Racial Bias in Bail Decisions*

David Arnold† Will Dobbie‡ Crystal S. Yang§

April 2018

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 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 use 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 OLS estimates are not biased by infra-marginality concerns. 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

[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.C for additional details on why the Knowles et al. (2001) and Anwar and Fang (2006) tests yield different results than our test.

prior literature has been critiqued for its reliance on restrictive assumptions about the unobserved risk 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 2018).[3]

In the first section of the paper, we formally develop two complementary estimators that use variation in the release tendencies of quasi-randomly assigned bail judges to identify the differences in pre-trial misconduct rates at the margin of release required for the Becker outcome test. Our first estimator uses the standard instrumental variables (IV) framework to identify differences in the local average treatment effects (LATEs) for white and black defendants near the margin of release. Though IV estimators are often criticized for the local nature of the estimates, we exploit the fact that the Becker test relies on (the differences between) exactly these kinds of local treatment effects to test for racial bias. In our context, our IV estimator measures the weighted average of racial bias across all bail judges with relatively few auxiliary assumptions, but at the potential cost that we cannot estimate judge-specific treatment effects and the weighting scheme underlying the IV estimator is not always policy relevant. In contrast, our second estimator uses the marginal treatment effects (MTE) framework developed by Heckman and Vytlacil (1999, 2005) to estimate judge-specific treatment effects for white and black defendants at the margin of release. Our MTE estimator therefore allows us to put equal weight on each judge in our sample, but with the estimation of the judge-specific estimates coming at the cost of additional auxiliary assumptions.

The second part of the paper tests for racial bias in bail setting using administrative court data from Miami and Philadelphia. We find evidence of significant racial bias against black defendants using both our IV and MTE estimators, ruling out statistical discrimination as the sole explanation for the racial disparities in bail. We find that marginally released white defendants are 22.2 to 23.1 percentage points more likely to be rearrested prior to disposition than marginally released black defendants using our IV and MTE estimators, respectively. Our estimates of racial bias are nearly identical if we account for other observable crime and defendant differences by race, suggesting

[3]See Dobbie et al. (2018), Gupta, Hansman, and Frenchman (2016), Leslie and Pope (2017), and Stevenson (2016) for evidence on the non-financial consequences of bail decisions.

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 results, naïve OLS estimates indicate, if anything, racial bias against white defendants, highlighting the importance of accounting for both infra-marginality and omitted variables when estimating racial 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, a result that is 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. (2018), 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. (2018) 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. In the context of bail decisions, Ayres and Waldfogel (1994) show that bail bond dealers in New Haven charge lower prices to minority defendants, and Bushway and Gelbach (2011) find evidence of racial bias among bail judges 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) and work using MTEs to extrapolate to other estimands of interest (e.g., Heckman and Vyltacil 2005, Heckman, Urzua, and Vytlacil 2006, Cornelissen et al. 2016). In recent work, Brinch, Mogstad, and Wiswall (2017) show that a discrete instrument can be used to identify marginal treatment effects using functional form assumptions. Kowalski (2016) similarly shows that it is possible to bound and estimate average treatment effects for always takers and never takers using functional form assumptions. Most recently, Mogstad, Santos, and Torgovitsky (2017) show that because a LATE generally places some restrictions on unknown marginal treatment effects, 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.

[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), 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.

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, 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 two complementary estimators to identify these race-specific treatment effects using the quasi-random assignment of cases to judges. Appendix B provides additional details and proofs.

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. 2018, Gupta et al. 2016, Leslie and Pope 2017, 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. In our setting, the bail judges we study in Philadelphia 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 a defendant 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 release for 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 the bail judge's sole task is to decide whether to release or detain a defendant given that this decision margin is the most important and consequential

(Kleinberg et al. 2018, Dobbie et al. 2018). In simplifying each judge's task to this single decision, we abstract away from the fact that bail judges may set different levels of monetary bail that take into account 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.

[5] In contrast to the two models we consider in this section, models of (accurate) 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 (accurate) statistical discrimination models, however, judges use race to form accurate predictions of risk, both on average and at the margin of release. As a result, neither form of (accurate) statistical discrimination will lead to marginal white defendants having a higher probability of misconduct than marginal black defendants.

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 weighted average of treatment effects for defendants of race r at the margin of release for judge j, α_r^j, for some weighting scheme, w^j, across all bail judges, $j = 1...J$, be given by:

$$\begin{aligned} \alpha_r^{*,w} &= \sum_{j=1}^{J} w^j \alpha_r^j \\ &= \sum_{j=1}^{J} w^j t_r^j \end{aligned} \tag{4}$$

where w^j are non-negative weights which sum to one that will be discussed in further detail below. 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, $\alpha_r^{*,w}$ represents a weighted average of the treatment effects for defendants of race r at the margin of release across all judges.

Following this notation, the average level of racial bias among bail judges, $D^{*,w}$, for the weighting scheme w^j is given by:

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

From Equation (4), we can express $D^{*,w}$ 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.

Standard OLS estimates will typically not recover unbiased estimates of the weighted average of racial bias, $D^{*,w}$, for two reasons. First, characteristics observable to the judge but not the econometrician may be correlated with pre-trial release, resulting in omitted variable bias when estimating the treatment effects for black and white defendants. The second, and more important, reason OLS estimates will not recover unbiased estimates of racial bias is that the average treatment effect identified by OLS will equal the treatment effect at the margin required by the outcome test unless one is willing to assume either identical risk distributions for black and white defendants or constant treatment effects across the entire distribution of both black and white defendants (e.g., Ayres 2002). Thus, even if the econometrician observes the full set of observables known to the bail judge, OLS estimates are still not sufficient to test for racial bias without extremely restrictive

assumptions.[6]

We therefore develop two complementary estimators for racial bias that use variation in the release tendencies of quasi-randomly assigned bail judges to identify differences in pre-trial misconduct rates at the margin of release. Our first estimator uses the standard IV framework to identify the difference in LATEs for white and black defendants near the margin of release. Our IV estimator allows us to estimate a weighted average of racial bias across bail judges with relatively few auxiliary assumptions, but with the caveats that we cannot estimate judge-specific treatment effects and the weighting scheme underlying the IV estimator may not be policy relevant. In contrast, our second estimator uses the MTE framework developed by Heckman and Vytlacil (1999, 2005) to estimate judge-specific treatment effects for white and black defendants at the margin of release, allowing us to choose our own weighting scheme when calculating racial bias in our data. In practice, we choose to impose equal weights on each judge – a parameter with a clear economic interpretation – meaning that our MTE estimates can be interpreted as the average level of bias across judges in our sample.

C.1. Setup

We first briefly review the econometric properties of a race-specific estimator that uses judge leniency as an instrumental variable for pre-trial release, baseline assumptions that underlie both our IV and MTE estimators. Section II.B provides empirical tests of each assumption.

Let Z_i be a scalar measure of the assigned judge's propensity for pre-trial release for defendant-case i 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 tendencies of judges. We calculate Z_i separately for white and black defendants to relax the standard monotonicity assumption that the judge ordering produced by the scalar Z_i is the same for both white and black defendants, implicitly allowing judges to exhibit different levels of racial bias.

Following Imbens and Angrist (1994), a race-specific 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, $Released_i$, is a nontrivial function of the instrument 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 $Released_i$.

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

[6] In Appendix C, we use a series of simple graphical examples to illustrate how a standard OLS estimator suffers from infra-marginality bias whenever there are differences in the risk distributions of black and white defendants. We then use a simple two-judge example to illustrate how a judge IV estimator can alleviate the infra-marginality bias.

pre-trial misconduct Y_i:

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

where $\mathbf{v}_i = \mathbf{U}_i + \varepsilon_i$ 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_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:

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

where $Released_i(z_j)$ equals 1 if for a given case, the defendant is released if assigned to judge j. Assumption 3 implies that for a given case, 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.

C.2. IV Estimator for Racial Bias

Given Assumptions 1-3, we now formally define our IV estimator for racial bias, provide conditions for consistency, and discuss the interpretation of the IV weights.

Defining our IV Estimator: Let the true IV-weighted level of racial bias, $D^{*,IV}$ be defined as:

$$\begin{aligned} D^{*,IV} &= \sum_{j=1}^{J} w^j \left(t_W^j - t_B^j \right) \\ &= \sum_{j=1}^{J} \lambda^j \left(t_W^j - t_B^j \right) \end{aligned} \tag{6}$$

where $w^j = \lambda^j$, the standard IV weights defined in Imbens and Angrist (1994).

Let our IV estimator that uses judge leniency as an instrumental variable for pre-trial release be defined as:

$$\begin{aligned} D^{IV} &= \alpha_W^{IV} - \alpha_B^{IV} \\ &= \sum_{j=1}^{J} \lambda_W^j \alpha_W^{j,j-1} - \sum_{j=1}^{J} \lambda_B^j \alpha_B^{j,j-1} \end{aligned} \tag{7}$$

where λ_r^j are again the standard IV weights and each pairwise treatment effect $\alpha_r^{j,j-1}$ captures the treatment effects of compliers within each j, $j-1$ pair. As we discuss in Appendix B, compliers for judge j and $j-1$ are individuals such that $\alpha_r^{j,j-1} \in (t_r^{j-1}, t_r^j]$.

Consistency of our IV Estimator: Our IV estimator D^{IV} provides a consistent estimate of $D^{*,IV}$ under two conditions: (1) Z_i is continuous and (2) λ_r^j is constant by race. The first condition 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, α_r^{IV}, approaches a weighted average of treatment effects for defendants at the margin of release. In Appendix B, we discuss the potential infra-marginality bias that may result if our instrument is discrete, as is the case in our data. With a discrete instrument, each defendant is no longer marginal to a particular judge and D^{IV} may no longer provide a consistent estimate of $D^{*,IV}$ if the distribution of white compliers differs from the distribution of black compliers. We show that the maximum infra-marginality bias of our IV estimator when the instrument is discrete is given by the following formula:

$$\max_j(\lambda^j)(\alpha^{max} - \alpha^{min})$$

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 potential for infra-marginality bias in our IV 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. In practice, we find that the maximum infra-marginality bias of our IV estimator D^{IV} from $D^{*,IV}$ is 1.1 percentage points in our setting.[7]

The second condition for consistency is that the weights on the pairwise LATEs must be equal across race. This equal weights assumption ensures that the race-specific IV estimates from Equation (7), α_W^{IV} and α_B^{IV}, provide the same weighted averages of $\alpha_W^{j,j-1}$ and $\alpha_B^{j,j-1}$. See Appendix B for proofs of consistency. In Appendix B, we empirically tests whether the IV weights λ_r^j are constant by race in our data, finding that the distributions of black and white IV weights are visually indistinguishable from each other and that a Kolmogorov-Smirnov test cannot reject the hypothesis that the two estimated distributions are drawn from the same continuous distribution (p-value =

[7] To better understand why the number of judges may affect the maximum infra-marginality 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, $t_W^j = t_B^j$, such that there is no racial bias, $D^{*,w} = 0$, under any weighting scheme w. 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 are the average treatment effects 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 ($\alpha_W^{IV} = 0.2$) that is larger in magnitude than the LATE for blacks ($\alpha_B^{IV} = 0.1$), causing us to estimate $D^{IV} = 0.1 > 0$. 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 $D^{*,w} > 0$. Under the worst-case scenario where we assume the maximum heterogeneity in treatment effects ($\alpha^{max} - \alpha^{min} = 1$), the maximum infra-marginality bias is $\max_j(\lambda^j) = 1$ because 100 percent of compliers fall within the two judges. In this case, 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 $\max_j(\lambda^j)$.

0.431). The IV weights for each judge-by-year cell are also highly correlated across race, with a regression of black IV weights for each judge-by-year cell on the white IV weight in the same cell yielding a coefficient equal to 1.028 (se = 0.033).

Interpretation of the IV Weights: As discussed above, our IV estimator yields a weighted average of racial bias across bail judges, where the weights λ^j are the standard IV weights defined in Imbens and Angrist (1994). If the IV weights are uncorrelated with the level of racial bias for a given judge, then our IV estimator will estimate the average level of discrimination across all bail judges. If the IV weights are correlated with the level of racial bias, however, then our IV estimator may under or overestimate the average level of racial bias across all bail judges, but may still be of policy relevance depending on the parameter of interest (e.g., an estimate of racial bias that puts more weights on judges with higher caseloads).

To better understand the economic interpretation of an IV-weighted estimate of racial bias, Appendix B investigates the relationship between our IV weights and judge-by-year characteristics. We find that our IV weights are positively correlated with both the number of cases in a judge-by-year cell and judge-by-year specific estimates of racial bias, implying that the IV-weighted estimate of racial bias may be larger than an equal-weighted estimate of racial bias. We return to this issue below when discussing the difference between our IV and MTE estimates.

C.3. MTE Estimator for Racial Bias

Finally, we formally define our MTE estimator of racial bias and provide conditions for consistency. Without loss of generality, we focus on an estimate of racial bias that places equal weight on each bail judge.

Defining our MTE Estimator: Let the true equal-weighted MTE estimate of racial bias, $D^{*,MTE}$ be defined as:

$$\begin{aligned} D^{*,MTE} &= \sum_{j=1}^{J} w^j \left(t_W^j - t_B^j \right) \\ &= \sum_{j=1}^{J} \frac{1}{J} \left(t_W^j - t_B^j \right) \end{aligned} \tag{8}$$

where $w^j = \frac{1}{J}$, such that $D^{*,MTE}$ can be interpreted as the average level of racial bias across judges.

Let our equal-weighted MTE estimator of racial bias, D^{MTE}, be defined as:

$$D^{MTE} = \sum_{j=1}^{J} \frac{1}{J} \left(MTE_W(p_W^j) - MTE_B(p_B^j) \right) \tag{9}$$

where p_r^j is the probability that judge j releases a defendant of race r calculated using only the variation in pre-trial release due to our judge leniency measure Z_i (i.e. judge j's race-specific

propensity score). $MTE_r(p_r^j)$ is the estimated MTE at the propensity score for judge j calculated separately for each defendant race r. In Appendix B, we show that $MTE_r(p_r^j) = \alpha_r^j$ when we map each judge j's release decision under our theory model to the MTE framework developed by Heckman and Vytlacil (2005).

Consistency of our MTE Estimator: Our MTE estimator D^{MTE} provides a consistent estimate of $D^{*,MTE}$ if the race-specific MTEs are identified over the entire support of the propensity score calculated using variation in Z_i. In practice, there are two conceptually different approaches to identifying the race-specific MTEs over the entire support of the propensity score. If Z_i is continuous, the local instrumental variables (LIV) estimand provides a consistent estimate of the MTE over the support of the propensity score with no additional assumptions (Heckman and Vytlacil 2005, Cornelissen et al. 2016). With a discrete instrument, however, our MTE estimator is only consistent under additional functional form restrictions that allow us to interpolate the MTEs between the values of the propensity score we observe in the data. Following Heckman and Vytlacil (2005) and Doyle (2007), we use a local polynomial function and information from the observed values of the propensity score to estimate the MTE curve over the full support of the propensity score. In other words, we implicitly assume that the functional form of the MTE curve is specified by a local polynomial function.

Following Cornelissen et al. (2016), we test our functional form assumption by comparing race-specific MTEs weighted by the standard IV weights to race-specific LATEs estimated using two-stage least squares. In line with our functional form assumption, we recover each nonparametric LATE using the appropriately weighted MTE up to sampling error. We also find similar MTE estimates under a range of different functional form assumptions, suggesting that our estimates are not particularly sensitive to the exact parametric restriction we choose. See Appendix B for details.

D. Discussion and Extensions

In this section, we discuss the interpretation of our test of racial bias under different assumptions and extensions.

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 police are more likely to rearrest black defendants conditional on having committed a new crime (Fryer 2016, Goncalves and Mello 2018). In this scenario, we will overestimate the probability of pre-trial misconduct for black versus white defendants at the margin and, as a result, underestimate the true amount of racial bias in bail setting. It is therefore possible that our estimates reflect a lower bound on the true amount of racial bias among bail judges to the extent that judges minimize new crime.[8]

[8] A related concern is that bail judges may be influenced by other court actors when making bail decisions, such as prosecutors or defense attorneys, who may themselves be racially biased against black defendants. In this scenario, it is possible that our empirical test identifies racial bias stemming from judges not overriding racially biased bail

Omitted Objectives for Release: We also assume that judges do not consider other objectives or outcomes, or what Kleinberg et al. (2018) 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. (2018), 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. (2018) 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, in particular the risk of new crime. 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."[9]

Racial 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.[10] Extending our model to incorporate these institutional details means

recommendations from these other court actors. Two pieces of evidence suggest a limited role for this possibility. First, we find substantial variation in pre-trial release tendencies across judges, a result that is inconsistent with the idea that judges "rubber-stamp" bail recommendations from other court actors. Second, we find that racial bias decreases with judge experience, a result that is inconsistent with other court actors driving the racial bias unless judge experience is correlated with the probability of overriding racially biased bail recommendations.

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

[10] While monetary bail is not meant to operate as a *sub rosa* vehicle for detaining defendants, several courts have held that there is no constitutional right to affordable bail. See, e.g., Brangan v. Commonwealth, 80 N.E.3d 949, 9960 (Mass. 2017) ("Bail that is beyond a defendant's reach is not prohibited. Where, based on the judge's consideration

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 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 bail (i.e., either ROR or non-monetary conditions) versus monetary bail has a larger impact on the probability of release for marginal black defendants.[11] 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 versus 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 versus 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.

Second, we directly estimate racial bias in the setting of non-monetary versus monetary bail to incorporate any additional bias stemming from this margin. We 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 similar estimates of racial bias when focusing on the non-monetary versus monetary bail decision.[12]

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, 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 at the margin for <u>all</u> reasons unrelated to true risk of pre-trial misconduct, including reasons related to seemingly non-race characteristics such as crime

of all the relevant circumstances, neither alternative nonfinancial conditions nor an amount the defendant can afford will adequately assure his appearance for trial, it is permissible to set bail at a higher amount, but no higher than necessary to ensure the defendant's appearance.")

[11] Dobbie et al. (2018) show that the assignment of ROR and non-monetary conditions have a statistically identical impact on defendant outcomes, including pre-trial misconduct. We therefore combine ROR and non-monetary conditions into a single category in our analysis.

[12] 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 22.2 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 19.1 percentage points more likely to be rearrested compared to marginal black defendants ($D^{IV} = \frac{0.085}{0.490} - \frac{-0.009}{0.511} = 0.191$).

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 in the spirit of DiNardo, Fortin, and Lemieux (1996) and Angrist and Fernández-Val (2013). This narrower test for racial bias relies on the assumption that judge preferences vary only by observable characteristics $\mathbf{X}_i$, i.e. $t_r^j(\mathbf{V}_i) = t_r^j(\mathbf{X}_i)$. We 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. We discuss these results in robustness checks below.

II. Data and Instrument Construction

This section summarizes the most relevant information regarding our administrative court data from Philadelphia and Miami-Dade, describes the construction of our judge leniency measure, and provides empirical support for the baseline assumptions required for our IV and MTE estimators of racial bias. Further details on the cleaning and coding of variables are contained in Appendix D.

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.[13]

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

[13] Appendix Table A3 presents results for marginal Hispanic white 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.

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 E, bail judges in Miami-Dade are assigned on a quasi-random basis only on the weekends. The final 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 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. In Philadelphia, released black defendants are 1.4 percentage points more likely to fail to appear in court compared to white defendants. Defining pre-trial misconduct as either failure to appear or rearrest in Philadelphia, and only rearrest in Miami, released black defendants are 4.1 percentage points more likely to commit any form of pre-trial misconduct compared to white defendants.[14]

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 as an instrument for release. In both Philadelphia and Miami-Dade, there are multiple bail judges serving at each point in time, 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,

[14]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.

bail judges in Miami-Dade rotate through the weekend felony and misdemeanor bail hearings. See Appendix E for additional details.

Following Dobbie et al. (2018), we construct our instrument using a residualized, leave-out judge leniency measure that accounts for the case assignment processes in Philadelphia and Miami-Dade. To construct this residualized judge leniency measure, we first regress pre-trial release decisions on an exhaustive set of court-by-time fixed effects, the level at which defendants are quasi-randomly assigned to judges in our setting. In Miami, these court-by-time fixed effects include 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 bail-day of week-by-bail shift fixed effects. We then use the residuals from this regression to calculate the leave-out mean judge release rate for each defendant. We calculate our instrument across all case types, but allow the instrument to vary across years and defendant race.[15]

Figure 1 presents the distribution of our residualized judge leniency measure for pre-trial release at the judge-by-year level for all defendants, white defendants, and black defendants. Our sample includes seven total bail judges in Philadelphia and 170 total bail judges in Miami-Dade. In Philadelphia, the average number of cases per judge is 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 cell including 179 cases. Controlling for the exhaustive set of court-by-time fixed effects, the judge release measure ranges from -0.283 to 0.253 with a standard deviation of 0.040. In other words, moving from the least to most lenient judge increases the probability of pre-trial release by 53.6 percentage points, a 76.8 percent change from the mean release rate of 69.8 percentage points.

C. Instrument Validity

Existence of First Stage: To examine the first-stage relationship between judge leniency (Z_{itj}) and whether a defendant is released pre-trial ($Released_{itj}$), we estimate the following equation for defendant-case i, assigned to judge j at time t using a linear probability model, estimated separately for white and black defendants:

$$Released_{itj} = \gamma_W Z_{itj} + \pi_W \mathbf{X}_{it} + v_{itj} \tag{10}$$

$$Released_{itj} = \gamma_B Z_{itj} + \pi_B \mathbf{X}_{it} + v_{itj} \tag{11}$$

where the vector $\mathbf{X}_{it}$ includes court-by-time fixed effects. The error term v_{itj} 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_{itj} are leave-out (jackknife) measures of judge leniency that are allowed to vary across years and defendant race.

[15] Our leave-out procedure is essentially a reduced-form version of jackknife IV, with the leave-out leniency measure 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 and court-by-time fixed effects. In unreported results, jackknife IV and LIML estimates using the full set of judge fixed effects as instruments yield similar results.

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 (10) and (11), 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.

Table 2 presents formal first stage results from Equations (10) and (11) 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 38.9 percentage points more likely to be released pre-trial. Judge leniency is also highly predictive of pre-trial release for both white and black defendants, with the first-stage coefficient being 0.360 and 0.415, respectively.[16]

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

[16]Consistent with prior work using judge stringency as an instrumental variable (e.g., Bhuller et al. 2016), 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, reducing the accuracy of our leave-out measure. Consistent with this explanation, we find first stage coefficients ranging from 0.6 to 0.7 in Monte Carlo simulations when judge tendencies are fixed over the course of the year, and first stage coefficients ranging from 0.2 to 0.4 when judge tendencies are allowed to change within each year. It is important to note that attenuation bias due to sampling variation in our leniency measure does not bias our estimates since it affects both the first stage and reduced form proportionally.

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. (2018) find that there are no independent effects of the money bail amount or the non-monetary bail conditions on defendant outcomes, and that bail judge assignment is uncorrelated with the assignment of public defenders and subsequent trial judges.

Monotonicity: The final condition needed for our IV and MTE estimators is that the impact of judge assignment on the probability of pre-trial release is monotonic across defendants of the same race. 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. The monotonicity assumption is required in order to identify and interpret our IV estimator as a well-defined LATE and to estimate marginal treatment effects using the standard local instrument variables (LIV) approach. See Angrist et al. (1996) and Heckman and Vytlacil (2005) for additional details. Importantly, we allow our judge leniency measure to vary by defendant race to allow for the possibility that the degree of racial bias varies across judges. In practice, we observe that judge behavior is only imperfectly monotonic with respect to race (see Appendix Figure A1), with a regression of the ranking of each judge's leniency measure for whites on the ranking of each judge's leniency measure for blacks yielding a coefficient equal to 0.827 (se=0.010). The non-monotonic behavior we observe with respect to race is driven by approximately 17.9 percent of judges who hear about 8.2 percent of all cases. Consistent with the monotonicity assumption within race, we find a strong first-stage relationship across various case and defendant types (see Appendix Table A2).[17]

[17] One specific concern is that lenient judges may be 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. (2018), 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

III. Results

In this section, we present our main results applying our empirical test for racial bias. We then show the robustness of our results to alternative specifications, before comparing 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 Test for Racial Bias

IV Estimates: We begin by presenting IV estimates of racial bias that rely on relatively few auxiliary assumptions, but with the caveat that the weighting scheme underlying the estimator may not always be policy relevant. We estimate these IV results using the following two-stage least squares specifications for defendant-case i assigned to judge j at time t, estimated separately for white and black defendants:

$$Y_{itj} = \alpha_W^{IV} Released_{itj} + \beta_W \mathbf{X}_{it} + \mathbf{v}_{itj} \tag{12}$$

$$Y_{itj} = \alpha_B^{IV} Released_{itj} + \beta_B \mathbf{X}_{it} + \mathbf{v}_{itj} \tag{13}$$

where Y_{itj} is the probability of pre-trial misconduct, as measured by the probability of rearrest prior to case disposition. The vector $\mathbf{X}_{it}$ includes court-by-time fixed effects and 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. As described previously, the error term $\mathbf{v}_{itj} = \mathbf{U}_{itj} + \varepsilon_{itj}$ consists of characteristics unobserved by the econometrician but observed by the judge, $\mathbf{U}_{itj}$, and idiosyncratic variation unobserved by both the econometrician and judge, ε_{itj}. We instrument for pre-trial release, $Released_{itj}$, with our measure of judge leniency, Z_{itj}, that is allowed to vary across years and defendant race. Robust standard errors are two-way clustered at the individual and judge-by-shift level.

Estimates from Equations (12) and (13) are presented in columns 1-2 of Table 4. Column 3 reports our IV estimate of racial bias $D^{IV} = \alpha_W^{IV} - \alpha_B^{IV}$. Panel A of Table 4 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 rearrested for a crime that includes a drug offense, 6.7 percent of defendants rearrested for a crime that includes

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. (2018), who find that judges more or less agree on how to rank-order defendants based on their observable characteristics.

a property offense, and 6.1 percent of defendants rearrested for a crime that includes a violent offense.[18]

We find convincing evidence of racial bias against black defendants using our IV estimator. We find that marginally released white defendants are 23.6 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 insignificant 1.4 percentage points (column 2). Taken together, these IV estimates imply that marginally released white defendants are 22.2 percentage points more likely to be rearrested prior to disposition than marginally released black defendants (column 3), consistent with racial bias against blacks (p-value = 0.027). Importantly, we can reject the null hypothesis of no racial bias even assuming the maximum infra-marginality bias in our IV estimator of 1.1 percentage points (see Appendix B).

In Panel B, we find suggestive evidence of racial bias against black defendants across all crime types, although the point estimates are too imprecise to make definitive conclusions. For example, we find that marginally released whites are about 8.0 percentage points more likely to be rearrested for a violent crime prior to disposition than marginally released blacks (p-value = 0.173). Marginally released white defendants are also 4.7 percentage points more likely to be rearrested for a drug crime prior to case disposition than marginally released black defendants (p-value = 0.430), and 16.3 percentage points more likely to be rearrested for a property crime (p-value = 0.025). These results suggest that judges are likely racially biased against black defendants even if they are most concerned about minimizing specific types of new crime, such as violent crimes.

MTE Estimates: Our second set of estimates comes from our MTE estimator that allows us to put equal weight on each judge in our sample, but at the cost of additional auxiliary assumptions. We estimate these MTE results using a two-step procedure. First, we estimate the entire distribution of MTEs using the derivative of residualized rearrest before case disposition, $\ddot{Y}_{itj}$, with respect to variation in the propensity score provided by our instrument, p_r^j, separately for white and black defendants:

$$MTE_W(p_W^j) = \frac{\partial}{\partial p_W^j}\mathbb{E}(\ddot{Y}_{itj}|p_W^j, W) \tag{14}$$

$$MTE_B(p_B^j) = \frac{\partial}{\partial p_B^j}\mathbb{E}(\ddot{Y}_{itj}|p_B^j, B) \tag{15}$$

where p_r^j is the propensity score for release for judge j and defendant race r and $\ddot{Y}_{itj}$ is rearrest residualized using the full set of court-by-time fixed effects and baseline crime and defendant controls, $\mathbf{X}_{it}$. Following Heckman, Urzua, and Vytlacil (2006) and Doyle (2007), we also residualize Z_{itj} and $Released_{itj}$ using $\mathbf{X}_{it}$. We then regress the residualized release variable on the residualized judge

[18]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.

leniency measure to calculate p_r^j, a race-specific propensity score. Next, we compute the numerical derivative of a local quadratic estimator relating $\ddot{Y}_{itj}$ to p_r^j to estimate race-specific MTEs. See Figure 2 for estimates of the full distribution of MTEs by defendant race.

Second, we use the race-specific MTEs to calculate the level of racial bias for each judge j. We calculate the average level of bias across all bail judges using a simple average of these judge-specific estimates:

$$\sum_{j=1}^{J} \frac{1}{J} \left(MTE_W(p_W^j) - MTE_B(p_B^j) \right) \tag{16}$$

We calculate standard errors by bootstrapping this two-step procedure at the judge-by-shift level. See Appendix B for additional details.

Estimates from Equations (14) and (15) are presented in columns 4-5 of Table 4, with column 6 reporting our MTE equal-weighted estimate of racial bias D^{MTE} from Equation (16). Consistent with our IV estimates, we find that marginally released white defendants are 24.9 percentage points more likely to be rearrested for any crime compared to marginally detained white defendants (column 4), while the effect of pre-trial release on rearrest rates for the marginally released black defendants is a statistically insignificant 1.7 percentage points (column 5). Our MTE estimates therefore imply that marginally released white defendants are 23.1 percentage points more likely to be rearrested prior to disposition than marginally released black defendants (column 6), consistent with racial bias against black defendants (p-value = 0.048).

In addition, Figure 2 shows that the MTEs for white defendants lie strictly above the MTEs for black defendants, implying that marginally released white defendants are riskier than marginally released black defendants at all points in the judge leniency distribution. In other words, the results from Figure 2 show that there is racial bias against black defendants at every part of the judge leniency distribution. These results, along with the fact that both IV and MTE approaches yield qualitatively similar estimates of racial bias, suggest that both the choice of IV weights and the additional parametric assumptions required to estimate the race-specific MTEs do not greatly affect our estimates of racial bias.

B. Robustness

Appendix Table A4 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 Philadelphia, we find that our estimates are qualitatively similar when we use a measure of pre-trial misconduct defined as failure to appear, or when we define pre-trial misconduct as either failure to appear or rearrest in Philadelphia, and only rearrest in Miami. To further explore the possibility that judges may only care about minimizing specific types of new crime, Appendix Table A5 presents estimates for a subset of more serious crime types for which estimates of social costs are available, such as assault and robbery, and weights each individual estimate of D^{IV} and D^{MTE} by

the corresponding social cost.[19] While our estimates become less precise given the infrequency of certain types of new crime, a social cost-weighted estimate of D^{IV} yields a lower bound of $8,637 and an upper bound of $20,658, while a social cost-weighted estimate of D^{MTE} yields a lower bound of $7,573 and an upper bound of $21,197, suggesting that marginally released white defendants generate larger social costs than marginally released black defendants.

Appendix Table A6 explores the sensitivity of our main results to a number of different specifications, where columns 1-5 report estimates of D^{IV} and columns 6-8 report estimates of D^{MTE}. Columns 1 and 6 drop 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 to explore whether judge preferences for non-race characteristics, such as crime type, can explain our main results (see Section I.D and Appendix B for details). Columns 3 and 7 presents results comparing outcomes for marginal non-Hispanic white defendants and marginal black defendants. Columns 4 and 8 presents results clustering more conservatively at the individual and judge level. Column 5 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.[20] 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.

C. Comparison to Other Outcome Tests

Appendix Tables A7-A9 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 estimates will not vary by defendant race. In contrast to our IV and MTE tests, however, standard OLS estimates suggest racial bias against white 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 also 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 and MTE estimates of racial bias are 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 et al. (2017) for additional details on these results.

[19] We exclude rearrest for crime types that are extremely rare, i.e. murder and rape. We also exclude rearrest for crime types that cannot be categorized into the listed categories.

[20] In these specifications, the coefficient on monetary bail amount is -0.002 (p-value = 0.500) for white defendants and -0.001 (p-value = 0.184) for black defendants, suggesting that monetary bail amount has no significant independent effect on pre-trial misconduct, consistent with findings reported in Dobbie et al. (2018).

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.[21] Racial animus may be a particular concern in bail setting due to the relatively low number of minority bail judges, the rapid-fire determination of bail decisions, and the lack of face-to-face contact between defendants and judges. Prior work has shown that it is exactly these types of settings where racial prejudice is most likely to translate into the disparate treatment of minorities (e.g., Greenwald et al. 2009).

One suggestive piece of evidence against this hypothesis is provided by the Anwar and Fang (2006) test of relative racial bias 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 and MTE 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.

B. Racially Biased Prediction Errors in Risk

A second explanation for our results is that bail 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 of release. 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.

[21] 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).

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 relative to whites (both Hispanic and non-Hispanic). To test this idea, Figure 3 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.[22] Results for each individual characteristic in our predicted risk measure are also presented in Appendix Table A10. 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 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 under 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

[22] 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. (2018), 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).

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 A4. 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 $\theta = 1.9$.[23] 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-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 A5), where we also tend to find strong evidence of racial bias.

A final piece of evidence in support of stereotyping comes from a comparison of the Hispanic and black distributions of risk relative to the non-Hispanic white distribution. Recall that we find no evidence of racial bias against Hispanic defendants (see Appendix Table A3). 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 A6). 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. One such 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 a 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 A7). In contrast, while it appears plausible

[23] Our estimate of θ is quantitatively similar to the magnitude of stereotypes in explaining investor overreaction to stock market news and the formation of credit cycles (Bordalo et al. forthcoming, Bordalo et al. 2017).

that race-based prediction errors will decrease with experience, there is no reason to believe that racial animus will change with experience.[24]

To test this idea, columns 1-4 of Table 5 presents our estimates of racial bias, D^{IV} and D^{MTE}, separately by court. Although we caution that there are likely many differences in the criminal justice systems of the two cities in our sample, one distinction is the degree to which bail judges specialize in conducting bail hearings. In Philadelphia, bail judges are full-time judges who specialize in setting bail 24 hours a day, seven days a week, hearing an average of 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 = 0.325 for IV, p-value = 0.442 for MTE). In Miami, we find that marginally released white defendants are 25.1 percentage points more likely to be rearrested using our IV estimator (p-value = 0.029) and 24.9 percentage points more likely to be rearrested using our MTE estimator (p-value = 0.040), compared to marginally released black defendants. In Philadelphia, we find no statistically significant evidence of racial bias under either our IV or MTE estimates, suggesting the possible importance of experience in alleviating any prediction errors.[25]

Columns 5-8 of Table 5 provide additional evidence on this issue by exploiting 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. Consistent with our across-court findings, we find suggestive evidence that inexperienced judges are more racially biased than experienced judges (p-value = 0.193 for IV, p-value= 0.095 for MTE). Among inexperienced judges, we find that marginally released white defendants are 48.7 percentage points more likely to be rearrested using our IV estimator (p-value = 0.040) and 51.0 percentage points more likely to be rearrested using our MTE estimator (p-value = 0.029), compared to marginally released black defendants. Among experienced judges, we find no statistically significant evidence of racial bias under either our IV or MTE estimates.

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

[24] 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.

[25] Our IV estimate of racial bias in Philadelphia should be interpreted with some caution given that we only observe seven judges for this city in our data. The maximum infra-marginality bias of our IV estimator in Philadelphia is 16.4 percentage points, compared to only 1.6 percentage points in Miami-Dade. We note, however, that there is no infra-marginality bias of our MTE estimator for either city if we have correctly specified the shape of the MTE function.

Kleinberg et al. (2018) 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, ruling out statistical discrimination as the sole explanation for the racial disparities in bail. Our estimates are nearly identical if we account for observable crime and defendant differences by race, indicating that our results cannot be explained by black-white differences in the probability of being arrested for certain types of crimes (e.g., the proportion of felonies versus misdemeanors) or black-white differences in defendant characteristics (e.g., the proportion of defendants with prior offenses versus no prior offenses).

We find several pieces of evidence consistent with our results being driven by racially biased prediction errors in risk, as opposed to racial animus among bail judges. First, we find that both white and black bail judges are racially biased against black defendants, a finding that is inconsistent with most models of racial animus. Second, we find that black defendants are sufficiently over-represented in the right tail of the predicted risk distribution to rationalize observed racial disparities in release rates under a theory of stereotyping. Finally, racial bias is significantly higher among both part-time and inexperienced judges, and descriptive evidence suggests that experienced judges can better predict misconduct risk for all defendants. Taken together, these results are most consistent with a model of 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. (2018), our findings also suggest that providing judges with data-based risk assessments may also 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] Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." ILR Review, 30(2): 175-187.

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

[4] Angrist, Joshua, and Iván Fernández-Val. 2013. "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework." In *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress, Volume III: Econometrics*, Dewatripont, M., Hansen, L., and S. Turnovsky (Eds) Econometric Society Monographs, 401-434.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

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

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

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

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

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

[17] 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.

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

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

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

[21] Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. American Economic Review, 94(4): 991-1013.

[22] 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.

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

[24] Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2017. "Diagnostic Expectations and Stock Returns." NBER Working Paper No. 23863.

[25] Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. Forthcoming. "Diagnostic Expectations and Credit Cycles." Journal of Finance.

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

[27] 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.

[28] Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Loken, and Magne Mogstad. 2016. "Incarceration, Recidivism and Employment." NBER Working Paper No. 22648.

[29] 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.

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

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

[32] 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.

[33] 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.

[34] Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2016. "From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions." Labour Economics, 41: 47-60.

[35] DiNardo, John, Nicole Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and The Distribution of Wages, 1973-1993: A Semi-Parametric Approach." Econometrica, 64(5): 1001-1045.

[36] 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.

[37] Dobbie, Will, Jacob Goldin, and Crystal Yang. 2018. "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." American Economic Review, 108(2): 201-240.

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

[39] 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.

[40] Evdokimov, Kirill S., and Michal Kolesár. 2018. "Inference in Instrumental Variables Analysis with Heterogeneous Treatment Effects." Unpublished Working Paper.

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

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

[43] 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.

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

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

[46] 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.

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

[48] 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.

[49] Goncalves, Felipe, and Steven Mello. 2018 "A Few Bad Apples? Racial Bias in Policing." Unpublished Working Paper.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

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

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

[57] Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. "Human Decisions and Machine Predictions." Quarterly Journal of Economics, 133(1): 237-293.

[58] 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.

[59] 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.

[60] Leslie, Emily, and Nolan G. Pope. 2017. "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from NYC Arraignments." Journal of Law and Economics, 60(3): 529-557.

[61] 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.

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

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

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

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

[66] 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.

[67] 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.

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

[69] 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.

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

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

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

[73] 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.

[74] 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.

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

[76] 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 (Phl only)	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, as described in the text. Information on race, gender, age, and criminal outcomes is derived from court records. Released is defined as being released at any point before trial. Detained is defined as never being released before trial. See Appendix D for additional details on the sample and variable construction.

Table 2: First Stage Results

	All Defendants		White		Black	
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-trial Release	0.405***	0.389***	0.373***	0.360***	0.434***	0.415***
	(0.027)	(0.025)	(0.036)	(0.032)	(0.036)	(0.033)
	[0.698]	[0.698]	[0.711]	[0.711]	[0.688]	[0.688]
Court x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	No	Yes	No	Yes	No	Yes
Observations	256,253	256,253	106,846	106,846	149,407	149,407

Note: This table reports the first-stage relationship between pre-trial release and judge leniency. 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, constructed separately by defendant race, following the procedure described in Section II.B. All regressions include court-by-time fixed effects. Baseline controls include 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, and other), crime severity (felony and misdemeanor), and indicators for any missing controls. 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***	−0.00005	−0.08593***	0.00004	−0.10379***	−0.00014
	(0.00235)	(0.00024)	(0.00325)	(0.00038)	(0.00323)	(0.00031)
Age at Bail Decision	−0.01725***	−0.00009	−0.02250***	−0.00015	−0.01512***	−0.00005
	(0.00086)	(0.00009)	(0.00127)	(0.00016)	(0.00104)	(0.00010)
Prior Offense in Past Year	−0.14922***	−0.00017	−0.16817***	0.00030	−0.13411***	−0.00044
	(0.00287)	(0.00028)	(0.00445)	(0.00046)	(0.00362)	(0.00036)
Arrested on Bail in Past Year	0.01066***	0.00004	0.01967***	−0.00166***	0.00495	0.00116***
	(0.00355)	(0.00034)	(0.00552)	(0.00057)	(0.00439)	(0.00042)
Failed to Appear in Court in Past Year	0.03318***	0.00012	0.03253***	0.00104**	0.03245***	−0.00047
	(0.00413)	(0.00025)	(0.00631)	(0.00043)	(0.00529)	(0.00031)
Number of Offenses	−0.02090***	−0.00001	−0.01829***	−0.00002	−0.02131***	0.00000
	(0.00053)	(0.00003)	(0.00085)	(0.00006)	(0.00063)	(0.00004)
Felony Offense	−0.17618***	−0.00003	−0.18817***	−0.00014	−0.16948***	0.00004
	(0.00257)	(0.00012)	(0.00397)	(0.00020)	(0.00323)	(0.00014)
Any Drug Offense	0.03514***	−0.00038	0.02558***	−0.00002	0.04069***	−0.00063*
	(0.00258)	(0.00026)	(0.00357)	(0.00039)	(0.00332)	(0.00032)
Any Property Offense	−0.04272***	−0.00013	−0.05560***	0.00009	−0.03188***	−0.00029
	(0.00285)	(0.00026)	(0.00388)	(0.00041)	(0.00354)	(0.00033)
Any Violent Offense	0.01640***	0.00028	0.07515***	0.00033	−0.02443***	0.00023
	(0.00389)	(0.00025)	(0.00497)	(0.00045)	(0.00429)	(0.00029)
Joint F-Test	[0.00000]	[0.60067]	[0.00000]	[0.21951]	[0.00000]	[0.08289]
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, constructed separately by defendant race, 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

	IV Results			MTE Results		
	White	Black	D^{IV}	White	Black	D^{MTE}
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)
Rearrest Prior to Disposition	0.236***	0.014	0.222**	0.249***	0.017	0.231**
	(0.073)	(0.070)	(0.101)	(0.084)	(0.080)	(0.117)
	[0.172]	[0.182]	–	[0.172]	[0.182]	–
Panel B: Rearrest by Crime Type						
Rearrest for Drug Crime	0.067	0.019	0.047	0.074	−0.024	0.097
	(0.043)	(0.043)	(0.060)	(0.048)	(0.054)	(0.074)
	[0.077]	[0.081]	–	[0.077]	[0.081]	–
Rearrest for Property Crime	0.158***	−0.005	0.163**	0.149**	0.043	0.106
	(0.057)	(0.047)	(0.073)	(0.066)	(0.053)	(0.084)
	[0.065]	[0.068]	–	[0.065]	[0.068]	–
Rearrest for Violent Crime	0.079**	−0.000	0.080	0.082*	−0.001	0.083
	(0.039)	(0.042)	(0.058)	(0.044)	(0.050)	(0.068)
	[0.047]	[0.071]	–	[0.047]	[0.071]	–
Observations	106,846	149,407	–	106,846	149,407	–

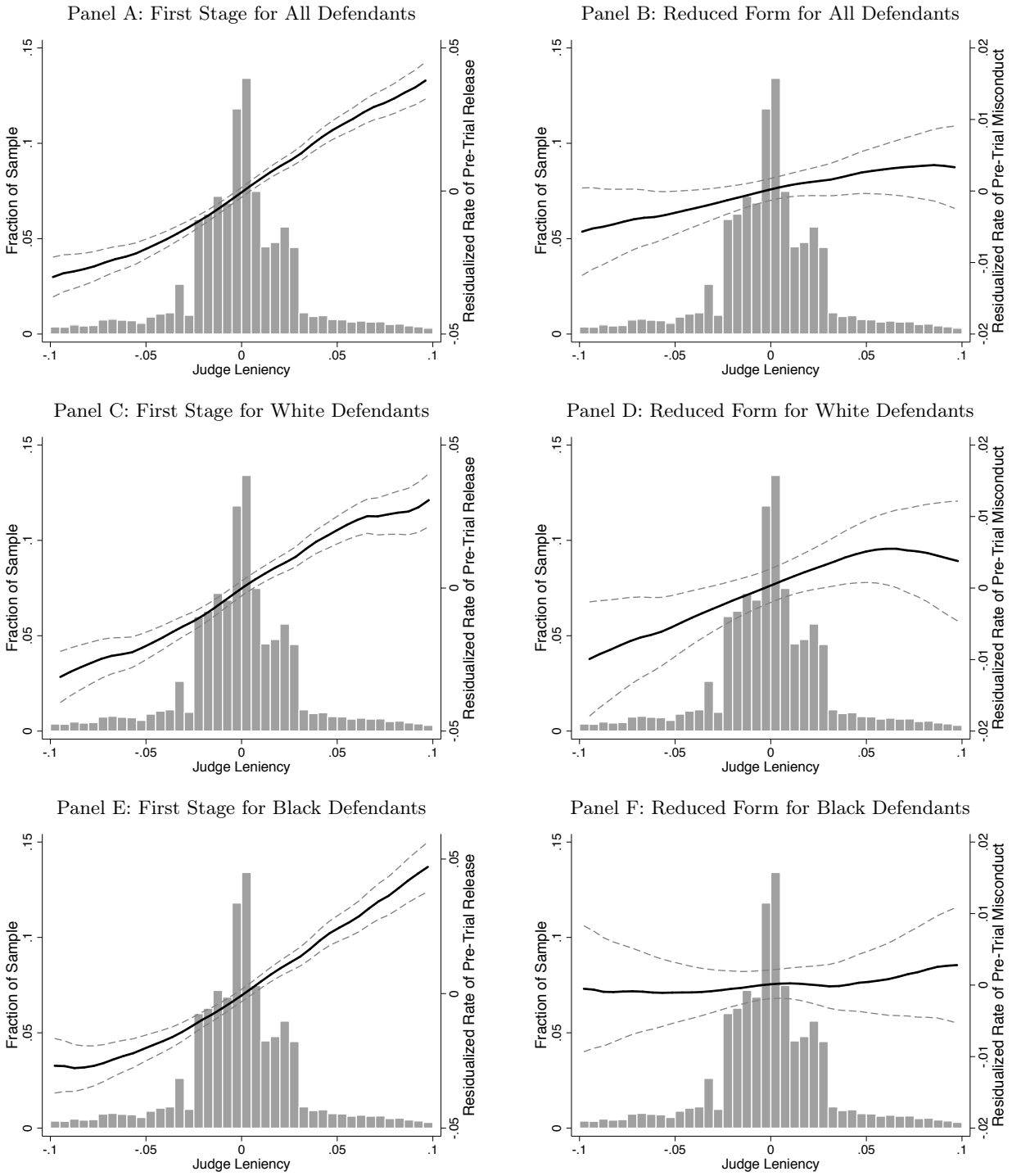
Note: This table reports estimates of racial bias in pre-trial release based on rearrest prior to case disposition. Columns 1-2 report two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race, while column 3 reports the difference between the white and black two-stage least squares coefficients, or D^{IV} as described in the text. Columns 1-3 use IV weights for each specification and report robust standard errors two-way clustered at the individual and judge-by-shift level in parentheses. Columns 4-5 report the average marginal treatment effect of the impact of pre-trial release on the probability of pre-trial misconduct separately by race, while column 6 reports the difference between the white and black MTE coefficients, or D^{MTE} as described in the text. Columns 4-6 use equal weights for each judge and report bootstrapped standard errors clustered at the judge-by-shift level in parentheses. All specifications control for court-by-time fixed effects and defendant 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. The sample means of the dependent variables are reported in brackets. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 5: Racial Bias in Pre-Trial Release by Judge Experience

	Judge Specialization				Judge Experience			
	Miami D^{IV}	Miami D^{MTE}	Phl D^{IV}	Phl D^{MTE}	Miami Low Exp D^{IV}	Miami Low Exp D^{MTE}	Miami High Exp D^{IV}	Miami High Exp D^{MTE}
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rearrest Prior to Disposition	0.251**	0.249**	0.040	0.078	0.487**	0.510**	0.144	0.036
	(0.114)	(0.121)	(0.184)	(0.195)	(0.237)	(0.233)	(0.178)	(0.164)
	[0.149]	[0.149]	[0.194]	[0.194]	[0.148]	[0.148]	[0.152]	[0.152]
Panel B: Rearrest by Crime Type								
Rearrest for Drug Crime	0.053	0.103	0.008	0.015	0.141	0.185	−0.013	0.006
	(0.066)	(0.077)	(0.138)	(0.150)	(0.119)	(0.138)	(0.101)	(0.110)
	[0.057]	[0.057]	[0.092]	[0.092]	[0.057]	[0.057]	[0.057]	[0.057]
Rearrest for Property Crime	0.196**	0.127	−0.031	−0.014	0.296**	0.293*	0.146	0.035
	(0.084)	(0.096)	(0.110)	(0.199)	(0.140)	(0.163)	(0.111)	(0.124)
	[0.078]	[0.078]	[0.060]	[0.060]	[0.078]	[0.078]	[0.079]	[0.079]
Rearrest for Violent Crime	0.082	0.079	0.065	0.066	0.204	0.218*	0.032	−0.036
	(0.065)	(0.075)	(0.115)	(0.119)	(0.134)	(0.119)	(0.100)	(0.099)
	[0.050]	[0.050]	[0.067]	[0.067]	[0.048]	[0.048]	[0.051]	[0.051]
Observations	93,417	93,417	162,836	162,836	47,692	47,692	45,725	45,725

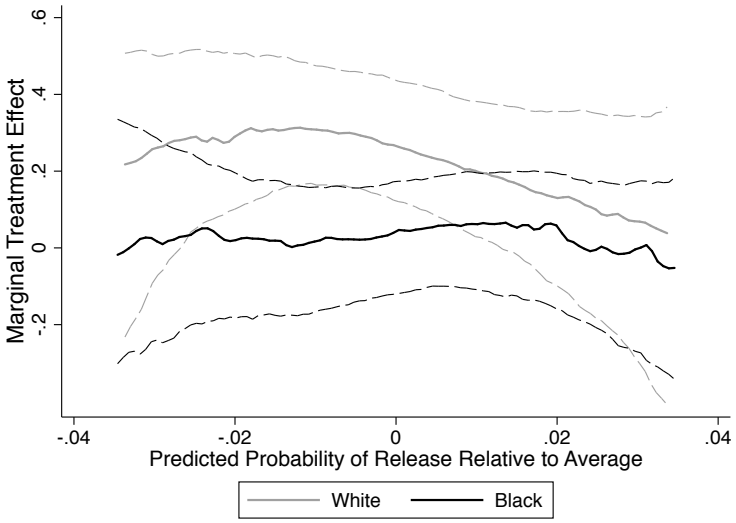
This table reports estimates of racial bias for different subgroups of judges. D^{IV} is the difference between the white and black two-stage least squares estimates coefficients of pre-trial release on pre-trial misconduct using IV weights. D^{MTE} is the difference between the average white and black MTE estimates of pre-trial release on pre-trial misconduct using equal weights by judge. Columns 1-2 report estimates for non-specialist bail judges in Miami-Dade. Columns 3-4 report estimates for specialist bail judges in Philadelphia. Columns 5-8 report estimates for non-specialist bail judges in Miami with below and above median years of experience. The sample is described in the notes to Table 1. The dependent variable is listed in each row. All specifications control for court-by-time fixed effects and defendant 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. The sample means of the dependent variables are reported in brackets. For IV specifications, robust standard errors two-way clustered at the individual and judge-by-shift level reported in parentheses. For MTE specifications, bootstrapped standard errors clustered at the judge-by-shift level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Figure 1: First Stage and Reduced Form Results



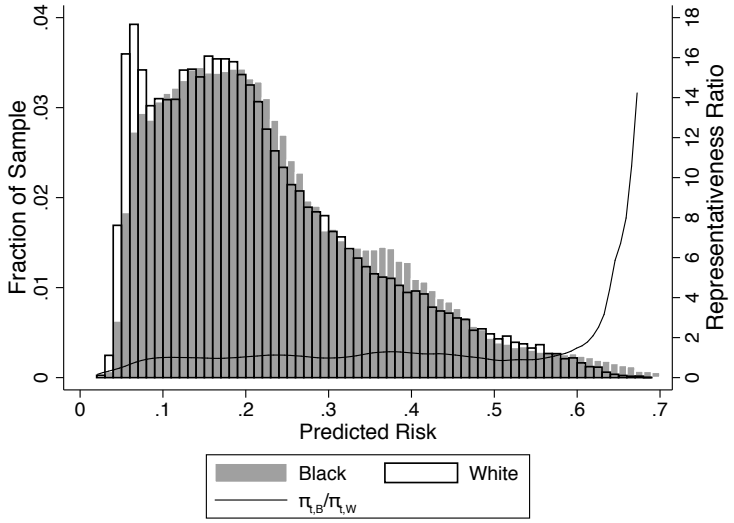
Note: These figures report the first stage and reduced form relationships between defendant outcomes and judge leniency. The regressions are estimated on the sample as described in the notes to Table 1. Judge leniency is estimated using data from other cases assigned to a bail judge in the same year, constructed separately by defendant race, following the procedure described in Section II.B. In the first stage regressions, the solid line is a local linear regression of pre-trial release on judge leniency. In the reduced form regressions, the solid line is a local linear regression of pre-trial misconduct on judge leniency. All regressions include court-by-time fixed effects and two-way cluster standard errors at the individual and judge-by-shift level.

Figure 2: Marginal Treatment Effects



Note: This figure reports the marginal treatment effects (MTEs) of pre-trial release on pre-trial rearrest separately by race. 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 = 0.030). 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 clustered at the judge-by-shift level. See the text for additional details.

Figure 3: Predicted Risk Distribution by Defendant Race



Note: This figure reports the predicted distribution of pre-trial misconduct risk separately by race. Pre-trial misconduct risk is estimated using the machine learning algorithm described in Appendix F. The solid line represents the representativeness ratio for black versus white defendants as described in the text, or the estimated misconduct risk for blacks divided by the estimated misconduct risk for whites. See the text for additional details.

Appendix A: Additional Results

Appendix Table A1: Racial Bias in the Assignment of Non-Monetary Bail

	White	Black	D^{IV}
Panel A: Pre-Trial Release	(1)	(2)	(3)
Pre-trial Release	0.490***	0.511***	−0.021
	(0.081)	(0.045)	(0.092)
	[0.711]	[0.688]	–
Panel B: Pre-Trial Misconduct			
Rearrest Prior to Disposition	0.085*	−0.009	0.094
	(0.050)	(0.039)	(0.065)
	[0.172]	[0.182]	–
Rearrest for Drug Crime	0.060**	−0.026	0.086**
	(0.030)	(0.026)	(0.041)
	[0.077]	[0.081]	–
Rearrest for Property Crime	0.087**	0.001	0.086*
	(0.037)	(0.029)	(0.048)
	[0.065]	[0.068]	–
Rearrest for Violent Crime	0.033	0.010	0.022
	(0.029)	(0.027)	(0.040)
	[0.047]	[0.071]	–
Observations	106,846	149,407	–

Note: This table reports estimates of the impact of assigning non-monetary bail (defined as both ROR and non-monetary conditions) versus monetary bail on pre-trial release (Panel A) and pre-trial misconduct (Panel B). Columns 1-2 report two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race, while column 3 reports the difference between the white and black two-stage least squares coefficients, or D^{IV} as described in the text. All specifications use IV weights for each specification and report robust standard errors two-way clustered at the individual and judge-by-shift level in parentheses. All specifications also control for court-by-time fixed effects and defendant 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. The sample means of the dependent variables are reported in brackets. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A2: 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	0.584***	0.204***	0.516***	0.364***	0.119***	0.452***	0.346***
	(0.042)	(0.035)	(0.046)	(0.048)	(0.041)	(0.038)	(0.028)
	[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 the first stage relationship between pre-trial release and judge leniency in different subsamples. 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, constructed separately by defendant race, following the procedure described in Section II.B. All regressions include court-by-time fixed effects and baseline controls for 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, and other), crime severity (felony and misdemeanor), and indicators for any missing controls. 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.

Appendix Table A3: White-Hispanic Bias in Pre-Trial Release

	IV Results			MTE Results		
	White	Hispanic	D^{IV}	White	Hispanic	D^{MTE}
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)
Rearrest Prior to Disposition	0.274***	0.246**	0.028	0.206**	0.261**	−0.055
	(0.099)	(0.119)	(0.147)	(0.093)	(0.130)	(0.161)
	[0.167]	[0.176]	–	[0.167]	[0.176]	–
Panel B: Rearrest by Crime Type						
Rearrest for Drug Crime	0.098	0.079	0.020	0.030	0.099	−0.069
	(0.067)	(0.068)	(0.093)	(0.061)	(0.072)	(0.098)
	[0.066]	[0.087]	–	[0.066]	[0.087]	–
Rearrest for Property Crime	0.117	0.211**	−0.094	0.112*	0.167	−0.055
	(0.073)	(0.102)	(0.125)	(0.068)	(0.115)	(0.133)
	[0.066]	[0.064]	–	[0.066]	[0.064]	–
Rearrest for Violent Crime	0.012	0.141**	−0.129	0.014	0.146*	−0.131
	(0.052)	(0.069)	(0.088)	(0.052)	(0.075)	(0.091)
	[0.043]	[0.052]	–	[0.043]	[0.052]	–
Observations	35,914	48,447	–	35,914	48,447	–

Note: This table reports estimates of white non-Hispanic versus white Hispanic bias in pre-trial release based on rearrest prior to case disposition. Columns 1-2 report two-stage least squares results of the impact of pre-trial release on the probability of pre-trial misconduct separately by race, while column 3 reports the difference between the white non-Hispanic and white Hispanic two-stage least squares coefficients, or D^{IV} as described in the text. Columns 1-3 use IV weights for each specification and report robust standard errors two-way clustered at the individual and judge-by-shift level in parentheses. Columns 4-5 report the average marginal treatment effect of the impact of pre-trial release on the probability of pre-trial misconduct separately by race, while column 6 reports the difference between the white non-Hispanic and white Hispanic MTE coefficients, or D^{MTE} as described in the text. Columns 4-6 use equal weights for each judge and report bootstrapped standard errors clustered at the judge-by-shift level in parentheses. All specifications control for court-by-time fixed effects and defendant 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. The sample means of the dependent variables are reported in brackets. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A4: Results for Other Definitions of Pre-Trial Misconduct

	Philadelphia		Miami		Pooled	
	D^{IV}	D^{MTE}	D^{IV}	D^{MTE}	D^{IV}	D^{MTE}
	(1)	(2)	(3)	(4)	(5)	(6)
Rearrest	0.045	0.078	0.263**	0.249**	0.222**	0.231**
	(0.183)	(0.194)	(0.115)	(0.121)	(0.101)	(0.117)
	[0.194]	[0.194]	[0.149]	[0.149]	[0.178]	[0.178]
FTA	−0.024	0.006	–	–	–	–
	(0.187)	(0.202)				
	[0.204]	[0.204]	–	–	–	–
FTA or Rearrest	0.008	0.042	0.263**	0.249**	0.208**	0.314*
	(0.209)	(0.221)	(0.115)	(0.121)	(0.102)	(0.189)
	[0.318]	[0.318]	[0.149]	[0.149]	[0.256]	[0.256]
Observations	162,836	162,836	93,417	93,417	256,253	256,253

Note: This table reports estimates of racial bias in pre-trial release based on rearrest prior to case disposition, FTA (available only in Philadelphia), and either rearrest or FTA. Columns 1-2 report two-stage least squares estimates of D^{IV} and MTE estimates of D^{MTE} for Philadelphia. Columns 3-4 report two-stage least squares estimates of D^{IV} and MTE estimates of D^{MTE} for Miami. Columns 5-6 report two-stage least squares estimates of D^{IV} and MTE estimates of D^{MTE} for the pooled sample. For IV specifications, robust standard errors two-way clustered at the individual and judge-by-shift level reported in parentheses. For MTE specifications, bootstrapped standard errors clustered at the judge-by-shift level are reported in parentheses. All specifications control for court-by-time fixed effects and defendant 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. The sample means of the dependent variables are reported in brackets. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A5: Social Cost of Crime Results

	D^{IV} Estimate	D^{MTE} Estimate	Lower Bound	Upper Bound
	(1)	(2)	(3)	(4)
Rearrest for Robbery	0.028	0.035	$73,196	$333,701
	(0.034)	(0.037)		
Rearrest for Assault	0.068	0.065	$41,046	$109,903
	(0.050)	(0.057)		
Rearrest for Burglary	0.047	0.018	$50,291	$50,291
	(0.048)	(0.058)		
Rearrest for Theft	0.118*	0.081	$9,598	$9,974
	(0.062)	(0.075)		
Rearrest for Drug	0.047	0.097	$2,544	$2,544
	(0.060)	(0.067)		
Rearrest for DUI	0.007	0.016	$25,842	$25,842
	(0.009)	(0.012)		

Note: This table reports the difference in two-stage least squares and marginal treatment effect 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. In column 1, robust standard errors two-way clustered at the individual and judge-by-shift level are reported in parentheses. In column 2, bootstrap standard errors clustered at the judge-by-shift level are reported in parentheses. All specifications control for court-by-time fixed effects and defendant 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. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A6: Robustness Results

	Estimates of D^{IV}					Estimates of D^{MTE}		
	Drop Impossible	Re-Weight Char.	Drop Hispanics	Cluster by Judge	Control Bail $	Drop Impossible	Drop Hispanics	Cluster by Judge
Panel A: Rearrest for All Crimes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rearrest Prior to Disposition	0.215**	0.238**	0.238**	0.222**	0.252**	0.221**	0.259*	0.231*
	(0.095)	(0.103)	(0.119)	(0.102)	(0.104)	(0.098)	(0.151)	(0.126)
Panel B: Rearrest by Crime Type								
Drug Crime	0.059	0.055	0.066	0.047	0.054	0.109	0.094	0.097
	(0.059)	(0.057)	(0.080)	(0.061)	(0.062)	(0.069)	(0.097)	(0.073)
Property Crime	0.150**	0.181**	0.105	0.163**	0.178**	0.096	0.074	0.106
	(0.066)	(0.076)	(0.083)	(0.077)	(0.076)	(0.074)	(0.106)	(0.097)
Violent Crime	0.059	0.103*	0.006	0.080	0.100	0.048	−0.004	0.083
	(0.054)	(0.060)	(0.069)	(0.064)	(0.062)	(0.062)	(0.088)	(0.070)
Observations	252,992	256,253	170,923	256,253	256,253	252,992	170,923	256,253

Note: This table reports robustness checks for our estimates of D^{IV} and D^{MTE}. Columns 1 and 6 drop the four percent of 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 following the procedure described in Appendix B. Columns 3 and 7 drop Hispanic whites from the sample. Column 4 reports standard errors clustered by defendant and judge. Column 5 instruments for monetary bail amount with a leave-out leniency measure constructed using monetary bail amount. Column 8 reports standard errors clustered by judge. All regressions include court-by-time fixed effects and baseline controls for 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, and other), crime severity (felony and misdemeanor), and indicators for any missing controls. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A7: 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			
Rearrest for Drug Crime	0.097***	0.103***	−0.006**
	(0.002)	(0.002)	(0.002)
	[0.077]	[0.081]	–
Rearrest for Property Crime	0.067***	0.073***	−0.006*
	(0.002)	(0.002)	(0.003)
	[0.065]	[0.068]	–
Rearrest for 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 reports OLS results of racial bias in pre-trial release based on rearrest prior to case disposition. The regressions are estimated on the sample as described in the notes to Table 1. Columns 1-2 report OLS estimates of the impact of pre-trial release on the probability of pre-trial misconduct separately by race, while column 3 reports the difference between the white and black OLS coefficients. 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 and defendant 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. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A8: Mean Pre-Trial Release and Misconduct Rates by Judge and Defendant Race

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

Note: This table presents mean 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 A9: p-values from Tests of Relative Racial Prejudice

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

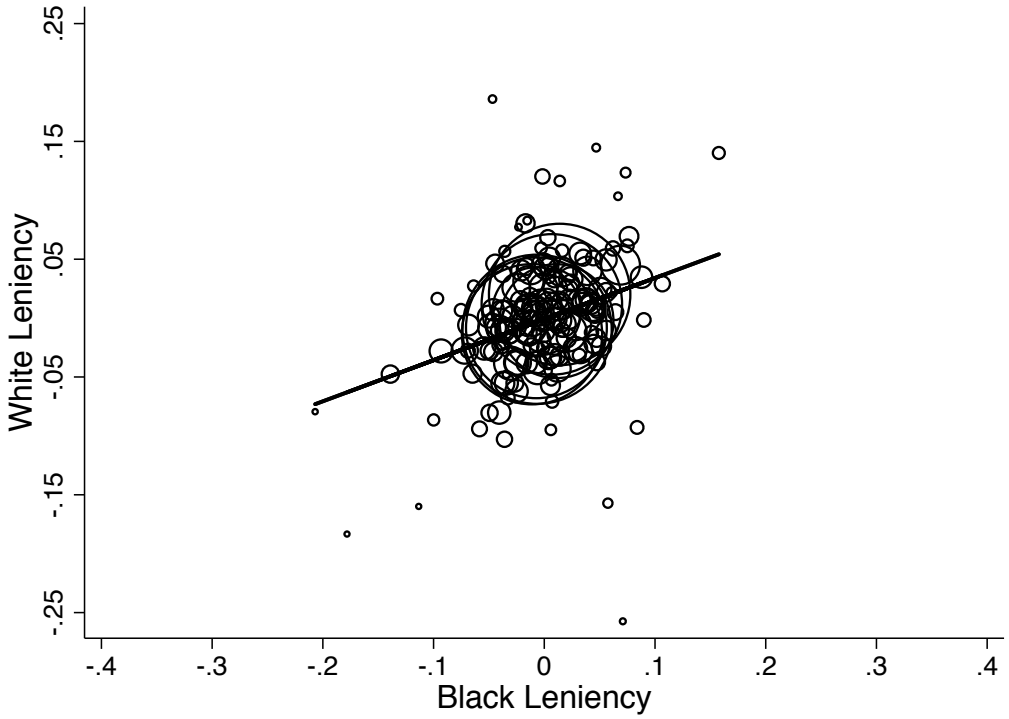
Note: This table replicates the Anwar and Fang (2006) test for pre-trial release rates and pre-trial misconduct rates. This table presents bootstrapped p-values testing for relative racial bias. The null hypothesis is rejected if white judges are more lenient on white defendants, and black judges are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

Appendix Table A10: 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

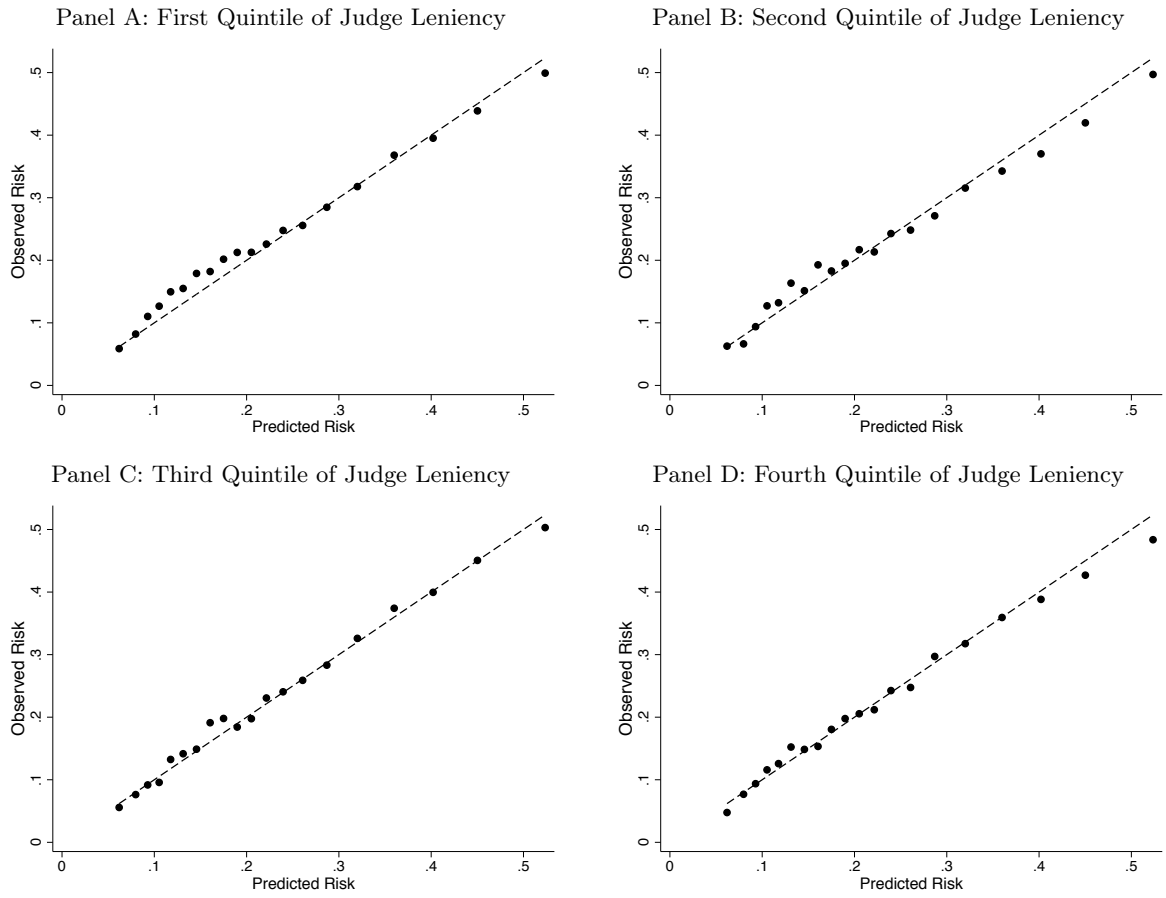
Note: This table reports the mean of the variable listed in the row given the defendant is black, divided by the mean of the variable listed in the row given the defendant is white. The sample is described in the notes to Table 1.

Appendix Figure A1: Judge Leniency by Race



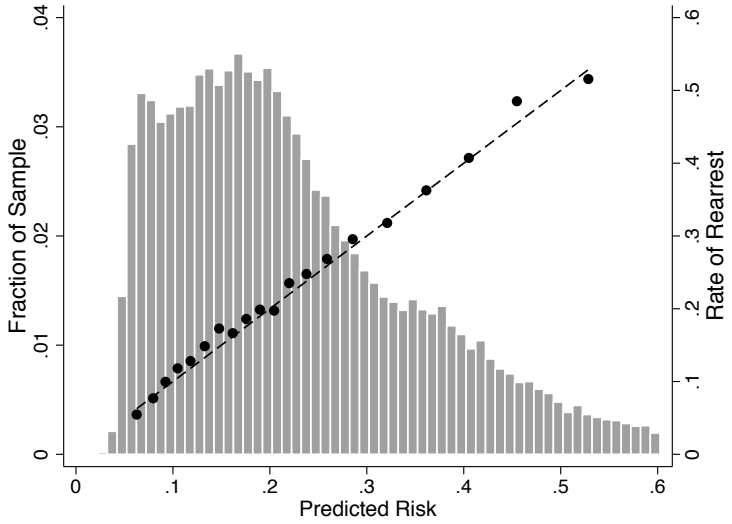
Note: These figures show the correlation between our residualized measure of judge leniency by defendant race over all available years of data. We also plot the linear best fit line estimated using OLS.

Appendix Figure A2: Predicted and Actual Risk by Judge Leniency



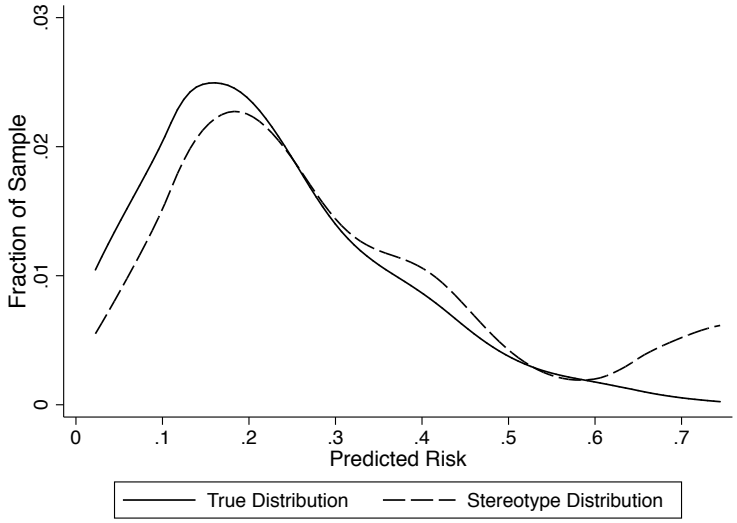
Note: These figures plot predicted pre-trial misconduct risk against actual pre-trial misconduct for different judge-leniency quintiles. Predicted risk is calculated using only cases from the most lenient quintile of judges and the machine learning algorithm described in Appendix F. See the text for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

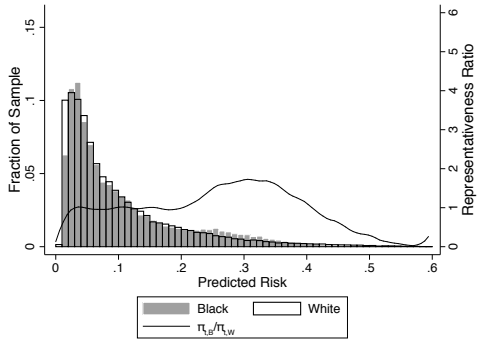
Appendix Figure A4: Stereotyped and True Distribution of Risk for Black Defendants



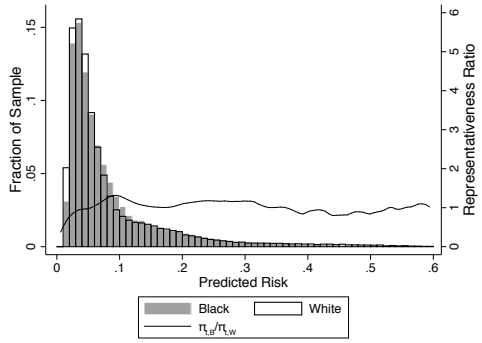
Note: This figure plots the true distribution of risk for black defendants alongside the perceived distribution of risk for black defendants. The stereotyped beliefs are generated by a representativeness-based discounting model with $\theta = 1.9$. This value of θ rationalizes an average release rate 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 A5: Crime-Specific Predicted Risk Distributions by Race

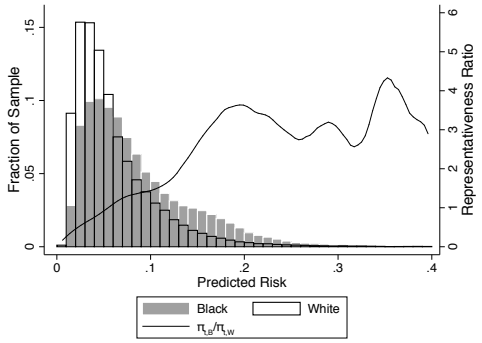
Panel A: Rearrest for Drug Crime



Panel B: Rearrest for Property Crime

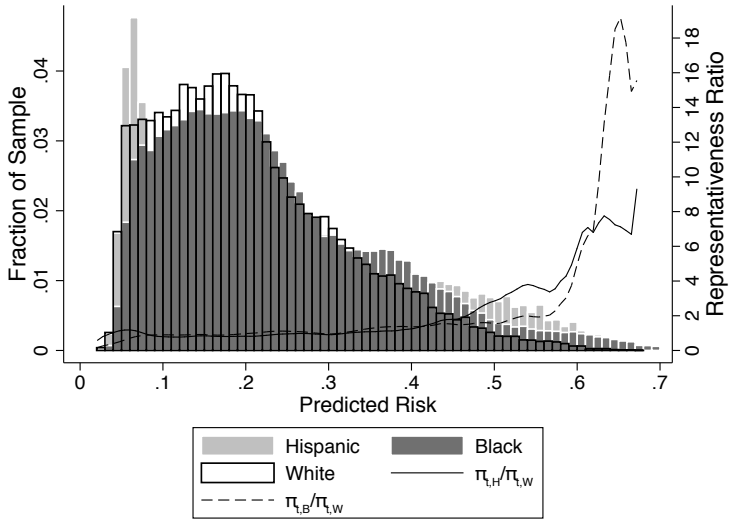


Panel C: Rearrest for Violent Crime



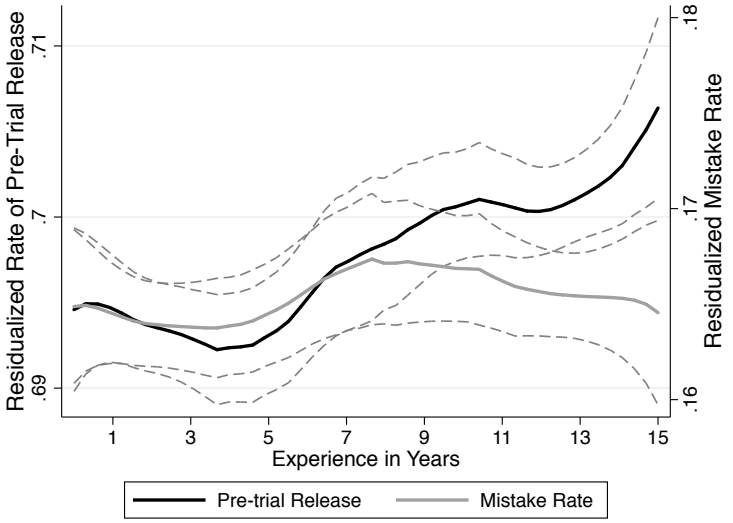
Note: These figures report the distribution of crime-specific risk separately by defendant race. Predicted risk is calculated using the machine learning algorithm described in Appendix F. The solid line in each figure represents the representativeness ratio for black versus white defendants. See the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

Appendix Figure A6: 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 A7: 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 for IV and MTE Estimators

This appendix reviews our empirical test for racial bias before providing additional details and proofs for both our IV and MTE estimation approaches. For completeness, we also include all relevant information from the main text in this appendix.

A. Overview

Recall that 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. Let the true weighted average of treatment effects for defendants of race r at the margin of release for judge j, α_r^j, for some weighting scheme, w^j, across all bail judges, $j = 1...J$, be given by:

$$\begin{aligned} \alpha_r^{*,w} &= \sum_{j=1}^{J} w^j \alpha_r^j \\ &= \sum_{j=1}^{J} w^j t_r^j \end{aligned} \tag{B.1}$$

where w^j are non-negative weights which sum to one that will be discussed in further detail below. Recall that, by definition, $\alpha_r^j = t_r^j$. Intuitively, $\alpha_r^{*,w}$ represents a weighted average of the treatment effects for defendants of race r at the margin of release across all judges.

Following this notation, the true average level of racial bias among bail judges, $D^{*,w}$, for the weighting scheme w^j is given by:

$$\begin{aligned} D^{*,w} &= \sum_{j=1}^{J} w^j \left(t_W^j - t_B^j \right) \\ &= \sum_{j=1}^{J} w^j t_W^j - \sum_{j=1}^{J} w^j t_B^j \\ &= \alpha_W^{*,w} - \alpha_B^{*,w} \end{aligned} \tag{B.2}$$

From Equation (B.1), we can express $D^{*,w}$ 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.

We develop two estimators for racial bias that use variation in the release tendencies of quasi-randomly assigned bail judges to identify differences in pre-trial misconduct rates at the margin of release. In theory, an estimator for $D^{*,w}$ should satisfy three criteria: (1) rely on minimal auxiliary assumptions to estimate judge-specific thresholds of release, t_r^j, (2) yield statistically precise estimates of the average level of bias, $D^{*,w}$, and (3) use a policy-relevant weighting scheme, w^j. In practice, however, no single estimator can accomplish all three criteria in our setting. The two-stage least squares IV estimator, for example, relies on relatively few auxiliary assumptions and provides

statistically precise estimates by giving greater weight to more precise LATEs, but the particular weighting of the pairwise LATEs may not always yield a policy-relevant estimate of racial bias. In contrast, a fully non-parametric approach where one reports each pairwise LATE separately and allows a researcher to choose a weighting scheme can yield a policy-relevant interpretation of racial bias with minimal assumptions, but often comes at the cost of statistical precision since any particular LATE is often estimated with considerable noise. The MTE framework developed by Heckman and Vytlacil (1999, 2005) provides a third option, allowing a researcher to estimate judge-specific treatment effects for white and black defendants at the margin of release and thus choose a weighting scheme, but with estimation of racial bias for each judge, and relatedly statistical precision, coming at the cost of additional auxiliary assumptions.

In this Appendix, we show that both IV and MTE estimators yield qualitatively similar estimates of the average level of racial bias in our setting, suggesting that neither the choice of IV weights nor the additional parametric assumptions required under our MTE approach greatly affect our estimates. In contrast, we show that the fully non-parametric approach yields uninformative estimates of the average level of racial bias due to very imprecise estimates of the individual pairwise LATEs.

B. Instrumental Variables Framework

Our first estimator uses IV weights, defined as $w^j = \lambda^j$, when estimating the weighted average level of bias, $D^{*,w}$. Recall that λ^j are the standard IV weights defined in Imbens and Angrist (1994). Our IV estimator allows us to estimate a weighted average of racial bias across bail judges with relatively few auxiliary assumptions, but with the caveats that we cannot estimate judge-specific treatment effects and the weighting scheme underlying the IV estimator may not be policy relevant. If the IV weights are uncorrelated with the level of racial bias for a given judge, then our IV estimator will estimate the average level of discrimination across all bail judges. If the IV weights are correlated with the level of racial bias, however, then our IV estimator may under or overestimate the average level of racial bias across all bail judges, but may still be of policy relevance depending on the parameter of interest (e.g., an estimate of racial bias that puts more weights on judges with higher caseloads).

In this subsection, we present a formal definition of the IV-weighted level of racial bias and our IV estimator, provide proofs for consistency, discuss tests of the identifying assumptions, the interpretation of the IV-weighted estimate, and the potential bias of our IV estimator from using a discrete instrument. We then consider a re-weighting procedure that accounts for judge bias on observable non-race characteristics.

B.1. Definition and Consistency of IV Estimator

Definition: Let the IV-weighted level of racial bias, $D^{*,IV}$ be defined as:

$$\begin{aligned} D^{*,IV} &= \sum_{j=1}^{J} w^j \left(t_W^j - t_B^j \right) \\ &= \sum_{j=1}^{J} \lambda^j \left(t_W^j - t_B^j \right) \end{aligned} \tag{B.3}$$

where $w^j = \lambda^j$, the instrumental variable weights defined in Imbens and Angrist (1994) and described in the main text.

Following the definition in the main text, let our IV estimator be defined as:

$$\begin{aligned} D^{IV} &= \alpha_W^{IV} - \alpha_B^{IV} \\ &= \sum_{j=1}^{J} \lambda_W^j \alpha_W^{j,j-1} - \sum_{j=1}^{J} \lambda_B^j \alpha_B^{j,j-1} \end{aligned} \tag{B.4}$$

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

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 the IV-weighted level of racial bias, $D^{*,IV}$.

First Condition for Consistency: The first condition for our IV estimator D^{IV} to provide a consistent estimate of $D^{*,IV}$ 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.

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

$$\alpha_r^j = \alpha_r(z = z_j) = \lim_{dz \to 0} \mathbb{E}[Y_i(1) - Y_i(0) | Released_i(z) - Released_i(z - dz) = 1] \tag{B.5}$$

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

$$\alpha_r = \int \lambda_r(z) \alpha_r(z) dz \tag{B.6}$$

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

$$\lambda_r(z) = \frac{\frac{\partial Released_r}{\partial z}(z) \cdot \int_z^{\bar{z}} (y - \mathbb{E}[z]) \cdot f_r^z(y) dy}{\int_{\underline{z}}^{\bar{z}} \frac{\partial Released_r}{\partial z}(v) \cdot \int_v^{\bar{z}} (y - \mathbb{E}[z]) \cdot f_r^z(y) dy dv} \tag{B.7}$$

where $\frac{\partial Released_r}{\partial z}$ is the derivative of the probability of release with respect to leniency and f_r^z is the probability density function of leniency. If $\frac{\partial Released_r}{\partial z} \geq 0$ 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. □

Second Condition for Consistency: The second condition for our IV estimator D^{IV} to provide a consistent estimate of $D^{*,IV}$ is that the weights λ_r^j must be equal across race. Equal weights ensure that the race-specific IV estimates from Equation (7) in the main text, 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:

$$D^{IV} = \sum_{j=1}^{J} \lambda^j (\alpha_W^{j,j-1} - \alpha_B^{j,j-1}) \tag{B.8}$$

Proof of Consistency: We combine these two conditions to establish the consistency of our IV estimator. Recall that our IV estimator D^{IV} provides a consistent estimate of racial bias $D^{*,IV}$ if (1) Z_i is continuous and (2) λ_r^j is constant by race.

To begin, we write D^{IV} as:

$$\begin{aligned} D^{IV} &= \alpha_W^{IV} - \alpha_B^{IV} \\ &= \sum_{j=1}^{J} \lambda_W^j \alpha_W^{j,j-1} - \sum_{j=1}^{J} \lambda_B^j \alpha_B^{j,j-1} \end{aligned} \tag{B.9}$$

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

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

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

$$\begin{aligned} D^{IV} &= \int \lambda(z) \left(\alpha_W(z) - \alpha_B(z) \right) dz \\ &= D^{*,IV} \end{aligned} \tag{B.11}$$

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 $D^{*,IV}$. □

B.2. Empirical Implementation

Testing the Equal Weights Assumption: A key assumption for the consistency of our IV estimator is that the IV weights are the same across race. Following Cornelissen et al. (2016), we calculate white and black IV weights for each judge-by-year cell by replacing the terms in Equation (B.7) with their sample analogues. Noting that our instrument is linear by construction and, as a result, that $\frac{\partial Released_r(z)}{\partial z}(z) = c$, we drop the term $\frac{\partial Released_r(z)}{\partial z}(z) = c$, as this appears in both the numerator and denominator of Equation (B.7). We then use kernel density methods to retrieve an estimate $\hat{f}_r^z$, which is the density of leniency for race r. With this estimate of the density of leniency for race r, we can plug in the sample analogue of $\mathbb{E}[z]$ and use numerical integration to estimate the remaining terms and estimate IV weights by race for each point in the distribution.

One implication of the equal weights assumption is that the distributions of black and white IV weights over the distribution of judge leniency are statistically identical. To implement this test, Appendix Figure B1 plots the IV weights for each judge-by-year cell, the level of our variation, by race. The distributions of black and white IV weights are visually indistinguishable from each other and a Kolmogorov-Smirnov test cannot reject the hypothesis that the two estimated distributions are drawn from the same continuous distribution (p-value = 0.431).

A second implication of the equal weights assumption is that the relationship between the black IV weights and the white IV weights should fit a 45-degree line up to sampling error. Appendix Figure B2 plots the black IV weights and the white IV weights for each judge-by-year cell, where we discretize the continuous weights to retrieve an estimate of the weights for each judge-by-year cell and then normalize the weights so that the weights sum to one (in the continuous version the weights integrate to one). The black and white IV weights for each judge-by-year cell are highly correlated across race. To formally test for violations of the equal weights assumption, we regress each black IV weight for each judge-by-year cell on the white IV weight for the same cell. This regression yields a coefficient on the white IV weight equal to 1.028 with a standard error of 0.033. Thus, both tests suggest that our assumption of equal IV weights by race is satisfied in the data.

Understanding the IV Weights: We now investigate the relationship between IV weights and judge characteristics to better understand the economic interpretation of an IV-weighted estimate of racial bias. Appendix Table B1 presents OLS estimates of IV weights in each judge-by-year cell on observable judge-by-year characteristics separately by race. The correlation between the IV weights and both average leniency and whether the judge is a minority is statistically zero in both the white and black distribution, with only a weak correlation between the IV weights and judge experience in a given year. Conversely, the IV weights are positively correlated with the number of cases in a judge-by-year cell and a judge being from Philadelphia (where each judge-by-year cell has more observations). These results suggest that the additional precision in our IV regressions comes, at least in part, from placing more weight on judge-by-year cells with more observations. The IV weights are also positively correlated with judge-by-year specific estimates of racial bias (estimated using the MTE approach discussed in Section D below), although not differentially by

defendant race. The positive correlation between the IV weights and the judge-by-year estimates of bias implies that the IV-weighted estimate of racial bias will be larger than an equal-weighted estimate of racial bias. All of our IV results should be interpreted with these correlations in mind.

Bounding the Maximum Bias of the IV Estimator with a Discrete Instrument: Our approach assumes continuity of the instrument Z_i. If the instrument is discrete, we can characterize the maximum potential bias of our IV estimator D^{IV} relative to $D^{*,IV}$, e.g. "infra-marginality bias."

Proposition B.2. If the instrumental variable weights are equal by race, the maximum bias of our IV estimator D^{IV} from $D^{*,IV}$ 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:

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

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

To prove that this proposition holds, we proceed in five steps. First, we show that $D^{*,IV}$ 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 $D^{*,IV}$ 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.

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:

$$\alpha_r^{j,j-1} \in (t_r^{j-1}, t_r^j] \tag{B.13}$$

Note that we can rewrite $D^{*,IV}$ as:

$$\begin{aligned}
D^{*,IV} &= \sum_{j=1}^{J} \lambda^j \left(t_W^j - t_B^j \right) \\
&= \sum_{j=1}^{J} \lambda^j \left(\alpha_W^{j,j-1} - \alpha_B^{j,j-1} \right) + \sum_{j=1}^{J} \lambda^j \left(t_W^j - \alpha_W^{j,j-1} \right) + \sum_{j=1}^{J} \lambda^j \left(\alpha_B^{j,j-1} - t_B^j \right) \\
&= D^{IV} + \underbrace{\sum_{j=1}^{J} \lambda^j \left(t_W^j - \alpha_W^{j,j-1} \right) + \sum_{j=1}^{J} \lambda^j \left(\alpha_B^{j,j-1} - t_B^j \right)}_{\text{infra-marginality bias}}
\end{aligned} \tag{B.14}$$

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 $D^{*,IV}$ and rearranging terms. The third line follows from assuming equal IV weights by race. Equation B.14 shows that $D^{*,IV}$ 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 $D^{*,IV}$. First, note that Equation (B.13) 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.14) to obtain an upper bound on $D^{*,IV}$:

$$\begin{aligned} D^{*,IV} &\leq D^{IV} + \sum_{j=1}^{J} \lambda^j \left(t_W^j - \alpha_W^{j,j-1}\right) \\ &< D^{IV} + \sum_{j=1}^{J} \lambda^j \left(t_W^j - t_W^{j-1}\right) \end{aligned} \tag{B.15}$$

where the second line follows from Equation (B.13) ($t_W^{j-1} < \alpha_W^{j,j-1}$).

Using similar logic, we can also derive a lower bound for $D^{*,IV}$. Equation (B.13) 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.14) to obtain a lower bound on $D^{*,IV}$:

$$\begin{aligned} D^{*,IV} &\geq D^{IV} + \sum_{j=1}^{J} \lambda^j \left(\alpha_B^{j,j-1} - t_B^j\right) \\ &= D^{IV} - \sum_{j=1}^{J} \lambda^j \left(t_B^j - \alpha_B^{j,j-1}\right) \\ &> D^{IV} - \sum_{j=1}^{J} \lambda^j \left(t_B^j - t_B^{j-1}\right) \end{aligned} \tag{B.16}$$

where again, the last line follows from Equation (B.13) ($t_B^{j-1} < \alpha_B^{j,j-1}$).

We can now bound $D^{*,IV}$ using Equation (B.16) and Equation (B.15):

$$D^{IV} - \sum_{j=1}^{J} \lambda^j \left(t_B^j - t_B^{j-1}\right) < D^{*,IV} < D^{IV} + \sum_{j=1}^{J} \lambda^j \left(t_W^j - t_W^{j-1}\right) \tag{B.17}$$

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 $D^{*,IV}$ 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 $\sum_{j=1}^{J} \left(t_r^j - t_r^{j-1}\right)$:

$$\max_j(\lambda^j) \sum_{j=1}^{J} \left(t_r^j - t_r^{j-1}\right) = \max_j(\lambda^j)(t_r^J - t_r^0) \tag{B.18}$$

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:

$$D^{IV} - \max_j(\lambda^j)(\alpha^{max} - \alpha^{min}) < D^{*,IV} < D^{IV} + \max_j(\lambda^j)(\alpha^{max} - \alpha^{min}) \tag{B.19}$$

which proves Proposition B.2. In other words, the maximum bias of our IV estimator D^{IV} from $D^{*,IV}$ 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.19) can be re-written as:

$$D^{IV} - \max_j(\lambda^j) < D^{*,IV} < D^{IV} + \max_j(\lambda^j) \tag{B.20}$$

Rearranging terms yields:

$$-\max_j(\lambda^j) < D^{*,IV} - D^{IV} < \max_j(\lambda^j) \tag{B.21}$$

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

To understand the intuition of our maximum bias formula, note that under Proposition B.2, the maximum bias of D^{IV} relative to $D^{*,IV}$ 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. 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 $D^{*,IV}$. 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 weights depend partially on the share of compliers between any two adjacent judges. 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.

We can now illustrate how we empirically estimate the maximum potential bias of our IV estimator from $D^{*,IV}$ 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. Imbens and Angrist (1994) show that the instrumental variables weights, λ^j, for a discrete multi-valued instrument are given by the following formula:

$$\lambda^j = \frac{(Pr(Released|z_j) - Pr(Released|z_{j-1})) \cdot \sum_{l=j}^{J} \pi^l (g(z_l) - \mathbb{E}[g(Z)])}{\sum_{m=1}^{J} (Pr(Released|z_m) - Pr(Released|z_{m-1})) \cdot \sum_{l=m}^{J} \pi^l (g(z_l) - \mathbb{E}[g(Z)])} \tag{B.22}$$

where π^l is the probability a defendant is assigned to judge l, $g(z_l)$ is a function of the instrument, and $Pr(Released|z_j)$ if the probability a defendant is released if assigned to judge j. While λ^j is not indexed by r, we estimate the weights completely separately by race. In both cases, we find the maximum weight to be the same. To proceed, we residualize both the endogenous variable *Released* and the judge leniency instrument using all exogenous regressors. An instrumental variables regression utilizing residualized variables yields a numerically identical estimate as the specification in the main text (Evdokimov and Kolesár 2018). To estimate the weight λ^j we simply replace each expression in Equation (B.22) with the empirical counterpart. Formally:

$$Pr(Released|z_j) - Pr(Released|z_{j-1}) = \mathbb{E}[\ddot{R}|\ddot{z}_j] - \mathbb{E}[\ddot{R}|\ddot{z}_{j-1}] \tag{B.23}$$

where $\ddot{R}$ is *Released* residualized by the exogenous regressors and $\ddot{z}_j$ is the residualized value of the instrument. Since we use residualized judge leniency as the instrument we replace $g(\ddot{z}_l) = \ddot{z}_l$. Lastly, we replace π^j and $\mathbb{E}[Z]$ with their empirical counterparts:

$$\hat{\pi}^j = \sum_{i=1}^{N} \frac{\mathbb{1}\{\ddot{Z}_i = \ddot{z}_j\}}{N} \tag{B.24}$$

$$\mathbb{E}[Z] = \frac{1}{N} \sum_{i=1}^{N} \ddot{Z}_i \tag{B.25}$$

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 $D^{*,IV}$. This procedure yields a maximum bias of 0.011 or 1.1 percentage points.

From Equation (B.20), we know:

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

Therefore, in our setting, $D^{*,IV}$ is bounded within 1.1 percentage points of our IV estimate for racial bias. □

B.3. Re-weighting Procedure to Allow Judge Preferences for Non-Race Characteristics

In this subsection, we show that a re-weighting procedure using our IV estimator 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:

$$\Psi(x) = \frac{Pr(W|x)Pr(B)}{Pr(B|x)Pr(W)} \tag{B.26}$$

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:

$$\alpha_r^{j,j-1}(x) = \mathbb{E}[Y_i(1) - Y_i(0)|R_i(z_j) - R_i(z_{j-1}) = 1|r_i = r, \mathbf{X}_i = x] \tag{B.27}$$

As noted by Fröhlich (2007) and discussed in Angrist and Fernández-Val (2013), the unconditional LATE can be expressed as:

$$\alpha_r^{j,j-1} = \sum_{x \in X} \alpha_r^{j,j-1}(x) \frac{Pr(Released|z_j, x, r) - Pr(Released|z_{j-1}, x, r)}{Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r)} P(x|r) \tag{B.28}$$

We assume:

$$\frac{Pr(Released|z_j, x, r) - Pr(Released|z_{j-1}, x, r)}{Pr(Released|z_j, r) - Pr(Released|z_{j-1}, r)} = \xi(x) \tag{B.29}$$

In words, while the first stage may vary based on covariates, it varies in the same way for white

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

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

where line 2 follows by plugging in the formula for $\Psi(x)$ and lines 3 and 5 follow from Bayes' rule. These steps closely follow DiNardo, Fortin, and Lemieux (1996), although our parameter of interest is a treatment effect rather than a distribution. Given that the weights for all white defendants are equal to 1, D^{IV} is given by:

$$D^{IV} = \sum_{j=1}^{J} \lambda^j \left(\sum_{x \in X} \xi(x)Pr(x|W) \left(\alpha_W^{j,j-1}(x) - \alpha_B^{j,j-1}(x) \right) \right) \tag{B.30}$$

□

C. Non-Parametric Pairwise LATE Framework

A second approach to estimating the average level of racial bias is to estimate each pairwise LATE separately and then impose the preferred weighting scheme across these non-parametric estimates. We consider, for example, an approach that places equal weight on each judge to estimate the average level of racial bias across judges all judges in the sample. This fully non-parametric approach can yield a policy-relevant interpretation of racial bias with minimal assumptions, but often comes at the cost of statistical precision since any particular LATE is often estimated with considerable noise.

In this subsection, we present a formal definition of the equal-weighted level of bias and our non-parametric estimator, provide proofs for consistency, and evaluate the feasibility of this non-parametric approach using Monte Carlo simulations.

C.1. Definition and Consistency of Pairwise LATE Estimator

Definition: Let the equal-weighted LATE estimate of racial bias based on the non-parametric pairwise estimates, $D^{*,PW}$ be defined as:

$$\begin{aligned} D^{*,PW} &= \sum_{j=1}^{J} w^j \left(t_W^j - t_B^j \right) \\ &= \sum_{j=1}^{J} \frac{1}{J} \left(t_W^j - t_B^j \right) \end{aligned} \tag{B.31}$$

where $w^j = \frac{1}{J}$, such that $D^{*,PW}$ can be interpreted as the average level of racial bias across judges – an estimate with clear economic interpretation.

Let the equal-weighted pairwise LATE estimator of racial bias, D^{PW}, be defined as:

$$D^{PW} = \sum_{j=1}^{J} \frac{1}{J} (\alpha_W^{j,j-1} - \alpha_B^{j,j-1}) \tag{B.32}$$

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

Conditions for Consistency: Following the proofs for the IV estimator, D^{PW} provides a consistent estimate of racial bias $D^{*,PW}$ if (1) Z_i is continuous and (2) w^j is constant by race, which is satisfied because the weights are chosen ex post to be equal ($w^j = \frac{1}{J}$).

C.2. Empirical Implementation

Estimating the Pairwise LATEs: We estimate non-parametric LATEs using the following Wald estimator for each pair of judges j and judge $j-1$:

$$\hat{\alpha}_r^{j,j-1} = \frac{\mathbb{E}[Y_i | Z_i = z_j, r] - \mathbb{E}[Y_i | Z_i = z_{j-1}, r]}{\mathbb{E}[Released_i | Z_i = z_j, r] - \mathbb{E}[Released_i | Z_i = z_{j-1}, r]} \tag{B.33}$$

where $\mathbb{E}[Y_i | Z_i = z_j, r]$ is the probability a defendant of race r assigned to judge j is rearrested and $\mathbb{E}[Released_i | Z_i = z_j, r]$ is the probability a defendant of race r assigned to judge j is released. Following the above discussion, our equal-weighted estimate of racial bias is equal to the simple difference between the average estimated pairwise LATE for white defendants and the average estimated pairwise LATE for black defendants.

Monte Carlo Simulation: As discussed above, a fully non-parametric approach can yield a policy-relevant interpretation of racial bias with minimal assumptions, but often comes at the cost of statistical precision since any particular LATE is often estimated with considerable noise. We therefore begin by examining the performance of our non-parametric estimator using Monte Carlo simulations. Specifically, we create a simulated 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 0.100. The probability that a released defendant is rearrested ($Y_i = 1$) conditional on release is equal to the risk of the released defendant.

In each draw of the simulated data, we estimate non-parametric LATEs using the Wald estimator described above. Our estimate of racial bias in each draw of the simulated data is equal to the difference between the average release threshold for white defendants and the average release threshold for black defendants. We repeat this entire process 500 times and plot the resulting estimates of the average level of racial bias across all bail judges.

Panel A of Appendix Figure B3 presents the results from this Monte Carlo exercise. The average level of racial bias across all simulations is equal to 0.125, close to the true level. However, the variance of the estimates is extremely large, with nearly 20 percent of the simulations yielding an estimate of racial bias that is greater than one in absolute value. The high variance in the estimates stems from weak first stages between judges that are very close in the leniency distribution. We conclude from this exercise that a fully non-parametric approach yields uninformative estimates of average racial bias in our setting, and do not explore this approach further.[26]

D. Marginal Treatment Effects Framework

Our final estimator uses the MTE framework developed by Heckman and Vytlacil (1999, 2005) to estimate the average level of bias, $D^{*,w}$, where we impose equal weights for each judge. The MTE framework allows us to estimate judge-specific treatment effects for white and black defendants at the margin of release and choose a weighting scheme across all judges, but with the identification and estimation of the judge-specific estimates, t_r^j, coming at the cost of additional auxiliary assumptions.

In this subsection, we present a formal definition of the equal-weighted level of bias and our MTE estimator, provide details on the mapping of the MTE framework to our test of racial bias, provide proofs for consistency, and discuss the details of the empirical implementation and tests of the parametric assumptions.

[26] In unreported results, we also examine the performance of a non-parametric estimator where estimates of α_r^j are formed using a Wald estimator between judge j to judge $j-k$, where $k > 1$. We find that increasing k decreases variance in the simulated estimates, but increases estimation bias, as judges further away in the distribution are used to estimate judge j's threshold. Even with relatively large k, we find the MTE procedure described in Section D is more precise than the pairwise LATE procedure.

D.1. Definition and Consistency of MTE Estimator

Definition: Following the discussion of the equal-weighted non-parametric estimator, let the equal-weighted MTE estimate of racial bias, $D^{*,MTE}$ be defined as:

$$\begin{aligned} D^{*,MTE} &= \sum_{j=1}^{J} w^j \left(t_W^j - t_B^j \right) \\ &= \sum_{j=1}^{J} \frac{1}{J} \left(t_W^j - t_B^j \right) \end{aligned} \tag{B.34}$$

where $w^j = \frac{1}{J}$, such that $D^{*,MTE}$ can again be interpreted as the average level of racial bias across judges.

Let our equal-weighted MTE estimator of racial bias, D^{MTE}, be defined as:

$$D^{MTE} = \sum_{j=1}^{J} \frac{1}{J} \left(MTE_W(p_W^j) - MTE_B(p_B^j) \right) \tag{B.35}$$

where p_r^j is the probability judge j releases a defendant of race r (i.e. judge j's propensity score) and $MTE_r(p_r^j)$ is the estimated MTE at the propensity score for judge j calculated separately for each defendant of race r.

MTE 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:

$$Released_i(z_j, r) = \mathbb{1}\{\mathbb{E}[\alpha_i | r] \leq t_r^j\} \tag{B.36}$$

where $Released_i(z_j, r)$ indicates the probability defendant i of race r is released by judge j, and 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:

$$\begin{aligned} Released_i(z_j, r) &= \mathbb{1}\{F_{\alpha,r}(\mathbb{E}[\alpha_i | r]) \leq F_{\alpha,r}(t_r^j)\} \\ &= \mathbb{1}\{U_{i,r} \leq p_r^j\} \end{aligned} \tag{B.37}$$

where $U_{i,r} \in [0, 1]$ by construction. In this latent-index model, defendants with $U_{i,r} \leq p_r^j$ are released, defendants with $U_{i,r} > p_r^j$ are detained, and defendants with $U_{i,r} = p_r^j$ 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:

$$MTE_r(u) = \mathbb{E}[\alpha_i | r, U_{i,r} = u] \tag{B.38}$$

where $\mathbb{E}[\alpha_i|r, U_{i,r} = p_r^j]$ denotes the treatment effect for a defendant of race r who is on the margin of release to a judge with propensity score equal p_r^j. 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 (B.38) 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:

$$\alpha_r^j = \mathbb{E}[\alpha_i|r, \mathbb{E}[\alpha_i|r] = t_r^j] \tag{B.39}$$

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:

$$\begin{aligned} \alpha_r^j &= \mathbb{E}[\alpha_i|r, \mathbb{E}[\alpha_i|r] = t_r^j] \\ &= \mathbb{E}[\alpha_i|r, U_{i,r} = p_r^j] \\ &= MTE_r(p_r^j) \end{aligned} \tag{B.40}$$

Equation (B.40) shows that we can use the race-specific MTEs to identify the race-specific treatment effect of each judge, α_r^j, and as a result, race-specific thresholds of release, t_r^j. We can then estimate the level of racial bias for each judge j, $t_W^j - t_B^j$. To see this, let judge j have a propensity score to release white defendants equal to p_W^j and a propensity to release black defendants equal to p_B^j. Given Equation (B.40), the level of racial bias for judge j is therefore equal to $MTE_W(p_W^j) - MTE_B(p_B^j)$. From these judge-specific estimates of racial bias, we can then ex post impose equal weights across judges to estimate D^{MTE}, the average level of racial bias.

Conditions for Consistency: In addition to the assumptions required for a causal interpretation of the IV estimator (existence, exclusion restriction, and monotonicity), our MTE estimator D^{MTE} provides a consistent estimate of $D^{*,MTE}$ if the race-specific MTEs are identified over the entire support of the propensity score calculated using variation in Z_i.

If Z_i is continuous, the local instrumental variables (LIV) estimand provides a consistent estimate of the MTE over the support of the propensity score with no additional assumptions (Heckman and Vytlacil 2005, Cornelissen et al. 2016). With a discrete instrument, however, our MTE estimator is only consistent under additional functional form restrictions that allow us to interpolate the MTEs between the values of the propensity score we observe in the data. In our MTE framework, if our specification of the MTE is flexible enough to capture the true shape of the MTE function, then there will be no infra-marginality bias. If the specification is too restrictive, then there may be misspecification bias in estimating the MTE.

Recall that our goal is to construct the average level of racial bias across judges:

$$\begin{aligned} D^{*,MTE} &= \sum_{j=1}^{J} \frac{1}{J} \left(t_W^j - t_B^j \right) \\ &= \sum_{j=1}^{J} \frac{1}{J} \left(\alpha_W^j - \alpha_B^j \right) \end{aligned} \tag{B.41}$$

With a continuous instrument, α_W^j and α_B^j are identified by evaluating $MTE(p_W^j)$ and $MTE(p_B^j)$. Heckman and Vytlacil (1999) show local instrumental variables (LIV) can be used to identify the MTE non-parametrically. With a discrete instrument, however, $MTE(p_r^j)$ is no longer non-parametrically identified.

Following Heckman and Vytlacil (2005) and Doyle (2007), we use a local polynomial function and information from the observed values of the propensity score to estimate the MTE curve over the full support of the propensity score. Specifically, we use a local quadratic estimator to approximate $\mathbb{E}[Y_i|p_r^j]$, and then estimate the MTE as the numerical derivative of the local quadratic function. In this estimation, we specify a bandwidth, and therefore use information from all judges in a given bandwidth to estimate the threshold for a given judge.

Let the estimated MTE be denoted by $\hat{MTE}(p_r^j)$. We can express our MTE estimator D^{MTE} as:

$$\begin{aligned} D^{MTE} = & \underbrace{\sum_{j=1}^{J} \frac{1}{J} \left(\hat{MTE}(p_W^j) - \hat{MTE}(p_B^j) \right)}_{\text{Estimated MTE}} + \\ & \underbrace{\sum_{j=1}^{J} \frac{1}{J} \left(MTE(p_W^j) - \hat{MTE}(p_W^j) \right) + \sum_{j=1}^{J} \frac{1}{J} \left(\hat{MTE}(p_B^j) - MTE(p_B^j) \right)}_{\text{infra-marginality bias}} \end{aligned} \tag{B.42}$$

In this case, infra-marginality bias arises because we allow for the possibility that the local quadratic function does not provide enough flexibility to accurately capture the shape of the MTE. If we assume our specification of the MTE is flexible enough to capture the shape of the MTE, then $\mathbb{E}[\hat{MTE}(p_r^j)] = MTE(p_r^j)$, indicating there is no infra-marginality bias. Therefore, if we correctly specify the form of the MTE function, then D^{MTE} provides a consistent estimate of $D^{*,MTE}$:

$$\begin{aligned} D^{MTE} &= \sum_{j=1}^{J} \frac{1}{J} \left(MTE_W(p_W^j) - MTE_B(p_B^j) \right) \\ &= \sum_{j=1}^{J} \frac{1}{J} \left(t_W^j - t_B^j \right) \\ &= D^{*,MTE} \end{aligned} \tag{B.43}$$

D.2. Empirical Implementation

Estimating the MTE Curve: We estimate D^{MTE} using a two-step procedure. First, we estimate the entire distribution of MTEs. To estimate each race-specific MTE, we estimate 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) separately for white and black defendants:

$$MTE_W(p_W^j) = \frac{\partial}{\partial p_W^j}\mathbb{E}(\ddot{Y}_i|p_W^j, W) \tag{B.44}$$

$$MTE_B(p_B^j) = \frac{\partial}{\partial p_B^j}\mathbb{E}(\ddot{Y}_i|p_B^j, B) \tag{B.45}$$

where p_r^j is the propensity score for release for judge j and defendant race r and $\ddot{Y}_i$ is rearrest residualized on all observables: an using exhaustive set of court-by-time fixed effects as well as our baseline crime and defendant controls: 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.

Following Heckman, Urzua, and Vytlacil (2006) and Doyle (2007), we begin by residualizing our measure of pre-trial misconduct, pre-trial release, and judge leniency using the full set of fixed effects and observables. We can then calculate the race-specific propensity score using a regression of the residualized release variable on our residualized measure of judge leniency, capturing only the variation in pre-trial release due to variation in the instrument.[27] Next, we compute the numerical derivative of a smoothed function relating residualized pre-trial misconduct to the race-specific propensity score. Specifically, we estimate the relationship between the residualized misconduct variable and the race-specific propensity score using a local quadratic estimator. We then compute the numerical derivative of the local quadratic estimator for each race separately to obtain the race-specific MTEs. In unreported results, we also find nearly identical results using alternative estimation procedures, such as the global polynomials used in Kowalski (2016).

Second, we use the race-specific MTE distributions to calculate the level of racial bias for each judge j. We aggregate these judge-specific estimates of racial bias to calculate an equal-weighted

[27] A common approach in the MTE literature is to exploit variation in the propensity score that arises from covariates. Many treatment effect parameters, such as the average treatment effect, rely on having wide support of the propensity score. However, in practice, it is difficult to identify such strong instruments, so researchers rely on utilizing variation driven by observables. In our setting, we rely on the continuity of the propensity score to estimate the MTE, but require no assumptions concerning the range of the propensity score. In particular, the treatment effects we are interested in are identified by variation in judge leniency by definition.

estimate of racial bias:

$$D^{MTE} = \sum_{j=1}^{J} \frac{1}{J} \left(MTE_W(p_W^j) - MTE_B(p_B^j) \right) \tag{B.46}$$

We calculate standard errors of this equal-weighted estimate by bootstrapping this two-step procedure 500 times at the judge-by-shift level.

Monte Carlo Simulation: To examine the performance of our MTE estimator, we again use a Monte Carlo simulation. Following the simulation used to test the non-parametric estimator, we create a simulated 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 0.100. The probability that a released defendant is rearrested ($Y_i = 1$) conditional on release is equal to the risk of the released defendant.

In each draw of the simulated data, 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. We repeat this entire process 500 times and plot the resulting estimates of the average level of racial bias across all bail judges.

Panel B of Appendix Figure B3 presents the results from this Monte Carlo exercise. The average level of racial bias across all simulations is equal to 0.090 with a standard deviation of only 0.051. In addition, the 10^{th} percentile of estimates is equal to 0.036 and the 90^{th} percentile equal to 0.143. These results stand in sharp contrast to the statistically uninformative results from our non-parametric estimator and suggest that, in practice, our MTE estimator is likely to yield statistically precise estimates of the average level of racial bias across all bail judges.

Testing the MTE Functional Form Assumption: Following Cornelissen et al. (2016), we test whether the MTE is misspecified by constructing a non-parametric IV estimate of racial bias by taking the correct weighted average of the MTE. Specifically, we re-estimate the IV weights from Equation (B.7), but substitute $p(z_j)$ in for z_j, given that we estimate the MTE curve over the distribution of the propensity score, and not the distribution of leniency. We denote these weights ω_r^{IV}. As shown in Heckman and Vytlacil (2005), the IV estimate, α_r^{IV} is related to the MTE_r by:

$$\alpha_r^{IV} = \int MTE_r(u)\omega_r^{IV}(u)du \tag{B.47}$$

Intuitively, the MTE approach relies on identifying the MTE curve. To do so, we must impose structure on the relationship between the propensity score and the outcome of interest. This implies

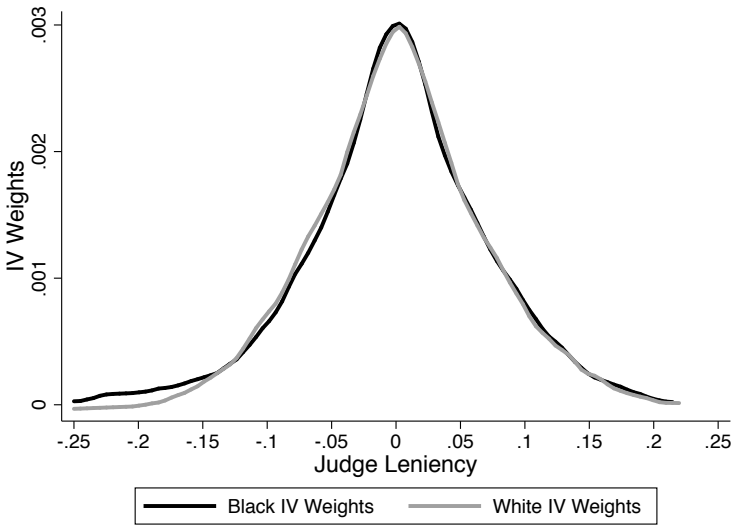
we also impose structure on the derivative of this relationship, which is equal to the MTE curve. If the structure does not bias our estimate of the MTE curve, then we should be able to construct the non-parametric IV by taking the weighted average of the MTE curve shown in Equation (B.47). However, if the estimated MTE is biased, then in general, the weighted average of the MTE will not be equal to the non-parametric IV estimate. We find that our MTE weighted by the IV weights is very close to the non-parametric IV estimate of racial bias. Specifically, the white IV estimate for the effect of release on rearrest is equal to 0.236, while the MTE weighted by the white IV weights yields an estimate of 0.261. Similarly, the black IV estimate for the effect of release on rearrest is equal to 0.014, while the MTE weighted by the black IV weights yields an estimate of 0.021. These results indicate that our MTE is likely to be correctly specified.

Appendix Table B1: Correlation between IV Weights and Observables

	White IV Weights x 100	Black IV Weights x 100
	(1)	(2)
Discrimination	0.424***	0.518***
	(0.066)	(0.062)
Philadelphia	0.117***	0.104***
	(0.016)	(0.016)
Case Load (100s)	0.004***	0.006***
	(0.001)	(0.001)
Average Leniency	0.044	0.000
	(0.055)	(0.054)
Experience	−0.000	0.002*
	(0.001)	(0.001)
Minority Judge	0.003	0.004
	(0.008)	(0.008)
Observations	552	552

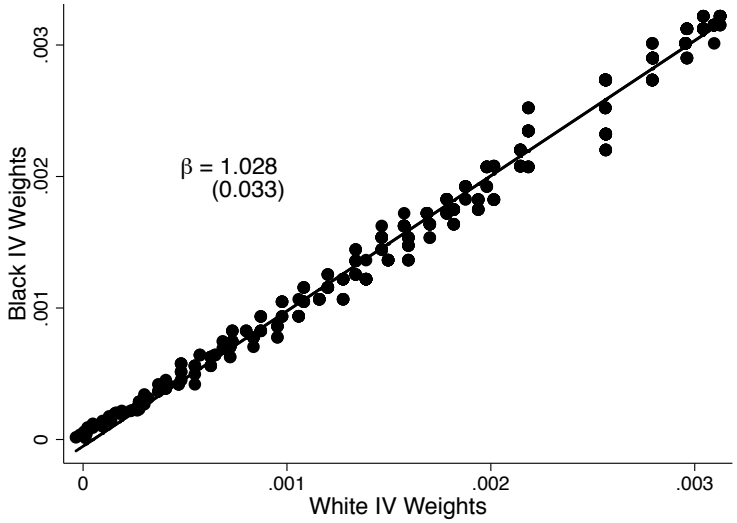
Note: This table estimates the relationship between instrumental variable weights assigned to a given judge-by-year cell on observables of the judge-by-year cell. To ease readability, the coefficients are multiplied by a 100. Column 1 presents results for IV weights calculated for white defendants. Column 2 presents results for IV weights calculated for black defendants. To compute the weight assigned to a judge-by-year cell, we first compute the continuous weights by constructing sample analogues to the terms which appear in Equation (B.7) following the procedure described in Cornelissen et al. (2016) and Appendix B. To move from the continuous weights to a weight associated with a given judge, we compute the average leniency of each judge-by-year cell in the data. We then compute the weight associated with the average leniency of the judge-by-year cell using the results from the continuous weights estimation. We divide the resulting weights by the sum total to ensure the discretized weights sum to one.

Appendix Figure B1: Distribution of IV Weights by Race Across Judge Leniency Distribution



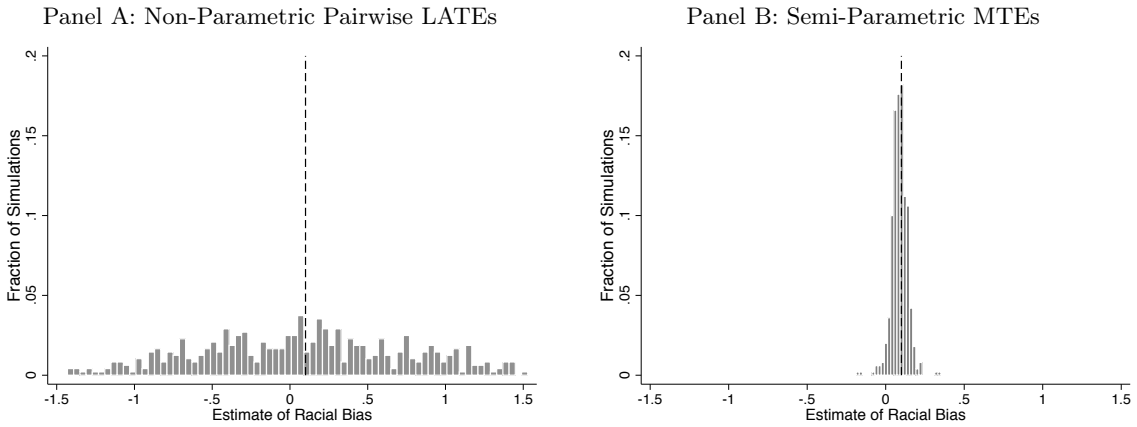
Note: This figure plots the instrumental variables weights over the distribution of judge leniency for both black and white defendants. To compute the instrumental variable weights, we first compute the continuous weights by constructing sample analogues to the terms which appear in Equation (B.7) following the procedure described in Cornelissen et al. (2016) and Appendix B. To move from the continuous weights to a weight associated with a given judge-by-year, we compute the average leniency of each judge-by-year cell in the data. We then compute the weight associated with the average leniency of the judge-by-year cell using the results from the continuous weights estimation. We divide the resulting weights by the sum total to ensure the discretized weights sum to one.

Appendix Figure B2: Correlation Between White IV Weights and Black IV Weights



Note: This figure plots the instrumental variables weight assigned to judge j in year t in the white leniency distribution vs. the instrumental variables weight assigned to judge j in year t in the black distribution. To compute the weight assigned to a judge-by-year cell, we first compute the continuous weights by constructing sample analogues to the terms which appear in Equation (B.7) following the procedure described in Cornelissen et al. (2016) and Appendix B. To move from the continuous weights to a weight associated with a given judge, we compute the average leniency of each judge-by-year cell in the data. We then compute the weight associated with the average leniency of the judge-by-year cell using the results from the continuous weights estimation. We divide the resulting weights by the sum total to ensure the discretized weights sum to one.

Appendix Figure B3: Monte Carlo Simulations of Racial Bias Estimators



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 0.100. 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. For each judge, we impose a 10 percentage point higher threshold for black defendants, so that the "true" level of racial bias in the simulated data is equal to 0.100. Panel A presents estimates from a non-parametric LATE procedure, where we form the Wald estimator between judge j and judge $j-1$ to estimate the release threshold for judge j. Panel B presents estimates from the MTE procedure. The estimate of racial bias is equal to the average estimated release threshold for white defendants minus the average estimated release threshold for black defendants across judges.

Appendix C: Simple Graphical Example

In this appendix, we use a series of simple graphical examples to illustrate how a judge IV estimator for racial bias improves upon the standard OLS estimator. We first consider the OLS estimator in settings with either a single race-neutral judge or a single racially biased judge, showing that the standard estimator suffers from infra-marginality bias whenever there are differences in the risk distributions of black and white defendants. We then use a simple two-judge example to illustrate how a judge IV estimator can alleviate the infra-marginality bias in both settings.

OLS Estimator: To illustrate the potential for infra-marginality bias when using a standard OLS estimator, we begin with the case of a single race-neutral judge. The judge perfectly observes risk and chooses the same threshold for white and black defendants, but the distributions of risk differ by defendant race. Panel A of Appendix Figure C1 illustrates such a case, where we assume that white defendants have more mass in the left tail of the risk distribution, i.e. that whites are, on average, less risky than blacks. Letting the vertical lines denote the judge's release threshold, standard OLS estimates of α_W and α_B measure the average risk of released defendants for white and black defendants, respectively. In the case illustrated in Panel A, the standard OLS estimator indicates that the judge is biased against white defendants, when, in reality, the judge is race-neutral.

To further illustrate this point, Panel B of Appendix Figure C1 considers the case of a single judge that is racially biased against black defendants. Once again, the distributions of risk differ by defendant race, but now the judge chooses different thresholds for white and black defendants. In the case illustrated in Panel B, white and black defendants have the exact same expected risk conditional on release. As a result, the standard OLS estimator indicates that the judge is race-neutral, when, in reality, the judge is biased against black defendants. Following the same logic, we could choose risk distributions and release thresholds such that the OLS estimator indicates racial bias against white defendants or racial bias against black defendants. In other words, the OLS estimator is uninformative about the extent of racial bias in bail decisions without strong assumptions about differences in the underlying distributions of risk by defendant race.

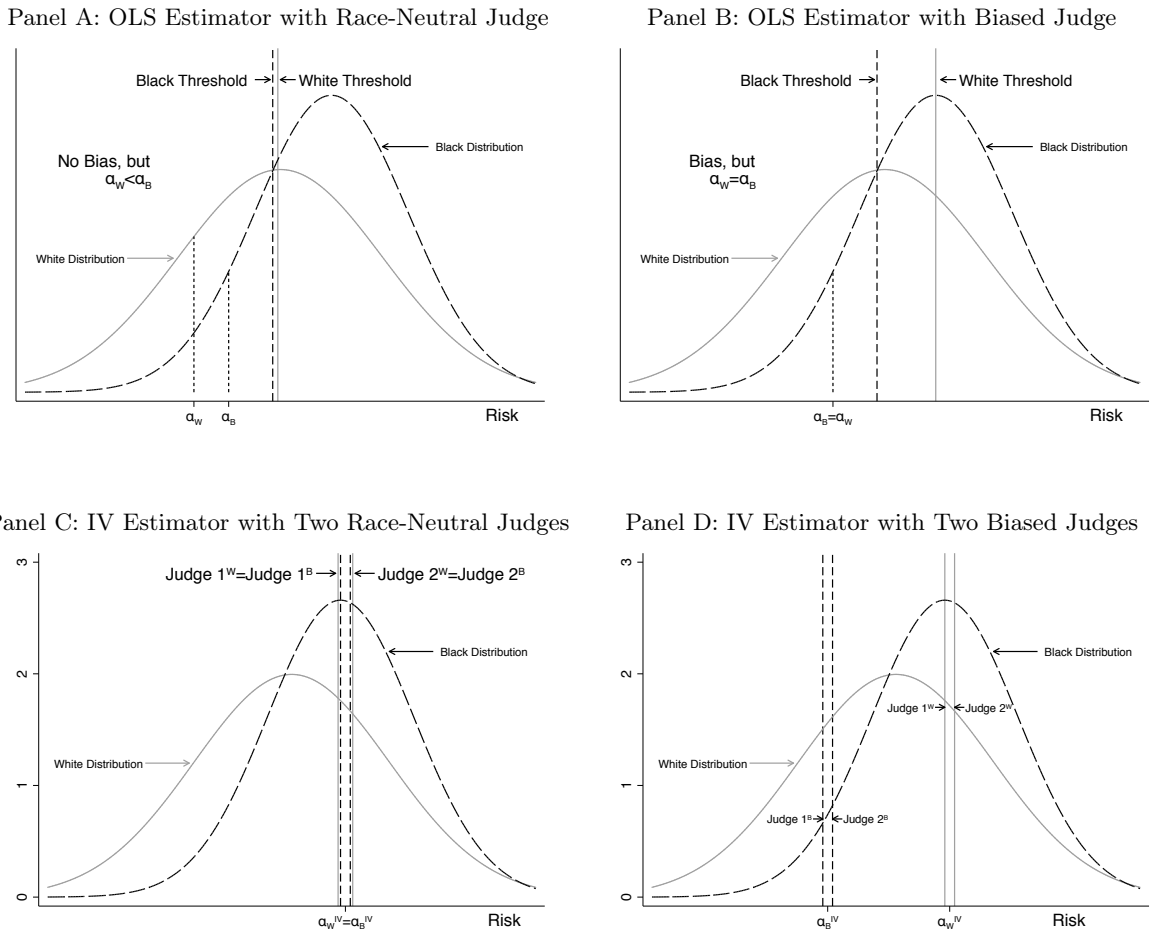
IV Estimators: We now illustrate how a judge IV estimator for racial bias can potentially solve this infra-marginality problem by focusing the analysis on defendants at the margin of release. We use a simple two-judge example to illustrate the intuition behind our approach, while maintaining our assumption that judges perfectly observe risk and that the distributions of risk differ by defendant race. Throughout, we assume that judge 2 is more lenient than judge 1.

Panel C of Appendix Figure C1 considers the case where both judges are race-neutral, such that both judges use the same thresholds of release for white and black defendants. In this case, an IV estimator using judge leniency as an instrument for pre-trial release will estimate the average risk for defendants who are released by the lenient judge but detained by the strict judge (i.e. the average risk of compliers), α_W^{IV} and α_B^{IV}. When the two judges are "close enough" in leniency, the IV estimator for racial bias will approximately estimate the risk of marginally released black defendants

and marginally released white defendants. Intuitively, the IV estimator measures misconduct risk only for defendants near the margin of release, essentially ignoring the risk of defendants who are infra-marginal to the judge thresholds. As our measure of judge leniency becomes more continuous, our IV estimator will consistently estimate racial bias as the difference between α_W^{IV} and α_B^{IV}. The IV estimator will therefore indicate that marginal black and marginal white defendants have similar misconduct rates, allowing us to correctly conclude that the judges are race-neutral.

To further illustrate this point, Panel D of Appendix Figure C1 considers the case where both judges are racially biased against black defendants, such that both judges have higher thresholds of release for white defendants relative to black defendants. Following the same logic as above, the IV estimator measures the pre-trial misconduct risk of marginally released white and black defendants, α_W^{IV} and α_B^{IV}, so long as the two judges are "close enough" in leniency. The IV estimator will therefore indicate that marginal black defendants are lower risk than marginal white defendants, allowing us to correctly conclude that judges are racially biased against black defendants.

Appendix Figure C1: Infra-marginality Bias with OLS and Judge IV Estimators



Note: These figures plot hypothetical risk distributions for white and black defendants. Panel A illustrates the OLS estimator with a race-neutral judge that chooses the same threshold for release for white and black defendants. Panel B illustrates the OLS estimator with a racially biased judge that chooses a higher threshold for release for white defendants compared to black defendants. Panel C illustrates the judge IV estimator with two race-neutral judges. Panel D illustrates the judge IV estimator with two racially biased judges.

Appendix D: 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. This variable is only available in Philadelphia.

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.

Failure to Appear in Court or Rearrest Prior to Disposition: An indicator for whether a defendant failed to appear in court or was rearrested in Philadelphia, and for whether a defendant was rearrested in Miami.

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 E: Institutional Details

The institutional details described in this Appendix follow directly from Dobbie et al. (2018). 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." In Philadelphia, it is well known that bail judges consider the risk of new crime when setting bail (see Goldkamp and Gottfredson 1988), and in fact, the Philadelphia bail guidelines are designed to "reduce the risk of releasing dangerous defendants into the community while ensuring that defendants who pose minimal risk are not confined to prison to await trial."[28]

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. As noted in Florida's bail statute, "[i]t is the intent of the Legislature that the primary consideration be the protection of the community from risk of physical harm to persons." Fla. Stat. Ann. §907.041(1).

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

[28]See https://www.courts.phila.gov/pdf/notices/2012/6-12-12-Notice-to-Bar-Proposed-Bail-Guidelines.pdf.

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 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. (2018) 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 3 that representativeness of blacks is strictly increasing in risk. Therefore, a representativeness-based discounting model will over-weight the right tail of risk for black defendants.

To compute the stereotyped distribution, we first assume a value of θ, and then compute $\pi_{t,r}$ for every risk category t and race r. We can then compute $\pi^{st}_{t,B}$ by plugging in the values for $\pi_{t,r}$ and the assumed value of θ into Equation (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 A4. The average risk in the stereotyped distribution is about 5.4 percentage points greater than the mean in the true distribution of risk.