If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions*

Alex Albright†

(Click here for most recent version)

September 3, 2019

Abstract

Predictive tools are touted as a way to reduce group disparities that come from human discretion. However, in practice, these tools provide recommendations to human decision-makers rather than overriding them entirely. I use data from Kentucky to show how imperfect compliance with tool recommendations can increase group disparities. A 2011 policy change in Kentucky set recommended defaults for judge bail decisions based on defendant risk levels. The policy caused an increase in raw racial disparities in initial bond, originally illustrated by Stevenson (2017). I show that this increase is not a simple consequence of different risk-based recommendations by race. Rather, I find judges are more likely to override tool recommendations (in favor of harsher bond conditions) for black defendants than similar white defendants. There are two forces behind this result. For one, judges in whiter counties comply with the new default more than judges in blacker counties. Second, even within judge and time, judges are more likely to override the recommended default for moderate risk black defendants than similar moderate risk white defendants. This result suggests that interaction with the same predictive score may lead to different predictions by race.

*I am grateful to Mandy Pallais, Ed Glaeser, Crystal Yang, Larry Katz, Alma Cohen, Louis Kaplow, Emma Harrington, Hannah Shaffer, Jon Roth, Megan Stevenson, Nathan Nunn, Scott Kominers, Alberto Alesina, Roland Fryer, Ashesh Rambachan, Michael Thaler, Moya Chin, Jennifer Doleac, Philip Marx, Jason Furman, Michael Norton, Karen Shen, Natalia Emanuel, Liz Engle, Namrata Narain, Evan Soltas, Emma Pierson, Kadeem Noray, John Tebes, Brian Wheaton, LJ Ristovska, Tianwang Liu, Sarah Armitage, Jack Feng, Frank Pinter, Tanaya Devi, and Mattie Toma for their comments and suggestions. Thanks to participants in the Harvard Graduate Student Workshop in Labor Economics, the Harvard Inequality and Social Policy Proseminar, and the Harvard Graduate Student Workshop in Political Economy for insightful feedback. I thank Daniel Sturtevant and the Kentucky Administrative Office of the Courts for providing me with the data used in this paper as well as Tara Blair, Christy May, and Kathy Schiflett for explaining institutional details about Kentucky Pretrial Services. Thanks to Brenda Piquet, Carolina Harvey, and Alicia McGovern for administrative assistance. I deeply appreciate the R community for developing the many open source packages used during this project. This research has been supported by a James M. and Cathleen D. Stone PhD Scholar fellowship from the Multidisciplinary Program in Inequality & Social Policy at Harvard University as well as a Terence M. Considine Fellowship in Law and Economics from the John M. Olin Center for Law, Economics, and Business at Harvard Law School.

†Economics Department, Harvard University, Cambridge, MA 02138. apalbright@g.harvard.edu.

1 Introduction

Algorithm optimists argue that predictive tools can weaken the role human biases play in high-stakes decision-making (Harris and Paul 2017; Kleinberg et al. 2018).[1] While skeptics worry that predictive tools are biased against disadvantaged groups,[2] recent research demonstrates that perfect adherence to algorithmic tools can decrease bias compared to human decision-makers (Kleinberg et al. 2017; Dobbie et al. 2018; Cowgill 2018a), yet perfect adherence is not the norm in policy environments (Main 2016; Hoffman, Kahn, and Li 2017; New 2015). Rather, prediction tools are introduced into systems ripe with human discretion – tools provide recommendations to human decision-makers rather than overriding them entirely. Despite the first-order importance of such discretion, there is little evidence on how imperfect compliance with predictive tool recommendations may complicate their effects on demographic disparities. Of particular policy interest is the following question: *are predictive recommendations followed similarly across racial groups?*

To address this question, I focus on the usage of risk assessment recommendations in judge bail decisions. Risk assessments are predictive tools generated based on individual-level characteristics and are used in bail, pretrial, or sentencing hearings in 49 of 50 US states (Traughber 2018).[3] Bail decisions determine conditions for release from jail before trial.[4] The most well-known conditions are monetary[5] – financial bond means that a defendant needs to provide some amount of money to be released from jail, while non-financial bond means that the defendant does not.[6] My decision to focus on bail decisions is due to both policy relevance as well as methodological advantages. Bail decisions directly affect pretrial detention, which has downstream effects on future outcomes such as likelihood of conviction.[7] Moreover, pretrial detainees "account for two-thirds of jail inmates and 95% of the growth in the jail population over the last 20 years" (Stevenson and Mayson 2018), making them a significant piece of the policy conversation on mass incarceration. Methodologically, bail is a promising environment for studying the introduction of risk assessments since bail decisions are made quickly (in a matter of minutes) and the legal objective is clear (Arnold, Dobbie, and Yang 2018).[8]

[1] High-stakes decisions are increasingly informed by data-driven prediction tools. Loan officers use credit scores to help make lending decisions, managers use predictions in making hiring decisions, and judges use risk assessments to set bail (Miller 2015; Einav, Jenkins, and Levin 2013; Mamalian 2011).

[2] Risk assessments in criminal justice have received ample public scrutiny on this front. See Angwin and Kirchner (2016), which motivated much of the current academic work on this topic. Michelle Alexander, author of *The New Jim Crow*, has even called risk assessment scores the "Newest Jim Crow" (Alexander 2018).

[3] Furthermore, risk assessments are used in the pretrial context in dozens of jurisdictions and at least six entire states. They are also used in sentencing in at least twenty-eight entire states (Doleac and Stevenson 2018).

[4] Despite the common usage of "pretrial" terminology, about 90-95% of cases are estimated to resolve in plea bargaining rather than trial (Devers 2011).

[5] Non-monetary conditions can include supervision or specific caveats such as no drinking.

[6] Empirically, in my setting, non-financial bond corresponds to a 96% chance of immediate release; financial bond corresponds to 20% chance of immediate release.

[7] For empirical evidence on the effects of pretrial detention, see Dobbie, Goldin, and Yang (2018) and Cowgill (2018b).

[8] The objective is much clearer than, say, in the example of sentencing. In the words of Arnold, Dobbie,

The policy change I focus on is House Bill 463 (HB463) in Kentucky, which set bail recommendations based on defendant risk levels and required judges to provide a reason for overriding a recommendation. Within 24 hours of defendant booking, Kentucky judges make an initial bond decision based on information shared by a Pretrial Services Officer. Kentucky is unique in that police have full authority to charge, meaning there is no prosecutorial review before the judge sets initial bond.[9] Before HB463, judicial consideration of a defendant's Kentucky Pretrial Risk Assessment level (low, moderate, or high) was optional in initial bond decisions (meaning many judges did not know defendants' risk levels).[10] However, when HB463 was enacted on June 8, 2011, judges became required to consider the risk level in their initial decision. More powerfully, judges were required to set non-financial bond for low and moderate risk defendants unless they give a reason for deviating from the recommendation.

Counter to the expectations of prediction advocates, racial disparities increase after the implementation of HB463, as first illustrated by Stevenson (2017). Rates of non-financial bond jump up discontinuously at the date of policy implementation, with white defendants experiencing larger gains. Given that black and white defendants differ in their underlying risk level distributions (black defendants are more likely to score at higher risk levels than white defendants), the increase in disparities could be a simple mechanical consequence of judges' following the recommendations at equal rates for white and black defendants. Using case-level data from the Kentucky Administrative Office of the Courts, I show that differences in risk levels alone cannot explain the increase in racial disparities.[11] *Rather, after HB463, I find that judges overrule the policy default more for black defendants.*[12] In fact, post-HB463 racial disparities are almost twice as large (8.9 percentage points) as they would have been had judges followed the recommendations without discretion (4.7 percentage points).

The next part of the paper then asks *why would recommendation adherence differ by defendant race*? Adjusting for features in the judge information set (such as charge and risk score component characteristics) does not explain the differential policy effect across defendant race. Instead, there are two forces behind differential adherence. First, a substantial part of the explanation is different responses by judge, which aligns with Stevenson (2017)'s prior work investigating pretrial release disparities. Recall that while HB463 was a state-wide policy change, judges each had discretion in when to deviate from the presumptive default.

and Yang (2018), the objective is "to set bail conditions that allow most defendants to be released while minimizing the risk of pre-trial misconduct."

[9] This institutional feature means makes interpreting judge decisions more straightforward in Kentucky than in other states (since prosecutor actions at this stage are nonexistent).

[10] There is more information about this decision in the later sections. The decision usually takes the form of a phone call between a pretrial officer and a judge, in which the pretrial officer relays relevant information about the defendant (age, name) and charges (description, class, level). Judges take such calls 1-4 times per day depending on the size of the county they work in. Ability to pay is not mandated to be in the call. Judges can ask questions and calls may differ by judge-pretrial officer pair.

[11] I show that differential effects remain when honing in on low and moderate risk defendants.

[12] Racial disparities for low and moderate risk defendants discontinuously increase after HB463 even though the policy recommended the same treatment (non-financial bond) for both racial groups with those risk levels.

Extreme spatial variation in percentages of black defendants across counties means that different policy responses across judges (who serve specific counties[13]) had large effects on aggregate racial disparities. In fact, allowing for heterogeneous policy responses by judge explains the vast majority of racial disparities observed among low risk defendants.[14] *This finding highlights the potential for geographic variation in policy responsiveness to widen disparities across demographic groups.* I discuss a few theories for why variation in judicial policy responsiveness may correlate with racial make-up of judges' defendant populations; future work will investigate this empirically using data on judge characteristics.[15]

Second, *even within judge and time, black moderate risk defendants are treated more harshly than similar white moderate risk defendants after but not before HB463.* After HB463, judges are 10% more likely to deviate from the non-financial bond recommendation for moderate black rather than similar moderate white defendants. (The same is not true for low risk defendants.) This is suggestive evidence that judges interpret risk score levels differently based on defendant race.[16] In terms of mechanism, if judges want to be more cautious than the policy default, they may shift away from non-financial bond for moderate risk defendants since they seem more ambiguous than low risk defendants. If judges have some sense of the underlying continuous risk score distribution and assume that moderate risk black men are higher risk than moderate risk white men, this could potentially explain the results. However, this explanation would require that judges shifted weight away from more detailed information (e.g., components of the risk score levels) to the combination of two heuristics: risk level and race. This is conceptually related to Kleinberg and Mullainathan (2019)'s theoretical finding that simplified prediction functions incentivize decision-makers to consider group membership information. Further work is required to better pin down this result and the underlying mechanism.

On the whole, the paper demonstrates how interactions between human discretion and prediction tool recommendations can yield unequal policy effects across racial groups. It contributes to the larger conversation on the societal effects of algorithms by addressing the gap between (1) academic literature that shows gains in social equality from algorithms assuming perfect compliance (Kleinberg et al. 2017; Dobbie et al. 2018; Cowgill 2018a) and (2) ample evidence that we are not in a world of perfect compliance (Main 2016; Hoffman, Kahn, and Li 2017; New 2015). Whether predictive tool policies alleviate or exacerbate existing demographic disparities depends on a number of factors: the status quo,[17] the

[13]Judges serve single counties or groups of counties depending on the county sizes.

[14]This result is highly reminiscent of Goncalves and Mello (2017)'s recent work which finds that a large share of the disparity in treatment of minority drivers by police officers "is due to the fact that minorities drive in areas where officers are less lenient to all motorists."

[15]The importance of decision-maker characteristics was recently illustrated by Bulman (2019); this paper showed that sheriff race is related to arrest rate racial differences. Specifically, he provided evidence that different crime type targeting could be the underlying mechanism.

[16]The result is consistent with the theory of disparate interactions, introduced by Green and Chen (2019). It is also consistent with Cowgill (2018b)'s findings that black defendants' outcomes are more sensitive to risk thresholds.

[17]To address the status quo means to speak to racial disparities in the absence of prediction tools – that is, much of the sociological and economics literature on discrimination. Related papers span intentional experimental studies (e.g., Pager, Bonikowski, and Western 2009) as well as natural experiments (e.g., Goldin

tools themselves,[18] the policies that guide their usage, and how decision-makers follow said policies. While I use this paper to focus on policy recommendation adherence, this work is inherently related to those other factors and their accompanying literatures.[19]

For one, this paper contributes to a quickly growing literature on risk assessment policies (Sloan, Naufal, and Caspers 2018; Stevenson 2017; Doleac and Stevenson 2018; Garrett and Monahan 2018; DeMichele et al. 2018). Most related to my paper is Stevenson (2017), which also focuses on HB463 in Kentucky. In her thorough investigation, Stevenson uses graphical time-trends to show that while HB463 had effects on bail setting behavior, the effects on pretrial release (after 3 days of booking) were minimal in comparison. She illustrates increases in racial disparities in bail setting and release after policy adoption. Digging deeper into the release disparities, she shows that this increase was primarily because judges in largely white counties responded more to the policy than judges in more racially mixed counties, which I find is true for non-financial bond setting as well. However, within the same county, she shows white and black defendants saw similar increases in release. The object of interest in my paper differs in two ways. Recall I am interested in whether judges' initial bond decisions deviated from the non-financial bond recommendation. This approach (1) focuses on the initial judge decisions themselves rather detention consequences[20] and (2) takes into account the associated risk score levels, which allows me to investigate deviations from the actual policy recommendation (set non-financial bond if low or moderate risk) as well as heterogeneity in results over risk levels (low, moderate, high).

A number of recent papers also focus on human usage of predictive tools. Green and Chen (2019) uses an Amazon Mechanical Turk experiment (rather than observational data on judge decisions) to investigate human interactions with risk scores. They find "risk assessments led to higher risk predictions about black defendants and lower risk predictions about white defendants."[21] Running an experiment with real judges, Skeem, Scurich, and Monahan (2019) finds that the same risk assessment information produces different judicial decisions based on socioeconomic class of the defendant. Using data from Broward county, Cowgill (2018b) finds that outcomes for black defendants are more sensitive to risk score thresholds than are outcomes for white defendants.[22] On the theoretical side, Kleinberg and Mullainathan (2019) show that simplified prediction

and Rouse 2000). Of particular relevance is the literature on race and judge decisions (Arnold, Dobbie, and Yang 2018; Abrams, Bertrand, and Mullainathan 2012; Cohen and Yang 2019).

[18]Kleinberg et al. (2017), Jung et al. (2017), Kleinberg et al. (2018), and Corbett-Davies et al. (2017) focus on the technical validation and generation of risk assessments.

[19]For a literature review on algorithmic fairness and economics, see Cowgill and Tucker (2019).

[20]Detention consequences are a function of the initial (and followup) judge decisions as well as ability to pay. See Appendix A for more on marrying bond decision and detention trends in Kentucky. Note that the first-stage decision is quick and economically meaningful for predicting immediate release – non-financial bond corresponds with 95.6% chance of immediate release while financial bond corresponds with a 20.4% chance of immediate release.

[21]This is conceptually related to the "shifting standards" model outlined by Biernat and Manis (1994).

[22]Cowgill (2018b)'s outcome data corresponds to length of jail stay rather than judge decision. Length of stay is a downstream consequence from the judicial decision-making itself. See Appendix A for more on that distinction.

functions (e.g., risk assessments) create incentives for decision-makers to consider group membership information. Very similar to my setting but focused on the labor market, Hoffman, Kahn, and Li (2017) look at how managers use discretion in following a score-based recommendation when making hiring decisions. Deviation from a default is a binary decision of interest for understanding the implications of providing experts (e.g., managers, judges) with prediction-based recommendations. I am able to build on Hoffman, Kahn, and Li (2017) by addressing the importance of demographics in investigating recommendation deviations.[23]

Finally, this paper joins a literature that highlights how spatial variation in decision-making plays a large role in explaining racial disparities. While the role of geography has been highlighted repeatedly in the context of health disparities (Chandra and Skinner 2003), this line of research has recently evolved in the context of policing. Goel et al. (2016) explain that much of racial disparities generated by Stop and Frisk were due to the highly localized nature of the policy. On a similar note, Goncalves and Mello (2017) demonstrate that racial disparities in speeding punishments are largely because minorities drive in areas where officers are less lenient overall. They argue that reallocating officers across locations could reduce the aggregate disparity in treatment.[24]

The remainder of the paper proceeds as follows. Section 2 provides a conceptual framework for understanding how discretion complicates the effects of predictive tool recommendations. Section 3 introduces the Kentucky pretrial environment. Section 4 describes the data. Section 5 presents the main results by empirically exploring disparities in deviation behavior. Section 6 discusses mechanisms behind variation in judge responsiveness and the lingering racial disparity for moderate risk defendants. Section 6 concludes and discusses avenues for future work.

2 Conceptual Framework

In order to discuss how discretion and predictive tools interact, I consider the salient example of risk scores in bail decisions that maps onto my empirical environment. In such contexts, risk scores do not mechanically determine final outcomes. Rather, they are decision-making aids that are provided to final human decision-makers – judges.[25] Risk score policies often set recommended actions for decision-makers based on risk score ranges. Whether these policies reduce or increase racial disparities depends on disparities in the pre-period and disparities in the post-period, which are a function of: risk score calculation, risk score recommendations, and how judges follow or deviate from those recommendations. I use this section to describe two types of deviations from risk score recommendations can create larger or smaller racial disparities than those generated by perfect compliance (i.e., no judicial discretion).

[23] They are unable to investigate results by demographics due to data limitations.

[24] However, this assumes that lenience is not endogenous to assigned area.

[25] Note that pretrial decisions can also be made by other criminal justice system actors such as magistrates.

In this set-up, judges can set either financial or non-financial bond. Financial bond is more restrictive so judges set it for defendants who are more likely to commit pretrial misconduct.[26] Consider a risk score policy that recommends 85% of black defendants receive non-financial bond and 90% of white defendants receive non-financial bond. (Assume the risk score policy is constructed to help increase non-financial bond rates and judges set non-financial bond at low rates before the policy.) Under perfect compliance, this would mean a racial disparity (in favor of whites) of 5 percentage points.[27] However, the actual observed disparity will be complicated by one or both of the following two types of deviations.

Deviation by Judge: Regardless of policy changes, judges retain individual discretion, meaning they can differ in how they respond to policy recommendations. Assume each judge picks some percent of the time they will follow the recommendation (a "compliance rate") and assume they follow this percentage equally by defendant race. Now, recall judges serve different populations of defendants, often based on geography. If judges with higher compliance rates tend to serve defendant populations with relatively higher percentages of black people, then racial disparities (in favor of whites) will be smaller than those generated by perfect compliance (since black defendants are more likely to face more compliant judges[28]). If judges with higher compliance rates tend to serve defendant populations with relatively higher percentages of white people, then racial disparities (in favor of whites) will be larger than those generated by perfect compliance.[29] In short, differences in deviations across judges (even if their deviation behavior is identical by defendant race) can meaningfully affect aggregate racial disparities due to the geography of judges.

Deviation by Defendant Race: The first deviation was across judges, assuming identical treatment of racial groups within judge. The second sort of deviation is different treatment of racial groups within judges. Judges may interpret risk score recommendations differently based on defendant race. In a decision-making system with simplified predictive scores, Kleinberg and Mullainathan (2019) explain there are additional incentives for decision-makers to consider group traits such as race. If judges treat black defendants as more risky than white defendants with identical risk scores, this yields larger racial disparities (in favor of whites) than those generated by perfect compliance.[30] If the opposite is true (as judges could, for instance, take into account the increased likelihood of low level arrests and subsequent convictions for black people[31]), then this yields smaller racial

[26] Assume they do not have information about defendants' abilities to pay.

[27] Again, whether this is an improvement or exacerbation of the status quo depends on the specific policy context (i.e., the racial disparities in the pre-policy period). For instance, Kleinberg et al. (2017) presents evidence that algorithmic prediction could reduce racial disparities in New York City pretrial decisions.

[28] Assuming the policy was designed to increase non-financial bond rates, more compliant judges, in this case, are more lenient judges.

[29] This is related to Goncalves and Mello (2017)'s finding that minorities drive in areas where officers are less lenient overall. Similarly, Goel et al. (2016) finds that much of the racial disparity in Stop and Frisk hit rates is explained by geography.

[30] This would be in line with both Skeem, Scurich, and Monahan (2019) and Green and Chen (2019).

[31] Risk scores do not currently attempt to take into account possible biases generated by criminal history data though this has been proposed by AI (2019).

disparities than those generated by perfect compliance.

In theory, these two behavioral deviations (deviations across judges independent of race or deviations within judge by race) could go in either direction and, thus, yield either larger or smaller racial disparities than those generated by perfect compliance. In this paper, I discuss the Kentucky pretrial system and show that both of these two forces pushed in the direction of larger racial disparities in favor of whites.[32]

3 Empirical Environment

In response to large increases in the incarcerated population between 2000 and 2010,[33] Kentucky House Bill 463 (HB463) went into effect on June 8, 2011. The law made pretrial risk assessment a mandatory part of bail decision-making and set the default decision for low or moderate risk defendants to be non-financial bond.[34] If judges wanted to defect from this recommendation, they had to provide a reason.[35] As such, HB463 mandated the use of risk levels and set a recommended default based on those levels.

3.1 Kentucky Pretrial Overview

Kentucky is well-known for its pretrial services for a few reasons. For one, it was the first state to ban commercial bail bonds in 1976.[36] Kentucky boasts one pretrial services agency that serves all 120 counties in the state, meaning that data management and collection is unified and well-organized. Unlike in other states, Kentucky Pretrial Services is part of the judicial branch; it is a state entity that works for the courts (and is state-funded).[37] While pretrial employees are housed in individual counties, they do not work for the individual counties.[38] Kentucky Pretrial Services even has a virtual tour of their pretrial services system online for other jurisdictions to use in ongoing bail reform efforts.[39]

During 2009-2013 in Kentucky, after a defendants is booked into jail, a pretrial services officer in that county uses the Kentucky Pretrial Risk Assessment framework (discussed in the next subsection) to calculate a risk level. Within 24 hours of booking, the officer presents information about the defendant and incident in a bail hearings with a judge.[40]

[32]See Figure 13 for empirical illustration.

[33]According to Stevenson (2017), between "2000 and 2010, Kentucky's incarcerated population – both jail and prison – grew by 45%, more than three times the U.S. average."

[34]See bullet 3 in Figure 1.

[35]In practice, this could be as simple as saying a few words (e.g. "flight risk") to the pretrial officer.

[36]It was one of four states with this ban as of 2018.

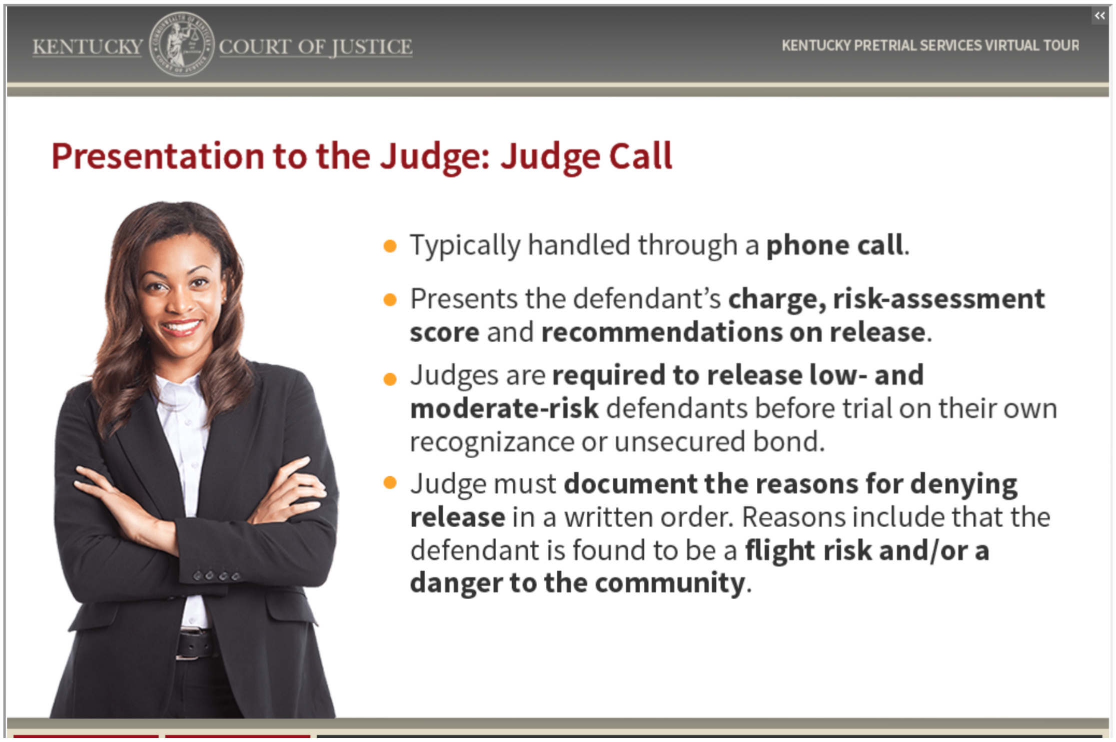
[37]Much of the information in the following paragraphs is from an interview with the Executive Officer of Kentucky Pretrial Services, Tara Blair.

[38]As of January 2019, there were about 251 employees in Pretrial Services in Kentucky. Approximately 202 employees are pretrial officers and/or supervisors and 49 are risk assessment specialists and/or coordinators.

[39]Kentucky was also the first jurisdiction to pilot the Public Safety Assessment (PSA) risk assessment.

[40]Appendix B for more information on judges.

Figure 1: Judge Call Information



The bail hearing usually is a phone conversation (see Figure 1) between a pretrial officer and a judge.[41]

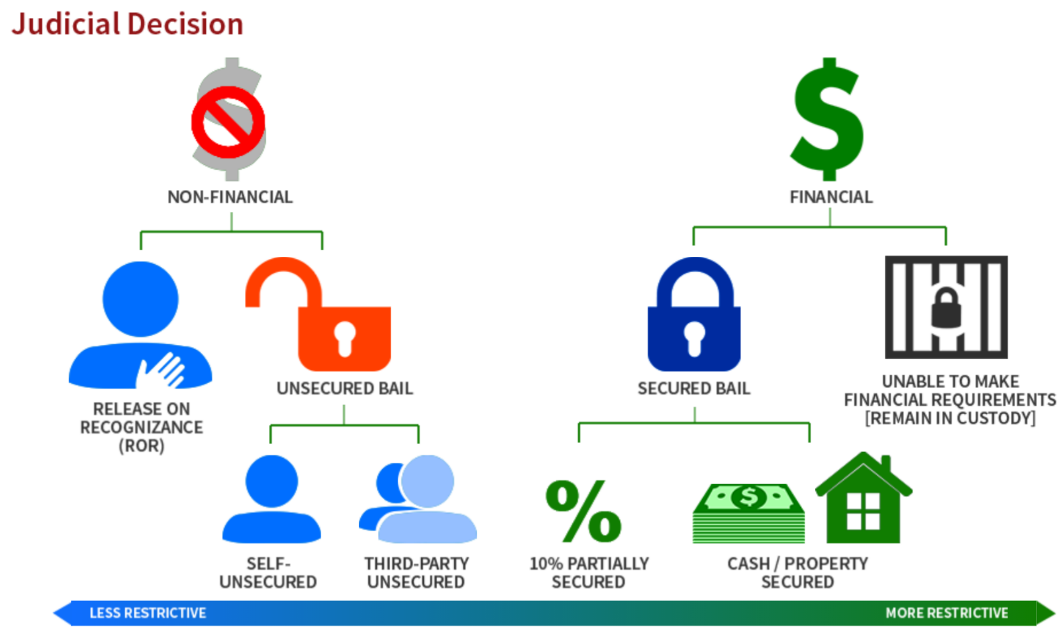
After receiving information about the defendant and case from the pretrial officer,[42] the judge decides on bond type and amount (if financial) as well as other conditions of release (such as supervision). In this paper, I focus on whether that initial bond decision was non-financial or financial since HB463 specifically recommended that judges set non-financial bond for low and moderate risk defendants.[43] While there are many smaller bail outcomes,

[41]This is abnormal in the US as most jurisdictions use in-person bail hearings. If pretrial officers and judges are in the same place, this could be an in-person meeting instead. The data does not specify whether initial bond decisions are via judge calls or not, so it is unclear to me how many initial bond decisions I observe are via phone calls. Kentucky has been using calls for pretrial services since 1976 – this is especially efficient in areas of the state where people are very spread out and, therefore, would mean significant time costs for in-person bail hearings.

[42]The eight example judge calls that available online on the Kentucky pretrial website include the following information: name, age, risk score information, list of charges, and incident description. The incident description quotes information from the police report. In Kentucky, police have full authority to charge; there is no prosecutorial review before the judge call. Note that while demographic information on race or gender can be missing explicitly in the call, these details are implicitly included. Gender is revealed through usage of pronouns (e.g. "he" and "she") when the pretrial officer discusses the defendant. Meanwhile, names (especially in combination with the county) can signal information about race. Moreover, race and ethnicity were on judge forms about cases during my studied time period of interest, meaning they could be explicitly observed when judges used said forms in their decision-making. (However, these details have since been removed from judge forms.)

[43]In the data, non-financial and financial bond correspond to a 95.6% or 20.4% chance of initial release

Figure 2: Bond Types



the key overarching decision is whether to set financial conditions or not, as illustrated by Figure 2.[44]

3.2 Kentucky Pretrial Risk Assessment

Kentucky has used a few different risk assessment tools over the years. At first, Kentucky used a six-question tool developed by the Vera Institute.[45] In 2006, Kentucky moved to its own Kentucky Pretrial Risk Assessment (KPRA) tool[46] – this is the tool used during my time period of interest. In June 2013, Kentucky began to use the Public Safety Assessment (PSA), which was developed by the Laura and John Arnold Foundation.[47]

(release on that bond), respectively. If the defendant has not posted bail within 24 hours of the initial decision, the pretrial officer informs the court and the judge can change the bond to increase the chance that they can be released pretrial. If the defendant remains detained pretrial, the next time bond could be reconsidered is usually first appearance (Stevenson 2017).

[44]Note that there is no non-refundable piece to bond (as there is in many states), so bond is fully returned to defendants at the disposition of the case regardless of outcome. All offenses are bailable except capital offenses in Kentucky, meaning judges can rarely simply detain defendants as they can in Washington DC or New Jersey (12-15% are denied bail in Washington D.C.) (Santo 2015).

[45]Information on the history of risk assessment in Kentucky is via communication with Executive Officer of Pretrial Tara Blair.

[46]The tool was created in-house, fitting a regression model to predict pretrial misconduct using the existing Kentucky data at the time.

[47]The PSA is used exclusively in the pretrial stage of the criminal justice system; its formula is open and meant to be shared publicly (Schuppe 2017a). It was initially developed in 2013 (and altered slightly in 2014) by investigating 746,525 cases in which defendants had been released pretrial (over 300 jurisdictions) to determine which defendant characteristics were most predictive of new crime, new violent crime, and

The KPRA tool is not a complex black-box machine learning tool.[48] Rather, it is a check-list tool that added up points based on Yes/No answers to a series of questions. It was modified slightly on March 18, 2011. Figure 3 documents the weights that various components are given in both the 7/1/09-3/17/11 and 3/18/11-6/30/13 version of the scores (Austin, Ocker, and Bhati 2010).

The factors in the KPRA are mostly criminal history elements (e.g., prior failure to appear, pending case) but there is also information about the current charge (e.g., whether the charge is a felony of class A, B, or C) and defendant personal history (e.g., verified local address, means of support).[49] To calculate a risk score level, the weights shown in Figure 3 are added up and then mapped to a low, moderate, or high score level. Before 3/18/11, totals of 0-5, 6-12, and 13-23 correspond to low, moderate, and high levels, respectively. As of 3/18/11, totals of 0-5, 6-13, and 14-24 correspond to low, moderate, and high levels, respectively. During the time period of interest (around the HB463 policy change), judges were informed of risk score levels rather than total number of points. The 2011 law did not introduce the KPRA levels for the first time; it mandated their consideration in bail decisions.

4 Data and Descriptive Statistics

I use data from the Kentucky Administrative Office of the Courts (KY AOC) on initial bond decisions for misdemeanors and felonies.[50] I consider all initial bond decisions about male defendants from July 1, 2009 to June 30, 2013 (the time period that featured the KPRA tool).[51] The final dataset consists of 383,080 initial bond decisions, which cover decisions

failure to appear pretrial (Laura & John Arnold Foundation 2013). As of late 2018, "over 40 jurisdictions have either adopted the PSA or is engaged in implementation with LJAF technical assistance" (John Arnold Foundation 2018).

[48]The Angwin and Kirchner (2016) article that generated lots of press about risk assessment scores was about a black-box machine learning tool called COMPAS.

[49]While most of these questions are straightforward given current charge and criminal history, items 2, 3, and 11 were more complicated to ascertain for Pretrial Services. Item 1 was a "yes" if at least five people (reached via the defendant's cell phone) were about to verify the defendant's local address and confirm they had lived in the are for the past twelve months. Item 2 was a "yes" if a defendant was one or more of the following: employed full-time, the primary caregiver of a child or disabled relative, a Social Security/disability recipient, employed part-time employee or a part-time student, a full-time student, retired, or living with someone who supported them. Item 11 was a "yes" if the defendant had 3 or more drug/alcohol-related convictions of drug/alcohol in last 5 years (a longer period was considered if the defendant had been incarcerated at some point).

[50]On a technical note, I use R Markdown for my data cleaning and analysis. The R packages I use are: Wickham, Chang, et al. (2018), Wickham (2018), Hlavac (2015), Dowle and Srinivasan (2018), Firke (2018), Wickham (2017), and Wickham, François, et al. (2018).

[51]Recall that Kentucky switched its risk score system to the PSA on July 1, 2013.

Figure 3: Weighting Rules for Kentucky Pretrial Risk Assessment

	Scoring Items	Current		Modified	
		Yes	No	Yes	No
1	Does the defendant have a verified local address and has the defendant lived in the area for the past twelve months?		1		2
2	Does the defendant have verified sufficient means of support?		1		1
3	Did a reference verify that he or she would be willing to attend court with the defendant or sign a surety bond?		1	Removed	
4	Is the defendant's current charge a Class A, B, or C Felony?	1		1	
5	Is the defendant charged with a new offense while there is a pending case?	5		7	
6	Does the defendant have an active warrant(s) for Failure to Appear prior to disposition? If no, does the defendant have a prior FTA for felony or misdemeanor?	4		2	
7	Does the defendant have prior FTA on his or her record for a criminal traffic violation?	1		1	
8	Does the defendant have prior misdemeanor convictions?	1		2	
9	Does the defendant have prior felony convictions?	1		1	
10	Does the defendant have prior violent crime convictions?	2		1	
11	Does the defendant have a history of drug/alcohol abuse?	2		2	
12	Does the defendant have a prior conviction for felony escape?	1		3	
13	Is the defendant currently on probation/ parole from a felony conviction?	2		1	

for 192,758 distinct defendants by 563 distinct judges.[52][53]

4.1 Charges

An important part of the initial decision is the set of charges brought against a defendant. Recall that I am focusing on misdemeanors and felonies.[54] In terms of charge severity, I plot the most severe (highest level and charge combination) charge for each initial decision in Figure 4. This illustrates that 79.5% top-charges are class A or class B misdemeanors, and 20.5% are class D felonies. For further context, Table 1 lists the most common charge for each level and class of offense. Only 8.5% of initial decisions include a top-charge that is a class A, B, or C felony. In terms of specific charge characteristics, 1.2% of initial decisions involved weapon-related charges, 4.9% of initial decisions involved violence-related charges, and 8.3% of initial decisions involved drug-related (excluding alcohol) charges.[55]

Table 1: Most Common Charges by Level and Class

Level	**Class**	**Most Common Charge**
Felony	A	Murder
Felony	B	Manufacturing Methamphetamine, 1st Offense
Felony	C	Burglary, 2nd Degree
Felony	D	Flagrant Non Support
Misdemeanor	A	Assault 4th Degree, Domestic Violence, Minor Injury
Misdemeanor	B	Operating Motor Vehicle Under the Influence of Alcohol/Drugs, .08, 1st Offense

4.2 Bond Types

Bond comprises conditions for release from jail. While there are a range of possible conditions, the most salient conditions are monetary. Recall from Figure 2 that the initial bond decision by the judge can be financial or non-financial. Financial bond means there are financial conditions that must be met before release; there are no such financial conditions

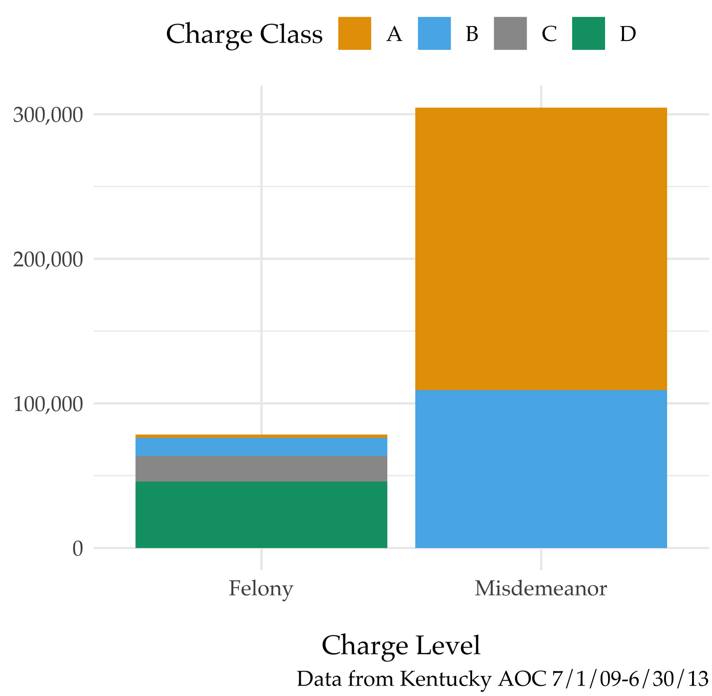
[52] I first consider all 1.56 million initial bond decisions for 7/1/09-12/30/17 and then subset to misdemeanor and felonies (with a known class) within the 7/1/09-6/30/13 time period with known age, gender, judge, race, risk level, and risk level components – this leads to a sample of 524,229 initial bond decisions. After subsetting to those decisions about male defendants, I have 383,080 initial bond decisions; this is 73% of the sample that includes both genders.

[53] For more on judge types see Appendix B.

[54] Most violation offenses, which are lower level, do not result in a bond hearing. In fact, they are so rarely associated with a bond hearing that if I don't mechanically exclude violation and other offenses, they comprise only 2% of the sample.

[55] I define weapon-related charges as those with descriptions including the words "gun", "firearm", or "weapon". I define violence-related charges as those with descriptions including the words "violence", "assault", "rape", or "murder". I define drug-related charges as those with descriptions including the words "cocaine", "heroin", "marijuana", "drug", or "meth", but excluding charges that include "under/infl" since those are agnostic to alcohol/drugs.

Figure 4: Top Charges by Level and Class



for non-financial bond. As such, non-financial bonds are less financially restrictive for defendants. Figure 5 shows the frequency of the specific types of initial bond outcomes. Bond can be refused only for capital crimes (e.g. murder) for Kentucky, so "no bond" is observed in only around 3.9% of observations. Bond is financial in 68.3% of the initial bond hearings and is mostly cash bond.[56] In 27.8% of initial bond hearings, the bond type is non-financial, which is pretty evenly split across release on recognizance, self-unsecured, and third-party unsecured bonds ("surety").[57][58]

4.3 Race

I am limiting my discussion to male defendants. In terms of race, defendants are white in 79.1% of these initial bond hearings and black in 20.6% of them.[59] For context, the 2017

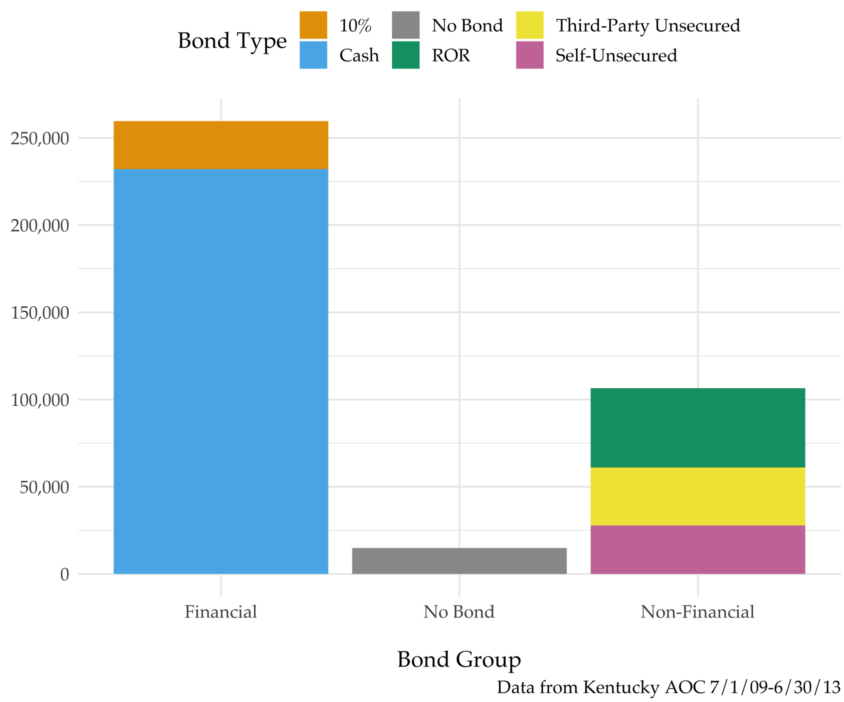
[56] Cash bond means the entire amount of the bail must be posted in cash. A 10% bond means that only 10% of the amount must be posted.

[57] Unsecured bond means that the defendant would owe some amount of money if the defendant fails to appear.

[58] This is consistent with Stevenson (2017)'s finding (using Kentucky data from a different time range) that "[i]f judges followed the recommendations associated with the risk assessment, 90% of defendants would be granted immediate non-financial release" but "[i]n practice, only 29% are released on non-monetary bond at the first bail-setting."

[59] Defendants are Asian in 0.24% of initial bond hearings.

Figure 5: Bond Outcomes by Group and Type



Kentucky state population was 87.8% white and 8.4% black.[60][61]

Racial composition varies spatially over the state of Kentucky. Figure 6 shows that while a handful of counties have over 30% black defendants, most counties have less than 5% black defendants. This variation is due to preexisting spatial racial segregation in the state. The choropleth in Figure 6 illustrates the variation across the state by coloring counties based on their percentages of black defendants. Christian, Jefferson, and Fayette counties are the counties with the highest percentages of black defendants in my data.[62] Meanwhile, most counties in the east are dark purple, meaning their percentages of black defendants are near zero.

4.3.1 Risk Scores and Race

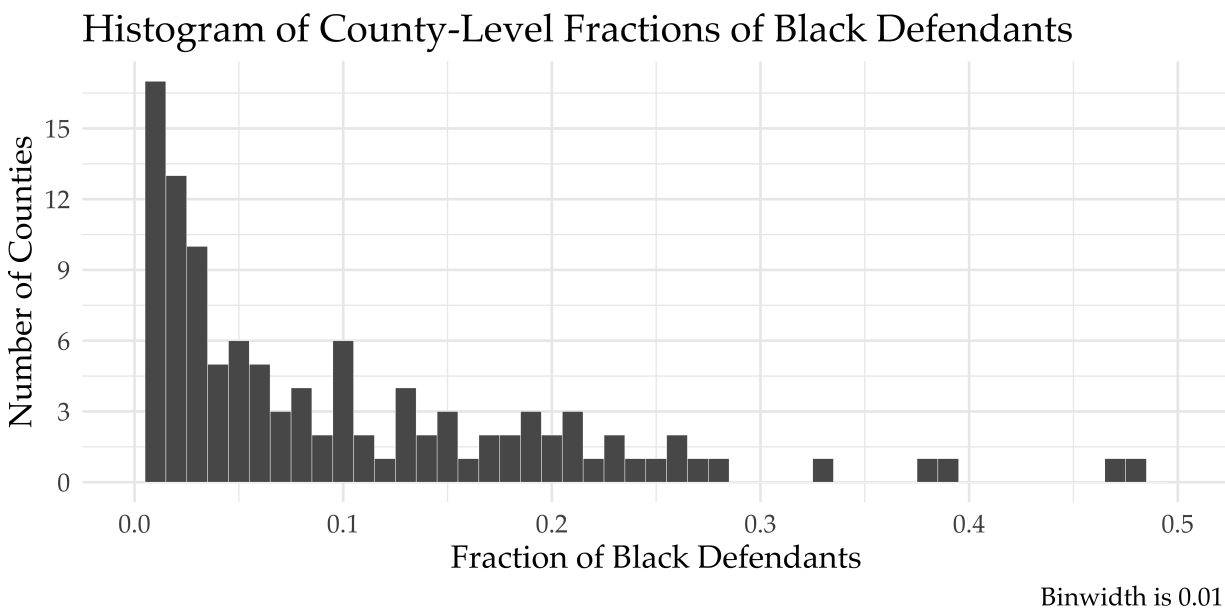
Risk assessment score distributions may differ across racial groups. This is the primary mechanism through which the current literature discusses how risk scores may impact racial disparities. In Angwin and Kirchner (2016)'s piece about the COMPAS algorithm, the

[60] This is from the Census QuickFacts data.

[61] In terms of ethnicity, defendants are recorded as Hispanic in only 2% of these initial bond hearings. In 70% of initial decisions, defendants are recorded as non-Hispanic and in 27.9% of initial decisions, ethnicity is recorded as unknown. Due to the small sample size, I will not be discussing disparities by ethnicity.

[62] For context, the largest cities in Kentucky are Louisville, located in Jefferson county, and Lexington, located in Fayette county.

Figure 6: County-Level Fractions of Black Defendants



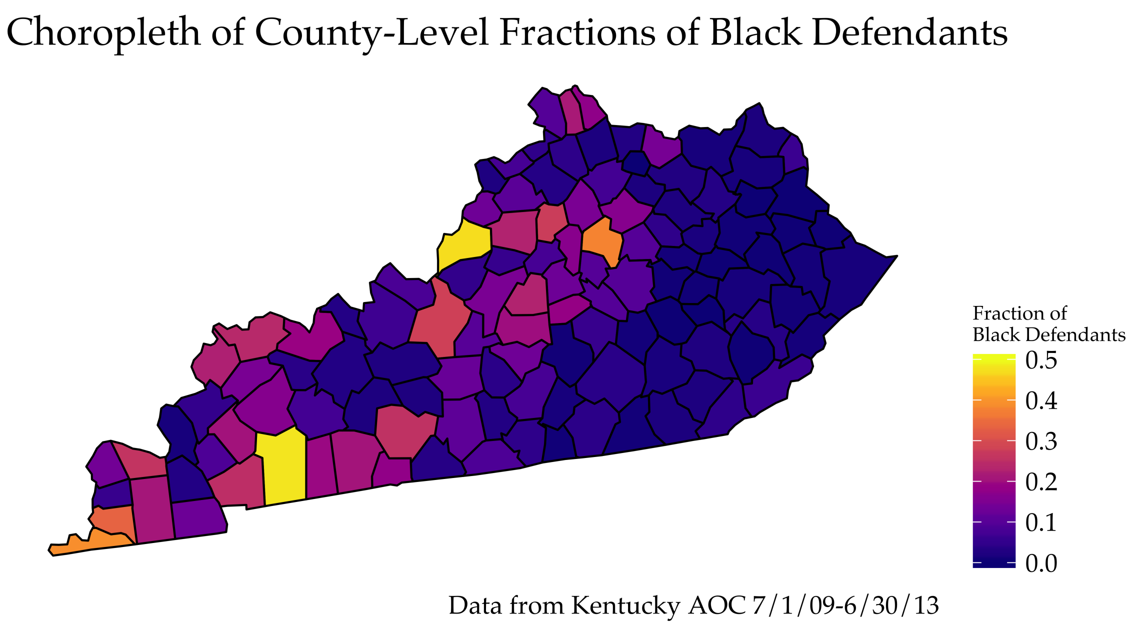
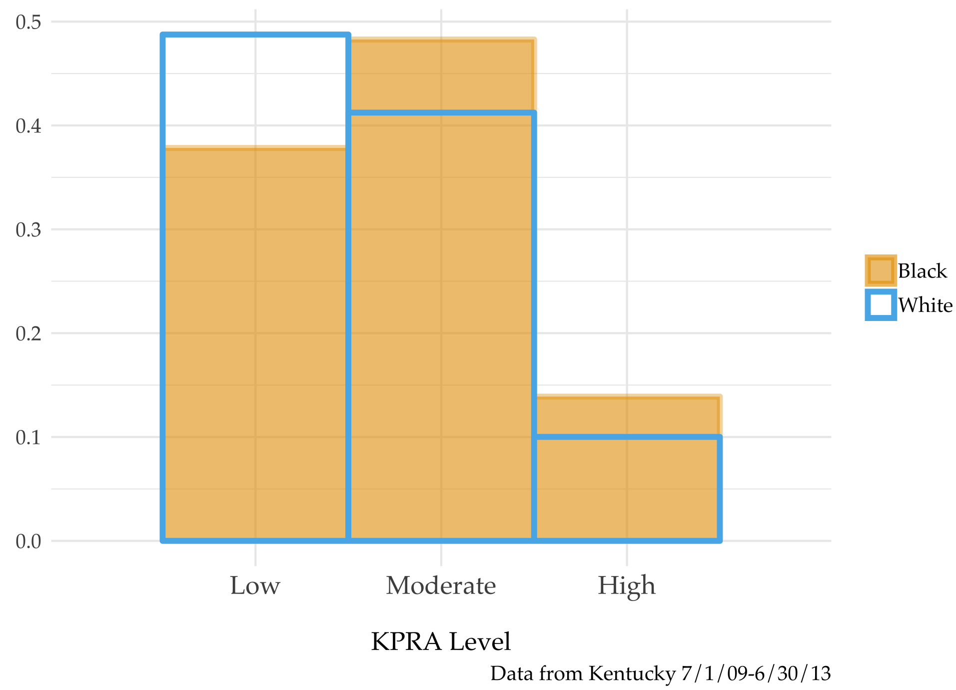


Figure 7: Risk Level Density by Race



score distributions by race are strikingly different – white defendants are notably skewed towards lower-risk categories (1 out of 10, in particular), while black defendants scores are evenly distributed across the full 1-10 range. Figure 7 compares the risk score densities for black and white defendants for KPRA risk assessment levels. The distributions are substantially more similar across races than in the case of COMPAS. However, Figure 7 does show that white defendants are more heavily skewed towards the lower-risk levels, while black defendants are more heavily skewed towards the higher-risk levels.

The three levels (low, moderate, and high) are what is communicated to judges during the time period of interest. However, as mentioned before, there are more specific raw scores calculated for each defendant, which are then converted into these coarse (low, moderate, high) categories. See Figure 8 for a more detailed comparison of score distributions.[63]

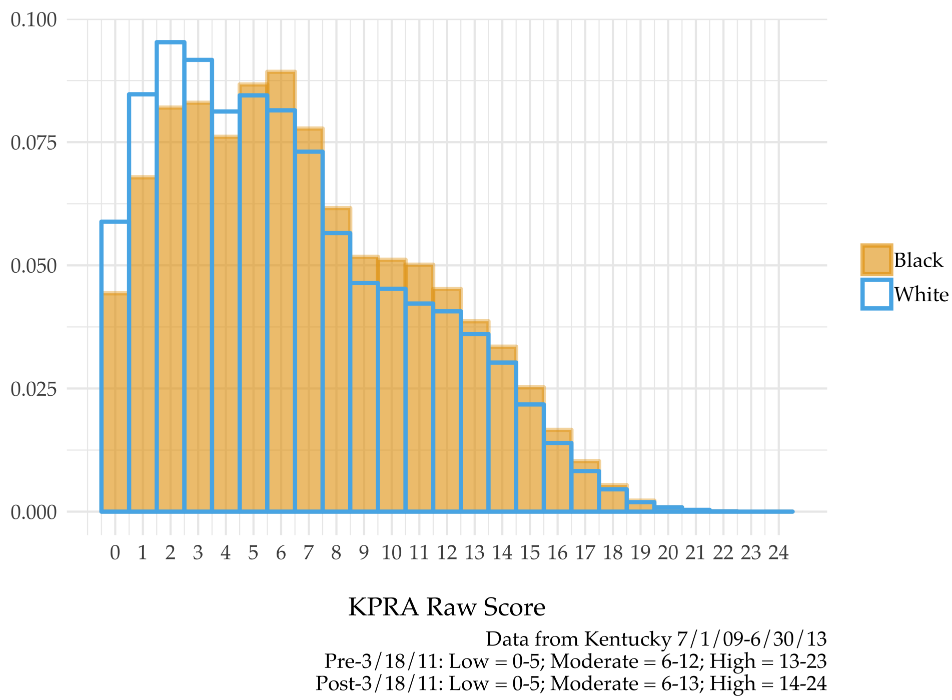
5 Disparities in Deviations

In this paper, I focus on how judges respond to predictive recommendations. Specifically, I focus on the decision to set non-financial bond since HB463 made non-financial bond the presumptive default for low and moderate risk defendants.[64] Judges had to give pretrial

[63] For the time period up until 3/17/11, the scores ranged from 0-23. After 3/18/11, the scores ranged from 0-24.

[64] Appendix A speaks to the realm of financial bond decisions as well as how these combine with non-financial bond decisions to explain trends in pretrial release on initial bond. That appendix builds on

Figure 8: Raw Risk Score Density by Race



officers a reason for defecting from this recommendation, as seen in Figure 1.[65]

Interpreting judge initial decisions is straight-forward in Kentucky since there is no prosecutorial review before the judge sets initial bond. This means that judge decisions are not framed or conditioned on prosecutor actions. Rather, Kentucky judges receive administrative information from pretrial officers. Some of this information is derived from the police, who have full authority to charge in Kentucky. Before addressing changes in judge behavior after HB463, it is first worth establishing that police did not discontinuously change their charging behavior at the time of this policy. Otherwise, any observed jumps in judge behavior could be partially due to changes in police behavior (i.e., discontinuous changes in charging). To track charging behavior, Figure 9 shows the percentage of top charges by level and class over time. There is no visual evidence that charging by police discontinuously changes at HB463 implementation.

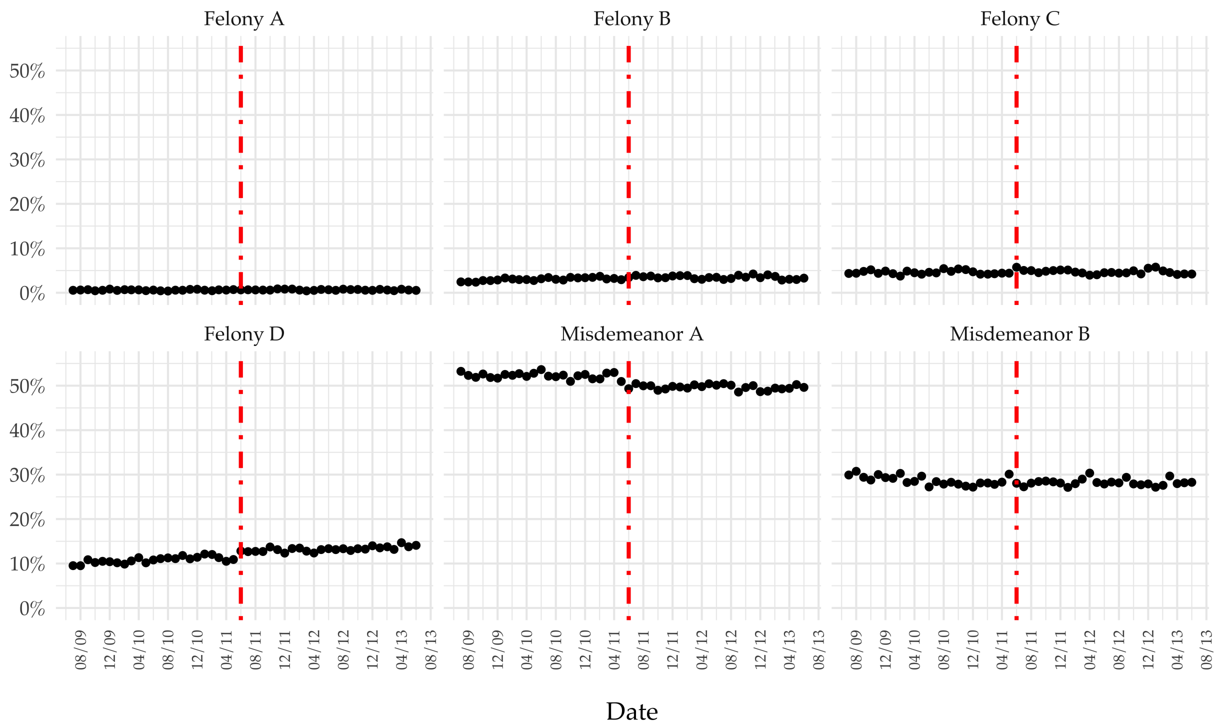
On the other hand, HB463 has a clear and immediate effect on judge decisions. Figure 10 shows the simplified effect of HB463 on non-financial bond outcomes across all defendants, black and white. The percentage of male defendants receiving non-financial bond jumps in a clearly discontinuous manner at the effective date of HB463, increasing from a pre-HB463 mean of 22.5% to a post-HB463 mean of 36.5%.

Black and white defendants did not equally benefit from HB463. Figure 11 presents an

Stevenson (2017)'s prior findings, marrying non-financial bond findings and pretrial detention trends after HB463.

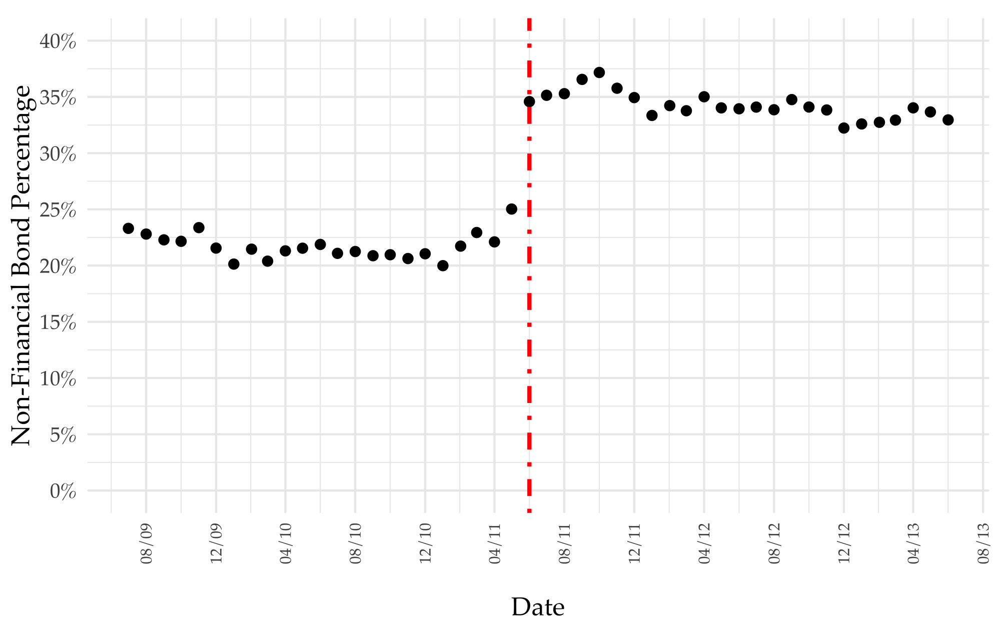
[65]However, the reason could be as simple as saying "flight risk," meaning the cost to deviation was not very high for judges.

Figure 9: Top Charge Percentages Before and After HB463



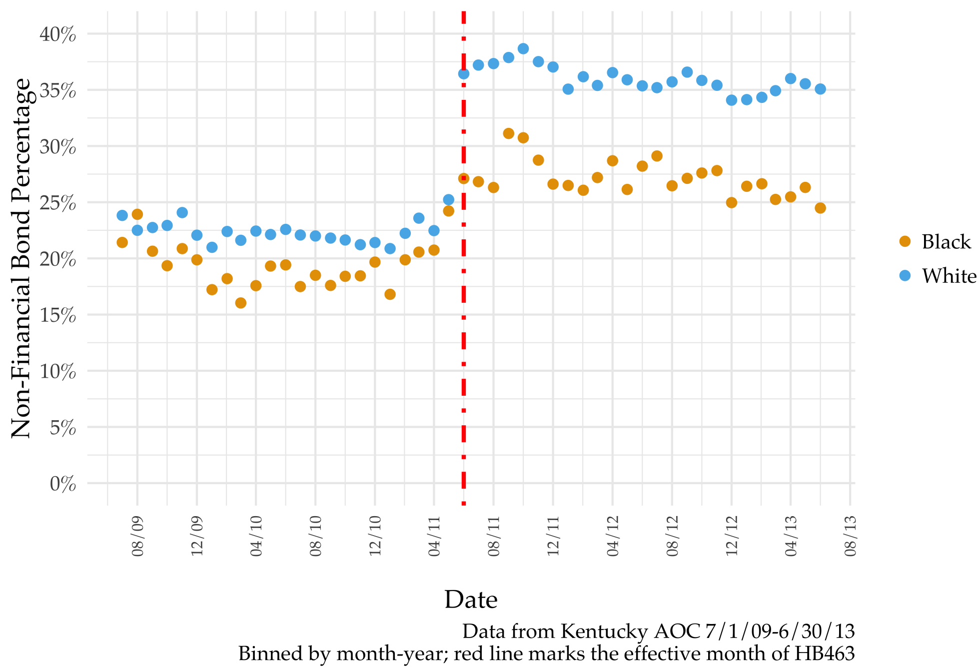
Data from Kentucky AOC 7/1/09-6/30/13
Binned by month-year; red line marks the effective month of HB463

Figure 10: Bond Outcomes Before and After HB463



Data from Kentucky AOC 7/1/09-6/30/13
Binned by month-year; red line marks the effective month of HB463

Figure 11: Bond Outcomes Before and After HB463 by Race



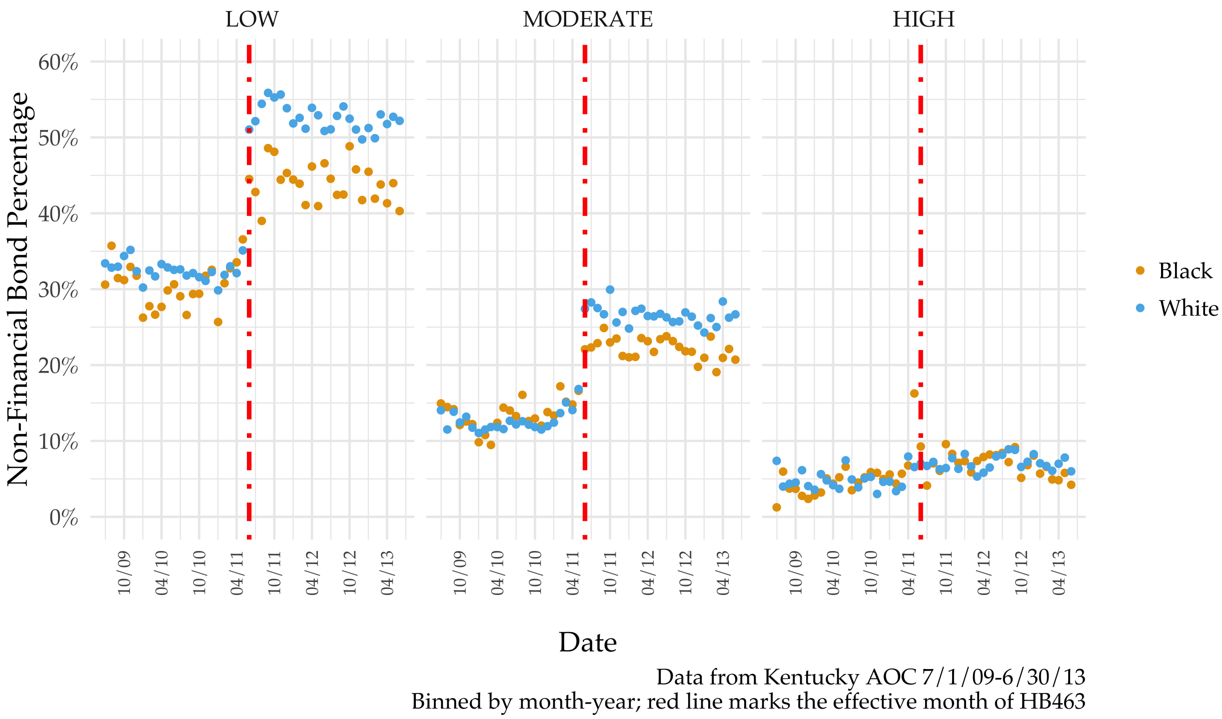
interrupted time-series, first presented by Stevenson (2017). The graph shows that rates of non-financial bond jump up discontinuously at the date of policy implementation but white defendants experience larger gains than black defendants. While both groups are more likely to receive non-financial bond after HB463, the gap between the two increases from 3 percentage points (in the pre-period) to around 8.9 percentage points (in the post-period).

Given that black and white defendants differ in their underlying risk level distributions (Figure 7 showed black defendants are more likely to be at higher risk levels than white defendants), this could be a natural consequence of judges' following the score recommendations (setting non-financial bond for low and moderate defendants) at equal rates for white and black defendants. In other words, it is possible that the judicial deviations look the same across racial groups but different risk level distributions cause the disparity increase visible in Figure 11. To address this possibility, I break the picture out by risk level in Figure 12. There are differential shifts in non-financial bond rates by race within the low and moderate risk levels. Therefore, the aggregate picture is not simply a consequence of black defendants' higher risk levels – rather, there are racial disparities in deviation from the recommendation.

Imperfect compliance with predictive recommendations after HB463 yields lower non-financial bond rates and larger racial disparities than those generated by perfect compliance.[66] Figure 13 compares the realized outcomes (observed data) and those mechanically

[66]Of course, perfect compliance would not be without costs; perfect compliance would likely generate larger pretrial misconduct rates.

Figure 12: Bond Outcomes Before and After HB463 by Race and Risk Level

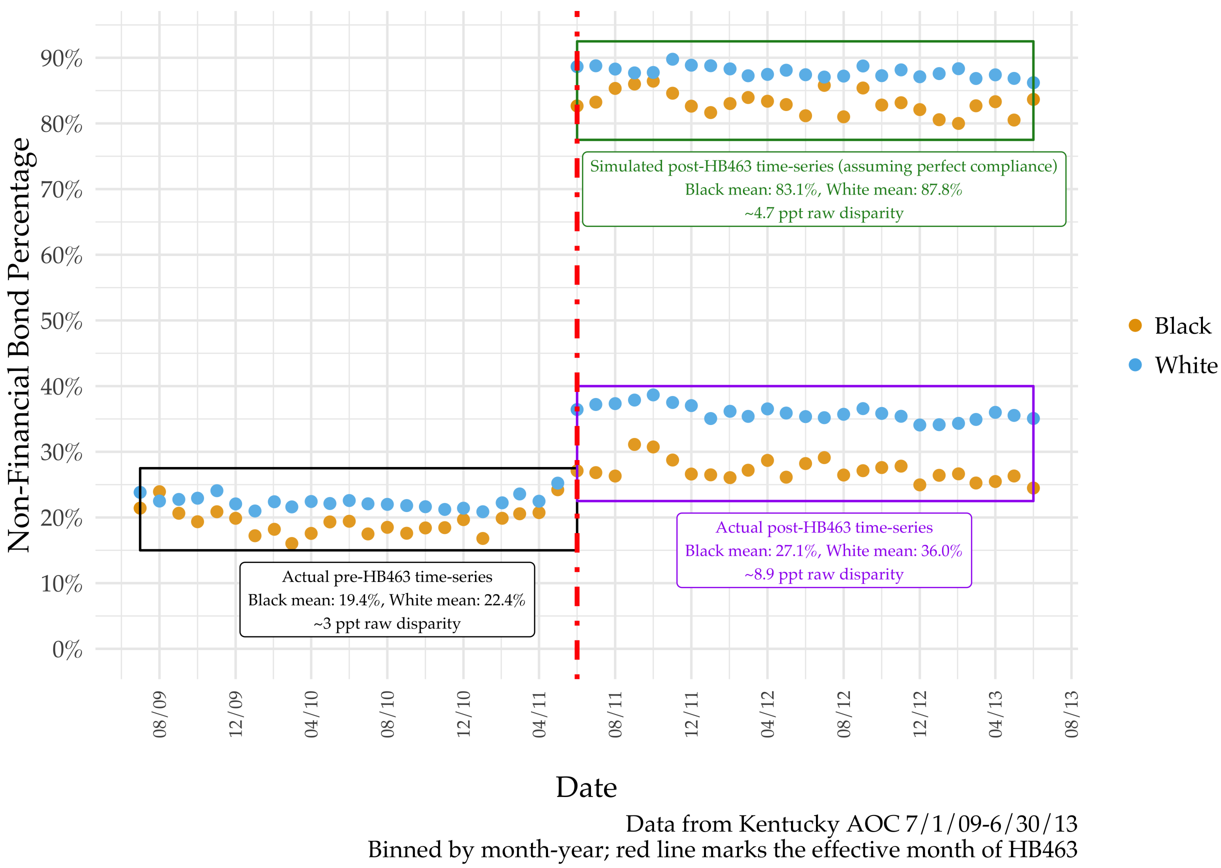


generated by perfect compliance with risk level recommendations (simulated data). Perfect compliance yields a much higher rate of non-financial bond for both groups (80%-90% instead of 25%-40%) as well as smaller racial disparities. Actual post-HB463 racial disparities are almost twice as large (8.9 percentage points) as they would have been had judges followed the recommendations without discretion (4.7 percentage points). Most of the observed increase in racial disparities is, therefore, due to imperfect compliance rather than the underlying risk levels.

In this empirical environment, imperfect compliance with predictive recommendations is responsible for more of the increase in racial disparities than are the recommendations themselves. The following sections investigate the underlying reasons for racial disparities in compliance, which are important to public policy as more jurisdictions encourage or mandate the use of predictive tools in bail decisions.[67]

[67]Pending a 2020 Referendum, California might abolish money bail and adopt a pretrial system largely based on risk assessments (McGough 2019). New Jersey has already significantly cut the prevalence of money bail and is using risk assessment predictions (Schuppe 2017b). (During the first month after 2017 bail reform in New Jersey, money bail was only set in 3 of 3,382 cases (Foderaro 2017).)

Figure 13: Real and Simulated Bond Outcomes Before and After HB463 by Race and Risk Level



5.1 Theoretical Framework and Empirical Methodology

Figure 12 demonstrates that the gaps in racial disparities in non-financial bond rates widen after HB463 for low and moderate risk defendants. While the risk levels themselves cannot be driving these results, the results could be driven by a myriad of other factors that are important to judges in bail decisions.[68] To motivate my empirical approach to identifying why there are these disparities in deviations, I provide a theoretical framework of judge bail decision-making and illustrate the equivalent empirical specifications.

5.1.1 Homogeneity in Policy Response

Assume judge j makes the binary decision to set non-financial bond ($b_{ictj} = 1$) or not ($b_{ictj} = 0$) for defendant i with charges c at time t. Since the probability of release without financial conditions is 95.6%, while probability of release drops meaningfully to 20.4% once any financial conditions are imposed,[69] I assume the judge interprets this binary decision as equivalent to the decision between releasing or detaining the defendant.[70] Following Arnold, Dobbie, and Yang (2018)'s framework, the judge will set non-financial bond for the defendant if and only if the expected cost of release is less than benefit. The cost can be conceptualized as the expected probability of pretrial misconduct, as perceived by judge j – that is, $E_j[p_{ic}]$.[71] I assume the benefit before HB463 is some fixed threshold ζ.[72] After HB463, there is a small cost to deviating from the presumptive default of non-financial bond for low and moderate risk defendants, η. By setting non-financial bond for low and moderate defendants, judges avoid this cost.

Given information about present charges κ_c, defendant characteristics δ_i (e.g., age, criminal history, etc.), defendant risk level $KPRA_i$, and defendant race $race_i$,[73] a judge will set non-financial bond if and only if:

$$E_{jt}[p_{ic}|\kappa_c, KPRA_i, \delta_i, race_i] < \zeta + \eta \times I[t \geq 6/8/11] \times I[KPRA_i \in \{Low, Moderate\}]$$

In this set-up, judges' unique decision thresholds all move to the same extent after HB463. That is, there is no variance in policy responsiveness across judges. This maps empirically onto estimating the following specification (I present results by each of the three $KPRA_i$ risk levels):

[68] In bail decisions in Kentucky, attorneys are not a part of the equation, so attorney quality is not a concern, as it would be for evaluating disparities in sentencing.

[69] Financial conditions often mean detention due to inability to pay.

[70] Other papers often focus on the release outcome rather than the bond decision. Doleac and Stevenson (2018) looks into release rates over the entire pretrial period while Dobbie, Goldin, and Yang (2018) focus on whether defendants were released within 3 days.

[71] Pretrial misconduct in this discussion contains both probability of new crime and failure to appear.

[72] In other words, I do not assume benefit to vary by case. However, it would be natural to extend this assumption, as Arnold, Dobbie, and Yang (2018) do.

[73] Race could be observed indirectly through names in calls or directly through forms or in-person meetings.

$$b_{ijct} = \alpha + \phi_1 HB463_t + \phi_2 Black_i + \phi_3(Black_i \times HB463_t) + \beta_1\kappa_c + \beta_2\delta_i + \omega_j + x_t + \epsilon_{ijct} \quad (1)$$

In this framework, b_{ijct} is a dummy variable that takes the value 1 if judge j set non-financial bond for defendant i with charges c during time t. $HB463_t$ is an indicator for if the decision takes place before or after the effective date of HB463. $Black_i$ is an indicator for if the defendant is black, and δ_i is a vector of defendant characteristics (age, criminal history variables, including dummies for prior FTAs, prior convictions, and pending cases). The vector of charge variables κ_c includes: (i) dummies for all combinations of charge levels (misdemeanor, felony, violation, other) and charge letter classes and (ii) dummies for if the charge description is related to drugs, weapons, or violence.[74][75] Given that judges are known to be heterogeneous in their decision-rules,[76] it is crucial to consider judge fixed-effects ω_j.[77] I also include month-year fixed effects x_t.

After attempting to approximate for the judge's information set, the coefficient of interest is ϕ_3 since this speaks to the change in the racial gap in non-financial bond that occurs after HB463.[78]

5.1.2 Heterogeneity in Policy Response

Judges were not regulated in their response to HB463 in Kentucky. As such, it would be more realistic to assume they varied in their costs of deviation η_j. Allowing for variation in policy responsiveness, the decision rule is subtly changed to the following:

[74] The weapon dummy is 1 when descriptions include the word "gun", "firearm", or "weapon". The violence dummy is 1 when descriptions include the word "violence", "assault", "rape", or "murder". The drug dummy is 1 when descriptions include the word "cocaine", "heroin", "marijuana", "drug", or "meth", but excluding charges that include "under/infl" since those are agnostic to alcohol/drugs.

[75] The reason for including these charge description dummies is that gun, violence, or drug-related offenses could be treated differently even if they share an offense level and class with a property crime offense. Without these variables, differences in charge specifics within charge severity bins by race for low risk defendants could drive observed disparities even after controlling for charge level and class.

[76] Thus the ability of researchers to exploit such variation with "judge designs" for causal inference.

[77] Moreover, recall the spatial variation in black defendants observed in Figure 6. Without fixed effects, I might be concerned that if judges in more populous counties, such as Christian and Fayette, are both harsher to everyone and working in counties where most black defendants are booked, then estimates of racial disparities will be biased upwards. For that reason, it is important to adjust for judge fixed-effects so that I compare bail decisions within given judges since I do not want differences that are stable within judges to drive results.

[78] In focusing on the interaction between race and time ($HB463_t$), the analysis relies on the assumption that any differences in important variables to the initial bond decision before and after HB463 are not statistically different by defendant race. (I.e., case characteristics before and after HB463 are not unbalanced by race.) My approach does not require that there are no differences in important variables by race across all time; this is similar to the assumptions required by Cohen and Yang (2018). While it seems unlikely that case characteristics discontinuously changed by race at the point of HB463, I should still prove this is the case empirically in future drafts.

$$E_{jt}[p_{ic}|\kappa_c, KPRA_i, \delta_i, race_i] < \zeta + \eta_j \times I[t \geq 6/8/11] \times I[KPRA_i \in \{Low, Moderate\}]$$

This then maps onto the following empirical specification:

$$b_{ijct} = \alpha + \phi_1 HB463_t + \phi_2 Black_i + \phi_3(Black_i \times HB463_t) + \beta_1\kappa_c + \beta_2\delta_i + \omega_{jt} + \epsilon_{ijct} \quad (2)$$

The single difference between equations 1 and 2 is that the latter includes time-varying (defined as month-year) judge fixed-effects, ω_{jt}. The comparison between the estimates of ϕ_3 in the two equations highlights the power of heterogeneous responses across judges to HB463 in driving changes in racial disparities – recall that Stevenson (2017) found that these were important in explaining pretrial detention disparities. Judge-specific responses are important to consider given the notable spatial variation in percentage of black defendants across the state. In other words, if judge responses are correlated with judge populations (fraction of black defendants), this could drive ϕ_3 in Equation 1 to be notably higher than ϕ_3 in Equation 2.

5.2 Empirical Results

I now estimate the specifications discussed above in order to see how much of ϕ_3 observed in raw Figure 12 is explained by differences in (i) judge information sets, (ii) variation in policy response across judges. Table 2 presents estimates for low risk defendants in columns 1 and 2, moderate risk defendants in columns 3 and 4, and high risk defendants in columns 5 and 6. The odd columns present results without any covariates, thus showing the regression equivalent of the visual trends in Figure 12. The even columns present results from specification 1.

Table 2 shows that after adjusting for judge information sets,[79] there are not observable racial disparities favoring white defendants in the pre-HB463 period. (The negative coefficient in the low specification in column (1) disappears once conditioning on the information set.) If anything, as seen in column 4, moderate risk black defendants are more likely (2.2 percentage points more) to receive non-financial bond than similar white defendants. However, in the post-period white defendants appear to be significantly advantaged after adjusting for the judge's information set. Low and moderate risk black defendants experienced 30% and 62% less of the short-term gains in non-financial bond setting than similar white defendants, respectively.[80]

[79]Specifically, in line with the prior subsection, I approximate judge information sets with the following covariates: defendant age, number of charges, top charge severity (level and class), characteristics (whether it is related to weapons, drugs, or violence), risk level components (see Figure 3 for full list), and separate fixed effects for judge and month-year.

[80]For low risk defendants, in column 2, the *Black* $\times$ *Post* coefficient is 4.8 percentage points and the *Post* coefficient is 15 percentage points. ($4.8/15 \approx .30$.) For moderate risk defendants, in column 4, the *Black* $\times$ *Post* coefficient is 4.9 percentage points and the *Post* coefficient is 7.9 percentage points. ($4.9/7.9 \approx .62$.)

Table 2: Disparities in Non-Financial Bond Deviations before/after HB463

	Dependent Variable = Non-Financial Bond					
	Low		Moderate		High	
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.022^{***}	0.004	0.007^{**}	0.022^{***}	0.002	0.005
	(0.004)	(0.006)	(0.003)	(0.005)	(0.004)	(0.005)
Post	0.201^{***}	0.150^{***}	0.138^{***}	0.079^{***}	0.021^{***}	0.039^{*}
	(0.003)	(0.020)	(0.002)	(0.020)	(0.003)	(0.020)
Black x Post	-0.065^{***}	-0.048^{***}	-0.051^{***}	-0.049^{***}	-0.004	-0.007
	(0.006)	(0.011)	(0.005)	(0.008)	(0.005)	(0.007)
Judge Info Sets?	No	Yes	No	Yes	No	Yes
Pre-White Mean	0.326	0.326	0.128	0.128	0.050	0.050
Cluster SE?	NA	Judge	NA	Judge	NA	Judge
N	178,238	178,238	163,479	163,479	41,363	41,363
R^2	0.039	0.220	0.027	0.144	0.002	0.068
Adjusted R^2	0.039	0.217	0.027	0.141	0.002	0.058

The dependent variable is a binary indicator for non-financial bond. I approximate judge information sets with the following covariates: defendant age, number of charges, top charge severity (level and class), characteristics (whether it is related to weapons, drugs, or violence), risk level components (see Figure 3 for full list). Even number specifications include time fixed-effects and judge fixed-effects with clustered errors by judge. I present OLS estimates. *** p<0.01; ** p<0.05; * p<0.1.

Table 3: Disparities in Non-Financial Bond Deviations before/after HB463

	Dependent Variable = Non-Financial Bond								
	Low			Moderate			High		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Black	-0.022^{***}	0.004	-0.011^{**}	0.007^{**}	0.022^{***}	0.006^{*}	0.002	0.005	0.006
	(0.004)	(0.006)	(0.005)	(0.003)	(0.005)	(0.003)	(0.004)	(0.005)	(0.005)
Post	0.201^{***}	0.150^{***}	0.145^{***}	0.138^{***}	0.079^{***}	0.081^{***}	0.021^{***}	0.039^{*}	0.042
	(0.003)	(0.020)	(0.021)	(0.002)	(0.020)	(0.025)	(0.003)	(0.020)	(0.027)
Black x Post	-0.065^{***}	-0.048^{***}	-0.012	-0.051^{***}	-0.049^{***}	-0.020^{***}	-0.004	-0.007	-0.007
	(0.006)	(0.011)	(0.008)	(0.005)	(0.008)	(0.006)	(0.005)	(0.007)	(0.007)
Judge Info Sets?	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Judge-Time FEs?	No	No	Yes	No	No	Yes	No	No	Yes
Pre-White Mean	0.326	0.326	0.326	0.128	0.128	0.128	0.050	0.050	0.050
Cluster SE?	NA	Judge	Judge	NA	Judge	Judge	NA	Judge	Judge
N	178,238	178,238	178,238	163,479	163,479	163,479	41,363	41,363	41,363
R^2	0.039	0.220	0.294	0.027	0.144	0.231	0.002	0.068	0.299
Adjusted R^2	0.039	0.217	0.242	0.027	0.141	0.166	0.002	0.058	0.091

The dependent variable is a binary indicator for non-financial bond. I approximate judge information sets with the following covariates: defendant age, number of charges, top charge severity (level and class), characteristics (whether it is related to weapons, drugs, or violence), risk level components (see Figure 3 for full list). Specifications (2), (5), (8) include time fixed-effects and judge fixed-effects with clustered errors by judge. Specifications (3), (6), (9) use time-varying (month-year) judge fixed-effects (as opposed to separate judge and time fixed-effects). I present OLS estimates. *** p<0.01; ** p<0.05; * p<0.1.

Different responses by judge are important to consider given that judges work within specific counties and there is notable spatial variation in percentage of black defendants across the state (see Figure 6). With Table 3, I test for whether the results on ϕ_3 from Table 2 are robust to allowing for time-varying judge fixed-effects; I also present the estimates for the main coefficient of interest visually using Figure 14.[81] In fact, the disparities for low risk defendants become indistinguishable from zero once judges are allowed to vary in their responsiveness.[82] The disparities in deviations for low risk defendants were driven by heterogeneous behavioral responses to HB463 across judges. However, moderate risk black defendants remain less likely than similar white defendants to receive non-financial bond even after allowing for time-varying judge fixed-effects – they experience a boost from the policy that is 25% less than the boost for similar white defendants.[83] In short, while the low risk defendant disparities are a consequence of variation in judge response, there are lingering unexplained results for moderate risk defendants.

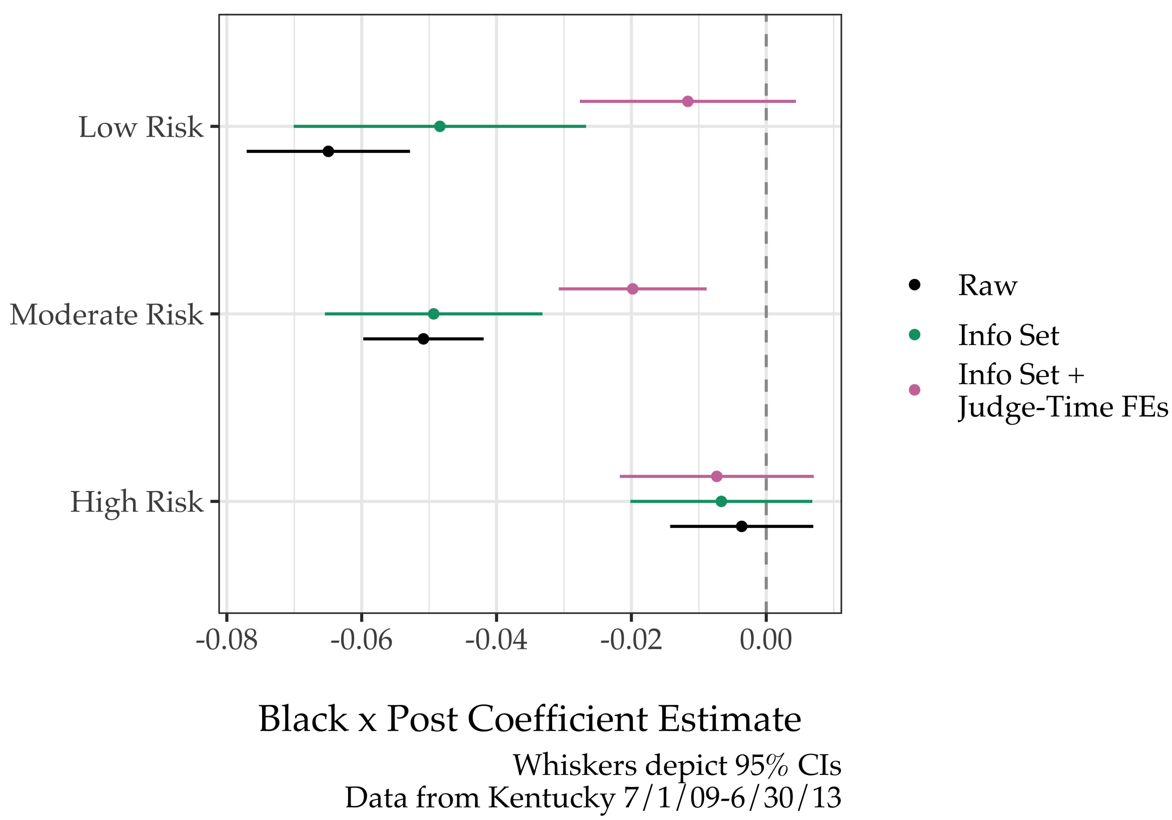
The results in Table 3 and Figure 14 are at odds with the often assumed story that score usage should necessarily equalize outcomes across racial groups with scores. Policy changes that are subject to judicial discretion may not be equally adopted across geographies. If

[81] Again, I approximate judge information sets with the following covariates: defendant age, number of charges, top charge severity (level and class), characteristics (whether it is related to weapons, drugs, or violence), risk level components (see Figure 3 for full list). Now, I use time-varying (month-year) judge fixed-effects (as opposed to separate judge and time fixed-effects).

[82] See the change from column 2 to column 3.

[83] In column 6, for moderate risk defendants, the *Black* $\times$ *Post* coefficient is 2 percentage points and the *Post* coefficient is 8.1 percentage points. ($2/8.1 \approx .25$.)

Figure 14: Coefficient Estimates across Specifications and Risk Levels



responsiveness is correlated with demographic features of the population, risk score policies which intend to render more similar decisions across races but within risk scores may lead to counterintuitive patterns. Moreover, even within judge-time, there is suggestive evidence that moderate risk levels may interact with race to produce different judicial decisions.

6 Mechanisms Discussion

6.1 Correlates of Judicial Responsiveness to HB463

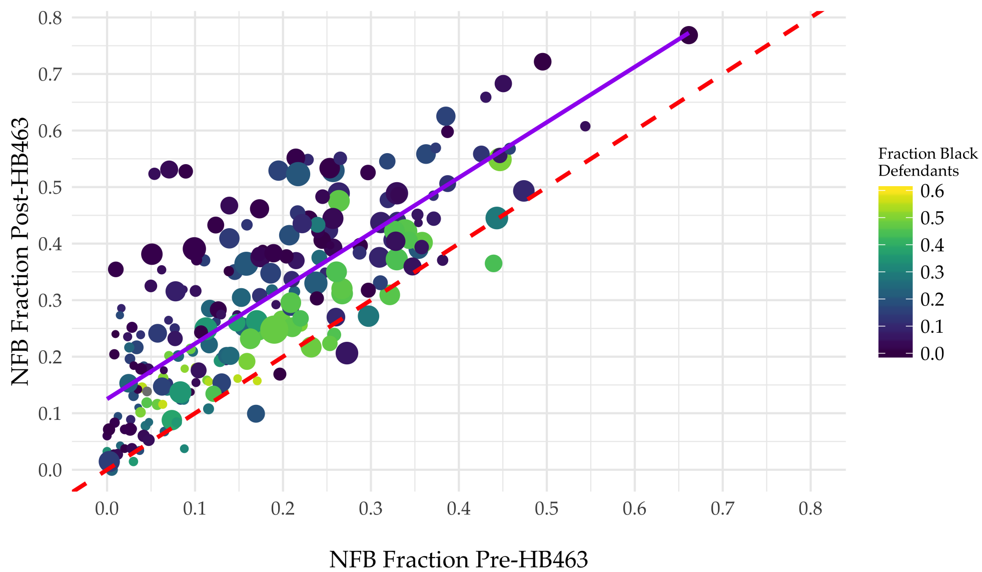
In this policy context, judge responsiveness is correlated with judge population (specifically, the fraction of black defendants). In order to provide exploratory evidence on what could be driving this reduced-form correlation, I investigate judge-level data on responses by focusing on the 233 judges who made at least 100 decisions before and after HB463. I first show visual evidence on the relationship between a judge's response and the population observed by that judge – specifically, I consider the fraction of initial bond decisions made that were about black defendants. I then suggest future work on this topic.

6.1.1 Further Reduced-Form Evidence

Figure 15 plots each judge as a point, whose size (area) corresponds to the number of decisions made over the whole time period, that demonstrates the judge's pre-HB463 and post-HB463 rates of non-financial bond decisions. The visual illustrates that the fraction of non-financial bond decisions increased after HB463, given that the overwhelming majority of points are above the 45 degree line, meaning most (though not all) judges were more likely to give non-financial bond after HB463. Interestingly, the regression line demonstrating the relationship between the pre- and post- rates is parallel to the 45 degree line, meaning that while the change in likelihood of non-financial bond is not meaningfully different dependent on the rate in the pre-period.

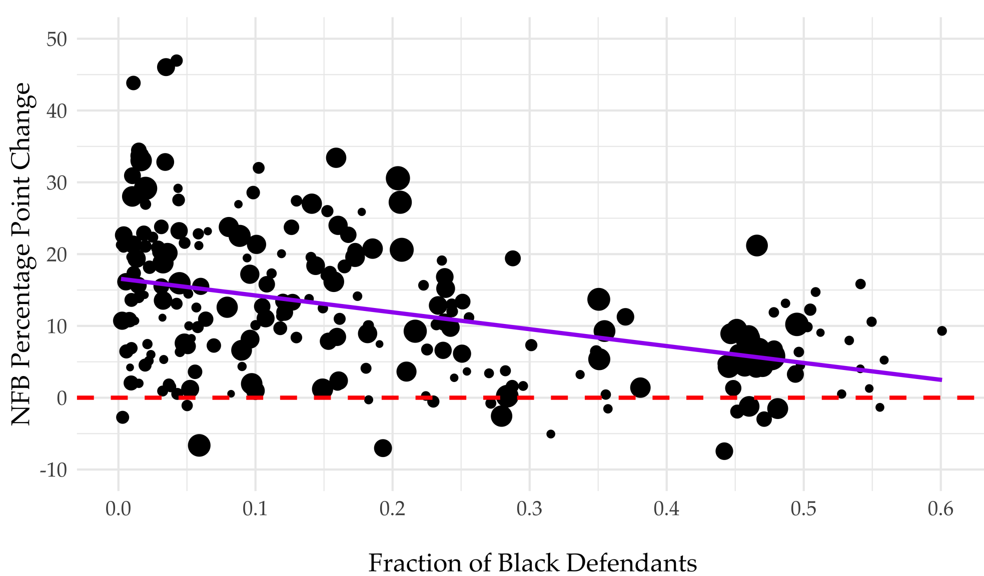
The color of the dots corresponds to the fraction of a judge's bail decisions that are for black defendants. The fact that the lighter yellow points are closer to the red line than the darker blue points are suggestive of the relationship between defendant population and responsiveness. Figure 16 then makes this relationship explicit by plotting each judge's response (fraction of non-financial bond decisions in the pre-period less fraction non-financial bond decisions in the post-period) by the judge's observed defendant population (fraction of decisions made about black defendants). The purple line displays a clear negative relationship between the two.

Figure 15: Judge Non-Financial Bond Rates Before and After HB463



Each point is one of 233 judges in Kentucky.
These are the judges who had at least 100 decisions in both time periods.
Area of points is scaled to the total number of decisions.
The red dashed line is a 45 degree line. The purple line is the regression line of post on pre fraction.
Data from Kentucky AOC 7/1/09-6/30/13

Figure 16: Judge Responses to HB463 and Judge Population



Each point is one of 233 judges in Kentucky.
These are the judges who had at least 100 decisions in both time periods.
Area of points is scaled to the total number of decisions.
The red dashed line marks a change of 0 ppts. The purple line is the regression line of change on fraction of black defendants.
Data from Kentucky AOC 7/1/09-6/30/13

6.1.2 Explaining *Why* Judge Responsiveness Correlates with Defendant Population

While there is a large body of empirical work examining differential treatment within place-time, there is limited work on explaining why policy response might be correlated with population demographics. Given the growing set of bail reform policies, it is important to understand why and how uneven take-up of policies occurs.

There are two main hypotheses for why judges may respond differently to policy reforms. For one, judges with more experience might be less likely to respond to policy changes. If judges who work in counties with higher fraction of black counties are more experienced (perhaps because these are larger counties and thus more competitive elections for judgeships) this could generate the observed relationship.[84] Second, we suspect judges who have made decisions that are associated with higher pretrial misconduct than others would respond less since they face a higher expected cost of release in changing their threshold. If judges who experience higher failure to appear or new criminal activity in their bail decisions work in the counties with more black defendants, this could generate the observed effect.

In future work, I plan to investigate these possible explanations. I plan to run a horserace with judicial experience, pretrial misconduct in the pre-period, and judge population demographics to test if the observed relationship between responsiveness and defendant population is explained by one of these two theories. This would be descriptive work that would be useful for putting structure on understanding judge willingness to adapt to policy suggestions.

6.2 Remaining Racial Disparities for Moderate Risk Defendants

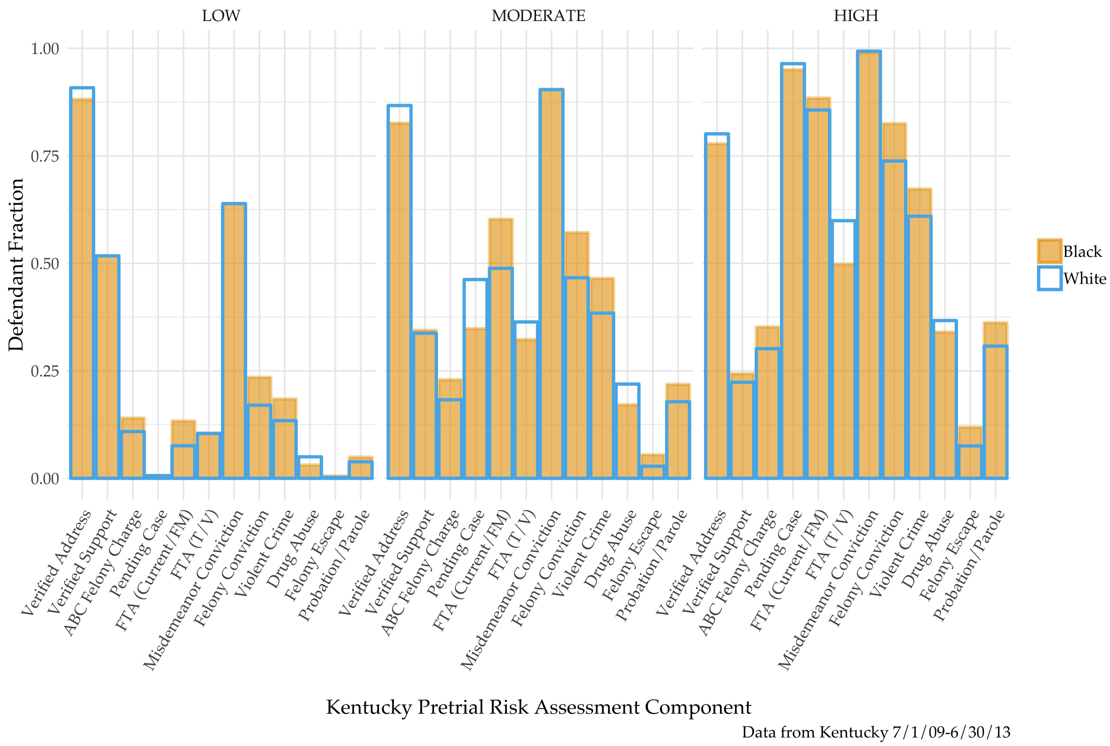
In understanding the mechanism behind the lingering disparities for moderate risk defendants, it is worth considering two possible explanations: differential weighting and disparate interactions. On differential weighting, it could be that there are criminal history characteristics already embedded in the Kentucky Pretrial Risk Assessment, as illustrated in Figure 3, that also make their way into the pretrial-judge conversation. If black and white men have different compositions of factors going into the same underlying risk score level, then this could explain why they are treated differently despite similar environments (judge-time), charge characteristics, and risk level. Figure 17 demonstrates that it is the case that component combinations look different across races within risk levels.[85]

Imagine that a judge considers a moderate risk black defendant and a moderate risk white defendant with similar charges. Despite their identical risk levels, the black defendant has an active warrant for a failure to appear (FTA) prior to disposition (factor 6 on the list), while a low risk white defendant does not. If a prior FTA is more likely to be relayed to the judge independently of the risk level (as occurs in one of the example judge calls on

[84]The experience explanation would tie into a model with different costs across judges η_j where costs are larger for more experienced judges.

[85]The ordering of components in Figure 17 matches the ordering in Figure 3.

Figure 17: Make-up of Components within Levels Across Race



the Kentucky online virtual tour), then that judge might deem that black defendant riskier than the white defendant even though the risk score already has taken this factor into account. In other words, score level and contributory factors (to the risk level) are strongly correlated but might not be identified as such, which could lead to double-counting in the vein of Enke and Zimmermann (2017). If this is the case, then allowing interaction terms between the time period and the risk components should reduce or eliminate the original result. Table 4 shows (see column 4) that despite inclusion of said interaction terms, the moderate risk result remains though it is slightly smaller in magnitude.

Differential weighting of components over time does not eliminate the moderate risk result, nor did different judge responses or judge information sets. As such, the moderate risk result must be a product of judges interpreting scores differently by race. Imagine judges have some sense of the underlying continuous risk score distribution and assume that moderate risk black men are still higher risk than moderate risk white men even though they are put in the same larger buckets. This could also explain the results and would accord well with recent findings by Green and Chen (2019) and Skeem, Scurich, and Monahan (2019).[86] Understanding this possible mechanism is broadly relevant to hiring, loan decisions, and other important high-stakes decisions.[87]

[86]However, more empirical work is necessary to pin down this result.
[87]For more on hiring, see Hoffman, Kahn, and Li (2017).

Table 4: Does Different Weighting of Components Explain the Moderate Result?

	Dependent Variable = Non-Financial Bond					
	Low		Moderate		High	
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.011^{**}	-0.012^{**}	0.006^{*}	0.005	0.006	0.006
	(0.005)	(0.005)	(0.003)	(0.003)	(0.005)	(0.005)
Post	0.145^{***}	0.106^{***}	0.081^{***}	0.106^{***}	0.042	0.035
	(0.021)	(0.022)	(0.025)	(0.026)	(0.027)	(0.065)
Black x Post	-0.012	-0.011	-0.020^{***}	-0.017^{***}	-0.007	-0.007
	(0.008)	(0.008)	(0.006)	(0.006)	(0.007)	(0.007)
Components x Post	No	Yes	No	Yes	No	Yes
Judge Info Sets?	Yes	Yes	Yes	Yes	Yes	Yes
Judge-Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Pre-White Mean	0.326	0.326	0.128	0.128	0.050	0.050
Cluster SE?	Judge	Judge	Judge	Judge	Judge	Judge
N	178,238	178,238	163,479	163,479	41,363	41,363
R^2	0.294	0.295	0.231	0.232	0.299	0.299
Adjusted R^2	0.242	0.242	0.166	0.167	0.091	0.091

The dependent variable is a binary indicator for non-financial bond. I approximate judge information sets with the following covariates: defendant age, number of charges, top charge severity (level and class), characteristics (whether it is related to weapons, drugs, or violence), risk level components (see Figure 3 for full list). Even specifications include interaction terms between all risk components and the post indicator. All specifications use time-varying (month-year) judge fixed-effects with clustered errors by judge. I present OLS estimates. *** p<0.01; ** p<0.05; * p<0.1.

7 Conclusion

As predictive tools continue to be integrated into high-stakes decisions, there is a growing need to understand how they are used by the human decision-makers (e.g., judges, loan officers, and hiring managers). While predictive tools often present recommendations, there is little oversight as to how decision-makers may overrule or follow them. I use this paper to show that, counter to intuition, the introduction of risk score recommendations can increase racial disparities for individuals with the same risk level.

This result is a consequence of two types of deviations by judges: across-judge and within-judge deviations. On the former, judges varied in their policy responsiveness; judges in whiter counties responded more to the new default (increasing their leniency) than judges in blacker counties. There is a striking correlation between a judge's response to the policy and a judge's defendant population. Given the growing set of bail reform policies, it is important to understand why this uneven take-up of policies occurs. Otherwise, similar patterns of judicial take-up could exacerbate racial disparities in other policy contexts.

Second, even within judge and time, I show judges are more likely to deviate from the recommended default for moderate risk black defendants than for similar moderate risk white defendants. This result suggests that interaction with the same predictive score may lead to different predictions by race, which warrants further investigation. Part of the public appeal of risk assessments is the movement towards a system that is more "objective" than the status quo (Harris and Paul 2017). However, if interpretation of the scores itself interacts with defendant race, the very judicial discretion that risk score proponents sought to reduce has simply been shifted to a later stage.

References

Abrams, David S, Marianne Bertrand, and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" *The Journal of Legal Studies* 41 (2). University of Chicago Press Chicago, IL: 347–83.

AI, Partnership on. 2019. "Report on Algorithmic Risk Assessment Tools in the Us Criminal Justice System."

Alexander, Michelle. 2018. "The Newest Jim Crow." Edited by The New York Times. https://www.nytimes.com/2018/11/08/opinion/sunday/criminal-justice-reforms-race-technology.html.

Angwin, Larson, Julia, and Lauren Kirchner. 2016. "Machine Bias: There's Software Used Across the Country to Predict Future Criminals and It's Biased Against Blacks." *ProPublica*.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics* 133 (4). Oxford University Press: 1885–1932.

Austin, James, Roger Ocker, and Avi Bhati. 2010. "Kentucky Pretrial Risk Assessment Instrument Validation." *Bureau of Justice Statistics. Grant*, nos. 2009-DB.

Biernat, Monica, and Melvin Manis. 1994. "Shifting Standards and Stereotype-Based Judgments." *Journal of Personality and Social Psychology* 66 (1). American Psychological Association: 5.

Bulman, George. 2019. "Law Enforcement Leaders and the Racial Composition of Arrests." *Economic Inquiry*. Wiley Online Library.

Chandra, Amitabh, and Jonathan Skinner. 2003. "Geography and Racial Health Disparities." National bureau of economic research.

Cohen, Alma, and Crystal Yang. 2018. "Judicial Politics and Sentencing Decisions." National Bureau of Economic Research.

Cohen, Alma, and Crystal S Yang. 2019. "Judicial Politics and Sentencing Decisions." *American Economic Journal: Economic Policy* 11 (1): 160–91.

Corbett-Davies, Sam, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. "Algorithmic Decision Making and the Cost of Fairness." In *Proceedings of the 23rd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 797–806. ACM.

Cowgill, Bo. 2018a. "Bias and Productivity in Humans and Algorithms: Theory and Evidence from Resume Screening." *Columbia Business School, Columbia University* 29.

———. 2018b. "The Impact of Algorithms on Judicial Discretion: Evidence from Regression Discontinuities."

Cowgill, Bo, and Catherine E Tucker. 2019. "Economics, Fairness and Algorithmic Bias." *Preparation for: Journal of Economic Perspectives*.

DeMichele, Matthew, Peter Baumgartner, Kelle Barrick, Megan Comfort, Samuel Scaggs, and Shilpi Misra. 2018. "What Do Criminal Justice Professionals Think About Risk Assessment at Pretrial?"

Devers, Lindsey. 2011. "Plea and Charge Bargaining." *Research Summary* 1.

Dobbie, Will, Jacob Goldin, and Crystal S Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108 (2): 201–40.

Dobbie, Will, Andres Liberman, Daniel Paravisini, and Vikram Pathania. 2018. "Measuring Bias in Consumer Lending." National Bureau of Economic Research.

Doleac, Jennifer, and Megan Stevenson. 2018. "The Roadblock to Reform." *American Constitution Society Research Report*.

Dowle, Matt, and Arun Srinivasan. 2018. *Data.table: Extension of 'Data.frame'*. https://CRAN.R-project.org/package=data.table.

Einav, Liran, Mark Jenkins, and Jonathan Levin. 2013. "The Impact of Credit Scoring on Consumer Lending." *The RAND Journal of Economics* 44 (2). Wiley Online Library: 249–74.

Enke, Benjamin, and Florian Zimmermann. 2017. "Correlation Neglect in Belief Formation." *The Review of Economic Studies* 86 (1). Oxford University Press: 313–32.

Firke, Sam. 2018. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Foderaro, Lisa W. 2017. "New Jersey Alters Its Bail System and Upends Legal Landscape." *The New York Times*.

Garrett, Brandon L, and John Monahan. 2018. "Judging Risk."

Goel, Sharad, Justin M Rao, Ravi Shroff, and others. 2016. "Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy." *The Annals of Applied Statistics* 10 (1). Institute of Mathematical Statistics: 365–94.

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of" Blind" Auditions on Female Musicians." *American Economic Review* 90 (4): 715–41.

Goncalves, Felipe, and Steven Mello. 2017. *A Few Bad Apples?: Racial Bias in Policing*. Industrial Relations Section, Princeton University.

Green, Ben, and Y Chen. 2019. "Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments." In *Proceedings of Conference on Fairness, Accountability, and Transparency*.

Harris, Kamala, and Rand Paul. 2017. "Pretrial Integrity and Safety Act of 2017." In *115th Congress*.

Hlavac, Marek. 2015. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. https://CRAN.R-project.org/package=stargazer.

Hoffman, Mitchell, Lisa B Kahn, and Danielle Li. 2017. "Discretion in Hiring." *The Quarterly Journal of Economics* 133 (2). Oxford University Press: 765–800.

John Arnold Foundation, Laura &. 2018. "Pretrial Justice." 2018. https://www.arnoldfoundation.org/initiative/criminal-justice/pretrial-justice/.

Jung, Jongbin, Connor Concannon, Ravi Shroff, Sharad Goel, and Daniel G Goldstein. 2017. "Simple Rules for Complex Decisions." *SSRN*.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human Decisions and Machine Predictions." *The Quarterly Journal of Economics* 133 (1). Oxford University Press: 237–93.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R Sunstein. 2018. "Discrimination in the Age of Algorithms." *Journal of Legal Analysis* 10. Narnia.

Kleinberg, Jon, and Sendhil Mullainathan. 2019. "Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability." National Bureau of Economic Research.

Laura & John Arnold Foundation. 2013. "Developing a National Model for Pretrial Risk."

Main, Frank. 2016. "Cook County Judges Not Following Bail Recommendations: Study." *Chicago Sun-Times*.

Mamalian, CA. 2011. "State of the Science of Pretrial Risk Assessment. Pretrial Justice Institute."

McGough, Michael. 2019. "The Fate of California's Cash Bail Industry Will Now Be Decided on the 2020 Ballot." *The Sacramento Bee*.

Miller, Claire Cain. 2015. "Can an Algorithm Hire Better Than a Human." *The New York Times* 25.

New, Joshua. 2015. "It's Humans, Not Algorithms, That Have a Bias Problem." *Center for Data Innovation*.

Pager, Devah, Bart Bonikowski, and Bruce Western. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74 (5). Sage Publications Sage CA: Los Angeles, CA: 777–99.

Santo, Alysia. 2015. "Kentucky's Protracted Struggle to Get Rid of Bail." *The Marshall Project*.

Schuppe, Jon. 2017a. "Post Bail." *NBC News*.

———. 2017b. "Post Bail." *NBC News*.

Skeem, Jennifer L, Nicholas Scurich, and John Monahan. 2019. "Impact of Risk Assessment on Judges' Fairness in Sentencing Relatively Poor Defendants." *Virginia Public Law and Legal Theory Research Paper*, nos. 2019-02.

Sloan, CarlyWill, George Naufal, and Heather Caspers. 2018. "The Effect of Risk Assessment Scores on Judicial Behavior and Defendant Outcomes." IZA Discussion Paper.

Stevenson, Megan. 2017. "Assessing Risk Assessment in Action." *Minnesota Law Review* 103.

Stevenson, Megan, and Sandra G Mayson. 2018. "Pretrial Detention and Bail." *Reforming Criminal Justice* 3: 21–47.

Traughber, Rachel. 2018. "Finding a Link to the Human in Algorithms Setting Justice." Edited by Harvard Gazette. https://news.harvard.edu/gazette/story/2018/05/grad-discovers-algorithms-in-justice-system-dont-always-compute/.

Wickham, Hadley. 2017. *Tidyverse: Easily Install and Load 'Tidyverse' Packages*. https://CRAN.R-project.org/package=tidyverse.

———. 2018. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, and Kara Woo. 2018. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2018. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Appendix

A Judge Decision Dimensions and Initial Release

A.1 Mapping Judge Decisions onto Release

There is a distinction between judge actions in terms of bond setting and outcomes such as pretrial detention. The set of judge actions for the purpose of this paper is simplified down to the binary decision between non-financial and financial bond since that was the main implication of the policy of interest, HB463. However, judge actions are much more complex than this simple binary decision. Bond is broader than simply setting a money amount – bond can include non-monetary conditions (e.g., no driving) on defendants. Furthermore, within financial and non-financial bond, there are a variety of amounts that can be picked for the money amount. (The non-financial bond types, surety and unsecured bond, only come into play if the defendant does not show up for court.)

The multi-dimensional judge decision then plays into whether a defendant is released on that initial bond. Release is a consequence of the initial bond but it is not explicitly set by Kentucky judges. In the aggregate, for initial bond, when a judge setting non-financial bond (nfb) or financial bond (fb), the following is true:

$$Pr(release) = Pr(release|fb)Pr(fb) + Pr(release|nfb)Pr(nfb) \tag{3}$$

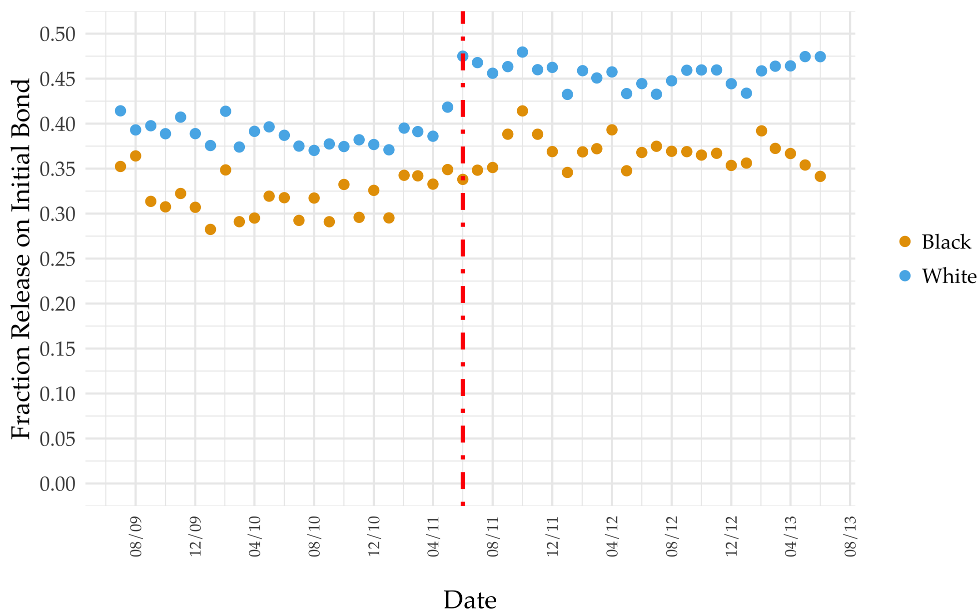
Note that since all decisions are either non-financial or financial bond, $Pr(fb) = 1 - Pr(nfb)$. While it is the case that HB463 significantly increased $Pr(nfb)$ and $Pr(release|fb) < Pr(release|nfb)$ since less financial conditions means higher probability of release, the change in $Pr(release)$ need not match that change in $Pr(nfb)$ if the probabilities of release conditional on the two bond groups are markedly changed as well. As a brief accounting exercise, see below for release probabilities decomposed for the pre- and post-HB463 periods.

Table 5: Bond and Release Probability Before and After HB463

Time Period	Pr(nfb)	Pr(release | nfb)	Pr(fb)	Pr(release | fb)	Pr(release)
Pre-HB463	0.22	0.965	0.78	0.22	0.37
Post-HB463	0.34	0.95	0.66	0.18	0.43

In short, decreasing likelihoods of release for both non-financial bond and financial bond recipients watered down the ultimate initial release gains from HB463.

Figure 18: Initial Release Before and After HB463 by Race



Data from Kentucky AOC 7/1/09-6/30/13
The red line marks the effective date of HB 463

A.1.1 Decomposing Immediate Release Disparities

For more on release and other HB463 outcomes, see the thorough account of Stevenson (2017). Figure 18 shows that HB463 led to a small increase in racial disparities in initial release (increased to 8.9 percentage points from around 7.1 percentage points). This is substantially smaller than the change in non-financial bond disparities.

The reasons that changes in racial disparities in initial release could be so different from those in non-financial bond should be clear from consideration of all possible moving parts in equation 3. In fact, the racial gap for $Pr(release|fb)$ decreased notably after HB463. Figure 19 shows that the gap in probability of release conditional on financial bond decreased meaningfully from 7.5 percentage points to 4.7 percentage points. This naturally leads to curiosity about how financial bond setting changed across the racial groups.

A.2 Financial Bond Setting

All financial bond amounts are not equally likely. There is significant bunching at certain round numbers. In fact, 75% of all financial bond amounts in the data are listed in the Table 6 below.

I use Figure 20 to plot out the density of financial bond amounts before and after HB463 (up to $10,000). Bonds of around $2,500 and below become less common relative to higher bonds, which makes sense as we'd think that the defendants given non-financial bond

Figure 19: Pr(Release | Financial Bond) Before and After HB463 by Race

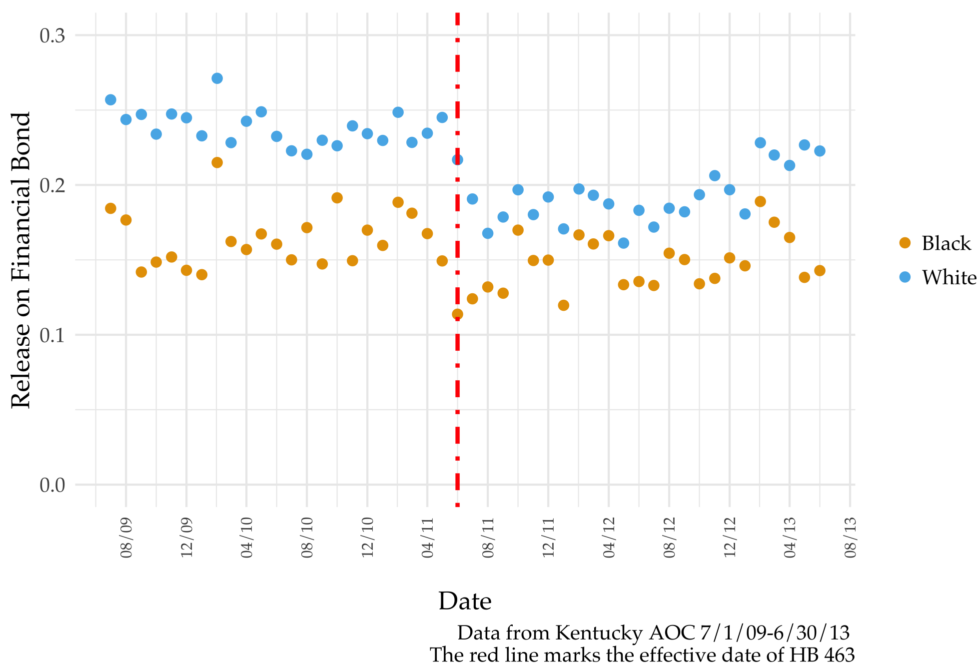
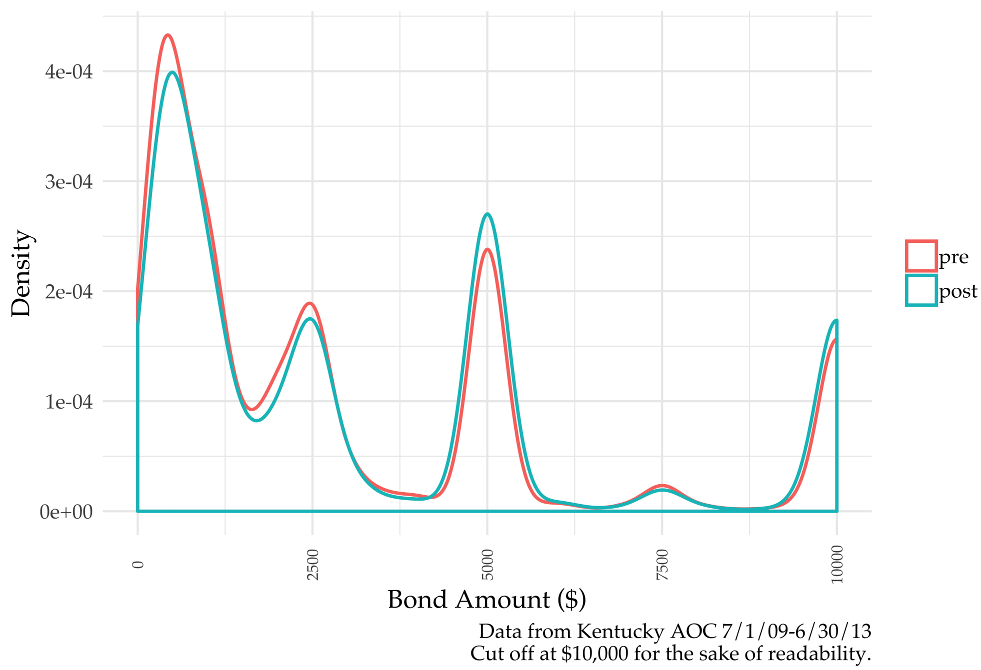


Table 6: Top 75 Percent of Financial Bond Amounts

Bond Amount	Observations	Pr(Release on Bond)
$250	10161	0.416
$500	32373	0.307
$1000	26302	0.265
$1500	6592	0.244
$2000	11477	0.284
$2500	23797	0.210
$5000	38185	0.127
$10000	25077	0.075
$20000	5143	0.042
$25000	9948	0.043
$50000	6224	0.023

Figure 20: Financial Bond Amount Densities Before and After HB463



now were likely to have lower bond levels had they been in the pre-period.

Since white defendants experience a larger dip in likelihood of release on financial bond than black defendants after HB463, we might think that white defendants experienced larger increases in financial bond than black defendants after HB463. In regressing dummies for bond under clear cut-points ($500, $1,000, $2,500, $5,000) on the post indicator and its interaction with race, we see that defendants are 2-6 percentage points less likely to receive low non-financial bond (depending on the cut-point definition) after HB463. If judges who set non-financial bond more in the post-period did so by moving defendants from low financial bond to non-financial, then we would expect black defendants to be more likely to receive low financial bond in the post period than white defendants (since white defendants were receiving non-financial bond more in the post-period). However, there is no significant difference in likelihood for black and white defendants after HB463 to receive most low financial bond definitions. If anything (for under $1,000 bond), blacks become less likely to receive low financial bond relative to whites. This suggests that judges did not simply substitute out low-financial bond for non-financial bond after HB463.

The results are consistent with a story of non-linearity in ability to pay financial bond. Note that black defendants are more likely to have lower bonds than white defendants in the pre-period, however, they are less likely to be released on said bonds. Given a shift to financial bonds that affected both racial groups, the behavioral response to this based on ability to pay necessarily depends on the initial level of the bond. It seems likely that since white defendants were less likely to be receiving smaller financial bonds to begin with, their response to a shift out in financial bonds was larger, thus giving us the picture of differential trends in Figure 19.

Table 7: Likelihood of Low Financial Bond

	Under \$500	Under \$1000	Under \$2500	Under \$5000
	(1)	(2)	(3)	(4)
Post	-0.033^{***}	-0.024^{***}	-0.058^{***}	-0.066^{***}
	(0.001)	(0.002)	(0.002)	(0.002)
Black	0.028^{***}	0.032^{***}	0.0003	-0.017^{***}
	(0.002)	(0.003)	(0.003)	(0.003)
Black x Post	-0.007^{**}	-0.014^{***}	-0.006	-0.002
	(0.003)	(0.004)	(0.005)	(0.005)
N	261,590	261,590	261,590	261,590
R^2	0.004	0.002	0.004	0.005
Adjusted R^2	0.004	0.002	0.003	0.005

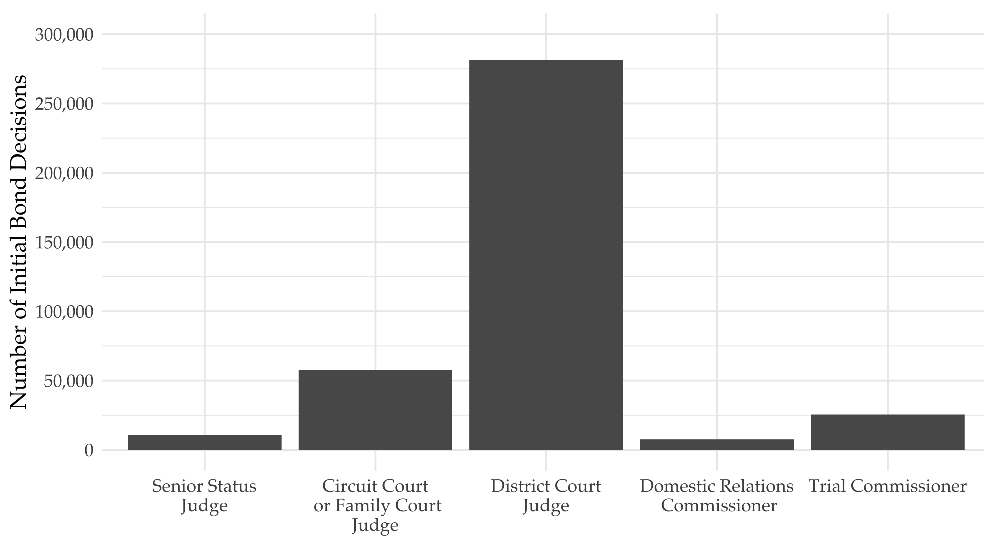
OLS estimates. *** p<0.01; ** p<0.05; * p<0.1.

B Note on Judge Types

In my data, 563 distinct judges make the 383,080 initial bond decisions of interest. There are five different types of judges. The judge type for an initial decision is partially determined by the type of case. Recall that the data covers felonies and misdemeanors, both of which are originally under the jurisdiction of District court judges. However, if a felony defendant is indicted, the case is under the jurisdiction of a Circuit court judge. Moreover, the Family Division of Circuit Court handles cases related to domestic violence, and child abuse and neglect.

What this means for initial bond decisions is that District judges make most of the decisions. Figure 21 shows that 73.5% of decisions are made by District Court Judges while another 15% are made by Circuit or Family Court Judges.[88] The remaining three types of judges fill in for other judges: trial commissioners fill in for District Court judges, Domestic Relations Commissioners fill in for Family Court Judges (if a given county doesn't have a Family Court judge), and Senior Status judges (who are retired) fill in as needed.

Figure 21: Initial Decision Count by Judge Type



Data from Kentucky AOC 7/1/09-6/30/13

[88]Judges in Kentucky are elected, while Commissioners are appointed by judges. District Court judges serve four-year terms, while Circuit Court and Family Court judges serve eight-year terms.