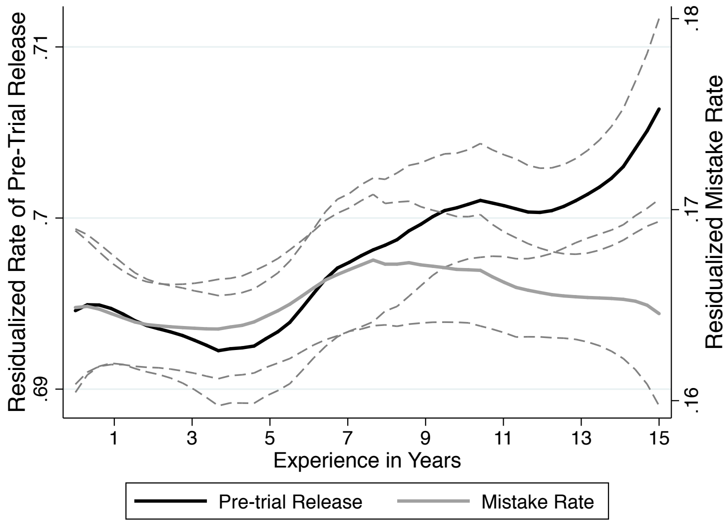
Note: These figures report the distribution of subgroup-specific risk separately by defendant race. Predicted risk is calculated using the machine learning algorithm described in Appendix F. The solid line in each figure represents the representativeness ratio for black versus white defendants. See the text for additional details.

Appendix Figure A8: Predicted Risk Distribution by Hispanic and Black versus White



Note: This figure reports the distribution of the risk of pre-trial misconduct separately by Hispanic, black, and white defendants. Predicted risk is calculated using the machine learning algorithm described in Appendix F. The dashed line represents the representativeness ratio for black versus white defendants and the solid line represents the representativeness ratio for Hispanic versus white defendants. See the text for additional details.

Appendix Figure A9: 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 pre-trial crime conditional on release (i.e. the mistake rate). Pre-trial release and pre-trial rearrest are both residualized using the full set of court-by-time fixed effects. See the text for additional details.

Appendix B: Proofs of Consistency of IV Estimator

A. Conditions for Consistency

Building on the standard IV framework, we now establish the two conditions under which our IV estimator for racial bias D^{IV} provides a consistent estimate of D^*.

First Condition for Consistency: 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 $|z_j - z_k| < \epsilon$.

Proposition B.1. 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.

Proposition B.1 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. $\square$

Second Condition for Consistency: 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 (8), 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:

$$D^{IV} = \sum_{j=1}^{J} \lambda^j (\alpha_W^{j,j-1} - \alpha_B^{j,j-1}) \tag{B.4}$$

Proof of Consistency: Proposition 3 in the main text combines these two conditions to establish the consistency of our IV estimator. Proposition 3 states that our IV estimator D^{IV} provides a consistent estimate of racial bias D^* if (1) Z_i is continuous and (2) λ_r^j is constant by race. 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:

$$\frac{Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)}{Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B)} = c \tag{B.5}$$

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.5) 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:

$$D^{IV} = \alpha_W^{IV} - \alpha_B^{IV} = \sum_{j=1}^{J} \lambda_W^j \alpha_W^{j,j-1} - \sum_{j=1}^{J} \lambda_B^j \alpha_B^{j,j-1} \tag{B.6}$$

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

$$D^{IV} = \sum_{j=1}^{J} \lambda^j \left(\alpha_W^{j,j-1} - \alpha_B^{j,j-1} \right) \tag{B.7}$$

Following Proposition B.1, as Z_i becomes continuously distributed, we can rewrite D^{IV} as:

$$D^{IV} = \int \lambda(z) \left(\alpha_W(z) - \alpha_B(z) \right) dz = D^* \tag{B.8}$$

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:

$$\frac{Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)}{Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B)} = c \tag{B.9}$$

where c is some constant.

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

$$\lambda_r^j = \frac{(P(Released|z_j, r) - P(Released|z_{j-1}, r)) \cdot \sum_{l=j}^{J} \pi_r^l (g(z_l) - \mathbb{E}[g(Z)])}{\sum_{m=1}^{J} (P(Released|z_m, r) - P(Released|z_{m-1}, r)) \cdot \sum_{l=m}^{J} \pi_r^l (g(z_l) - \mathbb{E}[g(Z)])} \tag{B.10}$$

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.5) 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.5).

Next, we prove that if the IV weights are constant by race, then Equation (B.5) holds. To do so, we prove the contrapositive statement. Suppose Equation (B.5) does not hold, so that there exists z_j and z_k such that:

$$\frac{Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)}{Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B)} = c_1 \tag{B.11}$$

$$\frac{Pr(Released|z_k, r = W) - Pr(Released|z_{k-1}, r = W)}{Pr(Released|z_k, r = B) - Pr(Released|z_{k-1}, r = B)} = c_2 \tag{B.12}$$

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:

$$\frac{\lambda_W^j}{\lambda_B^j} = \frac{1}{c_1} \frac{D_B}{D_W} \tag{B.13}$$

while

$$\frac{\lambda_W^k}{\lambda_B^k} = \frac{1}{c_2} \frac{D_B}{D_W} \tag{B.14}$$

where Equation (B.13) and Equation (B.14) follow by substituting Equation (B.11) and Equation (B.12) 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. $\square$

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 be given by a linear probability model, estimated separately by defendant race:

$$Released_i = \gamma_W Z_i + \pi_W \mathbf{X}_i + v_i \tag{B.15}$$

$$Released_i = \gamma_B Z_i + \pi_B \mathbf{X}_i + v_i \tag{B.16}$$

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

$$Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r) = \gamma_r(z_j - z_{j-1}) \tag{B.17}$$

Then, it is straightforward to show:

$$\frac{Pr(Released|z_j, r = W) - Pr(Released|z_{j-1}, r = W)}{Pr(Released|z_j, r = B) - Pr(Released|z_{j-1}, r = B)} = \frac{\gamma_W(z_j - z_{j-1})}{\gamma_B(z_j - z_{j-1})} = \frac{\gamma_W}{\gamma_B} \tag{B.18}$$

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

B. 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, D^*. As a result, 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 proceed, we map our model of racial bias to the MTE framework developed by Heckman and Vytlacil (1999, 2005). In Appendix E, we formally establish that the treatment effect for defendants on the margin of release to judge j (α_r^j) is equal to the MTE function evaluated at the propensity score of judge j:

$$\alpha_r^j = MTE_r(p_r^j) \tag{B.19}$$

where p_r^j denotes the probability judge j releases a defendant of race r. 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. 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:

$$MTE_r(u) \approx \sum_{j=1}^{J} \mathbb{1}\{u \in [p_r^{j-1}, p_r^j]\}[\theta_r^{1,j} + \theta_r^{2,j} u] \tag{B.20}$$

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 B.1. 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. (2017) 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 IV estimator for racial bias.

To show why Equation (B.20) 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:

$$LATE_r^{0,1} = \frac{1}{p_r^1 - p_r^0} \int_{p_r^0}^{p_r^1} MTE_r(u)du \tag{B.21}$$

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.20) 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

$$\tilde{\lambda}^j = \begin{cases} \frac{\lambda^1}{2} & j = 0 \\ \frac{\lambda^j + \lambda^{j+1}}{2} & j \in [1, J-1] \\ \frac{\lambda^J}{2} & j = J \end{cases} \tag{B.22}$$

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.20), D^{IV} is a consistent estimate of D^{**}, which is a weighted average of true racial bias.

C. Bounding Maximum Bias of IV Estimator with Discrete Instrument

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 B.2. If Proposition B.1 and Proposition 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:

$$\lambda^j = \frac{(z_j - z_{j-1}) \cdot \sum_{l=j}^{J} \pi^l (z_l - \mathbb{E}[Z])}{\sum_{m=1}^{J} (z_j - z_{j-1}) \cdot \sum_{l=m}^{J} \pi^l (z_l - \mathbb{E}[Z])} \tag{B.23}$$

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 show that the true level of racial bias is equal to D^{IV} plus a bias term, which we refer to as "infra-marginality bias." Second, we derive an upper bound for the bias term by replacing $\alpha_W^{j,j-1}$ with its minimum possible value for every judge j, and we derive a lower bound by 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.

Before proceeding to the proof, we note that a linear first stage is necessary in order to derive a closed-form solution of maximum bias with our IV strategy. See Appendix E for an alternative procedure that estimates upper and lower bounds on true racial bias which does not depend on a linear first stage.

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:

$$\alpha_r^{j,j-1} \in (t_r^{j-1}, t_r^j] \tag{B.24}$$

Note that we can rewrite D^* as:

$$\begin{aligned}
D^* &= \sum_{j=1}^{J} \lambda^j \left(t_W^j - t_B^j\right) \\
&= \sum_{j=1}^{J} \lambda^j \left(\alpha_W^{j,j-1} - \alpha_B^{j,j-1}\right) + \sum_{j=1}^{J} \lambda^j \left(t_W^j - \alpha_W^{j,j-1}\right) + \sum_{j=1}^{J} \lambda^j \left(\alpha_B^{j,j-1} - t_B^j\right) \\
&= D^{IV} + \underbrace{\sum_{j=1}^{J} \lambda^j \left(t_W^j - \alpha_W^{j,j-1}\right) + \sum_{j=1}^{J} \lambda^j \left(\alpha_B^{j,j-1} - t_B^j\right)}_{\text{infra-marginality bias}}
\end{aligned} \tag{B.25}$$

The second line follows from adding and subtracting $\sum_{j=1}^{J} \lambda^j \alpha_W^{j,j-1}$ and $\sum_{j=1}^{J} \lambda^j \alpha_B^{j,j-1}$ to D^* and rearranging terms. The third line follows from assuming equal IV weights by race. Equation B.25 shows that the true level of racial bias is equal to D^{IV} plus a bias term, which we refer to as "infra-marginality bias."

We will now derive an upper bound for D^*. First, note that Equation (B.24) implies $\alpha_B^{j,j-1} \leq t_B^j$. Therefore $\sum_{j=1}^{J} \lambda^j \left(\alpha_B^{j,j-1} - t_B^j\right) \leq 0$, given $\lambda^j \geq 0$ for all j. We can drop this term from Equation (B.25) to obtain an upper bound on D^*:

$$\begin{aligned}
D^* &\leq D^{IV} + \sum_{j=1}^{J} \lambda^j \left(t_W^j - \alpha_W^{j,j-1}\right) \\
&< D^{IV} + \sum_{j=1}^{J} \lambda^j \left(t_W^j - t_W^{j-1}\right)
\end{aligned} \tag{B.26}$$

where the second line follows from Equation (B.24) ($t_W^{j-1} < \alpha_W^{j,j-1}$).

Using similar logic, we can also derive a lower bound for D^*. Equation (B.24) implies $t_W^j \geq \alpha_W^{j,j-1}$. Therefore $\sum_{j=1}^{J} \lambda^j \left(t_W^j - \alpha_W^{j,j-1}\right) \geq 0$, given $\lambda^j \geq 0$ for all j. We can drop this term from

Equation (B.25) to obtain a lower bound on D^*:

$$\begin{aligned}
D^* &\geq D^{IV} + \sum_{j=1}^{J} \lambda^j \left(\alpha_B^{j,j-1} - t_B^j \right) \\
&= D^{IV} - \sum_{j=1}^{J} \lambda^j \left(t_B^j - \alpha_B^{j,j-1} \right) \\
&> D^{IV} - \sum_{j=1}^{J} \lambda^j \left(t_B^j - t_B^{j-1} \right) \qquad \text{(B.27)}
\end{aligned}$$

where again, the last line follows from Equation (B.24) ($t_B^{j-1} < \alpha_B^{j,j-1}$).

We can now bound D^* using Equation (B.27) and Equation (B.26):

$$D^{IV} - \sum_{j=1}^{J} \lambda^j \left(t_B^j - t_B^{j-1} \right) < D^* < D^{IV} + \sum_{j=1}^{J} \lambda^j \left(t_W^j - t_W^{j-1} \right) \qquad \text{(B.28)}$$

It is straightforward to see that the infra-marginality bias goes to zero as Z_i becomes continuous. Given that λ^j are non-negative weights which sum to one, $\sum_{j=1}^{J} \lambda^j \left(t_r^j - t_r^{j-1} \right) \leq \max_j (t_r^j - t_r^{j-1})$ (i.e. the average is less than the maximum). Therefore, if Z_i becomes continuous, then $t_r^j - t_r^{j-1} \to 0$ 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.

Note that $t_r^j - t_r^{j-1}$ is positive for all j, implying $\sum_{j=1}^{J} \lambda^j \left(t_r^j - t_r^{j-1} \right) \leq \max_j(\lambda^j) \sum_{j=1}^{J} \left(t_r^j - t_r^{j-1} \right)$, where $\max_j(\lambda^j)$ is the maximum weight across all judges. Given the recursive structure of $\sum_{j=1}^{J} \left(t_r^j - t_r^{j-1} \right)$:

$$\max_j(\lambda^j) \sum_{j=1}^{J} \left(t_r^j - t_r^{j-1} \right) = \max_j(\lambda^j)(t_r^J - t_r^0) \qquad \text{(B.29)}$$

Note that $t_r^J = \alpha_r^{max}$ (i.e. the largest treatment effect is associated with the most lenient judge) and $t_r^0 = \alpha_r^{min}$ (i.e. the smallest treatment effect is associated with the most strict judge). Therefore, letting α^{max} and α^{min} equal the maximum treatment effect and minimum treatment effect respectively across races, yields:

$$D^{IV} - \max_j(\lambda^j)(\alpha^{max} - \alpha^{min}) < D^* < D^{IV} + \max_j(\lambda^j)(\alpha^{max} - \alpha^{min}) \qquad \text{(B.30)}$$

which proves Proposition B.2. In other words, the maximum bias of our IV estimator D^{IV} from the true level of racial bias D^* is given by $\max_j(\lambda^j)(\alpha^{max} - \alpha^{min})$. □

Next, we simplify these bounds to retrieve estimable bounds. Note that $\alpha^{max} \leq 1$ and $\alpha^{min} \geq 0$ in theory, which implies $(\alpha^{max} - \alpha^{min}) \leq 1$. Therefore, the bounds in Equation (B.30) can be re-written as:

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

Rearranging terms yields:

$$-\max_j(\lambda^j) < D^* - D^{IV} < \max_j(\lambda^j) \tag{B.32}$$

Under this worst-case assumption, the maximum bias of our IV estimator D^{IV} from the true level of racial bias D^* is given by $\max_j(\lambda^j)$.

Intuition of Maximum Bias Formula: Under Proposition B.2, the maximum bias of D^{IV} relative to D^* decreases as (1) the heterogeneity in treatment effects among compliers decreases ($\alpha^{max} \rightarrow \alpha^{min}$) and (2) the maximum of the judge weights decreases ($\max_j(\lambda^j) \rightarrow 0$), as would occur when there are more judges distributed over the range of the instrument. 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 practice, we calculate the maximum bias of our estimator under the worst-case assumption of treatment effect heterogeneity (i.e. $\alpha^{max} - \alpha^{min} = 1$) (the maximum possible value). 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 $\max_j(\lambda^j)$.

In general, the IV weights, λ^j, will not be equal across judges. In particular, the IV weights depend in part on the probability a defendant is assigned a particular judge. Therefore, judges who handle different caseloads may have different weights. Also, the share of compliers between any two adjacent judges need not be the same. For example, if there are more infra-marginal defendants for lenient judges, then lenient judges will be given more weight in the estimation of racial bias. However, our bounding procedure of the maximum bias does not rely on any assumption about equal weights across judges. For example, consider an extreme case where although there are many judges, defendants are only infra-marginal to the most-strict and second most-strict judge. Then, the entire share of compliers will be defendants who are detained by the most-strict judge and released by the second most-strict judge. Therefore, the pairwise LATE for the most-strict judge and the second most-strict judge will receive the entire weight in estimating the effect of release on the probability of pre-trial misconduct. In this case, we would conclude that the maximum bias of our estimator is equal to one, and therefore, we would be unable to provide informative bounds on the true level of racial bias.

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 B.2. Again, because we do not 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^j_r are given by the following formula:

$$\lambda_r^j = \frac{(Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r)) \cdot \sum_{l=j}^{J} \pi_r^l (g(z_l) - \mathbb{E}[g(Z)])}{\sum_{m=1}^{J} (Pr(Released|z_m, r) - Pr(Released|z_{m-1}, r)) \cdot \sum_{l=m}^{J} \pi_r^l (g(z_l) - \mathbb{E}[g(Z)])} \tag{B.33}$$

As discussed in our proof of Proposition 3, 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 Equation (B.33) and simplifying yields:

$$\lambda^j = \frac{(z_j - z_{j-1}) \cdot \sum_{l=j}^{J} \pi^l (z_l - \mathbb{E}[Z])}{\sum_{m=1}^{J} (z_j - z_{j-1}) \cdot \sum_{l=m}^{J} \pi^l (z_l - \mathbb{E}[Z])} \tag{B.34}$$

We use Equation (B.34) to estimate the maximum bias of our estimator by replacing π^j and $\mathbb{E}[Z]$ with their empirical counterparts:

$$\hat{\pi}^j = \sum_{i=1}^{N} \frac{\mathbb{1}\{Z_i = z_j\}}{N} \tag{B.35}$$

$$\mathbb{E}[Z] = \frac{1}{N} \sum_{i=1}^{N} Z_i \tag{B.36}$$

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 of racial bias. This procedure yields a maximum bias of 0.011 or 1.1 percentage points.

From Equation (B.31), we know:

$$D^* < D^{IV} + \max_j(\lambda^j) = D^{IV} + 0.011$$
$$D^* > D^{IV} - \max_j(\lambda^j) = D^{IV} - 0.011$$

Therefore, in our setting, the true level of racial bias is bounded within 1.1 percentage points of our IV estimate for racial bias. $\square$

D. 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:

$$\Psi(x) = \frac{Pr(W|x)Pr(B)}{Pr(B|x)Pr(W)} \tag{B.37}$$

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:

$$\alpha_r^{j,j-1}(x) = \mathbb{E}[Y_i(1) - Y_i(0)|R_i(z_j) - R_i(z_{j-1}) = 1|r_i = r, \mathbf{X}_i = x] \tag{B.38}$$

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

$$\alpha_r^{j,j-1} = \sum_{x \in X} \alpha_r^{j,j-1}(x) \frac{Pr(Released|z_j, x, r) - Pr(Released|z_{j-1}, x, r)}{Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r)} P(x|r) \tag{B.39}$$

Given a linear first stage:

$$\frac{Pr(Released|z_j, x, r) - Pr(Released|z_{j-1}, x, r)}{Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r)} = 1 \tag{B.40}$$

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:

$$D^{IV} = \sum_{j=1}^{J} \lambda^j \left(\sum_{x \in X} Pr(x|W) \left(\alpha_W^{j,j-1}(x) - \alpha_B^{j,j-1}(x) \right) \right) \tag{B.41}$$

$\square$

Appendix C: Data Appendix

Judge Leniency: We calculate judge leniency as the leave-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 ever released prior to case disposition.

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 w/Conditions: An indicator for whether the defendant was released on non-monetary bail with conditions, 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 zero for defendants who receive non-monetary bail with conditions or ROR.

Race: Indicator for whether the defendant is black (versus non-black).

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 70 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.

Arrested on Bail in Past Year: An indicator for whether the defendant had been arrested while out on bail in the past year within the same county, set to missing for defendants who we cannot observe for a full year prior to their bail hearing.

Failed to Appear in Court in Past Year: An indicator for whether the defendant failed to appear in court while out on bail in the past year 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.

Any Drug Offense: An indicator for whether the defendant is charged with a drug offense.

Any DUI Offense: An indicator for whether the defendant is charged with a DUI offense.

Any Violent Offense: An indicator for whether the defendant is charged with a violent offense.

Any Property Offense: An indicator for whether the defendant is charged with a property offense.

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

Failure to Appear in Court: An indicator for whether the defendant failed to appear for a required court appearance, as proxied by the issuance of a bench warrant. This outcome is only available in Philadelphia.

Judge 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.

Judge 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

The institutional details described in this Appendix follow directly from Dobbie et al. (forthcoming). Like the federal government, both Pennsylvania and Florida grant a constitutional right to some form of bail for most defendants. For instance, Article I, §14 of the Pennsylvania Constitution states that "[a]ll prisoners shall be bailable by sufficient sureties, unless for capital offenses or for offenses for which the maximum sentence is life imprisonment or unless no condition or combination of conditions other than imprisonment will reasonably assure the safety of any person and the community...." Article I, §14 of the Florida Constitution states that "[u]nless charged with a capital offense or an offense punishable by life imprisonment...every person charged with a crime...shall be entitled to pretrial release on reasonable conditions."

Philadelphia County: In Philadelphia County, defendants are brought to one of six police stations immediately following their arrest, where they are interviewed by the city's Pre-Trial Services Bail Unit. The Philadelphia Bail Unit interviews all adults charged with offenses in Philadelphia through videoconference, collecting information on each defendant's charge severity, personal and financial history, family or community ties, and criminal history. The Bail Unit then uses this information to generate a release recommendation based on a four-by-ten grid of bail guidelines that is presented to the bail judge at the bail hearing. 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, defendants are brought in for a bail hearing. Bail hearings are conducted through videoconference by the bail judge on duty, with representatives from both the district attorney and local public defender's offices (or private defense counsel) present. However, while a defense attorney is present at the bail hearing, there is usually 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 a bail modification in subsequent court proceedings.

Under the Pennsylvania Rules of Criminal Procedure, "the bail authority shall consider all available information as that information is relevant to the defendant's appearance or nonappearance at subsequent proceedings, or compliance or noncompliance with the conditions of the bail bond," including information such as the nature of the offense, the defendant's employment status and relationships, and whether the defendant has a record of bail violations or flight. Pa. R. Crim. P. 523. In setting monetary bail, "[t]he amount of the monetary condition shall not be greater than is necessary to reasonably ensure the defendant's appearance and compliance with the conditions of the bail bond." Pa. R. Crim. P. 524. Under Pa. R. Crim. 526, a required condition of any bail bond is that the defendant "refrain from criminal activity."

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 be immediately released following arrest and booking by posting an amount designated by a standard bail schedule. The standard bail schedule ranks offenses according to their seriousness and assigns an amount of bond that must be posted before 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 through the standard bail schedule, with the other 70 percent of defendants attending a bail hearing (Goldkamp and Gottfredson 1988).

If a defendant is unable to post the standard bail amount 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 standard bail amount may be lowered, raised, or remain the same as the standard 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).

Under the Florida Rules of Criminal Procedure, "[t]he judicial officer shall impose the first ... conditions of release that will reasonably protect the community from risk of physical harm to persons, assure the presence of the accused at trial, or assure the integrity of the judicial process." Fl. R. Crim. P. 3.131.

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). 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.

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.

Appendix E: Relaxing the Monotonicity Assumption

In this section, we develop two alternative tests for racial bias that do not require judge leniency to be monotonic in race. Our first test estimates the level of racial bias using race-specific marginal treatment effects (MTEs) at different points in the leniency distribution. Our second test estimates bounds on the level of racial bias using non-parametric pairwise LATEs for judges at different points in the leniency distribution. In both cases, our estimate of racial bias imposes equal weights on each judge in our sample.

A. Parametric Estimates using MTEs

The first alternative test for racial bias relies on race-specific marginal treatment effects (MTEs). In our main IV estimation strategy, we assume that the support over the distribution of 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. Conversely, race-specific MTEs estimate the full distribution of treatment effects for defendants on the margin of release, rather than a weighted average as in the main results. If the MTE assumptions hold, these estimates allow us to (i) identify judge-specific estimates of racial bias and (ii) ex post apply any form of weights when calculating the average of racial bias across judges.

The MTE approach requires a different set of assumptions compared to our main IV estimation strategy. Most importantly, this approach allows us to relax the assumption that judge leniency is monotonic by race by estimating the MTEs using a race-specific leniency measure. As with our main estimation strategy, however, the MTE approach requires continuous support over the distribution of judge leniency. The MTE approach also requires an additional parametric assumption to identify the full distribution of treatment effects. Specifically, MTEs are computed by estimating the relationship between the outcome of interest (i.e. pre-trial crime) and the propensity score (i.e. the probability a judge releases a defendant) and then taking the derivative of the resulting function. As is common in the MTE literature, we assume that this relationship is well approximated by a local quadratic in a given bandwidth. We then estimate the MTE as the numerical derivative of this function.

Empirical Framework: To formally map our model of racial bias from the main text to the MTE framework developed by Heckman and Vytlacil (2005), we first characterize judge j's pre-trial release decision as:

$$Released_i(z_j) = \mathbb{1}\{\mathbb{E}[\alpha_i|r] \leq t_r^j\} \tag{E.1}$$

where $Released_i(z_j)$, α_i, and t_r^j are defined as in the main text. Let $F_{\alpha,r}$ be the cumulative density function of $\mathbb{E}[\alpha_i|r]$, which we assume is continuous on the interval $[0,1]$. Judge j's release decision can now be expressed as the following latent-index model:

$$Released_i(z_j) = \mathbb{1}\{F_{\alpha,r}(\mathbb{E}[\alpha_i|r]) \leq F_{\alpha,r}(t_r^j)\} = \mathbb{1}\{U_{i,r} \leq Pr(Released|z_j, r)\} \tag{E.2}$$

where $U_{i,r} \in [0,1]$ by construction. In this latent-index model, defendants with $U_{i,r} \leq Pr(Released|z_j, r)$

are released, defendants with $U_{i,r} > Pr(Released|z_j, r)$ are detained, and defendants with $U_{i,r} = Pr(Released|z_j, r)$ are on the margin of release for judge j.

Following Heckman and Vytlacil (2005), we define the race-specific marginal treatment effect as the treatment effect for defendants on the margin of release:

$$MTE_r(u) = \mathbb{E}[\alpha_i|r, U_{i,r} = u] \tag{E.3}$$

where $\mathbb{E}[\alpha_i|r, U_{i,r} = Pr(Released|z_j, r)]$ denotes the treatment effect for a defendant of race r who is on the margin of release to a judge with propensity score equal $Pr(Released|z_j, r)$. For simplicity, we denote judge j's propensity score as p_r^j.

Using the above framework, we can now describe how the race-specific MTEs defined by Equation (E.3) allow us estimate racial bias for each judge in our sample. First, recall that the estimand of interest is the treatment effect of pre-trial release for white and black defendants at the margin of release:

$$\alpha_r^j = \mathbb{E}[\alpha_i|r, \mathbb{E}[\alpha_i|r] = t_r^j] \tag{E.4}$$

Because $\mathbb{E}[\alpha_i|r] = t_r^j$ can be replaced with the equivalent condition, $U_{i,r} = p_r^j$, both of which state defendant i is marginal to judge j, we can equate α_r^j to the MTE function at p_r^j:

$$\alpha_r^j = \mathbb{E}[\alpha_i|r, \mathbb{E}[\alpha_i|r] = t_r^j] = \mathbb{E}[\alpha_i|r, U_{i,r} = p_r^j] = MTE_r(p_r^j) \tag{E.5}$$

Equation (E.5) shows that we can use the race-specific MTEs to identify the race-specific treatment effect of each judge, α_r^j. As a result, we can also use the race-specific MTEs to identify the level of racial bias for each judge. To see this, let judge j have a propensity score to release white equal to p_W^j and a propensity to release blacks equal to p_B^j. As in the main text, the level of racial bias for judge j is equal to $\alpha_W^j - \alpha_B^j$. Given Equation (E.5), the level of racial bias for judge j is therefore equal to $MTE_W(p_W^j) - MTE_B(p_B^j)$.

The above results mean that we can ex post apply any form of weights when calculating the average of racial bias across judges (as opposed to our main estimation strategy, which imposes IV weights to estimate a weighted average of the level of racial bias). In practice, we construct a simple average of racial bias across judges:

$$D^{MTE} = \frac{1}{J} \sum \left(MTE_W(p_W^j) - MTE_B(p_B^j) \right) \tag{E.6}$$

We also present the full distribution of race-specific MTE estimates to determine whether some judges are biased against black defendants while others are biased against white defendants. As we discuss below, we find that the white MTE lies strictly above the black MTE, indicating that all judges are biased against black defendants (though the level for a particular judge depends on where that judge falls in both the white and black distribution of leniency). The fact the white MTE lies strictly above the black MTE also implies that we would find positive racial bias for any weighted average of racial bias, not just for the simple average we calculate here or the IV weighted average in the main text.

Empirical Implementation: We estimate the race-specific MTEs by taking the derivative of our outcome measure (i.e. rearrest before case disposition) with respect to variation in the propensity score provided by our instrument (i.e. variation in the predicted probability of being released from the variation in judge leniency). In practice, we implement our estimation procedure in two steps. In the first step, we use our race-specific judge leniency measure to estimate the race-specific propensity score, capturing only the variation in pre-trial release due to the instrument. In the second step, we compute the numerical derivative of a smoothed function relating rearrest before disposition to the race-specific propensity score. Specifically, we first residualize the indicator for rearrest before disposition using court-by-time fixed effects to control for the randomization strata. We then estimate the relationship between the residualized rearrest variable and the propensity score using a local quadratic estimator. Finally, we compute the numerical derivative of the local quadratic estimator for each race separately to obtain the MTEs. Our two-step estimation procedure closely follows the approaches outlined in Heckman, Urzua, and Vytlacil (2006) and Doyle (2007), among many others.

Before turning to the MTE results, we first evaluate the small-sample properties of the race-specific MTEs in simulated data that approximates our setting. Specifically, we create a dataset with 170 judges, where each judge is assigned 500 cases with black defendants and 500 cases with white defendants. The latent risk of rearrest before disposition for each defendant is drawn from a uniform distribution between 0 and 1. Each judge releases defendants if and only if the risk of rearrest is less than his or her race-specific threshold. In the simulated data, each judge's threshold for white defendants is set to match the distribution of judge leniencies observed in the true data. For each judge, we then impose a 10 percentage point higher threshold for black defendants, so that the "true" level of racial bias in the simulated data is exactly equal to 0.1. Finally, we use the MTE estimation procedure outlined above to estimate both the race-specific MTEs and the average level of racial bias when each judge is weighted equally. Repeating the entire procedure 500 times, we estimate an average level of racial bias equal to 0.089 with a standard deviation of 0.034. In other words, we cannot reject that the estimated level of racial bias is equal to the "true" level of racial bias in the simulated data. See Appendix Figure E1 for the results from these simulations.

Results: Appendix Figure E2 presents the race-specific MTEs for our sample, where low propensity scores correspond to more strict judges and high propensity scores correspond to more lenient judges. Standard errors are estimated through a bootstrap procedure with 500 replications that clusters at the judge-shift level. We find that the white MTE lies strictly above the black MTE at all points in the leniency distribution, suggesting that we would find evidence of racial bias regardless of our choice of weighting scheme. Weighting each judge equally, we find an average level of racial bias equal to 0.325 (p-value=0.004). The similarity of our MTE results and main IV results suggests that violations of monotonicity and our choice of IV weights are not driving our findings.

B. Non-Parametric Bounds using Pairwise LATEs

The second alternative test for racial bias estimates non-parametric bounds on the level of racial bias using pairwise LATEs for judges at different points in the leniency distribution. At a high level, this approach is based on the idea that defendants released by a lenient judge but detained by a strict judge must have treatment effects between the release thresholds of these lenient and strict judges. As a result, the average of all compliers' treatment effects must be greater than the release threshold for the strict judge, but less than the threshold of release for the lenient judge. This observation allows us to bound the release threshold (t_r^j) for each judge by estimating pairwise LATEs, which in turn allows us to estimate bounds for the level of racial bias. To relax the monotonicity assumption, we also allow a judge's rank in the leniency distribution to vary by race. Additionally, unlike our main estimation strategy and the MTE approach, this approach does not require continuous support over the distribution of leniency.

Empirical Framework: To formally show how we use pairwise LATEs to estimate non-parametric bounds on the level of racial bias, first note that defendants released by judge j but detained by a more strict judge $j-k$ (where $k > 0$) must have treatment effects between the release thresholds for the pair of judges: $t_r^{j-k} < \mathbb{E}[\alpha_i|r] \leq t_r^j$. The pairwise LATE for judges j and $j-k$, $\alpha_r^{j,j-k}$, is therefore an average of these compliers' treatment effects. Because the average treatment effect of compliers must be lower than the highest treatment effect of compliers, it follows that:

$$t_r^j \geq \alpha_r^{j,j-k} \tag{E.7}$$

The pairwise LATE for judges j and $j-k$ therefore provides a lower bound for judge j's release threshold. Using identical logic, the pairwise LATE for judges j and $j+k$ provides an upper bound for judge j's release threshold:

$$t_r^j \leq \alpha_r^{j+k,j} \tag{E.8}$$

Next, recall that the level of racial bias for judge j is equal to $t_W^j - t_B^j$. We can therefore estimate a lower bound of racial bias for judge j by replacing t_W^j with a weakly lower quantity, i.e. the pairwise LATE for judge j and a more strict judge ($\alpha_W^{j,j-k}$) and replacing t_B^j with a weakly larger quantity, i.e. the pairwise LATE for judge j and a more lenient judge ($\alpha_B^{m(j)+k,m(j)}$), where $m(j)$ denotes judge j's rank in the black distribution of leniency. By allowing each judge's rank in the white distribution, j, to differ from that judge's rank in the black distribution, $m(j)$, judge leniency is not assumed to be monotonic by race.

By estimating a bound for each judge, we can ex post apply any form of weights when calculating the average of racial bias across judges (as opposed to our main estimation strategy, which imposes IV weights to estimate a weighted average of the level of racial bias). In practice, we construct a lower bound which is a simple average of racial bias across judges:

$$D^{PW,Lower} = \frac{1}{J} \sum \alpha_W^{j,j-k} - \alpha_B^{m(j)+k,m(j)} \tag{E.9}$$

Similarly, we can estimate an upper bound of racial bias for judge j by replacing t_W^j with a weakly larger quantity, i.e. the pairwise LATE for judge j and a more lenient judge ($\alpha_W^{j+k,j}$) and replacing t_B^j with a weakly smaller quantity, i.e. the pairwise LATE for judge j and a more strict judge ($\alpha_B^{m(j),m(j)-k}$), where $m(j)$ denotes judge j's rank in the black distribution of leniency. From this procedure, we construct an upper bound which is a simple average of racial bias across judges:

$$D^{PW,Upper} = \frac{1}{J} \sum \alpha_W^{j+k,j} - \alpha_B^{m(j),m(j)-k} \tag{E.10}$$

Note that in the Equations (E.9)-(E.10), lower and upper bounds are not identified for some judges. For example, there is no stricter judge for the most strict judge. Therefore, the lower bound is not identified for the most strict judge. We assume that the lower bound for this judge's threshold for release is equal to zero, the lowest possible value. Similarly, the upper bound for the most lenient judge's threshold for release is not identified. We assume this upper bound for this judge is equal to one, the highest possible value. The higher the value of k, the more judges for which upper and lower bounds are not identified in practice.

From these bounds, we know that the true level of racial bias D^{IV} must fall between $D^{PW,lower}$ and $D^{PW,upper}$. In summary, this non-parametric approach allows us to estimate bounds on the true level of racial bias while relaxing the monotonicity assumption, imposing equal weights on each judge in our sample ex post, and allowing the distribution of leniency to be discrete.

Empirical Implementation: We estimate the non-parametric bounds using a two-step procedure. First, we residualize rearrest prior to case disposition using court-by-time fixed effects as in the MTE specifications. Second, we form the following Wald estimator for each pair of judges j and judge $j-k$:

$$\alpha_r^{j,j-k} = \frac{\mathbb{E}[\tilde{Y}|Z=j,r] - \mathbb{E}[\tilde{Y}|Z=j-k,r]}{\mathbb{E}[Released|Z=j,r] - \mathbb{E}[Released|Z=j-k,r]} = \frac{\mathbb{E}[\tilde{Y}|Z=j,r] - \mathbb{E}[\tilde{Y}|Z=j-k,r]}{\gamma_r(z_j - z_{j-k})} \tag{E.11}$$

where $\tilde{Y}$ is our residualized outcome variable, $(z_j - z_{j-k})$ is the difference in leniency across judges j and judge $j-k$, and γ_r is the first stage regression coefficient. We plug in the appropriate values of $\alpha_r^{j,j-k}$ to Equations (E.9)-(E.10) to estimate lower and upper bounds for racial bias, respectively. In theory, smaller values of k will lead to sharper bounds, as the difference in release thresholds, and hence the number of compliers, is strictly increasing in k. In practice, however, smaller values of k will also lead exacerbate sampling error in small samples due to well-known problems with weak instruments. In other words, we will tend to estimate extremely imprecise bounds when judges adjacent in the leniency distribution tend to have very similar release rates, as any sampling error is "blown up" by a near-zero first stage. Given this tradeoff in the choice of k, we provide results for several different choices of k (i.e. 10, 15, and 20) and compare the bounds estimated from each choice.

Before turning to the bounding results, we again evaluate the small-sample properties of our estimation procedure using the simulated data described above. Standard errors are estimated through a bootstrap procedure with 500 replications which clusters at the judge-shift level. Columns 1 and 2 of Appendix Table E1 present bounds in these simulated data with a "true" level of bias of 0.1. As expected, the estimated bounds become more conservative, but also more precisely estimated, with larger values of k. For values of $k = \{10, 15, 20\}$, the average lower bounds, respectively, are $D^{PW,Lower} = \{0.035, -0.035, -0.106\}$ with standard deviations equal to $SD_{PW,Lower} = \{0.081, 0.064, 0.054\}$. The average upper bounds, respectively, are $D^{PW,Upper} = \{0.192, 0.233, 0.291\}$ with standard deviations equal to $SD_{PW,Upper} = \{0.080, 0.064, 0.059\}$. In unreported results, we also estimate upper and lower bounds for $k < 10$. However, for very low values of k, the estimates become so imprecise that we often estimate bounds outside of the theoretically possible levels of discrimination.

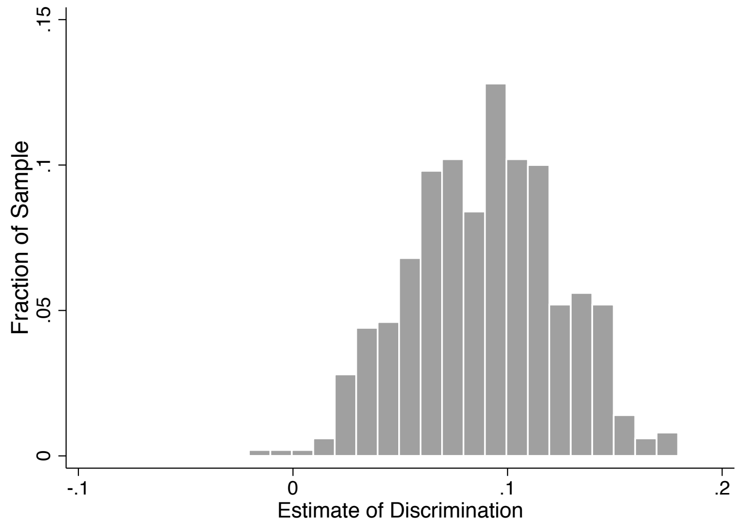
Results: Columns 3 and 4 of Appendix Table E1 present estimates of lower and upper bounds of racial bias in our sample. Following the simulation results, we present estimates for $k = \{10, 15, 20\}$ and ex post apply equal weights for each judge in our sample. Standard errors are again estimated through a bootstrap procedure with 500 replications which clusters at the judge-shift level. The lower bound ranges from 0.322 (p-value=0.334) for $k = 20$ to 0.881 (p-value=0.264) for $k = 10$, while the upper bound ranges from 0.548 (p-value = 0.128) for $k = 20$ to 0.994 (p-value = 0.186) for $k = 10$. We view these results as broadly consistent with both our main IV estimates and the MTE estimates discussed above, further suggesting that violations of monotonicity and our choice of IV weights are not driving our findings.

Appendix Table E1: Non-parametric Pairwise LATEs Bounds

	Simulated Data		Real Data	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
	(1)	(2)	(3)	(4)
k=10	0.035	0.192	0.881	0.994
	(0.081)	(0.080)	(0.789)	(0.751)
k=15	−0.035	0.233	0.417	0.587
	(0.064)	(0.064)	(0.457)	(0.464)
k=20	−0.106	0.291	0.322	0.548
	(0.054)	(0.059)	(0.344)	(0.360)

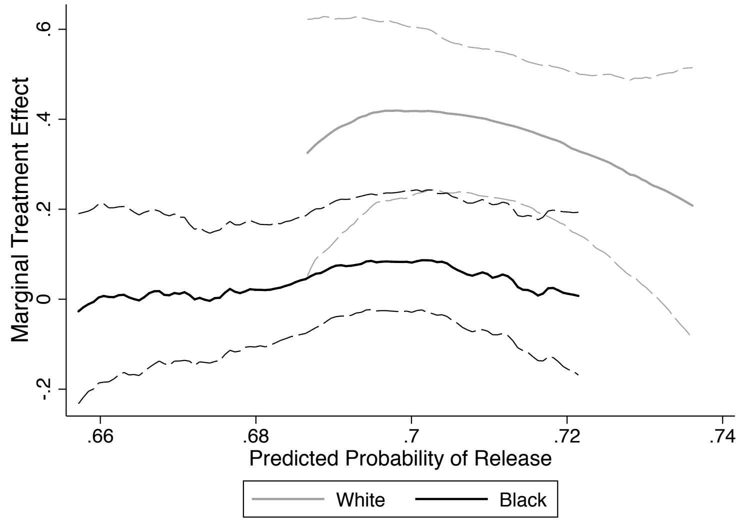
Note: This table reports upper and lower bounds on the level of racial bias. Columns 1-2 report results for simulated data with a "true" level of racial bias of 0.1. The simulated data include 170 judges, where each judge is assigned 500 black defendants and 500 white defendants. Defendant risk in the simulated data is drawn from a uniform distribution between 0 and 1. Judges release defendants if the risk is less than a judge-specific threshold, where the distribution of judge-specific threshold matches the empirical distribution of judge leniency. The standard deviation of the estimates across all simulations is presented in parentheses. Columns 3-4 reports results for the real data, where we estimate the non-parametric bounds developed in Appendix E. Standard errors are computed using 500 bootstrap replications which clusters at the judge shift level.

Appendix Figure E1: Distribution of Marginal Treatment Effects in Simulated Data



Note: This figure reports the distribution of estimated racial bias using a race-specific judge leniency measure in simulated data with a “true” level of racial bias of 0.1. The simulated data include 170 judges, where each judge is assigned 500 black defendants and 500 white defendants. Defendant risk in the simulated data is drawn from a uniform distribution between 0 and 1. Judges release defendants if the risk is less than a judge-specific threshold, where the distribution of judge-specific threshold matches the empirical distribution of judge leniency. See the Appendix text for additional details on the simulated data and the notes to Figure 3 for additional details on the MTE estimation procedure.

Appendix Figure E2: Marginal Treatment Effects with Race-Specific Judge Leniency



Note: This figure displays marginal treatment effects using a race-specific judge leniency measure. See the notes to Figure 3 for additional details on the MTE estimation procedure.

Appendix F: Model of Stereotypes

In this appendix, 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 using a machine learning algorithm that efficiently uses all observable crime and defendant characteristics. In short, we use a randomly-selected subset of the data to train the model using all individuals released on bail. In training the model, we must choose the shrinkage, the number of trees, and the depth of each tree. Following common practice, we choose the smallest shrinkage parameter (i.e. 0.005) that allows the training process to run in a reasonable time frame. We use a 5-fold cross validation on the training sample in order to choose the optimal number of trees for the predictions. The interaction depth is set to 5, which allows each tree to use at most 5 variables. Using the optimal number of trees from the cross validation step, predicted probabilities are then created for the full sample.

Following the construction of the continuous predicted risk variable, we 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. (forthcoming) and plot the relationship between predicted risk and true risk in the test sample. 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 A3). This is true for both white and black defendants. Therefore, we interpret the predicted distributions of risk based on observables as the true distributions of risk throughout.

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 70.7 percent if they were treated the same as white defendants. This implied release rate is lower than the true release rate of white defendants (71.2 percent), but higher than the true release rate for black defendants (68.9 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 2.7 percent expected probability of being rearrested before disposition while a defendant with $t = 100$ has a 74.5 percent probability of being rearrested before disposition.

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:

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

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 2 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 (F.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,

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

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 (F.2), from which we can back out the average release rate for black defendants from the stereotyped distribution.

We find that $\theta = 1.9$ rationalizes the average release rate for blacks we observe in the data (68.8 percent). That is, if judges use a representativeness-based discounting model with $\theta = 1.9$ to form perceptions of the risk distribution, we would expect judges to release 68.8 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 $\theta = 1.9$ alongside the true distribution of risk for blacks in Appendix Figure A5. The average risk in the stereotyped distribution is about 5.4 percentage points greater than the mean in the true distribution of risk.