What Can Plea Bargaining Teach Us About Racial Bias in Criminal Justice?

Andrew Jordan

October 19, 2020

Click Here for Latest Version

Abstract

I examine racial bias in the criminal justice system with a model of plea bargaining and data from Cook County, IL. Prosecutors offer plea bargains that screen against defendant risk aversion. They face a tradeoff between securing long sentences and avoiding trials. Stronger cases against defendants ease this tradeoff, allowing both long sentences and few trials. I derive regression-based tests for differences in prosecutor preferences and case strength and apply them to a large sample of felony cases in Chicago. Black defendants demand more trials and receive shorter sentences than nonblack defendants facing the same charges. Viewed through the lens of my model, these results suggest that criminal courts bring weaker cases against black defendants. Simulated Method of Moments estimation confirms this. It also finds that prosecutors value sentences against black defendants more.

I thank Derek Neal, Canice Prendergast, Alessandra Voena, Gadi Barlevy, Azeem Shaikh, and numerous seminar participants for invaluable advice and support. Lawrence Fox, Daniel Coyne, Sarah Staudt, Rob Warden, and Michael Cooney provided insight into the criminal justice system of Cook County. I further thank the Chicago Data Collaborative for data and Lindsey Simon for graphics.

1 Introduction

The US criminal justice system has highly unequal outcomes by race. It incarcerates black men at nearly six times the rate of white men (Bronson 2019). This disparity in outcomes must necessarily be caused by a difference between these two groups and/or their experiences in the system. However, understanding and documenting these differences is difficult because the criminal justice system is complex. Prosecutors, judges, police, and others employ both professional expertise and firsthand knowledge when making decisions about each case. As outside observers, researchers must be especially careful to distinguish disparities caused by case characteristics observed only by the decisionmakers from disparities caused by biased decisions. This effort requires assumptions, stated or unstated, about unobservable case characteristics and about the incentives and constraints that frame each decision. With these structures in place, it becomes possible to detect where and why differences in the treatment of black and nonblack defendants arise.[1]

This paper studies the question of racial disparities in the criminal justice system beginning with one crucial step: plea bargaining. In plea bargaining, defendants accused of a crime waive their constitutional right to a trial and plead guilty to the charges against them. In return, they receive a previously agreed-upon sentence that is lower than the sentence they would have faced had they gone to trial and lost. Both prosecutors and defendants have an incentive to plea bargain. Trials are costly to prosecutors because they consume court time and resources. Risk-averse defendants find trials undesirable because they carry high stakes. In practice, most cases are resolved by plea bargains. Only 11.9% of cases in my analysis data end in a trial.

Plea bargaining is a good entry point to studying the criminal justice system for four reasons. First, plea bargaining brings together many aspects of the criminal justice system. Defendants and prosecutors bargain on the basis of cases assembled by police and in the shadow of potential verdicts rendered by judges and juries. Plea bargaining can therefore be informative about much more than just the decisions of prosecutors and defendants. Second, plea bargaining combines bargaining and insurance, both of which economists have studied extensively. This paper significantly extends Bebchuk (1984), one of the first models to make this observation. Third, plea bargaining determines the outcome of most criminal cases. An empirical strategy that focused only on trial cases would be limited to a small and selected sample. Fourth, plea bargaining produces a rich set of outcomes available in public data. Economists since Becker (1957) have studied discrimination by examining the outcomes of potentially biased decisions. However, most settings studied by economists, especially in criminal justice, have involved a single binary choice and outcome. In plea bargaining, prosecutors make a continuous choice of what sentence to offer the defendant, and this produces

[1] I use the terms 'black' and 'nonblack' when referring to criminal defendants I study. This matches the race information available in my analysis data. See Section 5 for details.

both a binary outcome (whether the offer is accepted) and a continuous outcome (the defendant's eventual sentence). With careful application of economic principles, these outcomes can be informative about both the bias of the prosecutor and the strength of the case brought against the defendant.

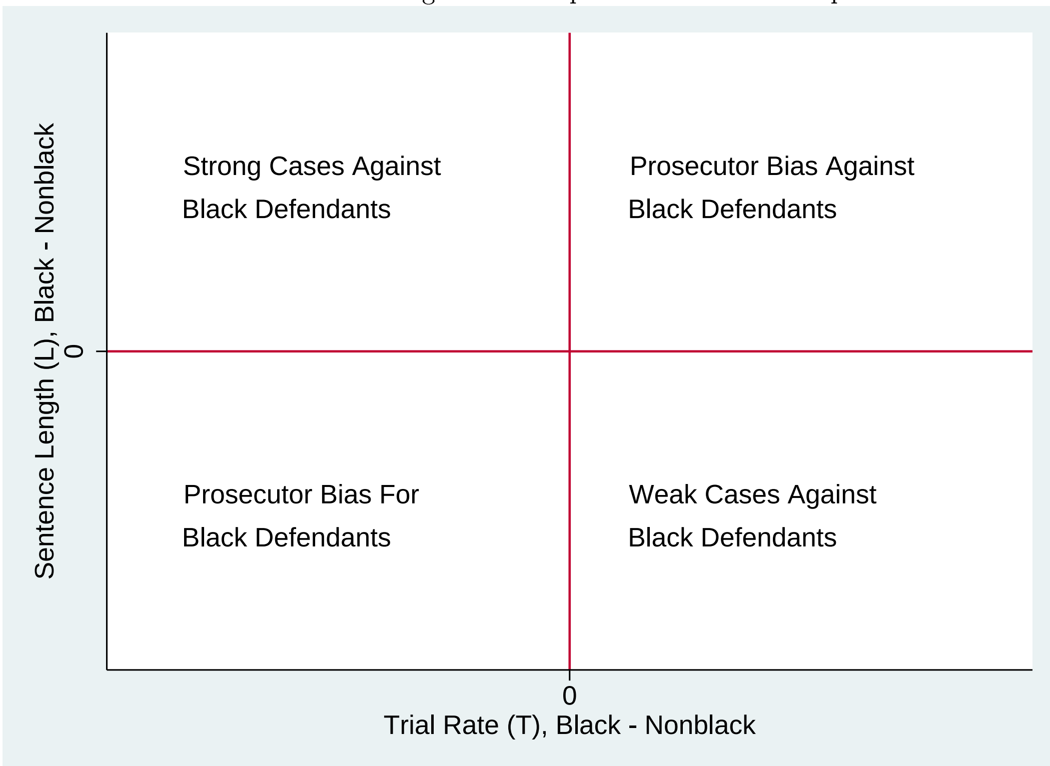
In Section 3, I describe a model of plea bargaining. After a defendant has been charged with a crime, the prosecutor handling their case must choose a plea bargain sentence to offer. My model is interested in two unobservables: how the prosecutor values securing a long sentence compared to the cost of a trial, k, and the probability a defendant would be convicted if their case went to trial, θ, which I also call case strength. I assume that the prosecutor and defendant have common knowledge of case strength.[2] Defendants in my model vary in their degree of risk aversion, ρ, and prosecutors only know the distribution of risk aversion, not the personal risk aversion of each defendant.

I call prosecutors who place more emphasis on securing long sentences against black defendants biased. As in the rest of the discrimination literature, it is important not to confuse the bias defined here with other meanings of the word. One could say there is bias against black defendants if they are systematically less likely to afford skilled lawyers, if police arrest them with less evidence, or if judges are more inclined to rule against them. However, these forms of bias, along with any others outside the narrow scope of prosecutor preferences, will instead enter my model in θ, the probability that the defendant is convicted at trial. I refer to differences in k as "prosecutor bias" and to θ simply as "case strength" because my empirical strategy will ultimately be able to isolate the effects of prosecutor preferences on outcomes, but it cannot disentangle the many factors that influence the probability of conviction at trial.

My model is fundamentally a screening model. Defendants accept or reject plea bargain offers according to a cutoff rule that depends on their personal risk aversion. Prosecutors pick offers that balance the benefit of securing long plea sentences against the cost of pushing more risk-tolerant defendants into trial. I solve this model to find the prosecutor's unique optimal plea bargain offer and the corresponding critical risk aversion for defendants. I use this solution to derive comparative statics between my two unobservables: probability of conviction, θ, and prosecutor preference parameter, k, and two observables: expected sentence length, L, and trial rate, T. I find that as the strength of the case against the defendant increases, two things happen. Sentence lengths rise, and trial rates generally fall. Intuitively, a higher probability of conviction at trial weakens the defendant's outside option while bargaining. This gives prosecutors more leverage. Prosecutors spend this leverage on both of the things they want in the negotiation, a longer sentence and a lower probability of trial. I find that increases in k also cause sentence lengths to rise, but now trial rates rise as well. Bias affects only prosecutor preferences. It does not change the defendant's bargaining position

[2] The assumption of common knowledge of case strength is motivated by the strong discovery provisions in Illinois criminal law. See Appendix B for details.

Figure 1: Interpretation of Racial Gaps



or the set of acceptable plea bargains, so it cannot give prosecutors any additional leverage. Therefore, prosecutors who pursue longer sentences out of bias must make a tradeoff and accept a higher probability of a costly trial.

This pattern of comparative statics suggests a strategy for learning about the unobservables in my model by studying the *joint* patterns of L and T. Figure 1 outlines the intuition behind this strategy. The x-axis is the trial rate, T, for black defendants minus the trial rate for nonblack defendants. When this difference $\Delta T > 0$, black defendants have more trials. The y-axis shows the difference in sentence length, L. When this difference $\Delta L > 0$, black defendants receive longer sentences than nonblack defendants. As discussed above, if cases against black defendants are systematically stronger, the data will show longer sentences and fewer trials for black defendants: $\Delta L > 0$ and $\Delta T < 0$. This corresponds to the upper-left quadrant of Figure 1. Likewise, if prosecutors are biased against black defendants, the data will show longer sentences and more trials for black defendants: $\Delta L > 0$ and $\Delta T < 0$. This corresponds to the upper-right quadrant of Figure 1. Similar arguments establish the lower quadrants. This correspondence is not immediate. It requires the assumption that other factors impacting the outcome of the case can be held relatively constant. It also requires more detail on what it means for cases to be "systematically" stronger or weaker, which I treat as a stochastic ordering of θ distributions conditional on race. Lastly, if there are simultaneously differences in case strength and prosecutor bias, this strategy will only detect whichever has the largest impact on ΔL and ΔT. I discuss these details and the proofs underlying Figure 1 in Section 4, and I consider some alternative interpretations of my findings in Section 7.

One significant advantage of this strategy is that it is relatively simple to apply to data. In a correctly-specified regression with L or T on the left hand side, the coefficient on an indicator for black will estimate ΔL or ΔT as in Figure 1. In Sections 5 and 6, I apply this idea to newly-collected administrative data from the Cook County Circuit Court. These data cover all felony cases in Cook County from 1984 to 2019. The long time period available, coupled with high case volumes in Cook County, allows me to analyze large samples composed exclusively of felony cases.[3] My administrative data are also very detailed. They are informative about the charges brought against the defendant, their criminal history in Cook County, precise sentencing outcomes, and other aspects of the case like judge assignment, public defender usage, and pretrial confinement. See Appendix C for details on the data and how I processed them.

With these data, I find that black defendants in Chicago receive somewhat shorter sentences than their nonblack counterparts and demand trials significantly more often. As interpreted by Figure 1, these results suggest that black defendants in Chicago face *weaker* cases than nonblack defendants. This fact, combined with the fact that black Chicagoans are vastly over-represented in the population of felony defendants, suggests that selection into the population of felony defendants contributes significantly to incarceration disparities. Selection rules that set a lower standard for cases against black suspects would explain both facts. In Jordan (2020), I explore this further using data from the felony review process, where the State's Attorney decides which arrests to pursue. I find evidence that the quality of cases against black suspects is lower than against nonblack suspects. These differences appear to be present starting at the time when police make their arrest decisions.

To further investigate the findings of the reduced-form exercise, I structurally estimate my plea bargaining model in Section 8. I use Simulated Method of Moments to match an 11-parameter version of my model to 16 empirical moments. My SMM specification simulates heterogeneity both in the defendants and in the judges those defendants are assigned to. My estimation produces a good fit to the empirical distribution of plea bargaining and sentencing outcomes. The estimated parameters confirm the reduced-form finding that black defendants face weaker cases. By simulating the full model, I can also conclude that black defendants also face some prosecutor bias.

I next use the estimated parameters to examine the prosecutor's sentencing policies separately by race. This exercise reveals the large impact of case strength differences on defendant outcomes relative to the

[3] In many criminal justice datasets, the bulk of the records are misdemeanors. These cases rarely result in prison sentences and never in sentences longer than one year. I wish to understand incarceration disparities, and my empirical strategy depends on variation in incarceration sentence length, so I require large numbers of felony cases.

small impact of prosecutor bias. I also use my structural estimation to evaluate counterfactuals. I compare a regime where judges adopt the sentencing attitudes of their more lenient peers to a regime where they instead adopt lenient conviction attitudes. I find that the former is most beneficial for defendants because it leads to substantial declines in sentences and modest declines in trial rate. The latter only reduces sentences slightly and increases trial rates. I also find that plea bargaining significantly increases average sentences compared to a hypothetical regime where all defendants must go to trial. This suggests that defendants are willing to pay large amounts of their own time, on average 6 months, to avoid the uncertainty of a trial.

The remainder of the paper is organized as follows: Section 2 reviews prior work. Section 3 presents my model of plea bargaining, its solution, and its comparative statics. Section 4 establishes a link between the conclusions of the model and what can be observed in data. Section 5 describes the Cook County Circuit Court and Cook County State's Attorney data. Section 6 presents my regression estimations of ΔT and ΔL. Section 7 considers alternative explanations for these results. Section 8 structurally estimates my model using SMM. Section 9 concludes.

2 Prior Work

The literature on legal settlement models begins with Landes (1971). It attempts to explain why most, but not all, court cases conclude with a pretrial settlements. Trials are costly and uncertain, so a settlement that avoids trial is typically the most efficient outcome. However, attempts to reach a mutually agreeable settlement can break down in the presence of asymmetric information. In Grossman and Katz (1983) or Bebchuk (1984), defendants are better-informed that prosecutors, so plea bargain offers serve a screening function. In Reinganum (1988), prosecutors are better-informed than defendants, so plea bargain offers serve a signaling function. In Priest and Klein (1984), both prosecutors and defendants have imperfect information and go to trial when their signals disagree. Daughety and Reinganum (2017) provide a review of recent work in settlement models. Most recent developments have focused on aspects of the legal system specific to civil cases. Recent theoretical extensions of criminal plea bargaining (Kim 2010, Lee 2014) have focused primarily on the question of the defendant's innocence. Silveira (2017) structurally estimates the Bebchuk model using data from North Carolina.

The model in this paper blends settlement models with economic models of racial bias. My model builds off of the settlement screening model of Bebchuk. Defendants have private information, and prosecutors know only the distribution of this information. Prosecutors make plea bargain offers that balance obtaining a long sentence against engaging in a costly trial. Trials happen because it is suboptimal for prosecutors to make plea bargain offers that are acceptable to the full distribution of defendants. In the Bebchuk model, defendants are risk neutral, and their private information is the probability of conviction at trial. In my model, I assume that the probability of conviction at trial is common knowledge, and each defendant's private information is their degree of risk aversion. This assumption is motivated by the strong discovery provisions in criminal law.[4] These compel both prosecution and defense to share all the evidence they intend to present. This choice also clarifies the discussion of how the probability of conviction influences sentencing and trial outcomes.

My model also advances the racial bias literature by adding a bias term to the prosecutor's preferences that distorts their tolerance for trials. This idea follows Becker (1957), who modeled bias as a coefficient in a decisionmaker's preferences that varies with the race of the subject and distorts their tolerance for unsuccessful outcomes. For example, if loan officers have a bias towards white applicants, they set a looser standard for white applicants. This leads to more defaults among accepted white applicants. A recent literature including Knowles, Persico, and Todd (KPT 2001), Anwar and Fang (2006), Antinovics and Knight (2009), and Simoiu, Corbett-Davies, and Goel (2017) applies this logic to the criminal justice setting, especially traffic stops. KPT propose an outcome test to indirectly asses whether police use the same rules to search white and black motorists by comparing the rates at which they find contraband, conditional on choosing to search for it. Ayres (2002) points out that outcome tests of this form have infra-marginality problems. Police could search all motorists, regardless of race, who have a minimum probability of carrying contraband, but they will still fail an outcome test if black motorists tend to just barely exceed this threshold while white motorists exceed it by much more. KPT avoid this problem with strong assumptions about the information structure of the game played by police and motorists, which ultimately result in all drivers carrying contraband with equal probability. Later papers showed how supplementary data, such as information about arresting officers or datasets from several jurisdictions, can be used to relax these assumptions. Arnold, Dobbie, and Yang (2018) show that using a valid instrument for the decision can help to alleviate the infra-marginality problem.[5] Arnold, Dobbie, and Hull (2020) study the bail release decisions of judges, using estimates of true and false negative rates in the release decision to detect unwarranted racial disparities in judges' decisions.

Finally, this study contributes to empirical studies of racial bias in criminal justice outcomes. Alesina and Ferrara (2014) find that capital sentences in cases with white victims and black defendants are more likely to be reversed on appeal. Rehavi and Starr (2014) find that race is a strong determinant of whether federal prosecutors file charges with a mandatory minimum, which in turn has a strong effect on final sentences. Several papers show that black defendants are less likely to plead guilty than white defendants (Albonetti

[4]Though the Bebchuk model is readily adapted to the criminal context, it was originally written to model civil settlements.

[5]This works because the set of compliers for the instrument is clustered around the decision threshold. If an instrument induces marginally more traffic searches, and the induced traffic searches are successful less often for black motorists, this implies that the decision threshold for black motorists was set below the one for white motorists.

1990, Frenzel and Ball 2008, Metcalfe and Chiricos 2018), but none of them consider this outcome through the lens of the economic discrimination literature.

3 Model

A defendant (she) charged with a crime is brought before a prosecutor (he). After a non-strategic period of discovery and pretrial motions, both sides observe an identical signal of the probability that the defendant is convicted if the case goes to trial, $\theta \in (0,1)$. They then play the following game.

Technologies

The case may be resolved in one of three ways:

- **Trial:** The case may go to trial. The outcome of the trial is random: the defendant will be convicted with probability θ and not convicted with probability $(1-\theta)$. If the defendant is convicted, she is given sentence $S > 0$. If the defendant is not convicted, she is given sentence 0. Going to trial incurs a cost $c > 0$ for the prosecutor regardless of the outcome.
- **Plea Bargain:** The prosecutor may offer the defendant a plea bargain sentence of $s > 0$. s is freely chosen by the prosecutor. This is a one-time take-it-or-leave-it offer. If the defendant accepts the offer, she is given sentence s. If the defendant rejects the offer, the prosecutor chooses to resolve the case either by trial or by dropping the charge. Plea bargains incur no additional cost to the prosecutor or the defendant.
- **Dropped Charge:** The prosecutor may drop all charges against the defendant. Neither party pays any costs, and the defendant is given sentence 0.

Preferences

Defendant

The defendant belongs to an observable demographic group $j \in \{b, nb\}$. Her payoff, regardless of demographic group, is decreasing in the sentence assigned. I will represent this sentence with the placeholder variable ς. I assume the defendant's utility is given by:

$$-\frac{\varsigma^{\rho}}{\rho}$$

The defendant's relative risk aversion is constant, and her coefficient of relative risk aversion is $1-\rho$. When $\rho=1$, she is risk neutral, and her risk aversion increases as ρ increases. I assume that ρ is private information drawn from a continuously differentiable distribution function $G(\rho)$ with support $[1,\infty)$ and associated density function $g(\rho)$. $G(\rho)$ does not depend on demographic group and satisfies the increasing hazard property. That is, $\frac{g(\rho)}{1-G(\rho)}$ is weakly increasing for all $\rho>1$. The prosecutor's marginal cost of increasing s is controlled by the hazard of $G(\rho)$, so this assumption ensures that the prosecutor's choice of s has a unique solution. The increasing hazard property holds for many common distributions such as normal, uniform, and gamma with shape parameter >1. I further assume that $\lim_{x\to 1} g(\rho)=0$. This ensures there is not a mass of risk-neutral defendants. With this assumption, the prosecutor's optimal choice of s cannot collapse to the defendant's expected trial sentence due to a critical mass of defendants where that is the optimal offer.

Prosecutor

The risk-neutral prosecutor has linear preferences that are decreasing in the cost of going to trial and increasing in the sentence assigned, ς. His payout is scaled by a preference parameter, k_j, that may depend on the demographic group j the defendant belongs to. When $k_b>k_{nb}$, I will say the prosecutor is biased against group b. If the prosecutor avoids trial, he receives:

$$k_j\varsigma$$

If the prosecutor goes to trial, he receives:

$$k_j\varsigma-c$$

The defendant knows both c andk_j.

Timing and Choices

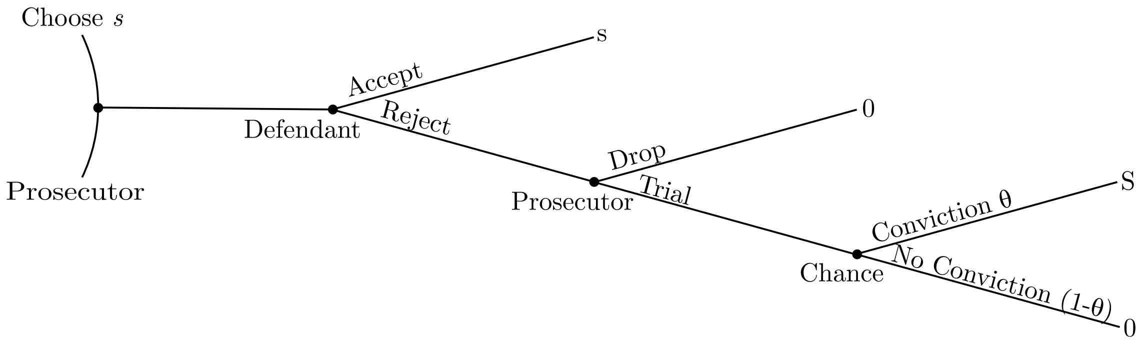
Figure 2 depicts the game played by the prosecutor and defendant.

1. The prosecutor chooses a value s to offer.

2. The defendant either rejects or accepts s. This may depend on the defendant's private information, ρ. I denote the defendant's choice by $\alpha(s,\rho)\in\{0,1\}$ where $\alpha(s,\rho)=0$ rejects the offer, and $\alpha(s,\rho)=1$ accepts the offer.

3. If the defendant accepts s, that is her sentence, and the game is over.

Figure 2: Game Tree



4. If the defendant rejects s, the prosecutor must choose whether to go to trial or drop the charges. I denote the prosecutor's choice by $\delta \in \{0, 1\}$ where $\delta = 0$ continues to trial and $\delta = 1$ drops charges.

5. If the prosecutor chooses to drop charges, the game is over, and the defendant receives a sentence of 0.

6. If the prosecutor chooses to go to trial, the trial proceeds, and the conviction outcome is determined by chance. If the defendant wins, the final sentence is 0. If the defendant loses, the final sentence is S.

Discussion of Model Assumptions

This model focuses on just one form of asymmetric information: the defendant's risk aversion parameter, ρ. Most notably, I assume common knowledge of θ, the probability of conviction. In Bebchuck's model, the analogue of θ in civil court is known only to the defendant, and the plaintiff makes settlement offers that screen against it. Common knowledge of θ is more plausible in the criminal setting due to the much stronger discovery provisions that ensure each side knows about all of the evidence available to the other. This assumption, while restrictive, allows my model to incorporate nonlinear preferences and generate comparative statics in θ. My model also assumes that the defendant has significant knowledge of the felony court system. She understands not just how a case progresses through the system, but also the prosecutor's values of k_j and c. This assumption simplifies the analysis of dropped cases by making it impossible for the prosecutor to credibly bluff about whether they would drop a case. It is plausible because the defendant is represented by a defense attorney who is playing a repeated game with the prosecutor and can share accumulated knowledge about the prosecutor with the defendant.

I bring risk aversion into the defendant's preferences with a disutility function that features marginally increasing disutility of time incarcerated. In a more complex specification, defendants could instead have a concave utility function over non-incarcerated time. In either specification, each additional day in prison hurts more than the previous. This ensures that defendants prefer short, certain prison spells over long, uncertain prison spells with the same expected incarceration time. I opt for the former specification of utility because it is more tractable. In particular, it avoids the need to specify the defendant's endowment of non-incarcerated time.

Agent Problems

I next describe each of the optimization problems solved by the agents in my game, working backwards from the end.

Prosecutor's Choice to Drop

I first define the problem for the last choice made in the game: the prosecutor's choice of whether or not to drop the charge. This problem takes as given his plea bargain offer, s, and is only made if the defendant has chosen not to accept the offer, $\alpha(s,\rho)=0$.

$$\max_{\delta\in\{0,1\}} (1-\delta)(\theta k_j S - c) + \delta * 0$$

I assume that when the payout to dropping the charge is equal to that of pursuing the case, the prosecutor will drop the charge. The solution to this problem is immediate. The prosecutor drops the case if his expected payout from a trial, $\theta k_j S$, is less than or equal to the cost of a trial, c. If the expected payout from trial instead exceeds the cost, the prosecutor will go to trial.

Defendant's Problem

I next define the problem for the second-to-last choice made in the game: the defendant's choice of whether accept the plea bargain. This problem takes as given the prosecutor's plea bargain offer, s, and anticipates his choice to drop, δ:

$$\min_{\alpha\in\{0,1\}} (1-\alpha)\left[\delta 0 + (1-\delta)\left(\theta\frac{S^\rho}{\rho} + (1-\theta)\frac{0^\rho}{\rho}\right)\right] + \alpha\frac{s^\rho}{\rho}$$

I assume that when the payout to insisting upon a trial is equal to that of accepting the plea bargain, the defendant will insist on trial. I address the solution to this problem below.

Prosecutor's Choice of Plea Bargain

Finally, I define the problem for the first choice made in the game. This is the prosecutor's choice of plea bargain offer s. This problem anticipates both the defendant's choice of $\alpha(s,\rho)$ and the prosecutor's choice of δ given the outcome of the game up to that point. The prosecutor does not know ρ, so he must integrate it out and consider the expectation of $\alpha(s,\rho)$ conditional on his choice of s:

$$\max_{s>0} \left(1 - E_\rho\left[\alpha(s,\rho)\,|s\right]\right)\left[\delta 0 + (1-\delta)\left(\theta k_j S - c\right)\right] + E_\rho\left[\alpha(s,\rho)\,|s\right] k_j s$$

Model Solution

Strategies

The prosecutor's strategy in this game is fully characterized by two choices. First, I define s^* as his optimal choice of s given the case characteristics (θ, S, k_j, c, and $G(\rho)$). Second, I define as δ^* his optimal choice of whether to drop the case in the third step given the characteristics of the case and the defendant's behavior up to that point.

The defendant's strategy is characterized by a single choice: whether to insist upon a trial given the prosecutor's offer of a plea bargain sentence, s, her private type, ρ, and the case characteristics. I define her optimal choice as $\alpha^*(s,\rho)$.

Given case characteristics, a subgame perfect equilibrium of the game is any triple $(s^*, \alpha^*(s^*,\rho), \delta^*)$ where δ^* is optimal for the prosecutor, $\alpha^*(s^*,\rho)$ is optimal for the defendant given δ^*, and s^* is optimal for the prosecutor given the defendant's response function $\alpha^*(s^*,\rho)$ and his own δ^*.[6]

Dropped Cases

First notice that $\delta^* = 1$ for *any* case where the prosecutor's expected trial payout is non-positive. Therefore $\theta k_j S - c \leq 0$ implies $\alpha^*(s,\rho) = 0$ for all ρ and s. If the expected trial payout is non-positive, the defendant always rejects the plea bargain, and the prosecutor always drops the charges. This outcome arises because the prosecutor cannot credibly threaten to take the case to trial. His expected utility at trial does not exceed his utility from dropping the case. Knowing this, the defendant can always force the prosecutor to drop the case by rejecting any plea bargains.

For the remainder of this discussion, I will restrict attention to cases where $\theta k_j S - c > 0$. In this event, $\delta^* = 0$ in any subgame perfect equilibrium because the prosecutor's utility at trial will always be higher than

[6] The equilibrium s^* will depend on the prosecutor's anticipation of $\alpha^*(s^*,\rho)$ and δ^*, and the equilibrium $\alpha^*(s^*,\rho)$ will depend on the defendant's anticipation of δ^*. For clarity, I do not include future decisions as arguments to any strategies.

his utility from dropping the case.

Defendant's Choice

Proposition 3.1: Take S, θ, and ρ as given. Suppose that the prosecutor offers $s < S$. Then, the defendant's optimal strategy follows a cutoff rule: For each s, there is a unique value $\hat{\rho}(s) = \frac{\ln \theta}{\ln s - \ln S}$ such that $\alpha^*(s, \rho) = 0$ if $\rho \leq \hat{\rho}(s)$ and $\alpha^*(s, \rho) = 1$ if $\rho > \hat{\rho}(s)$.

See Appendix A for proof.

Defendants participate in plea bargaining because they are risk averse, and plea bargaining allows them to avoid an uncertain trial. Given values for S and θ and an offer of s, only defendants with a sufficiently high aversion to risk will accept the offer. $\hat{\rho}(s)$ defines this critical value. If a defendant is at the margin of accepting an offer of s, increasing S or θ makes the trial less attractive, decreases $\hat{\rho}(s)$, and induces her to accept the offer.

Prosecutor's Choice of s

Given the defendant's strategy, I can rewrite the prosecutor's problem when $s < S$ as:

$$\max_{s} G(\hat{\rho}(s))(\theta k_j S - c) + (1 - G(\hat{\rho}(s)))\, k_j s$$

Proposition 3.2: Take S, θ, c, $G(\rho)$, j, and k_j as given. Then the prosecutor's strategy is given by a unique $s^* \in (\theta S, S)$.

See Appendix A for proof.

The prosecutor never offers a plea bargain sentence greater than S because the defendant will always reject that deal. This forces the prosecutor into a costly trial with expected payout of only θS. There is no cost to the prosecutor for offering a plea bargain in the range $(\theta S, S)$, and the worst that could happen is that the case goes to trial anyway. It therefore always benefits the prosecutor to offer a plea bargain that the defendant may accept, however unlikely that acceptance.

Likewise, the prosecutor never offers a plea bargain sentence less than θS because the defendant will always accept that deal. The defendant is risk averse, and θS is the certainty equivalent of a risk neutral defendant. If the prosecutor offers an even lower sentence than θS, he gives up sentence length with no compensating increase in the probability that the defendant will accept the deal.

Within $(\theta S, S)$, the prosecutor trades longer plea bargain sentences off against greater risk of the defendant rejecting the deal. When choosing s, the prosecutor also implicitly chooses $\hat{\rho}(s)$, the risk tolerance of the defendant who is indifferent between accepting and rejecting the plea bargain. I define $\rho^* = \hat{\rho}(s^*)$.

The prosecutor's central tradeoff is encapsulated in the first order condition to his choice of s:

$$(1 - G(\rho^*)) = g(\rho^*) \frac{\partial \hat{\rho}(s)}{\partial s}|_{s=s^*} [s^* - (\theta S - {}^c/_{k_j})]$$

The left hand side is the marginal benefit of increasing s. The prosecutor obtains a longer sentence for the $(1 - G(\rho^*))$ defendants who accept the plea bargain. The right hand side is the marginal cost of increasing s. The prosecutor shifts $g(\rho^*) \frac{\partial \hat{\rho}(s)}{\partial s}|_{s=s^*}$ defendants from accepting the plea bargain to demanding a trial, so he must trade off that many plea bargain payouts (s^*) for trial payouts $(\theta S - {}^c/_{k_j})$. Notice that k_j only appears relative to the trial cost c. All other payoff components in the prosecutor's problem are scaled by k_j, so his problem would be completely neutral in k_j without a trial cost that is constant in all cases.

Comparative Statics of $s^*$and ρ^*

I now describe how s^* and ρ^* change as underlying characteristics of the case change, holding all others fixed.

Proposition 3.3: As the prosecutor's bias increases, holding all else constant, plea sentence length increases and trial rate increases. That is, $\frac{\partial s^*}{\partial k_j} > 0$ and $\frac{\partial \rho^*}{\partial k_j} > 0$.

See Appendix A for proof.

An increase in k_j leads to a corresponding decrease in the prosecutor's perceived cost of trial, ${}^c/_{k_j}$. If the prosecutor perceives trials as less costly, he will be willing to trade longer plea bargain sentences for an increased risk of going to trial. Meanwhile, changing k_j has no effect on the defendant's cutoff rule, $\hat{\rho}(s)$. If a defendant would accept a certain plea deal from an unbiased prosecutor, she would accept exactly the same plea deal from a highly biased prosecutor. This means that the set of (s^*, ρ^*) pairs available to the prosecutor does not change with k_j. If he wants longer plea bargain sentences (a higher s^*), he *must* accept more trials (a higher ρ^*).

Proposition 3.4: As the probability of conviction at trial increases, plea sentence length increases. That is, $\frac{\partial s^*}{\partial \theta} > 0$. When $\theta \geq e^{-\rho^*}$, as the probability of conviction at trial increases, trial rate decreases. That is, $\frac{\partial \rho^*}{\partial \theta} < 0$.

See Appendix A for proof.

In contrast to k_j, a change in θ works primarily through the defendant's incentives. As the probability of conviction at trial increases, the defendant becomes more willing to accept a longer plea bargain sentence in order to avoid that trial. This changes the set of (s^*, ρ^*) pairs available to the prosecutor. As θ increases, he can hold s^*constant while decreasing ρ^* or hold ρ^* constant while increasing s^*. He will generally choose a convex combination of these two options, securing both longer sentences *and* fewer trials. Thus, a change in θ is not bound to the same tradeoff logic as a change in k_j. The latter is a change in preferences that does not grant the prosecutor any additional resources but merely shifts how he chooses to spend them. The former acts more like a windfall of income for the prosecutor that he may spend on both of the goods he likes.

The restriction that $\theta \geq e^{-\rho^*}$ is necessary to rule out the possibility that trial rate may act like an inferior good. When θ is very low, the prosecutor's case is very weak. The defendant on the margin of accepting his plea bargain offer has a very low risk tolerance. This low marginal risk tolerance allows the prosecutor to increase s^* a great deal and see only a small increase in the trial rate. Under these circumstances, granting the prosecutor greater bargaining leverage by increasing θ may cause him to increase s^* very aggressively and accept the small compensating increase in ρ^*. The assumption that $\theta \geq e^{-\rho^*}$ rules out this scenario regardless of the exact distribution of $G(\rho)$. It is a sufficient condition, but not a necessary one. Depending on $G(\rho)$, $\frac{\partial \rho^*}{\partial \theta} < 0$ may hold for some $\theta < e^{-\rho^*}$, but it will not hold in general as $\theta \to 0$.

A useful property of the $\theta \geq e^{-\rho^*}$ restriction is that it can be transformed, via $\rho^* = \hat{\rho}(s^*)$, into a form that is testable in data: $\frac{S}{s^*} < e$. The prosecutor's case must not be so weak that he offers plea sentence lengths that are very low compared to the defendant's potential punishment at trial. In Section 7, I test this form of the restriction in data from the Circuit Court of Cook County and find that it holds in 84.6% of cases. My results are not sensitive to the exclusion of the remaining cases.

Appendix A also calculates comparative statics for c and S. Increasing c directly increases perceived trial costs, c/k_j, so it has the opposite effects as an increase in k_j. A higher S increases both the expected sentence at trial and the variance associated with a trial. The increase in expected trial sentences makes all defendants more willing to accept long plea bargains. The increase in variance amplifies this effect for more risk averse defendants. The prosecutor capitalizes on this by greatly increasing the sentences he offers. Some risk-tolerant defendants will demand a trial, but the more risk-averse defendants will still accept. The overall effect is an increase in both plea bargain sentences and trial rates.

4 Empirical Implications

The comparative statics derived in Section 3 suggest a method for distinguishing differences in k_j from differences in θ. Increasing either leads to more punitive plea bargain offers, but they will have opposite effects on the the likelihood of trial. It may therefore be possible to learn something about how k_j and θ differ across races by examining the joint differences in s^* and ρ^*. Neither s^* nor ρ^* is observable, but I show in this section that their comparative statics are closely related to the comparative statics of average observed sentence length, L, and trial rate, T, in a population. I then explain how Figure 1 summarizes what can be learned from available data given the maintained assumptions of my model.

Distributions of θ

Although θ is known to both the prosecutor and the defendant, it is unknown to outside observers. I define $F(\theta|j)$ as the distribution of θ among defendants of race j who reach the plea bargaining stage.[7] Unlike prosecutor preferences, where there is a single k_b that can be directly compared to k_{nb}, it is not always possible to order $F(\theta|b)$ and $F(\theta|nb)$.[8] Therefore, I will restrict my attention to cases where $F(\theta|b)$ and $F(\theta|nb)$ can be compared using the usual stochastic order. That is, either $Pr\{\theta > x|b\} > Pr\{\theta > x|nb\}$ $\forall x \in (0,1)$, in which case I will say $F(\theta|b) >_{ST} F(\theta|nb)$, or $Pr\{\theta > x|b\} < Pr\{\theta > x|nb\}$ $\forall x \in (0,1)$, in which case I will say $F(\theta|b) <_{ST} F(\theta|nb)$.[9] If a group has the greater θ distribution, I will say that the cases against them are stronger, and if a group has the lesser θ distribution, I will say that the cases against them are weaker.

Stochastic order is a strong restriction, but it encompasses two key models about why θ may differ between black and nonblack defendants. First, consider a model of structural discrimination where black defendants face different probabilities of conviction due simply to their race. All defendants draw $\tilde{\theta}$ from a shared distribution. Nonblack defendants have probability of conviction $\theta = \tilde{\theta}$, but black defendants have probability of conviction $\theta = b\left(\tilde{\theta}\right)$ with $b\left(\tilde{\theta}\right) > \tilde{\theta}$ $\forall\tilde{\theta}$. In this simple model, $F(\theta|b) >_{ST} F(\theta|nb)$. For example, this model captures the behavior of a judge or juror who would convict a black defendant under circumstances where they would not convict a nonblack defendant.

Alternatively, consider a model of selection discrimination where the necessary θ to trigger arrest and criminal charges differs by race. That is, all potential defendants draw from the same distribution of θ, but the realized population of black defendants is those with $\theta > \hat{\theta}_b$, the realized population of nonblack defendants is those with $\theta > \hat{\theta}_{nb}$, and $\hat{\theta}_b < \hat{\theta}_{nb}$. In this simple model, $F(\theta|nb) >_{ST} F(\theta|b)$. For example, this model captures the behavior of a police officer who would arrest a black defendant under circumstances where they would not arrest a nonblack defendant.

Analogues for s^* and ρ^*

Neither s^* nor ρ^* is directly observable in typical court data. s^* describes plea bargain *offers* and is only observed when the defendant accepts an offer. ρ^* is private information of the defendant. In place of s^*, I

[7] These distributions may be shaped by the prosecutor's decision to drop cases. I consider the consequences of this in Section 7. This section and all of the empirical work guided by it consider the population of defendants *after* the prosecutor has dropped any cases he does not wish to pursue.

[8] Throughout I will maintain the abstraction of a monotlithic prosecutor with a pair of scalar k_j weights. The cases considered in my data were heard by many different prosecutors, but I do not know which prosecutor is assigned to which case. A dataset where this is known could plausibly learn about individual values for k_j.

[9] This order relationship is also known as First-Order Stochastic Dominance.

use a measure of average observed sentence length, including sentences from trials and counting not guilty verdicts as 0. I call this measure L:

$$L = (1 - G(\rho^*))\, s^* + G(\rho^*)\, \theta S$$

I show in Appendix A that L and s^* have the same comparative statics with respect to k_j and θ. That is, $\frac{\partial L}{\partial k_j} > 0$ and $\frac{\partial L}{\partial \theta} > 0$.[10]

In place of ρ^*, I use the rate at which defendants go to trial. I call this measure T:

$$T = G(\rho^*)$$

It is immediately apparent that for any parameter x, $\frac{\partial \rho^*}{\partial x} > 0 \Rightarrow \frac{\partial T}{\partial x} > 0$. Hence, $\frac{\partial T}{\partial k_j} > 0$ and $\frac{\partial T}{\partial \theta} < 0$.

Now consider sentence length and trial rate as functions of case strength: $L(\theta|j)$ and $T(\theta|j)$.[11] Within the set of non-dropped cases, and if there are no cases that are too weak, the above comparative statics hold everywhere, and $L(\theta|j, c, s)$ is strictly increasing in θ while $T(\theta|j, c, S)$ is strictly decreasing in θ. It is a general property of stochastic order that if $F(x) >_{ST} G(x)$, then for any strictly increasing (decreasing) and piecewise differentiable function $u(x)$: $E_F[u(x)]$ is greater than (less than) $E_G[u(x)]$. To fix ideas, I temporarily assume that case strength differences are the only differences by race, so $k_b = k_{nb}$. In this event, I can conclude $F(\theta|b) >_{ST} F(\theta|nb) \Rightarrow E[L(\theta)|b] > E[L(\theta)|nb]$ and $E[T(\theta)|b] < E[T(\theta)|nb]$.

If I focus instead on how sentence length and trial rate differ with prosecutor bias, I only need to order scalars rather than distributions. If $k_b > k_{nb}$, then $L(\theta|b) > L(\theta|nb)$ and $T(\theta|b) > T(\theta|nb)$ $\forall \theta$. Ignoring case strength differences for now by assuming$F(\theta|b) = F(\theta|nb)$. In this event, I can conclude: $k_b > k_{nb} \Rightarrow E[L(\theta)|b] > E[L(\theta)|nb]$ and $E[T(\theta)|b] > E[T(\theta)|nb]$.

Finally, I find it convenient to define $\Delta L = E[L(\theta)|b] - E[L(\theta)|nb]$ and $\Delta T = E[T(\theta)|b] - E[T(\theta)|nb]$. For example, if black defendants have, on average, longer sentences and more trials, then $\Delta L > 0$ and $\Delta T > 0$. These are the y-axis and x-axis of Figure 1, respectively. My regression analysis in Section 6 estimates ΔL and ΔT.

Results

The results above are informative about the data patterns I should expect to result from case strength differences or bias when everything else is held constant. When both bias and case strength are allowed

[10] The average observed plea bargain sentence, despite being the more natural analogue, does not necessarily have this property. k_j and θ also affect which plea bargains are rejected, so entirely excluding cases that go to trial interferes with properly measuring the effects of changing k_j and θ.

[11] I hold constant trial cost c, trial sentence S, and risk aversion distribution $G(\rho)$ throughout this analysis, but I suppress this notation.

to vary freely, their effects on ΔL or ΔT may be at odds. A state of the world with both prosecutor bias against black defendants and weak cases against black defendants will certainly see $\Delta T > 0$ because both forces tend to increase ΔT. However, ΔL may be zero if the positive effect from bias offsets the negative effect from weak cases. Even so, if I were to observe data where both ΔL and ΔT can be clearly signed, I can definitely learn something either about case strength or about prosecutor bias. For example, if $\Delta L > 0$ and $\Delta T < 0$, I know there must be strong cases against black defendants because bias alone could not have produced that data. The propositions below formalize this intuition.

Proposition 4.1: Consider the set of cases where $\theta k_j S - c > 0$. Suppose that c, S, and $G(\rho)$ are the same for all defendants. Further suppose that, $\theta \geq e^{-\rho^*}$ $\forall \{(\theta, j)\}$ in the support of $F(\theta|j)$. Lastly, assume $F(\theta|B)$ and $F(\theta|W)$ can be stochastically ordered. Then $\Delta L < 0$ and $\Delta T > 0$ implies that $F(\theta|nb) >_{ST} F(\theta|b)$. Likewise, if $\Delta L > 0$ and $\Delta T < 0$, then $F(\theta|b) >_{ST} F(\theta|nb)$.

Proof: Because c, S and $G(\rho)$ do not vary across group, the only candidates for explaining the cross-group differences in average plea bargain sentence length and average plea bargain acceptance rate are $F(\theta|j)$ and k_j.

Now, consider data where $\Delta L < 0$ and $\Delta T > 0$ and further suppose that $F(\theta|b) \geq_{ST} F(\theta|nb)$. If $k_b = k_{nb}$, this implies $\Delta L \geq 0$ and $\Delta T \leq 0$. Either constitutes a contradiction. Letting $k_b > k_{nb}$ will only exacerbate the first contradiction because L is increasing in k_j. Likewise, letting $k_b < k_{nb}$ will only exacerbate the second contradiction because T is increasing in k_j. It is therefore impossible to avoid a contradiction, and it must be the case that $F(\theta|nb) >_{ST} F(\theta|b)$. The arm of the proof where $\Delta L > 0$ and $\Delta T < 0$ is similar.

Proposition 4.2: Consider the set of cases where $\theta k_j S - c > 0$. Suppose that c, S, and $G(\rho)$ are the same for all defendants. Further suppose that, $\theta \geq e^{-\rho^*}$ $\forall \{(\theta, j)\}$ in the support of $F(\theta|j)$. Lastly, assume $F(\theta|b)$ and $F(\theta|nb)$ can be stochastically ordered. Then $\Delta L > 0$ and $\Delta T > 0$ implies that $k_b > k_{nb}$. Likewise, $\Delta L < 0$ and $\Delta T < 0$ implies that $k_b < k_{nb}$.

See Appendix A for proof.

Discussion

These theorems divide $(\Delta L, \Delta T)$ space into the four quadrants in Figure 1. It is important to note that, despite the appearance of Figure 1, this technique cannot provide information about the *magnitude* of any differences of bias or case strength, just about their presence. Proposition 4.1 says that if the data are anywhere in the top-left (bottom-right), I can conclude that cases against black defendants are stronger (weaker) than cases against their nonblack counterparts. Data in these quadrants, however, do not permit any conclusion about prosecutor bias, nor does data farther away from the origin necessarily imply that the differences are larger. Proposition 4.2 says that if the data are in the top-right (bottom-left), I can conclude that prosecutors are biased against (for) black defendants. Data in these quadrants, however, do not permit any conclusion about case strength.

The advantage to using these results to guide empirical work is that they do not require a research design that attempts to hold constant or structurally estimate k_j or θ. The disadvantage is that regression results will not always permit a conclusion about one of the parameters of interest. It it is also possible to see data where there are racial differences in *both* case strength and prosecutor bias, but their combined effect on ΔL and ΔT is such that one difference cannot be clearly signed, making it difficult to come to a clear conclusion.

The results in this section can cleanly associate racial differences in trial rate and sentence length with differences in case strength and prosecutor bias only by holding other parameters constant. In particular, prosecutor trial costs, potential sentences, and the distribution of risk aversion. In Section 7, I discuss the plausibility of those assumptions and which patterns of results could be explained by the failure of each.

5 Data

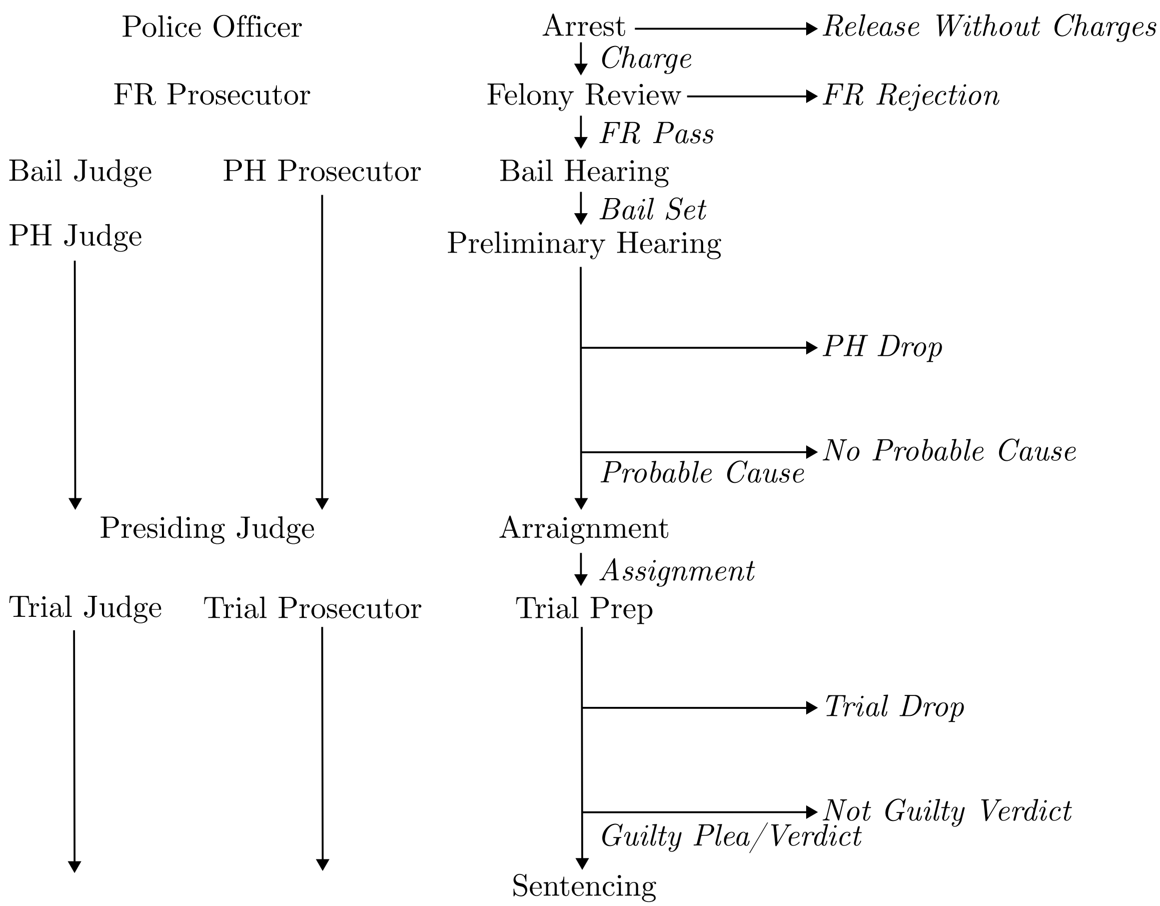
I next turn to a description of the administrative data court data I will use to estimate ΔL and ΔT.

Institutional Setting

My data come from the Criminal Division of the Cook County Circuit Court. Figure 3 gives a visual representation of how a criminal case progresses through that system from arrest to sentencing. The left side of Figure 3 tracks key state actors at each stage in the criminal case. The prosecutor in my model is the trial prosecutor who receives the case post-arraignment, after the charges have been set by the felony review and preliminary hearing prosecutors. This means that the prosecutor in my model takes the charges against the defendant as given. The right side of Figure 3 tracks ways for the case to proceed. Arrows pointing downwards indicate cases that continue within the system. Arrows pointing to the right indicate cases that exit the system. Appendix B discusses Figure 3, and the Cook County criminal justice system as a whole, in more detail.

My data begin at the arraignment stage when a defendant is formally charged and assigned to a trial courtroom and judge. See Appendix C for details on how I determine which cases were assigned to which judges. I only observe judge assignment for cases in Chicago, so I exclude cases from parts of Cook County outside of Chicago. Not every case in my data was assigned to a judge at random.

Figure 3: Felony Case Progression in Cook County



Circuit Court Data

My court data are electronic records of all unsealed cases heard by the Criminal Division of the Cook County Circuit Court from 1984 to 2019. I observe all charges filed in the case, including amended charges. I also observe the full disposition history of each case, which records all official actions in the courtroom (e.g. pleas, motions, continuances, orders, verdicts, and sentences). These data also have identifying and demographic information for each defendant.[12] The court stopped classifying defendants of Hispanic origin separately from white and black defendants around 2010, so I combine white and Hispanic origin defendants in my analysis into a single "nonblack" category. Most defendants have a unique, fingerprint-based ID number that I can use to follow them across cases and over time. If multiple cases against the same defendant that were initiated in overlapping time periods, I collapse them into a single case. This is because simultaneous cases

[12] The original data extract provided by the Clerk of Court did not include these fields for cases where the defendant was not convicted. I supplement the Clerk's extract with data obtained from scraping the Clerk's public computer system. This system is populated by the same database that generated my original extract.

are typically negotiated as a unit. I restrict my attention to defendants who were not already adults at the time my sample begins. This guarantees that I observe all of their felony cases in Cook County and can construct an accurate criminal history.

The data provided by the Clerk required substantial cleaning and reformatting before they were ready to be used in statistical analysis. See Appendix C for details of this process. In some cases, essential fields could not be recovered because they were never entered by the Clerk or contain invalid data. Sometimes I am able to plausibly impute this information. When fingerprint ID is missing, I impute on similarity in name and demographic information between cases. Conversely, I can fill in missing demographic information between cases that share a fingerprint ID.[13] This leaves just 6.5% of my total dataset that must be dropped due to missing or invalid fingerprint ID, demographic information, or charge information.

My two outcomes of interest are whether a case ended in a trial and the length of incarceration imposed. I mark a case as not ending in a trial if I see a guilty plea or a Supreme Court Rule 402 conference followed by a trial and a guilty verdict.[14] I measure incarceration as the nominal incarceration length announced by the judge, topcoded at 80 years. In relation to my model, this is s^* in the event of a plea bargain and S or 0 in the event of a trial. Prisoners in Illinois routinely receive day-for-day good time credit, so actual expected time served is half the time announced by the judge.[15] Some short incarceration sentences return the defendant to the county jail, but most send them to a state prison. I do not distinguish between these two types of incarceration. Defendants who awaited trial in jail typically receive a time served credit that reduces their post-conviction sentence by the amount of time spent in jail. Below, I consider an alternative sentence length measure that applies this credit. I record a sentence of 0 if the defendant is sentenced to probation or some other non-incarceration punishment.

The type of charges faced by the defendant is an important conditioning variable in my empirical model. I divide the space of criminal charges into two dimensions. The first is felony class. Illinois has 5 felony classes: X, 1, 2, 3, and 4. Felony class determines the range of sentences a judge may give if they find the defendant guilty, with X being the most severe and 4 being the least. The second dimension of a criminal charge is category. Guided by the criminal code of Illinois, I split charges into nine major categories: murder, sex crime, robbery, other violent, burglary, theft, drug, weapons, and other nonviolent. Some charges fall

[13] When demographic fields disagree among cases that share a fingerprint ID, I use the information that appears in the greatest number of cases associated with that ID.

[14] Illinois Supreme Court Rule 402 deals with plea bargain negotiations while the judge is present. These negotiations signal a "stipulated bench trial." Such trials can arise because the prosecutor is procedurally or politically unable to reduce charges in a way necessary to secure a satisfactory plea agreement. Alternatively, they can arise because the defendant does not want to formally admit guilt despite having no defense against the charges. In a stipulated bench trial: all parties meet to agree upon an outcome, the defendant stipulates to certain facts, the case goes to trial, and the judge quickly convicts on a reduced charge. This adjustment applies to 650 cases (0.2%) in my final estimation sample.

[15] Prisoners are not guaranteed this credit and may have it taken away if they misbehave. Illinois also has a Truth in Sentencing law that demand prisoners serve 85-100% of their nominal sentence for some serious crimes. Most of these are murder and sex crimes, which I exclude from my estimation sample.

Table 1: Sample Selection

Sample Selection Rule	Excluded Records	Percentage of Full Sample
Initiated in 1984-2016	77,810	7.4
Not missing essential data	61,799	5.9
Def. younger than 16 in 1984	351,397	33.6
Def. black, white, or Hispanic	2,154	0.2
Accepted charge category and no life/death sentence	73,655	7.1
Assigned in Chicago	108,490	10.4
Accepted conclusion	27,460	2.6
Final Sample	**342,616**	**32.8**

Notes: This table presents the number of records excluded from my sample with the sequential application of my 7 sample selection rules.

outside these categories (e.g. contempt of court, escape from prison, and aggravated DUI), and I do not include them in my estimation sample. A single case may carry multiple charges. When this happens, I condition on the most serious charge.[16]

Analysis Sample

I use the following selection rules to construct my final sample: (1) The case began between 1984 and 2016. Cases that began after 2016 have a much higher chance to still be ongoing. (2) No missing or invalid fingerprint ID, demographic, or charge data. (3) The defendant must have been at most 16 in 1984. This ensures that I can observe their full adult criminal history in Cook County. (4) The defendant's race was recorded as either black, white, or Hispanic. (5) The most serious charge is robbery, assault, burglary, theft, drug, weapon, or other nonviolent, and the defendant did not receive a life sentence or death sentence.[17] (6) The case was assigned to one of the primary courtrooms serving Chicago. This restriction is necessary to observe courtroom assignment. (7) The case ended in conviction, acquittal at trial, or dismissal by a judge. This excludes cases abandoned because the defendant died or fled, not yet closed, or with an ending that could otherwise not be parsed. My analysis sample also excludes cases dropped by the prosecution, which aligns it with the set of defendants considered by my model and in Propositions 4.1 and 4.2.

Table 1 presents the count of records each rule drops when they are applied sequentially. The sample rule with the largest impact is the restriction that defendants must be younger than 16 in 1984. This eliminates many cases in the 1980s and early 1990s. Most cases involve defendants younger than 25, so it is not as relevant in later years. I will find that a defendant's count of prior cases is a crucial conditioning variable. In the absence of outside information on each defendant's criminal record, this selection rule is necessary

[16] I define seriousness first by felony class, then by category, then by the order that they appear in the charging document.

[17] Life sentences and death sentences for crimes other than murder and rape are extremely rare. My estimation sample without that restriction would include only 67 more cases.

Table 2: Court Data Summary Statistics

Full Sample	Mean	SD	Black	Mean	SD	Nonblack	Mean	SD
Sentence (months)	22.8	43	Sentence (months)	23.7	43.5	Sentence (months)	19.2	41
% Plea Guilty	88.1		% Plea Guilty	87.7		% Plea Guilty	89.3	
% Convicted	94.9		% Convicted	95		% Convicted	94.9	
% Black	79.3							
Age at Arrest	24.5	6.4	Age at Arrest	24.4	6.4	Age at Arrest	24.9	6.3
% Male	90.5		% Male	90.5		% Male	90.2	
Prior Convictions	1.2	1.66	Prior Convictions	1.3	1.7	Prior Convictions	0.7	1.2
% Public Defender	65		% Public Defender	68.8		% Public Defender	50.2	
% Held in Custody	69.5		% Held in Custody	71.5		% Held in Custody	61.7	
% Class X Felony	14.4		% Class X Felony	15		% Class X Felony	12.1	
% Class 1 Felony	20.3		% Class 1 Felony	22.4		% Class 1 Felony	12.6	
% Class 2 Felony	27.2		% Class 2 Felony	26.4		% Class 2 Felony	30.2	
% Class 3 Felony	12.1		% Class 3 Felony	11.5		% Class 3 Felony	14.6	
% Class 4 Felony	26		% Class 4 Felony	24.8		% Class 4 Felony	30.6	

Notes: This table shows summary statistics, collectively and by race, for the 342,616 observations used in the analysis sample for the Cook County Court data.

despite its impact on sample size.

Table 2 presents summary statistics for my analysis sample in aggregate and separately by race. The average nominal sentence, including 0 month sentences, is about 1.9 years. This distribution is substantially skewed to the right. The median sentence is only 1/3 of a month. As in most criminal courts, the majority of defendants, 88.1%, plead guilty. My sample is 79.3% black, a much higher proportion than in the general population of Chicago, which is 30% black. Defendants in my sample are 90.5% male and generally young, with an average age of 24.5. Only 2.4% of defendants in my sample are over 40. Many defendants do not have the resources to make cash bail or pay a private attorney. 65% have a public defender appointed to represent them, and 69.5% spend at least some time after arraignment in jail.[18] These statistics are driven by black defendants, though many nonblack defendants also use public defenders or await trial in jail. Felonies of all 5 classes are well-represented. Black defendants are more likely to be charged with more serious felony classes like 1 and X. 37.4% of black defendants are charged with a Class X or 1 felony, while only 24.7% of nonblack defendants are. This will make any comparison between the groups that does not condition properly on felony class very misleading. Likewise, black defendants tend to have more prior convictions than nonblack defendants. First offenders are often afforded probation sentences, but this is rarely true of repeat offenders, so conditioning on this difference is crucial.

[18] Arraignment occurs at least a few days after the initial bail hearing, so all defendants would have had an opportunity to post bail by that time.

6 Regression Results

Regression Specification

In this section, I use my Cook County court data to investigate the correlation between defendant race and my two outcomes of interest using a linear regression specification:

$$Y_i = \alpha + \pi black_i + X_i\beta + \gamma_i^c\psi^c + \gamma_i^t\psi^t + \epsilon_i$$

Y_i is an outcome (either sentence length or an indicator for going to trial). $black_i$ is an indicator for whether the defendant is black. Thus, when $Y_i = L_i$, π will estimate ΔL, and when $Y_i = T_i$, π will estimate ΔT. X_i is a vector of controls (dummies for age and number of priors, plus indicators for male, multiple charges brought in the case, multiple defendants on the case, public defender appointed, and any jail time). γ_i^c is a vector of dummy variables that describe the charge. γ_i^t is a vector of dummy variables for the timing and courtroom of case assignment. ϵ_i is an error term that I assume is orthogonal to the other variables in my estimation model. In the context of my model, this assumption is equivalent to assuming that my controls capture all variation in case characteristics that my empirical strategy holds constant, such as the trial sentence S and trial cost c. This is a strong assumption, but it is aided by the detail of my control variable, as well as the fact that Illinois is a deterministic sentencing state where potential trial sentences are largely dictated by law rather than judicial discretion.

For my core results, I will present six iterative specifications. The first is a simple comparison of means between black and nonblack defendants. The second adds demographic covariates and some simple charge covariates. It also sets γ_i^t to a vector of dummies for year of assignment. The third adds fixed effects for the defendant's count of prior convictions. The fourth adds indicators for whether the defendant was represented by a public defender or awaited trial in jail. The fifth adds γ_i^c as a vector of charge fixed effects that interact crime category and felony class.[19] The sixth changes γ_t^t to a vector of dummies that interact year of assignment with courtroom of assignment.

All standard errors are heteroskedastically robust and clustered by assigned year and courtroom. I cluster at this level because courtroom assignment at a point in time determines both the specific judge and prosecutor who will handle the case. This may induce variation in θ and k at the courtroom level that could cause defendants assigned to the same courtroom at the same time to have correlated errors.

[19] Some of the resulting cells are sparse due to the criminal code. For example, there are almost no Class 4 robberies because most offenses classified under robbery are considered more serious than Class 4. In these cases, I combine adjacent classes into the same cell. In particular, I combine: robbery 2-4, burglary X-1, burglary 2-4, and theft X-1.

Table 3: Sentence Length

	Sentence	Sentence	Sentence	Sentence	Sentence	Sentence
Black	4.488***	4.692***	-0.997***	-1.669***	-1.127***	-1.429***
	(0.230)	(0.217)	(0.226)	(0.212)	(0.194)	(0.196)
Male		15.53***	8.240***	7.317***	5.606***	5.593***
		(0.211)	(0.188)	(0.177)	(0.162)	(0.162)
Multiple Defendants		7.389***	7.807***	6.988***	2.326***	2.372***
		(0.265)	(0.260)	(0.251)	(0.223)	(0.222)
Multiple Charges		17.17***	16.87***	14.95***	7.812***	7.801***
		(0.211)	(0.207)	(0.187)	(0.173)	(0.172)
Public Defender				1.291***	2.432***	3.007***
				(0.195)	(0.171)	(0.174)
Ever in Jail				16.50***	11.25***	11.39***
				(0.211)	(0.182)	(0.189)
Observations	342616	342616	342616	342616	342616	342616
Adjusted R^2	0.002	0.074	0.130	0.156	0.275	0.282
Prior Cond.	None	None	Yes	Yes	Yes	Yes
Charge Cond.	None	None	None	None	ClassXCat FE	ClassXCat FE
Assignment Cond.	None	Year FE	Year FE	Year FE	Year FE	CtrmXYear FE

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents the results of regressions of an indicator for black defendants on nominal sentence length in months. Beginning with the second column, covariates also include dummy variables for age and number of prior convictions. Standard errors are clustered at the courtroom-year level. See Section 5 for details about sample selection.

Aggregate Results

Table 3 shows the relationship between sentence length and race as I add conditioning variables. In a pure comparison of means, black defendants receive nominal sentences that are about 4.5 months longer than those of nonblack defendants. This changes very little in the second specification. The key covariate affecting ΔL is the count of prior convictions, which reverses the sentencing gap such that black defendants receive nominal sentences about one month *shorter* than those of nonblack defendants. This gap grows somewhat as I add additional case characteristics. In my most complete specification, black defendants receive sentences that are about 1.5 nominal months shorter than those of comparable nonblack defendants. With typical good time credit, that is 3 or fewer weeks of actual incarceration time.

Table 4 presents a similar progression for trial rate. I find that ΔT is positive for any specification and largest before the addition of charge effects. My most complete specification finds that black defendants are about 1.3% more likely to go to trial. This is a large difference in percentage terms because only 11.9% of cases go to trial overall. Given the low trial rate, a linear probability model may not be well specified, but

Table 4: Trial

	Trial	Trial	Trial	Trial	Trial	Trial
Black	0.0154***	0.0183***	0.0154***	0.0245***	0.0137***	0.0131***
	(0.00211)	(0.00193)	(0.00194)	(0.00203)	(0.00185)	(0.00169)
Male		0.0369***	0.0333***	0.0308***	0.0176***	0.0187***
		(0.00200)	(0.00201)	(0.00200)	(0.00195)	(0.00193)
Multiple Defendants		0.0399***	0.0398***	0.0398***	0.0344***	0.0328***
		(0.00179)	(0.00180)	(0.00180)	(0.00177)	(0.00174)
Multiple Charges		0.0241***	0.0237***	0.0241***	-0.0299***	-0.0313***
		(0.00173)	(0.00173)	(0.00173)	(0.00166)	(0.00161)
Public Defender				-0.0567***	-0.0458***	-0.0532***
				(0.00240)	(0.00224)	(0.00180)
Ever in Jail				-0.0126***	-0.0221***	-0.0229***
				(0.00184)	(0.00183)	(0.00174)
Observations	342616	342616	342616	342616	342616	342616
Adjusted R^2	0.000	0.018	0.019	0.026	0.056	0.078
Prior Cond.	None	None	Yes	Yes	Yes	Yes
Charge Cond.	None	None	None	None	ClassXCat FE	ClassXCat FE
Assignment Cond.	None	Year FE	Year FE	Year FE	Year FE	CtrmXYear FE

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents the results of regressions of an indicator for black defendants on whether the case ended in a trial. Beginning with the second column, covariates also include dummy variables for age and number of prior convictions. Standard errors are clustered at the courtroom-year level. See Section 5 for details about sample selection.

Table 5: Conviction

	Conviction	Conviction
Black	-0.00855	-0.00764***
	(0.00727)	(0.00116)
Observations	40911	342616
Adjusted R^2	0.121	0.039
Prior Cond.	Yes	Yes
Charge Cond.	ClassXCat FE	ClassXCat FE
Assignment Cond.	CtrmXYear FE	CtrmXYear FE
Sample	Trials	All Cases

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

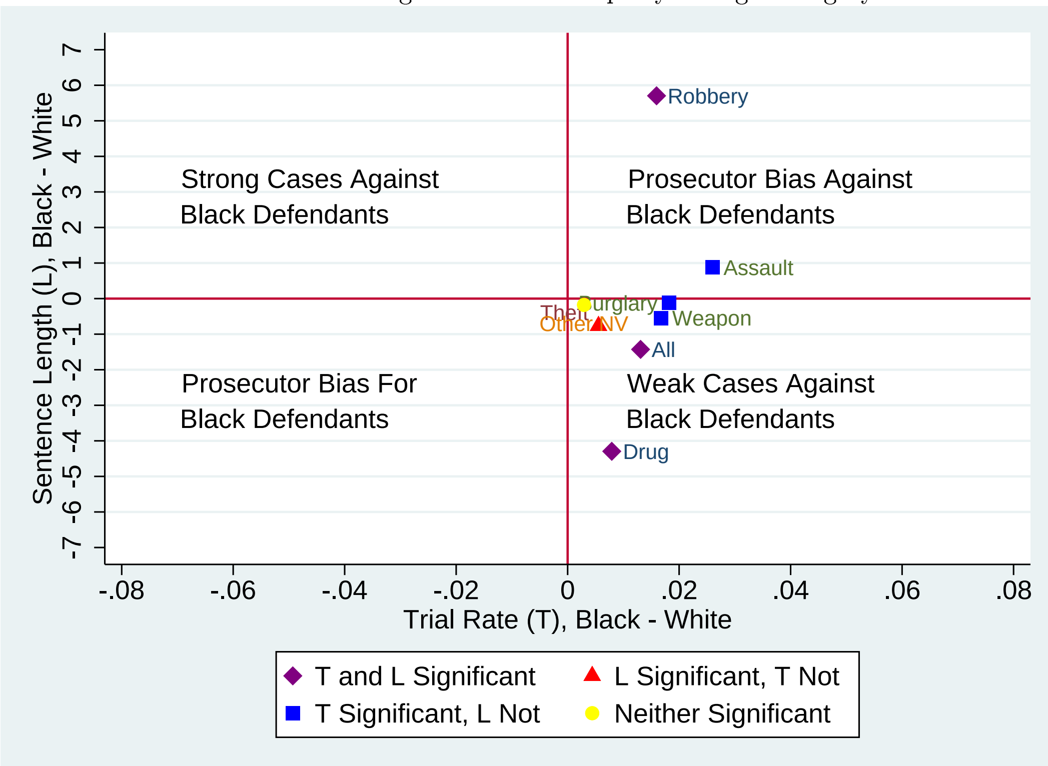
Notes: This table presents the results of regressions of an indicator for black defendants on whether the case ended in a conviction. Covariates also include dummy variables for age and number of prior convictions. Standard errors are clustered at the courtroom-year level. The first column restricts to cases that went to trial. See Section 5 for details about sample selection.

Table D1 shows that a logistic model also finds that black defendants are more likely to go to trial.

In the language of my model, these results indicate that $\Delta L < 0$ and $\Delta T > 0$. According to Proposition 4.1, if we assume that the θ distributions can be stochastically ordered, then black defendants must be the ones facing cases with a lower probability of conviction. The finding that black defendants face weaker cases may be surprising. The fact that African Americans are extremely over-represented in prisons has motivated the search for some structural disadvantage at court that makes them *more* likely to be convicted. I find instead that over-representation in prison is driven primarily by the composition of defendants who reach court. In a city that is 30% black, nearly 80% of my sample of felony defendants is black. This disparity, coupled with the fact that black defendants face weaker cases than their white counterparts, suggests that black Chicagoans may be charged with felonies on the basis of weaker evidence. I investigate this hypothesis further in Jordan (2020), which focuses on the decisions of Cook County's felony review prosecutors.

I also investigate the rate at which black defendants are actually convicted, though this analysis cannot speak directly to underlying θ distributions. The first column of Table 5 shows the racial difference in conviction rates for cases that reach trial, conditional on the covariates in my most complete regression specification. I find that black defendants are less likely to be convicted at trial, but the effect is not statistically significant. The second column of Table 5 looks at all defendants in my analysis sample, including the ones who accepted plea deals. I also find that black defendants in this sample are less likely to be convicted. These results are suggestive, but neither of these specifications are a direct test of the hypothesis that black defendants face weaker cases overall. The first focuses only on the selected sample of cases that proceed to trial, which may not be representative of the overall distribution. The second includes the fact that black defendants are less likely to accept plea bargains, which mechanically result in conviction.

Figure 4: Racial Gaps by Charge Category



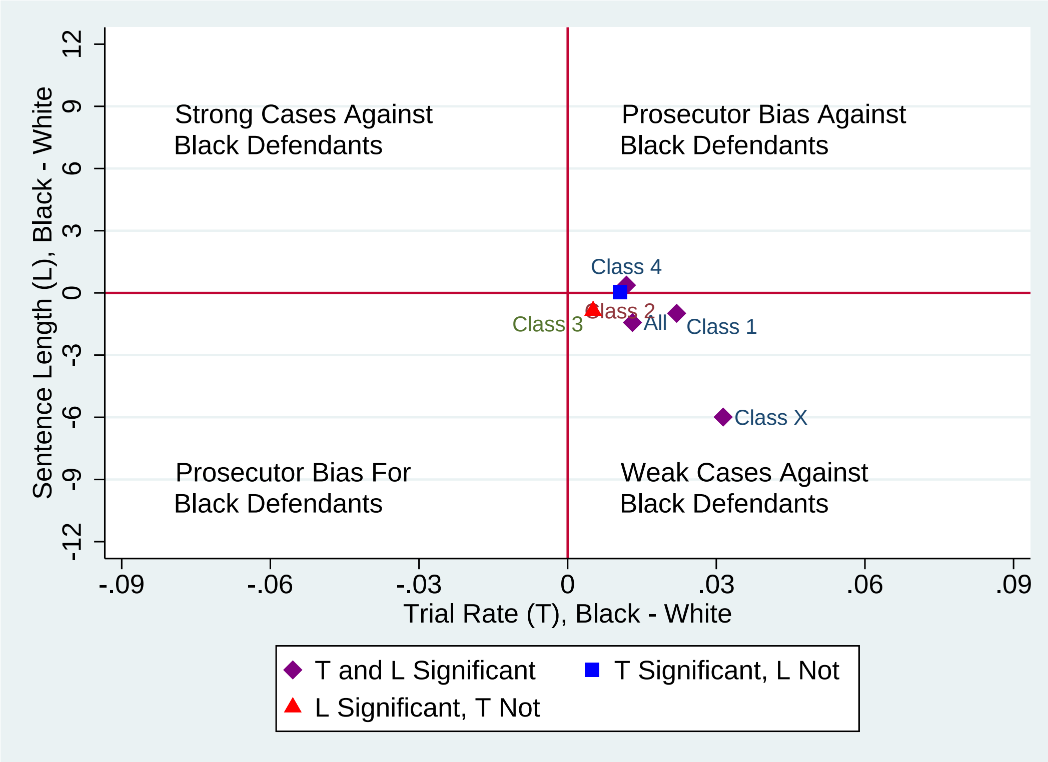
Notes: This figure plots the coefficients from regressions of an indicator for black defendants on whether the case ended in a trial (x-axis) and nominal sentence in months (y-axis). Each regression has covariates: dummy for quarter of case initiation, class of most seriouscharge, age, and number of prior convictions, plus indicators for whether the defendant is male, ever had a public defender, was ever held in custody awaiting trial, and whether the state filed multiple charges or included multiple defendants in the case. Standard errors are clustered at the courtroom-year level. See Section 5 for details about sample selection.

Heterogeneity

The aggregate results above potentially mask substantial heterogeneity across defendant subgroups. Figure 4 presents the results of several crime-category–specific regressions in the same $(\Delta T, \Delta L)$ space used in Figure 1. Purple diamonds are categories with significant effects on both dimensions. Blue squares are significant on only the plea deal acceptance dimension. Red triangles are categories significant only on the sentence length dimension. Yellow circles are significant in neither dimension. My data cluster around the bottom-right, but only one individual crime category shows significant evidence of weaker cases against black defendants: drug offenses.

Figure 5 divides the data by felony class instead of charge category. The results are split between the bottom-right and top-right quadrants. The sentencing gap within Class X is particularly large, which is not surprising given the fact that Class X felonies carry much longer sentences with no option of probation. Figure 6 divides the data by the number of prior convictions each defendant has. I find results that are in

Figure 5: Racial Gaps by Felony Class



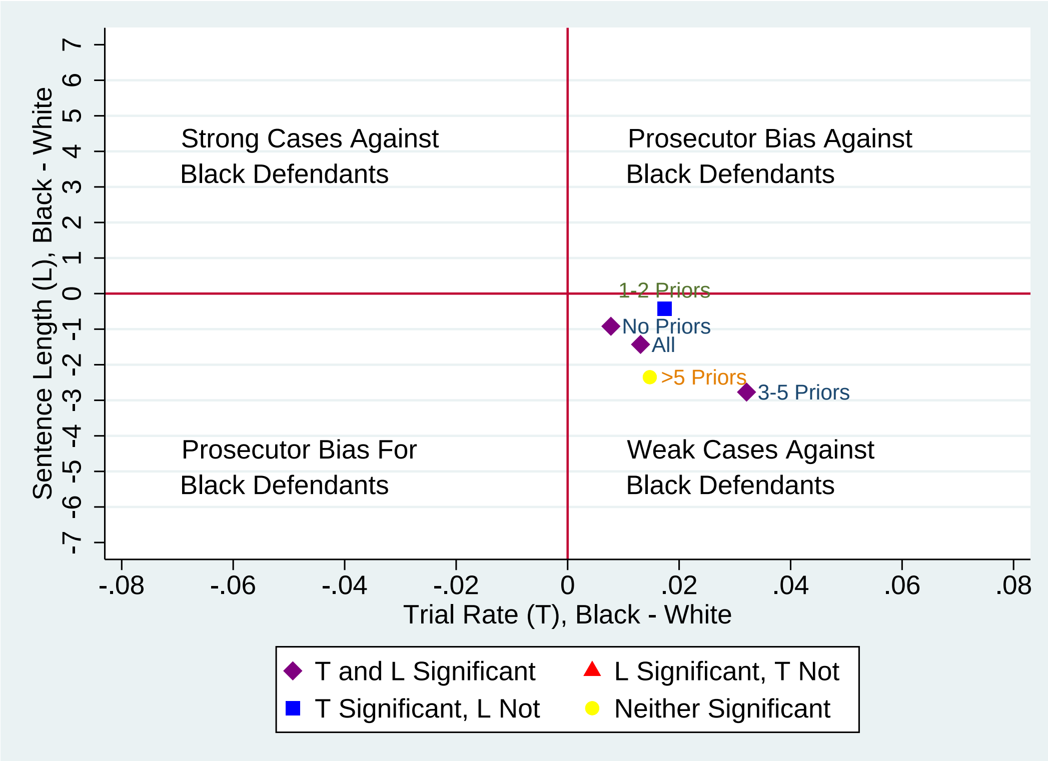
Notes: This figure plots the coefficients from regressions of an indicator for black defendants on whether the case ended in a trial (x-axis) and nominal sentence in months (y-axis), conditional on class of most serious charge. Each regression has covariates: dummy for quarter of case initiation, category of most serious charge, age, and number of prior convictions, plus indicators for whether the defendant is male, ever had a public defender, or was ever held in custody awaiting trial, and whether the state filed multiple charges or included multiple defendants in the case. Standard errors are clustered at the courtroom-year level. See Section 5 for details about sample selection.

the southeast quadrant and stronger for defendants who have many prior convictions.

Recall from my theoretical results in Section 4 that a finding of one type of racial disparity does not preclude the existence of the other. Even uniform and overwhelming evidence that black defendants face weaker cases would not rule out the possibility of prosecutor bias. When the two states coexist, regression evidence can reveal at most one. It is worth noting that not a single data point in Figures 4, 5, or 6 suggests that black defendants face stronger cases or are favored by prosecutors. Furthermore, some subsamples, such as robbery and Class 4 felonies, are in the region of $\Delta T, \Delta L$ space that implies prosecutor bias against black defendants.[20] I interpret this to mean that black defendants are probably subject to both weak sentences and prosecutor bias. However, the evidence for the former is stronger and more uniform. Furthermore, I do not have a clear theory as to why evidence of prosecutor bias would be concentrated in robberies and Class 4 felonies. This incomplete evidence partially motivates the structural estimation of my model in Section 8,

[20]These results remain statistically significant even after applying a Bonferroni correction for multiple hypothesis testing.

Figure 6: Racial Gaps by Prior Convictions



Notes: This figure plots the coefficients from regressions of an indicator for black defendants on whether the case ended in a trial (x-axis) and nominal sentence in months (y-axis), conditional on number of prior convictions. Each regression has covariates: dummy for quarter of case initiation, class of most serious charge, category of most serious charge, and age, plus indicators for whether the defendant is male, ever had a public defender, or was ever held in custody awaiting trial, and whether the state filed multiple charges or included multiple defendants in the case. Standard errors are clustered at the courtroom-year level. See Section 5 for details about sample selection.

which has the potential to reveal *both* prosecutor bias and differences in θ.

Figure D1 presents heterogeneity by year. Both ΔT and ΔL are more volatile in early sample years. Defendants in later years are more likely to have fully observed felony conviction records, so this may simply be noise driven by smaller sample sizes. After 2000, black defendants consistently have shorter sentences and more trials. Overall, I conclude that the differences documented in the aggregate regressions were likely driven by persistent forces rather than extreme factors confined to just one era.

Robustness

My primary sentence length outcome is only one way of interpreting the sentence given at court. Table 6 recreates the last column Table 3 using alternative measures of sentence length. The first calculates the marginal sentence length by subtracting any time served credits awarded at sentencing. This is the amount of time the defendant will actually spend in prison. The second calculates the total sentence length by

Table 6: Alternative Sentence Measures

	Marginal	Total
Black	-1.364***	-1.495***
	(0.173)	(0.198)
Male	4.422***	5.490***
	(0.142)	(0.165)
Public Defender	1.870***	2.914***
	(0.154)	(0.177)
Ever in Jail	8.341***	11.09***
	(0.177)	(0.191)
Multiple Defendants	1.196***	2.660***
	(0.203)	(0.227)
Multiple Charges	6.338***	8.054***
	(0.158)	(0.173)
Observations	342616	342616
Adjusted R^2	0.227	0.266
Charge Cond.	ClassXCat FE	ClassXCat FE
Assignment Cond.	CtrmXYear FE	CtrmXYear FE

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents the results of regressions of an indicator for black defendants on alternative sentence measures. Covariates also include dummy variables for age and number of prior convictions. Standard errors are clustered at the courtroom-year level. See Section 5 for details about sample selection.

including any time spent in jail even for defendants who were not eventually given an incarceration sentence. This is the amount of time the defendant spends incarcerated in any way, even if it is never labeled by the state as punishment. My measure of the sentencing gap does not react much to the choice of sentencing measure.[21]

My data include some serious crimes and cases with many charges, which can sometimes lead to extremely long nominal sentence lengths. I topcode sentence lengths at 80 years because this is a practical limit on how long someone can actually be incarcerated. Table D2 shows my primary sentencing result for different choices about topcoding, including dropping observations with unusually high sentence lengths. My estimates of the sentencing gap are robust to this choice.

7 Alternative Explanations

Differences in Other Parameters

Propositions 4.1 and 4.2 depend on holding constant c, S, and $G(\rho)$. I hold prosecutor trial costs c constant only for clarity. They are closely related to the bias term k_j, and my comparative statics could easily be rewritten in terms of the prosecutor's perceived trial cost, c/k_j. Thus, my model will never be able to distinguish racial patterns in prosecutor bias from racial patterns in prosecutor trial costs. Importantly, these would have to be *real* differences in the cost to try black and nonblack defendants. If prosecutors simply *perceive* a different cost to trying black defendants, than this is just a different way for the prosecutor's attitudes about race to enter the plea bargaining process. The clearest pattern in the data suggests that black and nonblack defendants differ in their θ distributions. This finding does not depend on the distinction between prosecutor bias and prosecutor trial cost.

I also hold constant S, the punishment the defendant would face if she lost at trial. It is difficult to directly observe S because so few cases go to trial, but it is still possible to condition on S using information about the charges brought against the defendant. Illinois has determinate criminal sentencing. Its criminal code lays out specific sentencing rules for each charge that judges are obligated to follow. Judges retain some discretion, but they have far less discretion than judges and parole boards in indeterminate sentencing states. As with c, even if S did vary by race, its comparative statics are such that racial differences in S could not explain the pattern of sentence lengths and trial rates observed in the data.

In Appendix A, I discuss a variation on my model that allows the risk tolerance distribution, $G(\rho)$, to

[21] I also considered a measure that treats a sentence to probation as a potential prison sentence that may or may not be realized depending on whether probation is violated. Thus, it measures a defendant who is sentenced to probation and later sent to prison for 2 years for a probation violation as having been given a 2-year sentence at trial. Under this measure, the sentencing gap shrinks to 0.9 months, reflecting the fact that white defendants are disproportionately likely to be sentenced to probation.

vary by race. I find that a group with higher overall risk tolerance will generate the same patterns in ΔT and ΔL as a group with lower values of θ. This could be an alternative explanation for the patterns I observe in the data. I am not aware of any study that quantifies the relative risk tolerance of black and nonblack defendants in the criminal justice context. However, empirical studies of risk attitudes in insurance and financial settings have not yielded clear evidence that African Americans are more tolerant of risk than their white counterparts (Halek and Eisenhauer 2001; Yao, Gutter, and Hanna 2005). It is possible that the process of selection into the criminal justice system could distort these characteristics, but this selection is likely to begin with people who are more tolerant of risk before moving on to people with less extreme risk attitudes. The fact that relatively more black Chicagoans end up in court will tend to moderate the risk tolerance of black defendants as compared to nonblack defendants.

Trial Rate as an Inferior Good

Proposition 3.4 depends on the assumption that $\theta \geq e^{-\rho^*}$. If this assumption fails, the marginal defendant may be sufficiently inelastic that the prosecutor responds to an increase in θ by increasing s^* so much that his trial rate also increases. In data where $\theta < e^{-\rho^*}$ for many cases, differences in θ may generate data that suggest differences in k instead. My results in Section 6 provide only limited evidence for differences in k, so this ambiguity is not an immediate concern. However, I test the condition to assess whether my theoretical results may be useful in other settings. Using the definition $\rho^* = \hat{\rho}(s^*)$, I can rewrite the condition as $\frac{S}{s^*} < e$. Figure D2 displays a histogram of $\frac{S}{s^*}$ for all non-trial cases in my estimation sample that resulted in an incarceration sentence.[22] Only 13.4% of these cases have a value of $\frac{S}{s^*} > e$, and I suspect that this is driven by imputation error for S.[23] To test robustness, Table D3 replicates my core results for a sample that excludes any case with a measured $\frac{S}{s^*} > e$. The results are very similar to the last columns in Tables 3 and 4.

Prosecutor Selection of Cases

In my definition of the defendant θ distributions, $F(\theta|b)$ and $F(\theta|nb)$, I was careful to note that these distributions are conditional on reaching the plea bargaining stage. Cases that do not meet the criterion

[22]For any case in real data, either S or s^* is counterfactual. A defendant cannot both accept a plea bargain and receive a trial sentence. However, S is a feature of the environment that is not dependent on any choices and closely tied to the charges brought against the defendant. Two cases with the same charge characteristics likely have close to the same value for S. Therefore, I impute S within felony class as the median sentence among cases lost at trial. I try several other specifications, including the mean of observed S, joint conditioning on felony class and category, and predicted values from a regression of observed S on covariates. These choices do not have a significant impact on my results.

[23]I find that 22.7% of cases have a value of $\frac{S}{s^*} < 1$. It is unlikely that defendants in my model, or any model where S is known before the trial, would ever accept a plea bargain for a longer sentence than what they would risk at trial. Likewise, of the cases for which $\frac{S}{s^*} > e$, more than half are extreme in that $\frac{S}{s^*} > 6$. These facts suggest that my imputation method for S does not work well for some cells in my data and creates a spread both below 1 and above e.

$\theta S k_j - c > 0$ are dropped before this stage. This drop decision depends on k_j, so prosecutor bias has an opportunity to shape defendant θ distributions. What appears to be variation in case strength could instead be prosecutor bias operating through a different vector.[24] Table D4 shows that race seems to matter in the prosecutor's decision to drop a case. However, I can bound the importance of this for my core results by considering a sample that includes all dropped cases as if they were plea bargains with a sentence of 0. Table D5 recreates the last columns of Tables 3 and 4 under this new sample, and my results change very little.

8 Structural Estimation

In this section, I use my Cook County court data to structurally estimate the parameters of the model described in Section 3. I do this with the Simulated Method of Moments (McFadden 1989), following the suggestions in Eisenhauer et al (2015). I have three goals in this estimation. First, my regression-based tests are not able to simultaneously detect differences in k and θ, but this is possible with structural estimation. Second, with estimates of k and θ in hand, I can determine how each contributes to observed differences in outcomes. Third, I can consider counterfactual scenarios.

Simulated Data Structure

To better capture underlying heterogeneity in θ and S, I simulate the assignment of defendants to judges in my data and allow judges to affect the probability of conviction and trial sentence of each case they hear.[25] I restrict attention to the 134 judges assigned at least 50 cases in my analysis dataset. In each simulated dataset, I assign defendants to judges in a way that matches the proportion of cases with defendants of each race assigned to each judge in the data.

The prosecutor bias parameters k_b and k_{nb} are the same for all defendants and judges in my simulated data. c only ever appears in a ratio with k_j, so I normalize it to 1. Each defendant draws a risk tolerance value ρ where $\rho = 1 + \tilde{\rho}$ and $\tilde{\rho}$ is drawn from $Gamma\left(\alpha_\rho, \beta_\rho\right)$ with $\alpha_\rho > 1$. This ensures the risk tolerance distribution will have the properties I assume about $G\left(\rho\right)$ in Section 3. Each defendant also draws a value S_{def} from a log-normal distribution with mean μ_S and standard deviation σ_S. Finally, defendants draw a case strength value θ_{def} from $U\left(\underline{\theta}_j, 1\right)$. Notice that if $\underline{\theta}_b < \underline{\theta}_{nb}$, this distribution for nonblack defendants

[24] Call the unconditional defendant θ distributions $\hat{F}\left(\theta|B\right)$ and $\hat{F}\left(\theta|W\right)$ and assume that $\hat{F}\left(\theta|B\right) = \hat{F}\left(\theta|W\right)$. If the prosecutor is biased against black defendants, he will be more willing to retain their cases rather than drop them. This will skew the defendant θ distributions such that $F\left(\theta|W\right) >_{ST} F\left(\theta|B\right)$. For this reason, Propositions 4.1 and 4.2 are not guaranteed to hold with $\hat{F}\left(\theta|B\right)$ and $\hat{F}\left(\theta|W\right)$ in the place of $F\left(\theta|B\right)$ and $F\left(\theta|W\right)$.

[25] I use judge assignment here rather than year-courtroom assignment because many year-coutroom cells are too small to ensure consistent calculation of moments. This, combined with the fact that I cannot directly observe prosecutor assignment, means I will focus only on heterogeneity generated by judge assignment and continue to treat my data as if they were assigned to a single prosecutor with a single value of k_b and k_{nb}.

will stochastically dominate the distribution for black defendants, and vice-versa.

Meanwhile, each judge draws two modifiers. First, a sentencing modifier, S_{jdg}, from a normal distribution with mean 1 and standard deviation ϵ_S, bounded below at 0. Second, a probability of conviction modifier, θ_{jdg}, from $Beta\left(\alpha_\theta, \beta_\theta\right)$. I calculate the final trial sentence as $S = S_{def}S_{jdg}$. The modal judge does not alter the defendant's sentencing draw, and more or less severe judges do so in a multiplicative way. Likewise, I calculate the final probability of conviction as $\theta = \theta_{def}\theta_{jdg}$. A judge with the maximum possible θ_{jdg} of 1 is the most severe in the sense that they give no "discount" to the defendant, and judges less inclined to convict give larger discounts. This multiplicative strategy for incorporating judge effects ensures that each simulated value of S is positive and each final value of θ is in the unit interval without having to impose artificial bounds.[26] Thus, my simulation procedure requires a vector of 11 parameters: $\{k_b, k_{nb}, \alpha_\rho, \beta_\rho, \underline{\theta}_b, \underline{\theta}_{nb}, \mu_S, \sigma_S, \epsilon_S, \alpha_\theta, \beta_\theta\}$.

I follow Eisenhauer et al (2015) to create my simulated data. Given a parameter vector, I simulate 20,000 nonblack defendants and 80,000 black defendants and calculate the moments of interest within that population. I repeat this procedure 20 times and capture the mean of my moments across simulations as a vector M_s.

Data and Moments

In constructing the data moments, I first residualize my data with respect to the dimensions of heterogeneity not explicitly considered in the simulation. This is all of the covarites included in the regressions in Section 6 except race and assigned-courtroom fixed effects. This means that, aside from the variation accounted for by judge assignment, my estimation procedure is attempting to simulate a pool of defendants with the average cross-race differences estimated in the regressions in Section 6.

I estimate 8 moments separately for black and nonblack defendants, for a total of 16. These are:

- Percentage of cases dropped
- Percentage of non-dropped cases that go to trial
- Percentage of cases that go to trial that result in a conviction
- Mean sentence in non-dropped cases
- Standard deviation of sentences in non-dropped cases
- Standard deviation of drop rate across judges

[26] In principle, the lower bound on the judge sentencing modifier distribution at 0 is artificial, but ϵ_S is always low enough that it does not actually apply.

Table 7: Structural Fit

Panel A	Black			Nonblack		
	Moment	Data	Model	Moment	Data	Model
	Drop Rate	0.044 (0.0003)	0.042	Drop Rate	0.057 (0.001)	0.057
	Trial Rate	0.121 (0.001)	0.121	Trial Rate	0.109 (0.001)	0.113
	Mean Sentence	22.5 (0.08)	22.6	Mean Sentence	23.5 (0.13)	23.2
	SD Sentence	37 (0.43)	37.8	SD Sentence	35 (0.71)	37.8
	Trial Conv. Rate	0.576 (0.003)	0.548	Trial Conv. Rate	0.574 (0.005)	0.563
	Judge Drop SD	0.015 (0.001)	0.017	Judge Drop SD	0.027 (0.002)	0.031
	Judge Trial SD	0.04 (0.001)	0.024	Judge Trial SD	0.046 (0.003)	0.043
	Judge Sent. SD	3.33 (0.19)	3.6	Judge Sent. SD	4.73 (0.23)	5.01
Panel B	Parameter	Value	SE	Parameter	Value	SE
	k_b	0.99	0.037	k_{nb}	0.78	0.029
	$\underline{\theta}_b$	0.244	0.018	$\underline{\theta}_{nb}$	0.283	0.019
	α_ρ	3.16	0.001	μ_S	2.69	0.007
	β_ρ	1.3	0.065	σ_S	1.13	0.006
	α_θ	30.9	0.448	ϵ_S	0.12	0.013
	β_θ	1.54	0.001			

Notes: This table displays the results of estimating the model in Section 3 using a Nelder-Mead algorithm to minimize the bootstrap-variance-weighted error in 16 moments between Cook County Court data and synthetic data created by 20 simulations of the model with 20,000 nonblack defendants and 80,000 black defendants. Panel A compares the data moments to the simulated moments. Bootstrap standard errors for the data moments in parentheses. Panel B shows the parameter estimates obtained from the estimation. Standard errors of parameter estimates calculated from numerical derivatives with step length ${}^{\log n}/_{n}$.

- Standard deviation of trial rate across judges
- Standard deviation of sentences across judges

I again follow Eisenhauer et al (2015) to estimate these moments in my data. To begin, I create 200 bootstrap samples from my data and calculate the moments of interest within each. I then save the average across these moments as a vector M_d. I also generate the VCV matrix of the moments across bootstrap replications and save the inverse of this matrix, W.

I use the Nelder-Mead algorithm to estimate my parameters. I minimize an objective that is the weighted difference of simulated and data parameters:

$$(M_s - M_d)\, W\, (M_s - M_d)'$$

Estimation Results

Panel A of Table 7 shows the data moments, the standard deviation of theses moments across bootstrap observations, and my fitted simulated moments. I provide the standard deviations to give a sense of the relative weight my optimization procedure placed on each moment. The standard deviations are quite small because the size of my dataset allows high precision when estimating the data moments. Overall the fit is satisfactory. I closely match empirical rates of dropped cases, rates of cases that proceed to trial, and average sentence length for both races. I slightly overestimate the variance of sentences, both within and across judges. I also underestimate the variance in trial rates across judges considering black defendants. The facts imply that there may be more variation in how specific judges treat (or are perceived by) defendants than my model allows for. I also underestimate trial conviction rates for both black and nonblack defendants and estimate a greater difference between the two than exists in the data.

Panel B of Table 7 lists my estimated parameter values and their standard errors.[27] The parameter estimates confirm the conclusion of Section 6 by setting the lower bound of the black θ distribution, $\underline{\theta}_b$, at 0.244, below $\underline{\theta}_{nb}$ at 0.283.[28] This means black defendants draw from an overall weaker distribution of cases. Interestingly, my model also finds that black defendants were subject to more prosecutor bias, $k_b = 0.99$, than nonblack defendants, $k_{nb} = 0.878$. The regression results in Section 6 found some evidence of this in particular subgroups, but the structural model was necessary to reveal the disparity in aggregate data.

As expected, I estimate a sentencing distribution for defendants with a long right tail. I also estimate an extreme distribution of $\theta_{jdg}$where most values are close to 1. The standard deviation of θ_{jdg} is 0.037 while the standard deviations of $\underline{\theta}_b$ and $\underline{\theta}_{nb}$ are 0.22 and 0.21, respectively . Thus, most variation in probability of conviction is attached to the defendant rather than the judge. The same is also true of sentence length, as can be seen by comparing $\sigma_s = 1.12$ to $\epsilon_S = 0.12$.

The top panel of Figure 7 presents the prosecutor's plea sentence offers for various values of θ when $S = 36$. The prosecutor's plea sentence policy for black defendants is given by the solid line, and his policy for nonblack defendants by the dotted line. All other parameters are as estimated in Table 7. The offers for nonblack and black defendants with the same θ are nearly identical, though the nonblack line is slightly lower than the black line. This difference in levels reflects the differences in k_j estimated in Table 7. Instead, the most important difference between nonblack and black defendants is that the solid line extends into lower values of θ, reflecting the difference in case strength distributions estimated in Table 7. The bottom panel of Figure 7 presents the trial rate induced by the plea bargain offers in the top panel. Here the line for black defendants is distinctly above the line for nonblack defendants. This indicates that trial rates are quite sensitive to plea bargain offers at the optimum, so prosecutor bias will affect trial rates more than it

[27]To calculate standard errors, I begin by finding numerical derivatives of each moment with respect to each parameter (Judd 1998). I use a stepsize of $n/\log n$ (Hong, Mahajan, and Nekipelov 2015). This procedure yields a Jacobian matrix, D. Following Duffie and Singleton (1993), I estimate the VCV matrix of my parameters as:

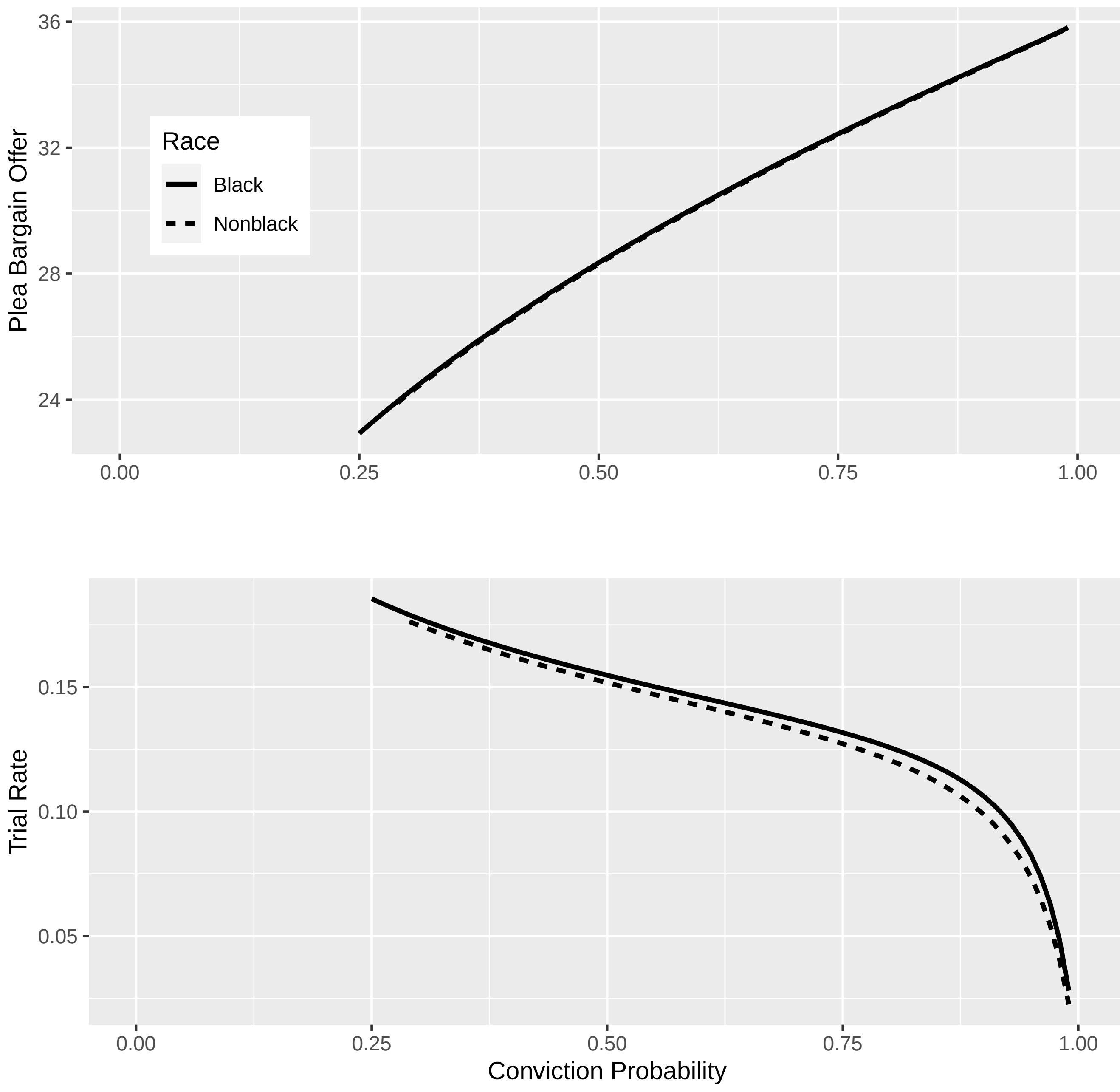
$$Q = \left(1 + \frac{1}{20}\right)\left[D'WD\right]^{-1}$$

The standard errors reported in Panel B of Table 7 are the square root of the diagonal of Q.

[28]A heuristic statistical test of $\underline{\theta}_b - \underline{\theta}_{nb}$ based on their individual standard errors calls significance into question. However, the covariance of $\underline{\theta}_b$ and $\underline{\theta}_{nb}$ in Q is large and positive, so the standard error of $\underline{\theta}_b - \underline{\theta}_{nb}$ is only 0.01.

affects sentences. Again, the black trial rates extend into lower values of θ, increasing the mean trial rate over all black defendants.

Figure 7: Plea Bargain Offer and Trial Rate Functions by Race



Notes: This figure shows the response of selected simulated moments to changes in the race-specific values of k and distributions of θ when $S = 36$. The vertical line denotes the value of the parameter found by my estimation procedure. The horizontal line denotes the value of the moment in the data. The solid curve holds constant all other parameter values at those estimated in Table 7. The dotted and dashed curves set the value of a selected parameter above and below its optimal value, respectively.

Figures D3, D4, D5, and D6 present the response of selected moments to changes in simulation parameters. Each line holds all other parameters constant and varies one parameter in the region around its estimated value. The solid lines set other parameters to their optimal values. The dotted and dashed lines set a related parameter to a value that is lower or higher than its optimal value, respectively. These figures select combinations of parameters and moments where changing the parameter has an obvious and monotonic effect on the simulated moment. This is not a formal proof of identification, but it provides intuition about the connections between particular moments and parameters in the model.

Counterfactual Analysis

Policies that attempt to directly influence or equalize θ distributions or values of k have the appeal of being conceptually simple within my theoretical framework, but they may be prohibitively difficult to implement in a real court system. If the party responsible for enforcing the policy cannot observe θ with the same accuracy as the parties actually involved in the case, they will have difficulty determining if certain cases are particularly weak or certain plea bargains are influenced by bias.

Instead I consider four straightforward policy changes that do not have obvious consequences for racial outcome disparities. The first is a policy that reduces sentencing severity across all crimes. I parameterize this as a 20% reduction in μ_S. The second is a policy that standardizes judge sentencing decisions at a level consistent with the most lenient judges. I model this as fixing S_{jdg} at $1 - 2\epsilon_S = 0.76$. The third is a similar policy that targets probability of conviction rather than sentencing policy. It fixes $\theta_{jdg} = 0.88$, two standard deviations below its mean. Last is a policy that forbids plea bargaining and forces all cases to go to trial. While these policies cannot directly address underlying inequities in the courts, it may still be interesting to know what effect, if any, they have on disparities in outcomes.[29]

Table 8 presents black and nonblack trial rate and sentencing distribution information for the baseline specification and my four counterfactual scenarios. To keep a consistent measure of trial rate and sentencing gaps across policies that change the levels of these measures, I also present the ratio of black to nonblack in each. I find first that reducing the mean of the sentencing distribution, predictably, reduces average realized sentences. It also reduces the rate at which defendants demand trial by lowering the stakes of those trials. It has only a small effect on trial rate and sentencing gaps, increasing (moving away from 1) both. When potential sentences are less extreme, it appears that outcomes for black defendants better reflect the fact that weaker cases are being brought against them.

The counterfactual scenario where judges agree on a lenient sentencing policy looks similar to the policy

[29]It is also worth considering whether it is *desirable* to close an outcome gap, especially the sentencing outcome gap, with such a policy. Black defendants would continue to face weaker cases, but they would no longer be partially compensated for this with shorter expected sentences.

Table 8: Counterfactual Analysis

Moment	Baseline	Red. Sentence	Low S Judges	Low θ Judges	No Pleas
Black Trial Rate	12.1%	10.9	11.5	12.9	100%
Nonblack Trial Rate	11.3%	10.1	10.7	12.2	100%
Ratio	1.071	1.079	1.075	1.057	1
Black Sentence Mean	22.6	14	17.6	22	17.1
Nonblack Sentence Mean	23.2	14.5	18.2	22.7	17.8
Ratio	0.974	0.965	0.967	0.969	0.961

Notes: This table displays the impact of two counterfactual exercises on the trial rates and sentence distribution (in months) of black and nonblack defendants. The Reduced Sentence counterfactual reduces μ_S by 20%. The Low S Judges counterfactual sets all judge sentencing effects to $S_{jdg} = 0.76$. The Low θ Judges counterfactual sets all judge probability of conviction effects to $\theta_{jdg} = 0.88$. The No Pleas counterfactual forces all cases to go to trial.

that reduces μ_S, though not identical because both the mechanism and magnitude are different.[30] The scenario where judges agree on a lenient conviction policy looks substantially different. Sentences fall when all judges adopt the same θ_{jdg} as a particularly lenient colleague, but they fall by much less than in the equivalent sentencing reform. This suggests that the variation in conviction attitudes currently present in the judiciary does not have a large impact on realized sentences. Rather than decrease, as in the sentencing reform, trial rates increase. This is consistent with the discussion in this paper about the consequences of a lower probability of conviction at trial. These counterfactuals emphasize that judges differ in (at least) two dimensions, sentencing leniency and conviction leniency, and these dimensions have different effects on the outcomes of interest. Given that conviction leniency can do less to reduce sentences within the current range of judge attitudes and tends to increase trials, policies emphasizing lower sentences—whether by statute or judicial discretion—will benefit defendants the most.

In the third column of Table 8, I find that eliminating plea bargaining would substantially decrease average sentences. This is consistent with the basic mechanism of plea bargaining, where the defendant insures themselves against an uncertain outcome at the cost of a longer expected sentence. However, the effect is not uniform across black and nonblack defendants. I find that sentences fall relatively more for black defendants in the absence of plea bargaining, thus increasing the sentencing gap. This is because prosecutors in my model can only exercise their bias through plea bargaining. I estimate that $k_b > k_{nb}$, so allowing prosecutor to plea bargain puts relatively more upward pressure on sentences for black defendants.

Note that, despite substantially reducing average sentences, a policy that forbids plea bargaining can never benefit defendants in this model. Each defendant who opts for a plea bargain weakly preferred that bargain to a trial, even if the expected sentence was higher. Defendants with different values of ρ vary in

[30]Because S_{def} is distributed log-normal, an 80% reduction in μ_S functions differently from the judge-based reduction, which is an 76% reduction in the ex-post draws from that log-normal. The former will do more to tame the right tail of the sentencing distribution than the latter. In terms of policy, this could correspond to a sentencing reform package that moderates or eliminates policies that add significant penalties to serious crimes, such as weapon or repeat-offender enhancements.

how much they value plea bargaining, but this counterfactual states that many defendants willing to increase their expected sentence by as much as 6 nominal months in order to avoid a trial. It is important to keep this in mind when considering the sentence length and trial gaps I have documented. Black defendants are partially compensated for the weak cases against them with somewhat lower sentences. However, their increased trial rate works in the opposite direction. It indicates that more black defendants are receiving the undesirable outcome of a trial when a mutually-agreeable plea bargain was available.

9 Conclusion

In this paper I used a model of plea bargaining to understand more about bias in the criminal justice system. My model explores two unobservable reasons why criminal justice outcomes might differ: case strength and prosecutor bias. From the comparative statics of that model, I derived a method for distinguishing between the two using court data. Greater case strength in a population allows prosecutors to secure both longer plea bargain sentences and fewer trials. Bias forces prosecutors to exchange longer sentences for more trials. I derived precise conditions under which this intuition applies to regression-based tests for differences in prosecutor bias and case strength.

I then investigated outcome patterns in felony court data from Chicago. I found that black defendants tend to receive shorter sentences and demand more trials. According to my model, this indicates that they face weaker cases overall. A structural estimation of my model confirmed the conclusion of the regression results and found additionally that black defendants are subject to prosecutor bias, though this has a relatively small effect on their outcomes.

My finding that black defendants face weaker cases, combined with the sample fact that they are far over-represented in the population of felony defendants relative to the population of Cook County, indicates that disparities at court are driven by selection into court. In Jordan (2020), I investigate this hypothesis further using data on felony review, a key step in the selection process for felony defendants. In that paper, I document three facts about the population of potential defendants: (1) The conviction rate disparity between black and nonblack defendants is largest at time of arrest. (2) Felony reviewers are more likely to approve charges against nonblack suspects. (3) Felony reviewers perform worse when classifying cases with black suspects. They are more likely to reject charges that could have resulted in convictions and more likely to approve charges that will not result in a conviction. This may be related to the quality of evidence presented by police in cases with black suspects. The results of Jordan (2020) and this paper motivate further research into the arrest and evidence-gathering practices of police.

References

Albonetti, C. A. 1990 "Race and the Probability of Pleading Guilty." *Journal of Quantitative Criminology* 6(3): 315-334.

Alesina, A., & La Ferrara, E. 2014. "A Test of Racial Bias in Capital Sentencing." *American Economic Review* 104(11): 3397-3433.

Antinovics, K. & Knight, B. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." *The Review of Economics and Statistics* 91(1): 163-177.

Anwar, S. & Fang, H,. 2006. "An Alternative Test of Racial Prejudice." *American Economic Review* 96(1): 127-151.

Arnold, D., Dobbie, W., & Hull, P. 2020. "Measuring Racial Discrimination in Bail Decisions." *NBER Working Paper No. 26999.*

Arnold, D., Dobbie, W. & Yang, C. 2018 "Racial Bias in Bail Decisions." *Quarterly Journal of Economics* 133(4): 1885-1932.

Ayres, I. 2002. "Outcome Tests of Racial Disparities in Police Practices." *Justice Research and Policy* 4(1-2): 131-142.

Bebchuk, L. 1984. "Litigation and Settlement Under Imperfect Information." *The RAND Journal of Economics* 15(3): 404-415.

Becker, G. 1957. The Economics of Discrimination. The University of Chicago Press.

Bronson, J. & Carson, A.E. 2019. "Prisoners 2017". Bureau of Justice Statistics publication.

Daughety, A., & Reinganum, J. 2017. "Settlement and Trial." in The Oxford Handbook of Law and Economics: Volume 3: Public Law and Legal Institutions. Oxford University Press.

Duffie, D. & Singleton, K. 1993. "Simulated Moments Estimation of Markov Models of Asset Prices." *Econometrica.* 61(4): 929-952.

Eisenhauer, P., Heckman, J., & Mosso, S. 2015 "Estimation of Dynamic Discrete Choice Models by Maximum Likelihood and the Simulated Method of Moments." *International Economic Review.* 56(2): 331-357.

Frenzel, E.D. & Ball, J.D. 2008. "Effects of Individual Characteristics on Plea Negotiations Under Sentencing Guidelines." *Journal of Ethnicity in Criminal Justice* 5(4): 59-82.

Grossman, G., & Katz, M. 1983. "Plea Bargaining and Social Welfare." *The American Economic Review* 73(4): 749-757.

Halek, M. & Eisenhauer, J. 2001. "Demography of Risk Aversion." *The Journal of Risk and Insurance* 68(1): 1-24.

Hong, H., Mahajan A., and Nekipelov D. 2015. "Extremum Estimation and Numerical Derivatives." *Journal* *of Econometrics* 188: 250-263.

Jordan, A. 2020. "Racial Patterns in Approval of Felony Charges." Working Paper.

Judd, K. 1998. Numerical Methods in Economics. MIT Press.

Kim, J. Y. 2010. "Credible Plea Bargaining." *European Journal of Law and Economics* 29(3): 279-293.

Knowles, J., Persico, N., & Todd, P. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109(1): 203-229.

Landes, W. M. 1971. "An Economic Analysis of the Courts." *Journal of Law Economics* 14(1): 61-108.

Lee, S. 2014. "Plea Bargaining: on the Selection of Jury Trials." *Economic Theory* 57(1): 59-88.

McFadden, D. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica* 57(5): 995-1026.

Metcalfe, C., & Chiricos, T. 2018. "Race, Plea, and Charge Reduction: An Assessment of Racial Disparities in the Plea Process." *Justice Quarterly* 35(2): 223-253.

Neal, D., & Rick, A. 2016. "The Prison Boom and Sentencing Policy." *The Journal of Legal Studies* 45(1): 1-41.

Prentice, R. 1974. "A Log Gamma Model and It's Maximum Likelihood Estimation." *Biometrika* 61 (3): 539-544.

Priest, G. L., & Klein, B. 1984. "The Selection of Disputes for Litigation." *The Journal of Legal Studies* 13(1): 1-55.

Rehavi, M. M., & Starr, S. B. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy* 122(6): 1320-1354.

Reinganum, J. F. 1988. "Plea Bargaining and Prosecutorial Discretion." *The American Economic Review* 78(4): 713-728.

Silveira, B. S. 2017. "Bargaining with Asymmetric Information: An Empirical Study of Plea Negotiations." *Econometrica* 85(2): 419-452.

Simoiu, C., Corbett-Davies, S., & Goel, S. 2017. "The Problem of Infra-Marginality in Outcome Tests for Discrimination." *The Annals of Applied Statistics* 11(3): 1193-1216.

Sutton, J. R. 2013. "Structural Bias in the Sentencing of Felony Defendants." *Social Science Research* 42(5): 1207-1221.

United States Department of Justice. 2017 "Investigation of the Chicago Police Department."

Various Authors. 2017. "Defending Illinois Criminal Cases: Pretrial Issues, Guilty Pleas, and Defenses." Illinois Institute of Continuing Legal Education (IICLE).

Yao, R., Gutter, M., & Hanna, S. 2005. "The Financial Tolerance of Blacks, Hispanics, and Whites." *Financial Counseling and Planning* 16(1): 51-62.

Appendix A: Mathematical Analysis

Proof of Proposition 3.1

Using the fact that $\delta^* = 0$, I can rewrite the defendant's problem as:

$$\min_{\alpha\in\{0,1\}} (1-\alpha)\left(\theta\frac{S^\rho}{\rho}\right) + \alpha\frac{s^\rho}{\rho}$$

Her objective is maximized by the following policy:

$$\alpha^* = \begin{cases} 0 & s^\rho > \theta S^\rho \\ 1 & s^\rho \leq \theta S^\rho \end{cases}$$

Rearranging to express this choice as a direct function of the defendant's risk tolerance type, I find the threshold value of:

$$\hat{\rho}(s) = \frac{\ln\theta}{\ln s - \ln S}$$

θ is a probability that lies in $(0,1)$, so $\ln\theta < 0$ and $\ln s - \ln S < 0$ by assumption. Therefore, $\hat{\rho}(s) > 0$ and $\hat{\rho}(s)$ is strictly increasing in s. Therefore, $\rho < \hat{\rho}(s)$ is equivalent to $s^\rho > \theta S^\rho$ and leads the defendant to reject the plea deal. $\rho \geq \hat{\rho}(s)$ is equivalent to $s^\rho \leq \theta S^\rho$ and leads the defendant to accept the plea deal.

Proof of Proposition 3.2

First, suppose the prosecutor chooses an $\tilde{s} > S$. The defendant will never accept this plea bargain, so the prosecutor will always receive $\theta k_j S - c$. However, for any $\hat{s} \in (\theta S, S)$, the prosecutor's objective evaluated at $\hat{s}$ is a convex combination of $\theta k_j S - c$ and $k_j\hat{s} > \theta k_j S - c$. Therefore, a choice of $\tilde{s} > S$ is not optimal for the prosecutor.

Second, suppose the prosecutor chooses an $\tilde{s} < \theta S$. Note that $\hat{\rho}(\theta S) = 1$ and $\hat{\rho}(\tilde{s}) < 1$, so $G(\hat{\rho}(\theta S)) = G(\hat{\rho}(\tilde{s})) = 0$. Therefore, the value of the prosecutor's objective when he offers $\tilde{s}$ is $k_j\tilde{s}$, but if he were to offer θS instead, the value of his objective would be $k_j\theta S > k_j\tilde{s}$. Therefore, offering $\tilde{s} < \theta S$ can never be optimal.

Using the fact that $\delta^* = 1$ and incorporating the defendant's cutoff strategy, I can rewrite the prosecutor's choice of plea bargain when $s < S$ as:

$$\max_{s\in(\theta S,S)} G(\hat{\rho}(s))(\theta k_j S - c) + (1 - G(\hat{\rho}(s)))\, k_j s$$

The FOC of this optimization problem defines an optimal choice s^* and the critical ρ induced by this choice $\rho^* = \hat{\rho}(s^*)$:

$$(1 - G(\rho^*))\, k_j - g(\rho^*) \frac{\partial \hat{\rho}(s)}{\partial s}|_{s=s^*} k_j s^* + g(\rho^*) \frac{\partial \hat{\rho}(s)}{\partial s}|_{s=s^*} (\theta k_j S - c) = 0$$

Where $\frac{\partial \hat{\rho}(s)}{\partial s}|_{s=s^*} = \frac{-\ln\theta}{(\ln(s^*) - \ln S)^2 s^*} = \frac{-\rho^*}{(\ln(s^*) - \ln S) s^*}$. Moving the second two terms to the right of 0 and cross-multiplying yields:

$$\frac{s^*(\ln S - \ln(s^*))}{s^* - \theta S + c/k_j} = \frac{g(\rho^*)}{1 - G(\rho^*)}\rho^*$$

Notice that the RHS of this expression is non-negative for all ρ^* and increasing in ρ^*. Because ρ^* is increasing in s^*, the RHS is increasing in s^*.

Next, observe that:

$$\begin{aligned} \frac{\partial LHS}{\partial s^*} &= \frac{(s^* - \theta S + c/k_j)(\ln S - \ln(s^*) - 1) - s^*(\ln S - \ln(s^*))}{(s^* - \theta S + c/k_j)^2} \\ &= \frac{(c/k_j - \theta S)(\ln S - \ln(s^*)) - (s - \theta S + c/k_J)}{(s^* - \theta S + c/k_j)^2} \end{aligned}$$

Recall that due to the prosecutor's ability to drop the case, $\theta S k_j - c > 0$, which implies $c/k_j - \theta S < 0$. Therefore, the LHS is is strictly decreasing in s^*.

When $s^* = \theta S$, the RHS smoothly approaches 0, and the LHS is positive. Furthermore when $s^* = S$, the LHS is 0. In conclusion, the LHS begins above the RHS and descends monotonically to 0 while the RHS rises monotonically from 0. Therefore, the two must intersect once and only once in the interval $(\theta S, S)$ at the prosecutor's unique, optimal choice of plea bargain sentence.

Sentencing Comparative Statics (Proofs of Propositions 3.3 and 3.4)

When calculating effects on s^* and ρ^* it will be convenient to explicitly write the LHS and RHS each in terms of exclusively ρ^* or s^*:

$$LHS_s = \frac{s^*(\ln S - \ln(s^*))}{(s^* - \theta S + c/k_j)}$$

$$RHS_s = \frac{g\left(\frac{\ln\theta}{\ln(s^*) - \ln S}\right)}{1 - G\left(\frac{\ln\theta}{\ln(s^*) - \ln S}\right)} \frac{\ln\theta}{\ln(s^*) - \ln S}$$

$$LHS_\rho = \frac{-\theta^{1/\rho^*} S^{\ln\theta}/\rho*}{\left(\theta^{1/\rho^*} S - \theta S + {}^{c}/_{k_j}\right)}$$

$$RHS_\rho = \frac{g\left(\rho^*\right)}{1 - G\left(\rho^*\right)}\rho^*$$

The main text establishes that $\frac{\partial LHS_s}{\partial s^*} < 0$. Observe that:

$$\left(\theta^{1/\rho^*} S - \theta S + {}^{c}/_{k_j}\right)^2 \frac{\partial LHS_\rho}{\partial \rho^*} = \left(\theta^{1/\rho^*} S - \theta S + {}^{c}/_{k_j}\right)\left(\theta^{1/\rho^*} S \frac{\left(\ln\theta\right)^2}{\left(\rho^*\right)^3} + \theta^{1/\rho^*} S \frac{\ln\theta}{\left(\rho^*\right)^2}\right)$$
$$+ \theta^{1/\rho^*} S \frac{\left(\ln\theta\right)^2}{\left(\rho^*\right)^3} \theta^{1/\rho^*} S$$

which simplifies to:

$$\theta^{1/\rho^*} S \frac{\ln\theta}{\left(\rho^*\right)^2}\left[\left({}^{c}/_{k_j} - \theta S\right)\frac{\ln\theta}{\rho^*} + {}^{c}/_{k_j} - \theta S + \theta^{1/\rho^*} S\right] < 0$$

Using the facts that ${}^{c}/_{k_j} - \theta S + \theta^{1/\rho^*} S > 0$ from the FOC and that ${}^{c}/_{k_j} - \theta S < 0$ from the condition on cases the prosecutor does not drop.

$\hat{\rho}\left(s\right)$

First I establish basic facts about the function that defines the cutoff value of ρ. Note that $\theta \in (0, 1)$, so $\ln\theta < 0$, and recall that $s^* < S$.

$$\frac{\partial\hat{\rho}\left(s\right)}{\partial s} = \frac{-\ln\theta}{\left(\ln\left(s\right) - \ln S\right)^2}\frac{1}{s} > 0$$

$$\frac{\partial\hat{\rho}\left(s\right)}{\partial S} = \frac{\ln\theta}{\left(\ln\left(s\right) - \ln S\right)^2}\frac{1}{S} < 0$$

$$\frac{\partial\hat{\rho}\left(s\right)}{\partial\theta} = \frac{1}{\left(\ln\left(s\right) - \ln S\right)}\frac{1}{\theta} < 0$$

Relative Prosecutor Trial Cost ${}^{c}/_{k_j}$

Taking derivatives

$$\frac{\partial LHS_s}{\partial\, {}^{c}/_{k_j}} = \frac{s^*\left(\ln\left(s^*\right) - \ln S\right)}{\left(s^* - \theta S + {}^{c}/_{k_j}\right)^2} < 0$$

$$\frac{\partial RHS_s}{\partial c/k_j} = 0$$

$$\frac{\partial LHS_\rho}{\partial c/k_j} = \frac{\theta^{1/\rho^*} S^{\ln\theta/\rho^*}}{\left(\theta^{1/\rho^*}(S+\gamma) - \theta S + c/k_j\right)^2} < 0$$

$$\frac{\partial RHS_\rho}{\partial c/k_j} = 0$$

Then by the implicit function theorem, noting that the denominators are negative:

$$\frac{\partial s^*}{\partial c/k_j} = \frac{-\left(\frac{\partial LHS_s}{\partial c/k_j} - \frac{\partial RHS_s}{\partial c/k_j}\right)}{\left(\frac{\partial LHS_s}{\partial s^*} - \frac{\partial RHS_s}{\partial s^*}\right)} < 0$$

$$\frac{\partial \rho^*}{\partial c/k_j} = \frac{-\left(\frac{\partial LHS_\rho}{\partial c/k_j} - \frac{\partial RHS_\rho}{\partial c/k_j}\right)}{\left(\frac{\partial LHS_\rho}{\partial \rho^*} - \frac{\partial RHS_\rho}{\partial \rho^*}\right)} < 0$$

A clear consequence of this is that $\frac{\partial s^*}{\partial k_j} > 0$ and $\frac{\partial \rho^*}{\partial k_j} > 0$.

Probability of Conviction θ

Taking derivatives

$$\frac{\partial LHS_s}{\partial \theta} = \frac{s^*\left(\ln S - \ln(s^*)\right) S}{\left(s^* - \theta S + c/k_j\right)^2} > 0$$

$$\frac{\partial RHS_s}{\partial \theta} = \frac{\partial RHS_\rho}{\partial \rho^*}\frac{\partial \hat{\rho}(s)}{\partial \theta} < 0$$

$$\frac{\partial LHS_\rho}{\partial \theta} = \frac{\theta^{1/\rho^*}\frac{S}{\rho^*}\left[\left(\theta S - c/k_j\right)\frac{1}{\theta}\left(1 + \frac{\ln\theta}{\rho^*}\right) - \theta^{\frac{1-\rho^*}{\rho^*}} S - S\ln\theta\right]}{\left(\theta^{1/\rho^*} S - \theta S + c/k_j\right)^2} < 0$$

When $\theta \geq e^{-\rho^*}$. To see this, change variables to $\theta = e^{-y}$ and call the expression in the brackets in the numerator $\Phi(y)$:

$$\Phi(y) = S\left[y - e^{y\left(1-\frac{1}{\rho^*}\right)} + 1 - \frac{y}{\rho^*}\right] - \frac{c}{k_j}\frac{e^y}{\rho^*}\left(\rho^* - y\right)$$

Notice that the last term is non-positive for any $y \leq \rho^*$ (which is analogous to $\theta \geq e^{-\rho^*}$). The first term is 0 when evaluated at $y = 0$, and its derivative with respect to y is:

$$\left(1 - e^{y\left(1-\frac{1}{\rho^*}\right)}\right)\left(1 - \frac{1}{\rho^*}\right) < 0$$

Therefore, $\Phi(0) < 0$, and increasing y from 0 cannot turn $\Phi(y)$ positive, at least so long as $y \leq \rho^*$. Finally, the sign of $\frac{\partial LHS_\rho}{\partial \theta}|_{\theta=e^{-y}}$ is determined by $\Phi(y)$.

$$\frac{\partial RHS_\rho}{\partial \theta} = 0$$

Then by the implicit function theorem, noting that the denominators are negative:

$$\frac{\partial s^*}{\partial \theta} = \frac{-\left(\frac{\partial LHS_s}{\partial \theta} - \frac{\partial RHS_s}{\partial \theta}\right)}{\left(\frac{\partial LHS_s}{\partial s^*} - \frac{\partial RHS_s}{\partial s^*}\right)} > 0$$

$$\frac{\partial \rho^*}{\partial \theta} = \frac{-\left(\frac{\partial LHS_\rho}{\partial \theta} - \frac{\partial RHS_\rho}{\partial \theta}\right)}{\left(\frac{\partial LHS_\rho}{\partial \rho^*} - \frac{\partial RHS_\rho}{\partial \rho^*}\right)} < 0$$

Statutory Sentence S

Taking derivatives

$$\frac{\partial LHS_s}{\partial S} = \frac{\left(s^* - \theta S + {}^{c}/_{k_j}\right)\frac{s^*}{S} + s^*\left(\ln S - \ln\left(s^*\right)\right)\theta}{\left(s^* - \theta S + {}^{c}/_{k_j}\right)^2} > 0$$

$$\frac{\partial RHS_s}{\partial S} = \frac{\partial RHS_\rho}{\partial \rho^*}\frac{\partial \hat{\rho}(s)}{\partial S} < 0$$

$$\frac{\partial LHS_\rho}{\partial S} = \frac{-\left(\theta^{1/\rho^*} S - \theta S + {}^{c}/_{k_j}\right)\theta^{1/\rho^*}\ln\theta/\rho^* + \theta^{1/\rho^*}\ln\theta/\rho^* S\left(\theta^{1/\rho^*} - \theta\right)}{\left(\theta^{1/\rho^*} S - \theta S + {}^{c}/_{k_j}\right)^2} > 0$$

$$\frac{\partial RHS_\rho}{\partial S} = 0$$

Then by the implicit function theorem, noting that the denominators are negative:

$$\frac{\partial s^*}{\partial S} = \frac{-\left(\frac{\partial LHS_s}{\partial S} - \frac{\partial RHS_s}{\partial S}\right)}{\left(\frac{\partial LHS_s}{\partial s^*} - \frac{\partial RHS_s}{\partial s^*}\right)} > 0$$

$$\frac{\partial \rho^*}{\partial S} = \frac{-\left(\frac{\partial LHS_\rho}{\partial S} - \frac{\partial RHS_\rho}{\partial S}\right)}{\left(\frac{\partial LHS_\rho}{\partial \rho^*} - \frac{\partial RHS_\rho}{\partial \rho^*}\right)} > 0$$

Comparative Statics of Observed Sentences

Probability of Conviction θ

Observe that the derivative of observed sentences with respect to θ is:

$$\frac{\partial L}{\partial \theta} = (1 - G(\rho^*)) \frac{\partial s^*}{\partial \theta} + G(\rho^*) S - (s^* - \theta S) g(\rho^*) \frac{\partial \rho^*}{\partial \theta}$$

Using the fact that $s^* \geq \theta S$ and $\frac{\partial \rho^*}{\partial \theta} < 0$, this is always positive.

Prosecutor Bias k_j

Define the prosecutor's value function as:

$$V = G\left(\frac{\ln \theta}{\ln s^* - \ln S}\right)(\theta S k_j - c) + \left(1 - G\left(\frac{\ln \theta}{\ln s^* - \ln S}\right)\right) s^* k_j$$

So by the Envelope Theorem:

$$\frac{\partial V}{\partial k_j} = G(\rho^*) \theta S + (1 - G(\rho^*)) s^*$$

Then note that observed sentences relate to V as:

$$L = \frac{V + G\left(\frac{\ln \theta}{\ln s^* - \ln S}\right) c}{k_j} = \frac{\partial V}{\partial k_j}$$

Therefore

$$\frac{\partial L}{\partial k_j} = \frac{k_j\left(\frac{\partial V}{\partial k_j} + g\left(\frac{\ln \theta}{\ln s^* - \ln S}\right)\frac{\partial \rho^*}{\partial k_j} c\right) - \left(V + G\left(\frac{\ln \theta}{\ln s^* - \ln S}\right) c\right)}{k_j^2} = g(\rho^*) \frac{\partial \rho^*}{\partial k_j} \frac{c}{k_j} > 0$$

Proof of Proposition 4.2

Because c, S, and $G(\rho)$ do not vary across group, the only candidates for explaining the cross-group differences in average plea bargain sentence length and average plea bargain acceptance rate are $F(\theta|j)$ and k_j.

First, suppose that $F(\theta|b)$ first-order stochastic dominates $F(\theta|nb)$. Assume that $k_b = k_{nb}$. Then,

$\Delta L > 0$. Likewise, because ρ^* is decreasing in θ, $\Delta T < 0$. This yields a contradiction. Letting $k_b < k_{nb}$ will only exacerbate the contradiction because T is increasing in k_j. However, for k_b sufficiently larger than k_{nb}, the effect of prosecutor bias on T can offset that of case strength, giving $\Delta T > 0$. Meanwhile, the effect of prosecutor bias on L would only reinforce the effect of case strength.

Next, suppose that $F(\theta|b)$ first-order stochastic dominates $F(\theta|nb)$. Assume that $k_b = k_{nb}$. Then, $\Delta L < 0$ and $\Delta T > 0$. This yields a contradiction. Letting $k_b < k_{nb}$ will only exacerbate the contradiction because L is increasing in k_j. However, for k_b sufficiently larger than k_{nb}, the effect of prosecutor bias on L can offset the effect of case strength, giving $\Delta L > 0$. Meanwhile, the effect of prosecutor bias on T would only reinforce the effect of case strength.

The second arm of the proof is similar.

Differences in Risk Aversion

Consider a simple parameterization that captures race-specific differences in the risk aversion distribution $G(\rho)$. Let $G_j(\rho) = G(\rho + \psi_j)$ so that groups with a higher value of ψ_j are more likely to draw a low value of ρ, which corresponds to higher risk tolerance. Notice that the defendant's cutoff rule does not depend on ψ_j, so the only difference in equilibrium behavior comes through the prosecutor's choice of s^*, which is determined by:

$$\frac{s^*(\ln S - \ln(s^*))}{s^* - \theta S + {}^{c}/{}_{k_j}} = \frac{g(\rho^* + \psi_j)}{1 - G(\rho^* + \psi_j)}\rho^*$$

Using the comparative static framework from earlier in this appendix, clearly $\frac{\partial LHS_s}{\partial \psi_j} = \frac{\partial LHS_\rho}{\partial \psi_j} = 0$. From the assumption that $G(\rho)$ has an increasing hazard rate, I can conclude that $\frac{\partial RHS_s}{\partial \psi_j} > 0$ and $\frac{\partial RHS_\rho}{\partial \psi_j} > 0$. Consequentially, $\frac{\partial s^*}{\partial \psi_j} < 0$ and $\frac{\partial \rho^*}{\partial \psi_j} < 0$. The first comparative static is intuitive. As the typical defendant in a given group becomes more tolerant of risk, the prosecutor must moderate the sentences he offers or else face a higher rate of expensive trials. The second comparative static states that the marginal defendant is now more risk tolerant, and therefore that some types of defendants who would have ordinarily demanded a trial no longer do.

This does not, however, imply that the *trial rate*, T, is decreasing in υ_j. The acceptance margin has shifted, but the distribution of ρ to compare it against has shifted as well. Using that $T = G(\rho^* + \psi_j)$, I can calculate:

$$\frac{\partial T}{\partial \psi_j} = g(\rho^* + \psi_j)\left(\frac{\partial \rho^*}{\partial \psi_j} + 1\right)$$

I next argue that $0 > \frac{\partial \rho^*}{\partial \psi_j} > -1$. That is, the shift in ρ^* induced by changing ψ_j is dominated by the direct change to the risk tolerance distribution and hence to the trial rate.

$$\frac{\partial \rho^*}{\partial \psi_j} = \frac{\frac{\partial RHS_\rho}{\partial \psi_j}}{\frac{\partial LHS_\rho}{\partial \rho^*} - \frac{\partial RHS_\rho}{\partial \rho^*}}$$

The numerator can be written as:

$$\frac{\partial}{\partial x}\left[\frac{g(x)}{1-G(x)}\right]\rho^*$$

And the denominator as

$$\frac{\partial LHS_\rho}{\partial \rho^*} - \frac{\partial}{\partial x}\left[\frac{g(x)}{1-G(x)}\right]\rho^* - \frac{g(\rho^*+\psi_j)}{1-G(\rho^*+\psi_j)} < -\frac{\partial}{\partial x}\left[\frac{g(x)}{1-g(x)}\right]\rho^*$$

The inequality holds because $\frac{\partial LHS_\rho}{\partial \rho^*} < 0$. These facts imply that $0 > \frac{\partial \rho^*}{\partial \psi_j} > -1$. Thus $\frac{\partial T}{\partial \psi_j} = g(\rho^* + \psi_j)\left(\frac{\partial \rho^*}{\partial \psi_j} + 1\right) > 0$.

Intuitively, the change in ρ^* induced by an increase in ψ_j is moderated by other forces in the prosecutor's problem. Therefore, he only partially counteracts the increase in overall risk tolerance for that group. The overall effect is still to increase trials, albeit less than if the prosecutor had not reacted at all.

Given that $\frac{\partial T}{\partial \psi_j} > 0$ and $\frac{\partial s^*}{\partial \psi_j} < 0$ and recalling $s^* > \theta S$, it is straightforward that

$$\frac{\partial L}{\partial \psi_j} = \frac{\partial T}{\partial \psi_j}\theta S - \frac{\partial T}{\partial \psi_j}s^* + (1-T)\frac{\partial s^*}{\partial \psi_j} = \frac{\partial T}{\partial \psi_j}(\theta S - s^*) + (1-T)\frac{\partial s^*}{\partial \psi_j} < 0$$

This shows that shifts in the risk tolerance distribution must move trial rates and sentence lengths in opposite directions. Therefore, cross-race differences of this form could serve as an alternative explanation to empirical results related to differences in θ but not differences in k_j.

Appendix B: Felony Court in Cook County

The information in this appendix is drawn from private conversations with defense attorneys, prosecutors, and judges with extensive experience in the Criminal Division of the Cook County Circuit Court, supplemented with information in IICLE (2017).

Arrest and Charging

Felony cases begin when the suspect is arrested by police. This is shown in the first row of Figure 3. Many agencies have arresting authority within Cook County, including the Chicago Police Department, the police departments of suburban areas within Cook County, the Cook County Sheriff's Office, and the Illinois State Police. Following an arrest, if the police wish to press charges, the arresting officer completes an arrest report that lists the evidence against the suspect. This arrest report can be thought of as a first draft of the charges against the defendant.

Before becoming a formal felony charge, the case must first be approved by the Cook County State's Attorney (SA). This is shown in the second row of Figure 3. The SA maintains a dedicated Felony Review Unit for this purpose, comprised of a relatively small group of Assistant State's Attorneys (ASAs). The Felony Review Unit is on call 24 hours a day to review and approve charges. In simple cases, they do this over the phone, but in more complex cases, they will travel to the relevant police station to examine the case directly. Though the police may technically override the decision of the Felony Review ASA, this is very rare. A felony case that lacks the support of the State's Attorney's Office will almost certainly fail to reach conviction. One can think of the Felony Review ASA as responsible for editing the draft charges brought in the police report, bringing them in line with the law and rejecting them when the evidence is too weak.

Bail and Preliminary Hearings

Following felony review, the case moves through two hearings. The first is a bail hearing, which takes place as soon as possible. This is shown in the third row of Figure 3. At this hearing, a bail judge listens to arguments for whether a defendant should be permitted to post a bail bond, and if they are, the amount of that bond. Bail bonds call for the defendant to give the court a specified amount of money as collateral. In return, the defendant is permitted to leave jail while awaiting trial. If the defendant does not return to court when required, they forfeit the collateral and may face additional charges. At the conclusion of court proceedings, the bond is refunded to the defendant, though some fines and costs may be deducted directly from this amount. Illinois only permits individuals to post bail. A defendant may be bailed out by a friend or family member but not by a commercial enterprise like a bail bondsman.

The second hearing is a preliminary hearing where the state shows probable cause. It may take place up to two months after arrest. This is shown in the fourth row of Figure 3. The purpose of this hearing is for the SA to show probable cause that the defendant committed a crime. This is a standard of proof well below the "shadow of a doubt" standard used at trial. Probable cause can be shown either in a preliminary hearing or in a grand jury. A preliminary hearing is adversarial. The defendant or their attorney has the opportunity to present their side of the case to a judge. Preliminary hearings produce a charging document known as an information. A grand jury is not adversarial, so it is typically easier for the SA to establish probable cause. However, grand juries produce a charging document known as an indictment. Unlike an information, an indictment cannot be readily altered by the prosecutor, though the prosecutor can still convict a defendant of a "lesser included offense" of the charges listed in the indictment. At this stage, the prosecutor may also opt to abandon the case. This is not the drop decision I observe in my court data. Cases dropped at or before this point will not appear at all in my data.

Bail and preliminary hearings are the responsibility of the SA's Preliminary Hearings Unit. One can think of the Preliminary Hearing ASA as responsible for publishing the charges that were drafted by the Felony Review ASA.

Arraignment and Assignment

After the SA has established probable cause, the case proceeds to the office of the Presiding Judge of the Criminal Division for arraignment and assignment. This is shown in the fifth row of Figure 3. At the arraignment, the Presiding Judge reads the charging document produced in the probable cause stage and asks the defendant how they plead. At this stage, nearly all defendants plead not guilty.

If the crime was committed in one of the suburban areas surrounding Chicago but still part of Cook County, the case is sent to the Municipal courthouse for that area. If the crime was committed in Chicago, the case is next assigned to a judge in the Criminal Division. This is shown in the sixth row of Figure 3. Typically, this assignment is made by a randomization program maintained by the Clerk of Court. This program randomly assigns the case to one of the permanent judges on the court. Exceptions to this assignment mechanism include: defendants already on probation are assigned to the judge who gave them probation; certain defendants may be diverted to problem solving courts (e.g. drug treatment, mental health, and veterans); some judges handle exclusively drug cases; some particularly difficult or sensitive cases may be diverted to more experienced judges.

Defendants have the right to make one Motion to Substitute Judge within 10 days of the initial judge assignment. The court automatically grants this first motion but will only grant subsequent motions if the defense is able to prove that the assigned judge showed "animosity, hostility, ill will, or distrust" towards the defendant. I observe all SOJ motions in my data. Anecdotally, defense attorneys use this automatic SOJ to avoid particularly strict judges.

Simultaneously with its assignment to a judge, the case is assigned to an ASA in the Criminal Prosecutions Bureau who will oversee it until its completion. This ASA is typically the prosecutor posted to the courtroom where the assigned judge hears cases. This is the prosecutor whose problem is captured in my model. Importantly, by the time the case reaches the hands of the Criminal Prosecutions ASA, three others have contributed to the charging document: the arresting officer, the Felony Review ASA, and the Preliminary

Hearing ASA. Further more, if that charging document is an indictment, it belongs to the grand jury in the sense that the prosecutor may not freely alter it. Thus, I model my prosecutor as taking the charges against the defendant as given.

Discovery

Illinois has strong discovery rules intended to "promote the search for the truth and to eliminate surprise as a trial tactic." Upon the request of the defense, prosecutors are required to disclose:

- The identities and criminal records of all intended witnesses and their statements
- All statements made by the defendant
- Minutes from any grand jury proceedings
- Expert witness reports and medical/scientific test results
- Documentary, physical, and surveillance evidence
- Any material favorable to the defense, even if the prosecution does not intend to present it at trial

Likewise, the defense is required to give access to the defendant for medical and other tests, disclose the results of medical and scientific reports, and announce all intended defenses.

Trial and Pleas

Following arraignment, the prosecution and defense meet periodically in court. This is shown in the space between the sixth and seventh rows of Figure 3. It is not unusual for these meetings to consist solely of a request for postponement. Defendants have a constitutional right to a speedy trial, but it is rarely binding because any postponement requested or agreed to by the defense does not count against the speedy trial clock. During the trial preparation, both sides file motions, review case materials, and prepare their arguments for trial. The prosecutor may opt to abandon the case, and this is the drop decision captured in my model. The defendant has the right to demand a jury trial, though the vast majority of trials carried out in Cook County are bench trials presided over only by the judge. During the trial, the prosecutor presents the state's case, followed by the defense, followed by a ruling of guilty or not guilty on each charge.

At any point during this process, the defendant may change their plea to guilty, almost always after having arranged an informal deal with the prosecutor to drop or amend certain charges and arrange a specific sentence for the remaining charges. No plea deal can be carried out without the assent of the judge.

They must ultimately be the one to enact the agreed-upon sentence. However, judges rarely object to the terms of a deal agreed to by the prosecution and defense.

Sentencing and Punishment

If the defendant pleads or is found guilty, the case proceeds to a sentencing hearing. This is shown in the seventh row of Figure 3. The purpose of this hearing is for both sides of the case to present evidence towards various factors in aggravation and in mitigation that may influence the length of the sentence. In the case of a plea bargain, the defendant typically waives this hearing as both sides have already agreed on a sentence. At the conclusion of the hearing, the judge passes a sentence that may include prison time, probation, alternative punishments like “boot camp,” or fines.

If the defendant is sentenced to probation, she must meet regularly with a probation officer and hold to the terms of their probation. These terms vary from case to case (e.g. remaining within the state or county, staying away from criminal associates, or submitting to regular drug tests). During this period, the probation officer may allege to the court that the defendant has violated the terms of her probation, at which point the case is returned to the judge who assigned the initial sentence of probation. If the defendant is found to have violated the terms of her probation, the judge may send her to prison for the original charge. The proceedings of this hearing to determine violation of probation are recorded as an extension of the original case.

If the defendant is sentenced to prison, she is transferred to a state facility to serve the sentence. Effective sentences may be shorter than recorded sentences both because prisoners receive credit for time served in prison while awaiting trial and because of the system of good time credits in Illinois. Prior to 1998, all prisoners in Illinois received sentence credit for good behavior at a rate of one day per day, meaning that a well-behaved prisoner served an effective sentence that was 50% of the nominal sentence, absent further adjustments for time served awaiting trial. Once in prison, defendants may receive additional credit for completing rehabilitation and training programs or educational degrees, and up to 6 months discretionary credit from the warden. In 1998, Illinois passed a Truth in Sentencing law that reduced good time credits for sentences related to certain charges. Prisoners serving a sentence for murder are now required to serve 100% of the nominal sentence, and prisoners for a number of other crimes (primarily sex crimes and violent crime resulting in great bodily harm to the victim) are required to serve 85% of the nominal sentence.

Following completion of a prison sentence, defendants are subject to 2-3 years of Mandatory Supervised Release (MSR). Much like probation, a person under MSR must adhere to certain conditions and may be returned to prison if found to have violated them. Illinois does not have a parole system. Prisoners serve a sentence of a determinate length, albeit with many ways to reduce the length of this sentence. There is no mechanism to release them before this sentence is over.

Appendix C: Cook County Circuit Court Data

The Structure of the Raw Data

The files provided by the Cook County Clerk of Court are a snapshot of the court's felony case database as of June 2019. The data were saved as fixed-width text files, which I converted to the Stata .dta format using provided record layout files. Some of the raw data files contained non-ASCII characters that impeded Stata from properly reading and interpreting the data. I removed these characters from the raw files before processing.

The Clerk's data are structured as a relational database. This database has three tables: Root, Charge, and Disposition. The tables are linked by unique case ID numbers, and in the case of Charge and Disposition, charge numbers. The Root table contains exactly one record per case. It stores information that does not vary over time, such as the defendant's name and address, demographic information, case initiation date, and various identification numbers.[31] It also contains some information that may vary over time (attorney, custody status, court dates, etc.). These fields were likely updated as the case proceeded in order to provide current information to the Clerk. Thus, these fields will generally reflect the status of cases at the time of their conclusion. I do not use any time-varying information from the Root table.

The Charge table contains one record per charge, and hence potentially many records per case. Within a case, each charge is identified by a unique charge number. The information in this file is based on the charging document (indictment or information) that initiated the felony case. Charges are described in three ways. First, the data record a reference to the Criminal Code of Illinois. For example, "18-2 (A) (2)" refers to Article 18 (Robbery), Chapter 2 (Armed Robbery), Part A (Definition), Subpart 2 (Weapon other than firearm). Second, the data record a written description of the charge. These typically paraphrase the code section quoted, e.g. "ARMED ROBBERY WEAP OTH FIREARM." The length of these descriptions is constrained, and they do not follow a consistent abbreviation scheme. Third, the data record a charge sequence code that assigns a number to the charge, though this mapping is not one-to-one. The felony class of each charge is listed in a separate field. Each charge record also indicates whether a charge was a amended and if so, all information about the amended charge.

[31] In addition to the unique case ID number, most records also contain a finger-print based "Internal Rapsheet" or "IR" number, which I use to link records belonging to the same person. "Central Booking" or "CB" numbers link cases to arrests. "Records Division" or "RD" numbers link cases to police reports. Some cases also have Illinois Department of Investigation or Federal Department of Investigation identifiers.

The Disposition table contains one record per disposition, and hence almost always includes many records per case. Within a case, each disposition is identified by a unique disposition number. These numbers proceed in chronological order. “Disposition” in this context refers to anything the judge records on the “half sheet” provided to the Clerk after each court session. These include any official actions, rulings, verdicts, or orders made by the judge, motions filed by attorneys, and pleas made by defendants.[32] Most dispositions are routine and so can be represented with a numerical “disposition code.” When the judge makes a special order or needs to provide additional information, this is recorded in the “free description” field of the data. Disposition histories in my data begin when the defendant is arraigned (See Appendix B for details). They may contain dispositions made after the initial judgment and sentence if any such dispositions were made (for example in probation hearings or appeals).

Each disposition also includes fields indicating which judge recorded the disposition and in which courtroom. Both judges and courtrooms are represented by numerical codes, and the Clerk did not provide any correspondence between these codes and actual judges and courtrooms. Disposition records include fields to record sentence lengths and fines for dispositions dealing with sentencing.[33] When a disposition record references a record in some other table, most often a charge or a bond, the exact record is indicated by the “CB reference” field. For example, a verdict finding the defendant guilty of charge 1 but not charge 2 would read as a guilty disposition referring to “C001” and a not guilty disposition referring to “C002.” When a disposition refers to all charges, the CB reference field reads “CALL.” This allows me to link records in the Charge table directly to records in the Disposition table. Finally, any dispositions where the case is being transferred or court business has concluded for the day has a separate field that indicates the courtroom where the case will appear next. This is most useful for identifying initial case assignment.

Deciphering Fields

Many aspects of the court data are doubly-encoded. The raw data stores them only as numbers, but even after the numbers are interpreted, specific legal knowledge is required to fully understand them. I focus first on the process of deciphering numerical encodings. The Clerk provided explicit data dictionaries for some numerical encodings: race, charging document type, and disposition code. The Clerk also confirmed that the final two digits of the ID variable enumerate defendants in a multi-defendant case. Thus, two records that match ID up to this point are associated with codefendants. The code section field of the Charge table can be deciphered using the IL Criminal Code itself. The charge sequence field can be deciphered by careful comparison to the code section and written description fields, but this process does not scale well. Judge

[32]“Disposition” more commonly refers exclusively to how a criminal charge is resolved.

[33]Confusingly, the fine field is sometimes used to encode information about bond amounts and hence must be analyzed carefully.

numbers can be matched to judge names by comparing the electronic version of the case given to me by the Clerk to the representations of the same case available in the Clerk's office (either on paper or via a dedicated computer terminal), which use judge names in place of the numerical code. Courtroom numbers can be matched to physical courtrooms via a similar process, though the courtroom numbers in the electronic data actually refer to "calls." Calls are groups of cases that are typically associated with a single judge and may have some other distinguishing feature, such as being scheduled in the evenings or diverted towards drug treatment. When a judge changes physical courtrooms, their call typically follows them. This means that the courtroom/call numbers in the Clerk's electronic files are generally more informative than physical courtroom numbers.

The two major legal deciphering tasks are categorizing dispositions and charges. With the aid of a lawyer with criminal defense experience in Chicago, I grouped dispositions into categories based on the descriptions associated with the codes. Examples of categories include: sentenced to incarceration, case dismissed, and public defender appointed. My goal was to identify only dispositions relevant to various data processing steps described below, so I did not categorize every disposition. I do attempt to exhaustively categorize charges in a manner approximating the FBI's UCR charge categorization scheme. This could primarily be done at the Article level in the IL Criminal Code, with a few categories requiring me to make distinctions at the Chapter level. Furthermore, I did not need to extract felony class information from the detailed charge information because it was already provided separately by the Clerk. Thus, my crime categorization algorithm works by parsing the code section strings in the Charge data to extract Article and Chapter information and recombining that information to form UCR-like categories.

Supplementing Defendant Information

The electronic data files provided by the Clerk of Court redacted defendant personal and demographic information in cases that, in the estimation of the Clerk, ended in a non-conviction. Other information, such as charges and dispositions, was unaffected. Nevertheless, this meant that a large and very selected subset of my initial dataset was lacking race information and thus could not be included in my analysis. I addressed this problem by supplementing the Clerk's electronic data files with information collected by the Chicago Data Collaborative (CDC). The CDC data were scraped from web forms populated by the same database given to me by the Clerk. However, the Clerk did not redact any information in this setting. Among records where both my data and the CDC data have defendant information, they match perfectly, even on name and address. This is why I am confident that the two datasets don't just carry the same *information* but are in fact drawing from precisely the same source database. I ultimately use defendant information provided by the CDC for N records in my final estimation sample.

Aggregation to Case-Level Records

The complex structure of the raw data as provided by the Clerk is not suitable to most forms of statistical analysis, which assume that each observation in the data can be written as a single vector of a constant length. This requires me to aggregate the Charge and Disposition tables from several records per case to a single record per case. I aggregate the Charge table by recording information about the most serious charge (as described in Section 5 of the text) and noting whether the defendant faced more than one charge.

In most cases, I aggregate the Disposition table by searching for *events* within the disposition histories from arraignment to initial sentencing. For example, to construct the indicator variable for public defender, I search the disposition history for disposition code 901, which decodes to "Public Defender Appointed." Any case with that code has the public defender indicator set to 1, and all other cases have it set to 0. Some defendants with the public defender indicator set to 1 may have, at some other time, hired a private attorney to handle their case. I do not attempt to differentiate defendants on that basis. I use a similar strategy to construct my indicators for guilty pleas, dismissal, defendant flight or death, and awaiting trial in jail.[34]

Sentencing Information

The sentencing information contained in disposition histories is too complex to allow for aggregation using an event-based strategy. It must be carefully aggregated across both time and charges. Disposition histories frequently include sentencing information from hearings that occurred after the initial sentencing, most often probation violation hearings. I therefore restrict my attention to the *first* day that I observe any sentencing disposition for any charge. Initial sentences are almost always handed down on a single day, and this strategy avoids erroneously including later sentencing dispositions.[35] This initial sentencing date is the endpoint of the case for most of my purposes, and my event-based aggregation strategy ignores most events that occur after this date.

I gather all sentencing dispositions made on the initial sentencing date, as well as dispositions indicating that the defendant is to get sentencing credit for time served in jail while awaiting trial and any dispositions indicating that sentences are to be served consecutively. I aggregate probation sentences as if they are to be served concurrently. I aggregate incarceration sentences as if they are to be served concurrently unless

[34] In light of the fact that disposition histories begin at arraignment, often at least a month after the initial arrest, I am not concerned about picking up any defendants who were jailed only before their bail hearing.

[35] In probation hearings, sentences are often passed on charge numbers that are unrelated to the initial charging document. A defendant who initially convicted of only the second charge in his charging document, may be sentenced to prison in a probation violation hearing based on a placeholder "charge 1." An approach that treated each charge separately would conflate the two sentences.

I observe a disposition that says they are to be served consecutively. When a defendant is given both a probation sentence and an incarceration sentence, I treat the incarceration sentence as their primary sentencing outcome, but I make a note of the probation sentence.

This information is sufficient to construct my main sentencing outcome, which I call "nominal incarceration." I can also construct "marginal incarceration" by subtracting any sentencing credit from the nominal incarceration sentence. I construct total incarceration by treating measured jail time as incarceration time when a defendant did *not* get an incarceration sentence.[36] Finally, I construct "realized incarceration" by considering cases with probation violation hearings and repeating my sentencing information procedure on the disposition histories following the initial sentencing date, then treating any resulting incarceration as if had been given at the initial sentencing.

Related Cases

The state sometimes wishes to charge a person with a set of crimes that cannot be contained in a single charging document, most frequently because some of the crimes are not part of the same "course of action" but sometimes because a defendant is arrested for a new crime while awaiting trial for the first. This can generate sets of criminal cases that are, for all practical purposes, treated as a single case. They are assigned to the same judge, negotiated as a unit, and sentenced at the same time. However, the cases have distinct ID numbers and are not explicitly linked in the Clerk's electronic records. If I ignored this feature of the data, I would risk mismeasuring case outcomes, especially when a plea bargain calls for a defendant to plead guilty to charges in only one case, causing the other cases to be dropped entirely. This is properly measured as a single plea bargain, but a naive approach would measure it as a plea bargain and a dropped case.

To address this problem, I first assemble sets of cases for combination. I combine cases if they have the same defendant and they overlap in the period between arraignment and first sentencing. I must then condense the information contained in this set of cases into a single vector of information. For some variables, primarily dates and event indicators, I assign to the combined case the minimum or maximum value across the set of cases, as appropriate. For all event indicators, I take the maximum. If a defendant has a public defender disposition in one case, the indicator should be turned on for the combined case, etc. All other variables I treat as sets. For example, *all* of the sentencing information in the combined case comes from the case that has the most severe sentence. All charge information comes from the case with the most severe charge, etc. This prevents nonsensical combinations of information while still accounting for scenarios where a defendant's plea bargain in one case was influenced by charges brought in a different case, even when those

[36] When sentencing credit is not available, I measure jail time as the sum of the time between dispositions where the defendant was observed to be in jail. Most court appearances include a disposition indicating whether the defendant was held in custody or on bond at the time.

charges are eventually dropped.

Judge Assignment

It is possible to use the information in the Clerk's electronic files both to determine which judge most cases were assigned to and to restrict to a subset of cases that were assigned randomly. This fact plays a small role in my plea bargain analysis but is generally important for future work, so I document it here.

Assignment to Judges

As noted above, when a disposition indicates that a case is to be transferred or concludes court business for a day, the record for that disposition includes information about the call under which it is scheduled to resume. This is also true of all dispositions that assign a case to a particular call. This is the best available measure of *assignment* and is superior to simply observing the call of future dispositions in the case. It is not uncommon for defendants to immediately respond to judge assignment with a Substitution of Judge motion.[37] Defendants who do so are immediately reassigned. This means that the defendant's initial assignment, the true point of randomization, sometimes *only* exists as an assignment disposition and cannot be observed from future calls.

This property of the data makes the relationship between cases and calls obvious, and I rely on it alone when forming my clustering variable in Section 6. To use a judge stringency measure, however, it is also necessary to establish which calls receive random assignments as well as the relationship between calls and judges. According to Lawrence Fox, a retired judge in the court, an established judge typically has only one call throughout their career, even if they move physical courtrooms. However, new judges frequently begin their careers as "floaters" who tend to the calls of other judges when they are absent from the court. Judges may also change calls if they make a major career change (e.g. moving from the Chicago courthouse to one in the suburbs). A small number of judges (Judge Fox included) also maintain a separate call for specialty courts, such as those for drug abusers or veterans. I focus on a set of primary calls. These are the calls located at the main court building that are most frequently assigned cases at arraignment. They take numbers 17XX where XX is between 02 and 34.[38] Judge Fox confirmed that these are the relevant calls and the ones receiving assignments from the general pool of cases.

The relationship between judges and calls is stable, but it is not one-to-one. Retiring judges hand off their calls to acceding judges. Judges who take a long leave of absence yield their call to another judge

[37] Defendants in Illinois are afforded one such motion "by right," meaning that they do not have to provide any reason for the substitution. Multiple sources familiar with the courts in Chicago have confirmed that defense attorneys use this right strategically and file Substitution of Judge motions when their case is assigned to a particularly harsh judge.

[38] 1701 is the Presiding Judge's call. 1749 is another administrative call. 1735-48 are used for various ancillary calls, most often dedicated drug courts.

for that time. Fortunately, the disposition histories themselves can be treated as a dataset that establishes which call (if any) a judge was hearing on any given day. I first count the number of dispositions each judge heard on each call and assign a judge to a call if they made the most dispositions on that call in that day. It is rare in practice to see a call with dispositions from more than one judge within a single day, so this step is straightforward. Then, to account for very brief absences, in which a floater judge may fill in for a permanent judge, I use information from other days to assign a permanent judge to each day. If a judge: (1) Held the call for the most days in the surrounding four week period (two before and two after); (2) Held the call for at least one day in the prior two weeks and at least one day in the subsequent two weeks; and (3) Held the call for a majority of days in the surrounding year (calculated *either* January to December *or* July to June), then I assign them to that call in that day. This produces a result that aligns with what insiders have told me about call assignment: judges hold their calls for a contiguous period of several years before handing them off to another judge. In rare cases, these tenures are interrupted for a period of a few months.

Random Assignment

Just because a case was assigned to a call that *may* receive random assignment does not mean that a case was randomly assigned. In conversations with court insiders, they listed three primary reasons why a case may not be randomly assigned: (1) The defendant already had a pending matter before the court. (2) The case was assigned to a drug or specialty treatment court. (3) The case was a high-profile "heater" intentionally assigned to a more experienced judge. The first reason is by far the most common. If the defendant in any new case was already awaiting trial or on probation, the case is automatically assigned to the judge overseeing the pending case. I address this problem by excluding from my randomized sample any case that was initiated as its defendant was awaiting trial or within the probation period of a prior case. As added insurance, I exclude *any* cases where a defendant is re-assigned to the same judge within 4.5 years of his last assignment to that judge. Specialty courts are managed from a list of calls distinct from the primary calls considered above. If news materials or interviews with court insiders indicate that a judge managed a specialty court at any time, I exclude all of their cases for good measure.[39] My sample does not include murder or sexual assault cases, which likely eliminates most heaters from my dataset. Large cases may be sufficiently complex or public to be designated a heater, so I also do not consider a case randomly assigned if it had more than 4 defendants.

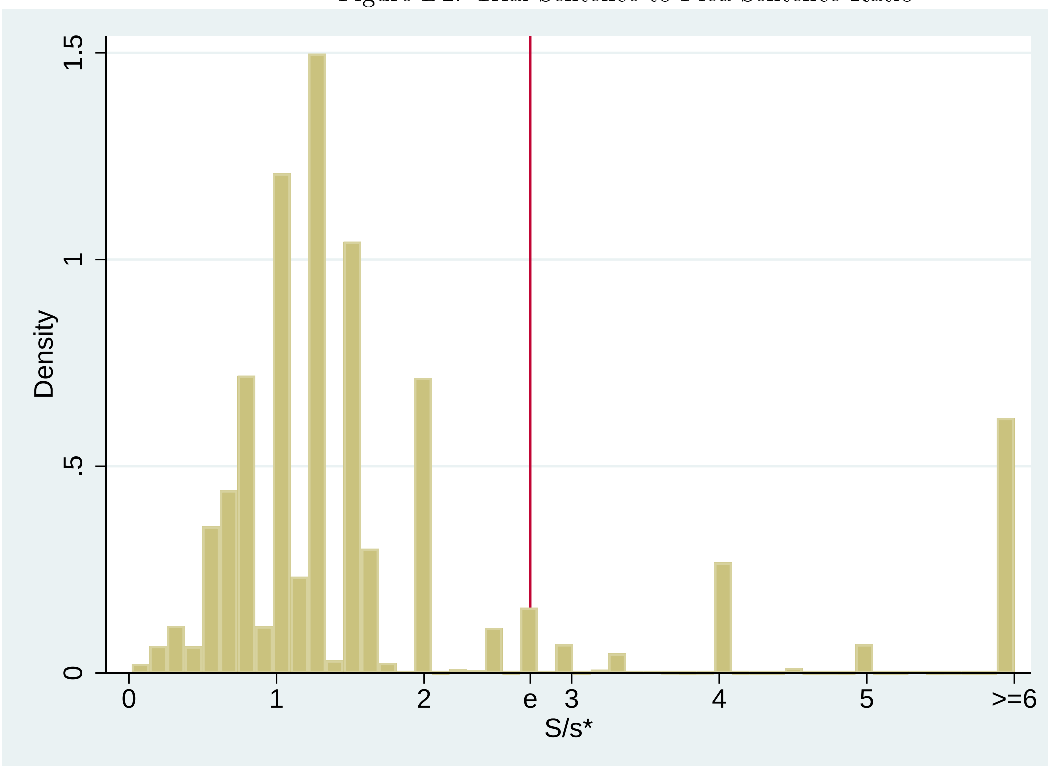
[39] It is not practical to exclude judges who headed drug courts at any time in their career. Throughout the 1990s, almost all new judges spent one to two years in night narcotics courts before being promoted to a permanent call.

Appendix D: Additional Empirical Results

Figure D1: Racial Differences Over Time

Notes: This figure plots the coefficients from regressions of an indicator for black defendants on nominal sentence length (upper panel) and whether a case ended in trial (lower panel), conditional on year of case initiation. Each regression has covariates: class of most serious charge, category of most serious charge, age, and number of prior convictions, plus indicators for whether the defendant is male, ever had a public defender, or was ever held in custody awaiting trial, and whether the state filed multiple charges or included multiple defendants in the case. Dashed lines present 95% confidence intervals clustered at the courtroom-year level. I exclude years with fewer than 500 observations, which affects only 1984. See Section 5 for more details about sample selection.

Figure D2: Trial Sentence to Plea Sentence Ratio



Notes: This figure shows a histogram of estimated values for S/s^* among cases that ended with a plea bargain. For these cases, S is estimated within felony class cells as the median sentence given to defendants convicted at trial.

Table D1: Trial Logit

	Trial	Trial	Trial	Trial
reg_trial				
Black	0.152***	0.241***	0.144***	0.127***
	(0.0135)	(0.0141)	(0.0148)	(0.0152)
Male		0.336***	0.204***	0.220***
		(0.0208)	(0.0214)	(0.0215)
Public Defender		-0.526***	-0.434***	-0.459***
		(0.0114)	(0.0117)	(0.0121)
Ever in Jail		-0.134***	-0.238***	-0.249***
		(0.0131)	(0.0137)	(0.0139)
Multiple Defendants		0.361***	0.326***	0.319***
		(0.0121)	(0.0128)	(0.0129)
Multiple Charges		0.243***	-0.288***	-0.299***
		(0.0112)	(0.0130)	(0.0131)
Observations	342616	342616	342616	342614
Charge Cond.	None	None	ClassXCat FE	ClassXCat FE
Assignment Cond.	None	None	None	Ctrm and Year FE

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents the results of logistic models using an indicator for black defendants as the independent variable and whether the case ended in a trial as the outcome variable. Beginning with the second column, covariates also include dummy variables for age and number of prior convictions. See Section 5 for details about sample selection.

Table D2: Alternative Topcodes

	60 Months	40 Months	20 Months	Censored at 80
Black	-1.467***	-1.506***	-1.556***	-1.567***
	(0.192)	(0.181)	(0.160)	(0.188)
Male	5.540***	5.408***	5.055***	5.458***
	(0.158)	(0.153)	(0.142)	(0.154)
Public Defender	2.973***	2.885***	2.659***	2.896***
	(0.169)	(0.158)	(0.137)	(0.167)
Ever in Jail	11.33***	11.16***	10.71***	11.23***
	(0.186)	(0.178)	(0.160)	(0.182)
Multiple Defendants	2.272***	2.002***	1.443***	2.025***
	(0.215)	(0.197)	(0.164)	(0.208)
Multiple Charges	7.722***	7.485***	6.848***	7.600***
	(0.168)	(0.160)	(0.142)	(0.165)
Observations	342616	342616	342616	342547
Adjusted R^2	0.296	0.325	0.383	0.303
Charge Cond.	ClassXCat FE	ClassXCat FE	ClassXCat FE	ClassXCat FE
Assignment Cond.	CtrmXYear FE	CtrmXYear FE	CtrmXYear FE	CtrmXYear FE

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents the results of regressions of an indicator for black defendants on nominal sentences in months under different topcoding rules. Covariates also include dummy variables for age and number of prior convictions. Standard errors are clustered at the courtroom-year level. See Section 5 for details about sample selection.

Table D3: No High s^*/s

	Sentence	Trial
Black	-1.496***	0.0140***
	(0.200)	(0.00175)
Male	5.552***	0.0188***
	(0.165)	(0.00201)
Public Defender	3.077***	-0.0556***
	(0.180)	(0.00188)
Ever in Jail	11.95***	-0.0215***
	(0.197)	(0.00178)
Multiple Defendants	2.503***	0.0340***
	(0.230)	(0.00182)
Multiple Charges	8.052***	-0.0330***
	(0.177)	(0.00166)
Observations	327333	327333
Adjusted R^2	0.289	0.081
Charge Cond.	ClassXCat FE	ClassXCat FE
Assignment Cond.	CtrmXYear FE	CtrmXYear FE

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents the results of regressions of an indicator for black defendants on both nominal sentences in months and whether the case ended in a trial. Covariates also include dummy variables for age and number of prior convictions. Standard errors are clustered at the courtroom-year level. Cases with an estimate of $s^*/s > e$ are excluded. See Section 5 for details about sample selection.

Table D4: Dropped Cases

	Case Dropped
Black	-0.0126***
	(0.00117)
Male	0.00446***
	(0.00127)
Public Defender	-0.0339***
	(0.00119)
Ever in Jail	-0.0323***
	(0.00126)
Multiple Defendants	0.00720***
	(0.00104)
Multiple Charges	-0.0227***
	(0.000916)
Observations	359348
Adjusted R^2	0.038
Charge Cond.	ClassXCat FE
Assignment Cond.	CtrmXYear FE

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents the results of a regression of an indicator for black defendants on whether the case was dropped by the prosecution. Covariates also include dummy variables for age and number of prior convictions. Standard errors are clustered at the courtroom-year level. See Section 5 for details about sample selection. This sample does not apply the standard restriction against cases dropped by the proseuction.

Table D5: Imputing Dropped Cases

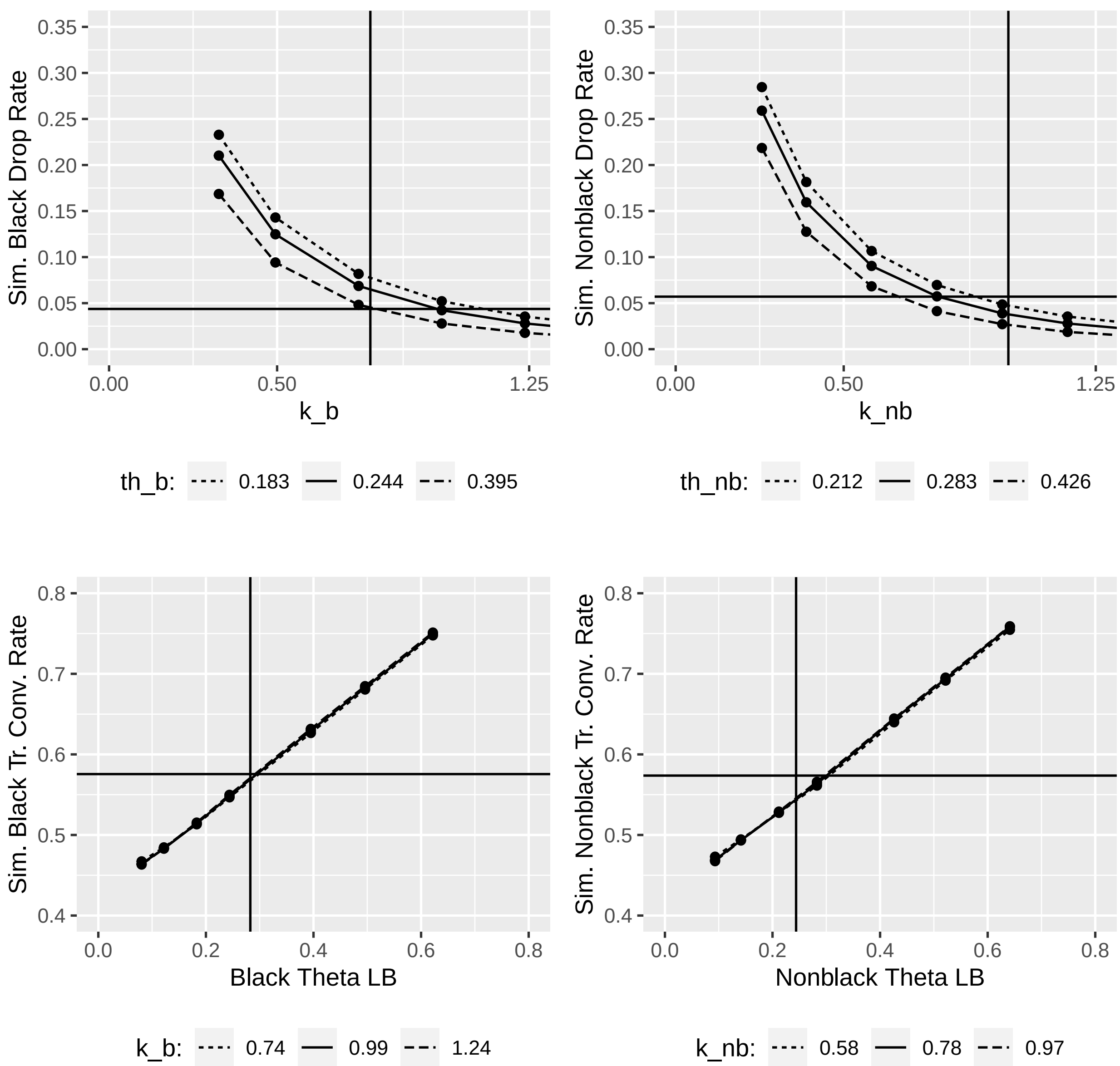
	Sentence	Trial
Black	-1.361***	0.0137***
	(0.188)	(0.00159)
Male	5.213***	0.0165***
	(0.155)	(0.00184)
Public Defender	3.403***	-0.0450***
	(0.165)	(0.00166)
Ever in Jail	11.16***	-0.0174***
	(0.183)	(0.00160)
Multiple Defendants	2.142***	0.0301***
	(0.216)	(0.00167)
Multiple Charges	7.896***	-0.0266***
	(0.168)	(0.00153)
Observations	359348	359348
Adjusted R^2	0.271	0.071
Charge Cond.	ClassXCat FE	ClassXCat FE
Assignment Cond.	CtrmXYear FE	CtrmXYear FE

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

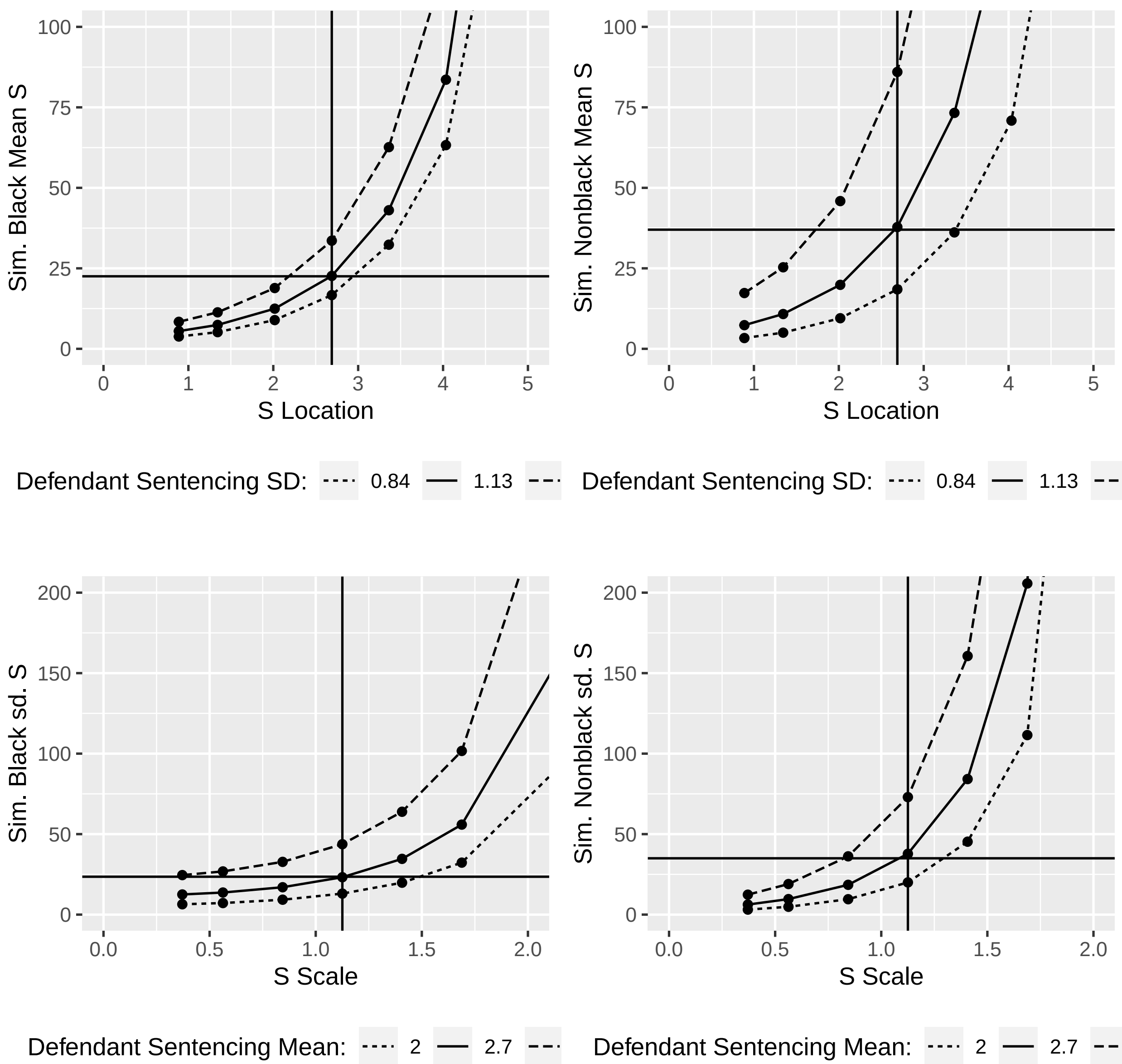
Notes: This table presents the results of regressions of an indicator for black defendants on both nominal sentences in months and whether the case ended in a trial. The both outcomes are set to 0 for cases dropped by the prosecution. Covariates also include dummy variables for age and number of prior convictions. Standard errors are clustered at the courtroom-year level. See Section 5 for details about sample selection. This sample does not apply the standard restriction against cases dropped by the prosecution.

Figure D3: Simulated Moment Responses to Parameter Shifts Pt. 1



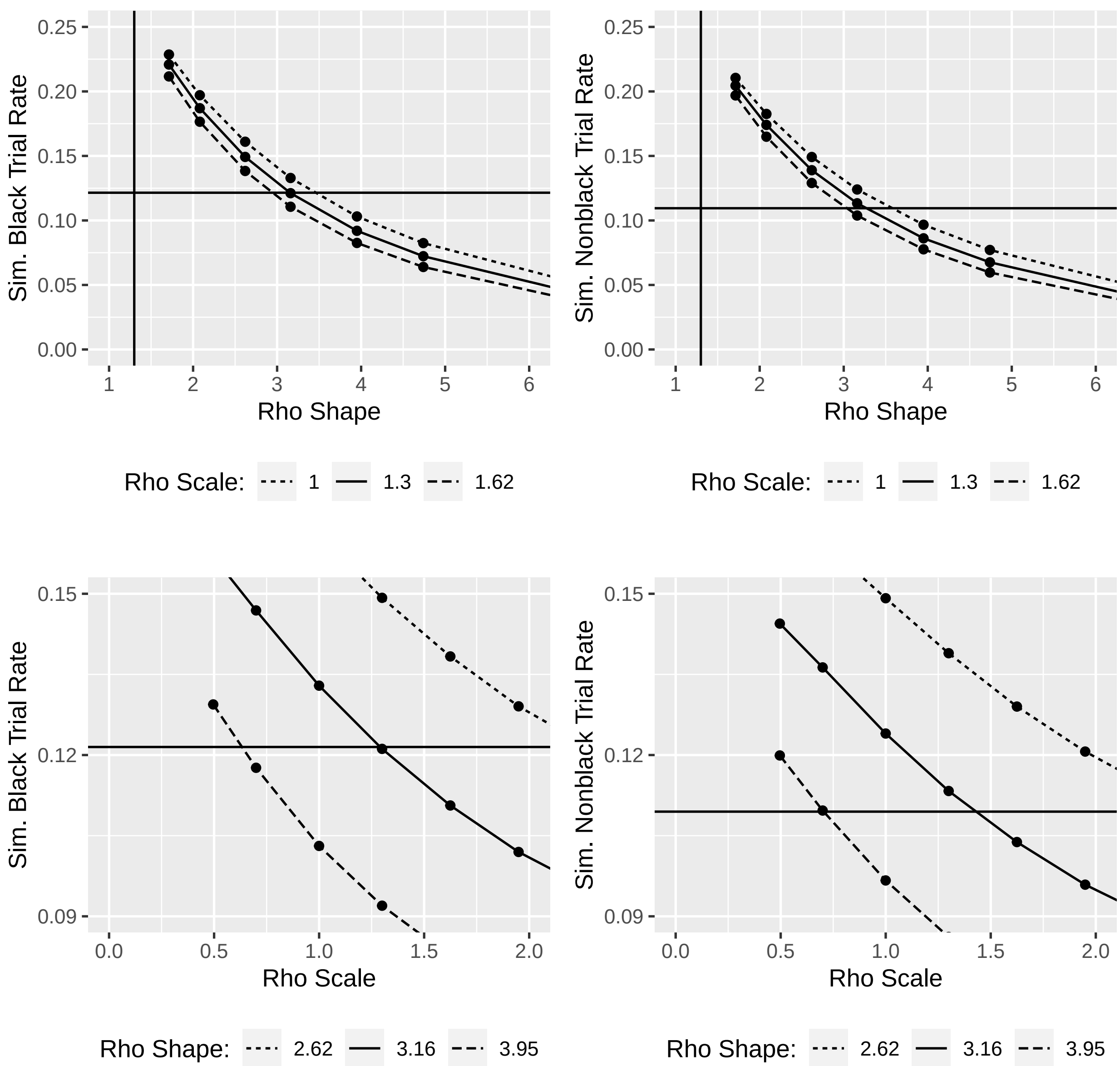
Notes: This figure shows the response of selected simulated moments to changes in the race-specific values of k and defendant distributions of θ. The vertical line denotes the value of the parameter found by my estimation procedure. The horizontal line denotes the value of the moment in the data. The solid curve holds constant all other parameter values at those estimated in Table 7. The dotted and dashed curves set the value of a selected parameter above and below its optimal value, respectively.

Figure D4: Simulated Moment Responses to Parameter Shifts Pt. 2



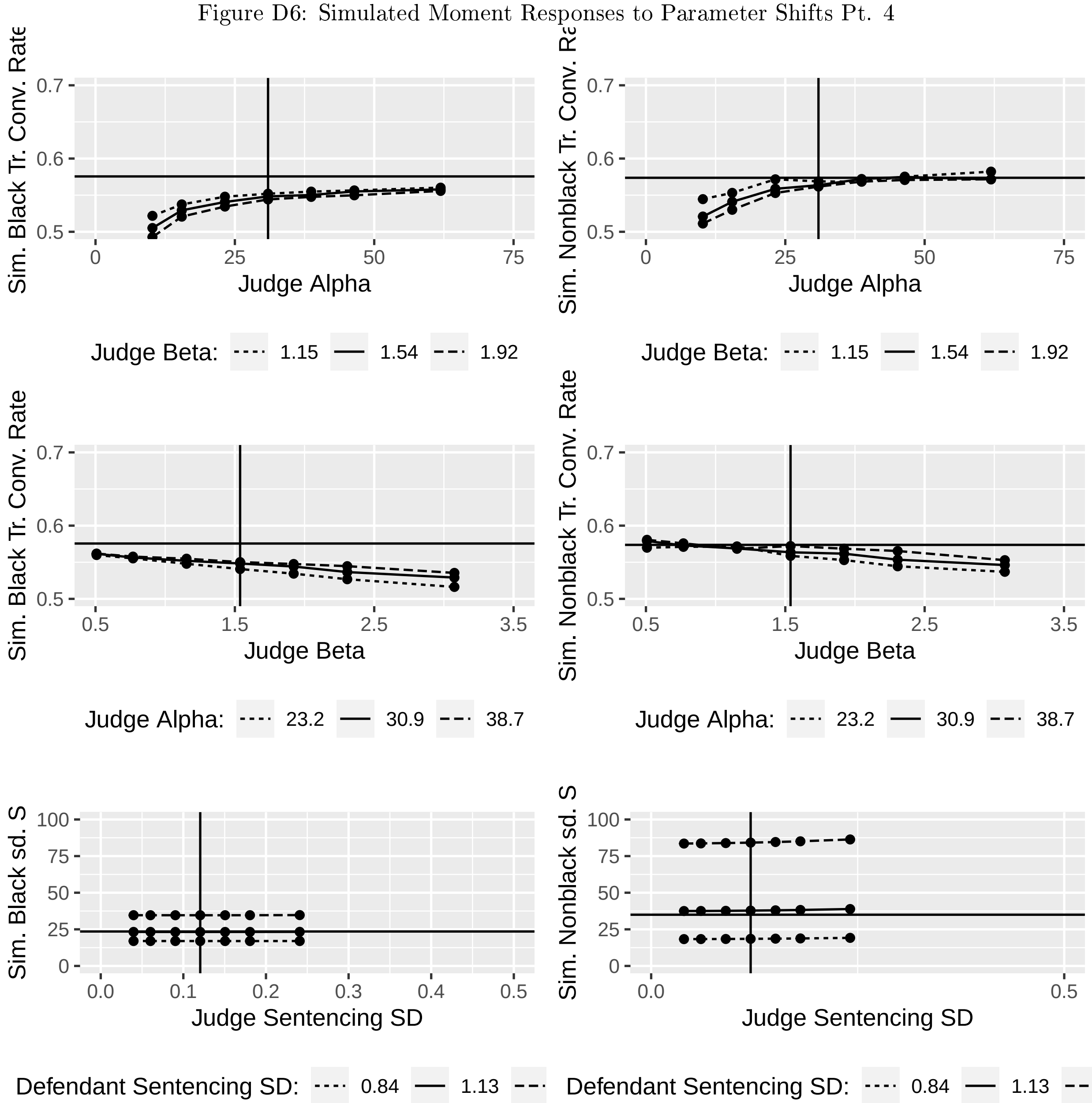
Notes: This figure shows the response of selected simulated moments to changes in the distribution of the defendant's contribution to S. The vertical line denotes the value of the parameter found by my estimation procedure. The horizontal line denotes the value of the moment in the data. The solid curve holds constant all other parameter values at those estimated in Table 7. The dotted and dashed curves set the value of a selected parameter above and below its optimal value, respectively.

Figure D5: Simulated Moment Responses to Parameter Shifts Pt. 3



Notes: This figure shows the response of selected simulated moments to changes in the distribution of ρ. The vertical line denotes the value of the parameter found by my estimation procedure. The horizontal line denotes the value of the moment in the data. The solid curve holds constant all other parameter values at those estimated in Table 7. The dotted and dashed curves set the value of a selected parameter above and below its optimal value, respectively.

Figure D6: Simulated Moment Responses to Parameter Shifts Pt. 4



Notes: This figure shows the response of selected simulated moments to changes in the distribution of judge effects on θ and S. The vertical line denotes the value of the parameter found by my estimation procedure. The horizontal line denotes the value of the moment in the data. The solid curve holds constant all other parameter values at those estimated in Table 7. The dotted and dashed curves set the value of a selected parameter above and below its fitted value, respectively.