BARGAINING WITH ASYMMETRIC INFORMATION: AN EMPIRICAL STUDY OF PLEA NEGOTIATIONS

BERNARDO S. SILVEIRA

OLIN BUSINESS SCHOOL
WASHINGTON UNIVERSITY IN ST. LOUIS

SEPTEMBER, 2016

Abstract. This paper empirically investigates how sentences to be assigned at trial impact plea bargaining. The analysis is based on the model of bargaining with asymmetric information by Bebchuk (1984). I provide conditions for the non-parametric identification of the model, propose a consistent non-parametric estimator, and implement it using data on criminal cases from North Carolina. Employing the estimated model, I evaluate how different sentencing reforms affect the outcome of criminal cases. My results indicate that lower mandatory minimum sentences could greatly reduce the total amount of incarceration time assigned by the courts, but may increase conviction rates. In contrast, the broader use of non-incarceration sentences for less serious crimes reduces the number of incarceration convictions, but has a very small effect over the total assigned incarceration time. I also consider the effects of a ban on plea bargains. Depending on the case characteristics, over 20 percent of the defendants who currently receive incarceration sentences would be acquitted if plea bargains were forbidden.
Keywords: Settlement models, plea bargaining, non-parametric identification, non-parametric estimation, bargaining models, structural estimation.

I am grateful to Alessandro Lizzeri for invaluable support, encouragement and advice. I also benefited from guidance by and insightful conversations with John Asker, Isabelle Perrigne and Quang Vuong. I would like to thank Andrew Daughety, Sandy Gordon, Alejandro Gutierrez-Li, Barton Hamilton, Jakub Kastl, Fernanda de Leon, Konrad Menzel, Nicola Persico, Luigi Pistaferri, Jennifer Reinganum, Yasutora Watanabe, and seminar and conference participants at Berkeley, Caltech, Carnegie Mellon, Cornell, FGV-EESP, ITAM, Maryland, Microsoft Research, Minnesota, NYU, Penn, Princeton, PUC-Rio, Rice, Stanford, Toronto, UCLA, USC, Vanderbilt, WUSTL, Wisconsin, the 2013 American Law and Economics Association Meeting, the 2013 Econometric Society Summer Meeting, the 2013 Southern Economic Association Meeting and the 2013 New Directions in Applied Economics Meeting. I would also like to thank Brad Fowler, from the North Carolina Administrative Office for the Courts, and Cathy Moss, from the Publications Division at the North Carolina Secretary of State's Office, for help in obtaining the data.
Address: One Brookings Drive, Campus Box 1133, St. Louis, MO 63130, silveira@wustl.edu.

1. Introduction

This paper analyzes how the harshness of sentences to be assigned at trial affects plea bargaining. A plea bargain is an agreement between the prosecutor and the defendant through which the latter pleads guilty of a crime without a trial. In return, the prosecutor reduces or dismisses some of the charges against the defendant, or recommends the court to assign a relatively short sentence. The vast majority of criminal cases in the United States are resolved by such an agreement.[1] In my analysis I build upon a bargaining model by Bebchuk (1984) and establish sufficient conditions for its non-parametric identification. I then propose a non-parametric estimator and implement it using data from North Carolina state courts.

The results presented here are relevant for the debate on sentencing reform. Due mainly to the high incarceration rates in the United States, a discussion has developed over whether current sentencing guidelines should be made more lenient. However, an assessment of what the optimal guidelines should be requires a better understanding of the effects of different sentencing reforms on the outcomes of prosecuted cases. This task is complicated by the prevalence of plea bargaining in the American justice system. Given that the vast majority of cases are settled, it is likely that the impact of any sentencing policy intervention depends largely on how it affects plea-bargaining, rather than on the outcome of cases actually brought to trial.

In the model, a prosecutor and a defendant bargain over the outcome of a case. The prosecutor offers the defendant to settle for a sentence. If there is no agreement, the case proceeds to trial, which is costly to both parties. The defendant is better informed than the prosecutor about the probability of being found guilty at trial, so bargaining takes place under asymmetric information. The sentence to be assigned in the event of a trial conviction is common knowledge.

Although the defendant's private information is continuously distributed, her action space consists simply of accepting or rejecting the prosecutor's offer. The lack of a one-to-one mapping between defendants' types and actions poses a challenge to the identification of the model. To overcome that challenge, I exploit information on the distribution of defendants' types conveyed by the prosecutor's offers. Specifically, there is a one-to-one mapping between the trial sentence and the prosecutor's optimal offer. If such a mapping is known by the econometrician, the distribution

[1] In 2001, 94% of Federal cases were resolved by plea bargain (Fischer, 2003). Numbers for state cases are similar. In 2000, 95% of all felony convictions in state courts were the result of a plea agreement (Durose and Langan, 2003).

of defendants' types can be recovered, using the first-order condition for the prosecutor's optimization problem. The identification of the prosecutor's optimal offer as a function of the trial sentence, however, is complicated by a selection problem. Indeed, the prosecutor's offer is observed only if plea bargaining is successful, while the trial sentence is reported only for cases resulting in a trial conviction. Nevertheless, by employing exogenous variation in the trial sentences, I show how to recover the prosecutor's optimal offer function from the available data. My empirical application uses cross-judge heterogeneity in sentencing patterns as a source of such exogenous variation in sentences.

Based on the identification strategy, I propose a non-parametric estimation procedure, which I implement using data on cases filed in the North Carolina Superior Courts between 1996 and 2009. The estimated model fits the main features of the data well. It accurately reproduces the observed settlement and conviction rates, as well as the average length of assigned sentences. I employ the estimated model to evaluate the effects of a number of policy interventions on the final outcome of criminal cases. My results indicate that a decrease in mandatory minimum sentences, which I express in the model by shortening the potential trial sentences in all cases, greatly reduces the total amount of incarceration time assigned, but slightly raises the proportion of cases that result in a conviction. Because of the latter effect, such intervention may actually raise incarceration rates in the short run. Another intervention, the broader use of non-incarceration sentences for mild cases, has little effect on the total assignment of incarceration time, but may reduce conviction rates significantly. In a third experiment, I assess the impact of eliminating plea bargains. I find that a large proportion of the defendants that currently settle their cases for incarceration sentences would be acquitted at trial if bargaining was forbidden. Depending on the characteristics of the case, such proportion is above 20 percent.

The paper is organized as follows: Section 2 discusses the contribution of this study to the literature. Section 3 describes the institutional setting and the data. In Section 4, I present the theoretical model on which I base the empirical analysis. The structural model is then presented in Section 5, followed by the empirical results, in Section 6. Section 7 contains the policy experiments, and Section 8 concludes. In the Appendix, I show details concerning the structural model and the estimation procedure. In an online Appendix, I present further empirical results and details on the manipulation of the data.

2. Related Literature

This paper contributes to a vast Law and Economics literature on settlement.[2] Most papers in that literature use game-theoretic models to investigate the litigation and resolution of civil disputes, but the same framework is readily adaptable to the analysis of plea bargaining.[3] Several studies explore the empirical implications of settlement models using data on criminal cases, with a particular interest in how the severity of potential trial sentences affects plea bargaining.[4] My investigation differs from these papers in that I develop and implement a framework for the structural analysis of data. This approach has several advantages over those of previous studies. First, I quantify the relationship between potential trial sentences and plea-bargaining results. Also, I recover objects that are not directly observable, such as the full distribution of sentences to be assigned at trial. Most importantly, I conduct policy experiments to evaluate the impact of different interventions on the justice system.

In that sense, my study is related to a group of papers that conduct the structural estimation of settlement models using data on civil cases—notably, Waldfogel (1995), Sieg (2000), Watanabe (2009) and Merlo and Tang (2015). My paper differs from those in three important aspects: First, my focus is on criminal, rather than civil, cases. Second, I conduct a careful identification analysis, showing what can be learned from the data using only restrictions derived from the theoretical model, and imposing minimal parametric assumptions on the model's primitives. Out of the papers mentioned above, only Merlo and Tang (2015) present a non-parametric identification analysis. Third, and due mainly to my focus on plea bargaining, the model estimated here departs from the ones used in previous papers. Waldfogel (1995) does not account for variation across cases in the size of trial awards, which would be the civil cases' equivalent to trial sentences. Such lack of variation prevents the investigation of how awards affect settlement decisions, which is the centerpiece of my analysis. Watanabe (2009) focuses on dynamic aspects of the bargaining process which are especially relevant in the context of tort cases. His model, for example, emphasizes

[2]The literature offers two basic explanations for why settlements fail, resulting in a costly trial. One is that agents bargain under asymmetric information (classic contributions include Bebchuk (1984); Reinganum and Wilde (1986); Nalebuff (1987); and Kennan and Wilson (1993)). The other is that agents are mutually optimistic—that is, they have divergent priors on the distribution of trial outcomes (see, for example, Priest and Klein, 1984). It is beyond the scope of my paper to compare the merits of these branches of literature. See Daughety and Reinganum (2012) for a detailed review.
[3]Theoretical papers that apply settlement models to the criminal context include Rubbinfeld and Sappington (1987), Reinganum (1988), Baker and Mezzetti (2001) and Bjerk (2007).
[4]See, for example, Elder (1989), LaCasse and Payne (1999), Boylan (2012) and Kuziemko (2006).

the differences between cases settled with and without the filing of a lawsuit, whereas the latter group of cases has no clear equivalent in the criminal justice system. The model estimated by Sieg (2000) is probably the closest to the one studied here, but, as the other papers mentioned above, some of its features make it more appropriate for the analysis of civil cases. Specifically, that model does not allow for changes in the distribution of trial awards to affect the defendant's win rate, an object of great interest in the discussion of sentencing reform. Moreover, following Nalebuff (1987), Sieg's model considers the scenario in which plaintiffs cannot take a case to trial unless their expected utility of doing so is positive. In contrast, I assume that prosecutors can commit to take cases to trial. For reasons explained later in the paper, the commitment assumption is plausible in the particular context of criminal cases. It is also worth noticing that, in the models by Waldfogel (1995), Watanabe (2009) and Merlo and Tang (2015), trials occur because agents are mutually optimistic—whereas in the present paper bargaining failures occur due to asymmetric information.

More generally, my paper fits into the empirical literature on bargaining. Besides the literature on settlement mentioned above, pioneer studies on the structural estimation of bargaining games include Eckstein and Wolpin (1995) and Merlo (1997). Recent years have seen a renewed interest in the empirical analysis of such games, especially in the presence of asymmetric information. See, for example, Merlo, Ortalo-Magné and Rust (2015), Keniston (2011) and Larsen (2014) on the estimation of bilateral trade models, and Bagwell, Staiger and Yurukoglu (2015) on tariff bargaining.

My paper also contributes to a growing empirical literature on political incentives in the justice system. Studies such as Huber and Gordon (2004), Gordon and Huber (2007), Lim (2013) and Lim, Silveira and Snyder (2016) evaluate how elections affect the sentencing behavior of trial judges. But measuring such behavior based only on observed sentences is challenging since numerous cases are resolved by plea bargain. The empirical framework presented here recovers the full distribution of trial sentences, which could be used to assess the harshness of different judges in future studies of judicial behavior.

3. Data and Institutional Details

3.1. Plea bargaining in North Carolina. I use data on criminal cases prosecuted in the North Carolina Superior Courts—the highest of the general trial courts in that state's justice system. The Superior Courts have exclusive jurisdiction over all felony cases, as well as over civil cases involving large amounts of money and misdemeanor and infraction cases appealed from a decision in the lower District Courts.

The state is divided into eight Superior Court divisions, and each division is further divided into districts for electoral and administrative purposes. There are 46 such districts statewide. A unique feature of the North Carolina Superior Courts is the rotation of judges. About 90 judges are elected for eight-year terms by voters in their districts, and may be re-elected indefinitely.[5] Although each judge belongs to a district, the state constitution mandates judges to rotate among the districts within their divisions on a regular basis. The rotation schedule for judges in the whole state is determined by the Administrative Office for the Courts. According to such schedule, judges rotate from one district to another every six months.

All Superior Court cases brought to trial are decided by a jury of 12, who determines whether the defendant is guilty. If the defendant is convicted, the judge decides on a sentence. The judge must follow a set of sentencing guidelines—i.e., a sentence is selected from a predetermined range, which depends on the severity of the crime and on the defendant's previous criminal record.[6] The sentence may include incarceration time and alternative punishments, such as fines, community service and probation.

The vast majority of criminal cases in North Carolina are resolved by plea bargain (see below for descriptive statistics on case resolution). Settling the case prevents the prosecutor and the defendant from incurring the costs of a trial. Moreover plea bargains typically involve a reduced sentence, relative to the punishment that would be assigned in the event of a trial conviction. In exchange for a guilty plea, the prosecutor can recommend a specific sentence to the judge, in a process known as sentence bargaining. Judges normally follow the prosecutors' recommendations, although they are not required to do so. Even in cases in which the judge is unlikely to accept a sentence recommendation, the prosecutor and the defendant can engage in charge

[5]A few judges may be appointed by the Governor for five-year terms.

[6]The judge sets only the minimum sentence length. The maximum length is determined according to a formula, and is roughly 120 percent of the minimum. There is no parole under the structured sentencing system in North Carolina—that is, offenders must serve at least their minimum sentence in incarceration. Whether the offender serves more than the minimum is determined by the Department of Corrections, not by the judge. For most of the period covered by this study, according to the North Carolina Sentencing and Policy Advisory Commission, convicted felons served on average 109 percent of their minimum incarceration sentences. For details, see http://www.nccourts.org/Courts/CRS/Councils/spac/Publication/Projections/Adult.asp. In the remainder of the paper, I refer to the minimum sentence simply as the sentence.

bargaining—that is, conditional on a guilty plea, the prosecutor agrees to reduce or dismiss some of the charges against the defendant.[7]

The typical timeline of a felony case is as follows: Law enforcement initiates the charges with an arrest. Within 96 hours of the arrest, a first appearance hearing takes place, in which a District Court judge schedules a probable cause hearing. The latter is held within 15 working days of the first appearance, and determines whether the state has sufficient evidence against the defendant. If the state finds sufficient evidence, the defendant is indicted (formally accused), and the case proceeds to the Superior Court.[8] Within 60 days of indictment, an administrative hearing with the Superior Court judge occurs. Most prosecutors make a plea offer prior to this hearing. At end of the administrative hearing phase the defendant must enter a plea. If the defendant pleads not guilty, the case proceeds to trial.

There are several potential sources of asymmetric information in plea bargaining. First, defendants are privately informed about their involvement with the alleged crime and the existence of incriminating or exculpatory evidence, such as possible witnesses and alibis. On the other side, prosecutors have private information about the strength of their cases against the defendants and the evidence that they already gathered. A mechanism that mitigates these informational asymmetries is pre-trial discovery, which requires the parties to share evidence with each other during negotiations. North Carolina's discovery policies grant to defendants access to most of the prosecutors' files, and figure among the broadest criminal discovery rights and duties in the country (Moore, 2012).[9] But discovery does not completely eliminate asymmetric information. In particular, defendants might still be better informed about the existence of evidence not yet gathered by prosecutors. Accordingly, the bargaining model presented in Section 4 admits private information on the defendant's side.

3.2. **Data.** In my analysis I use case-level data provided by the North Carolina Administrative Office of the Courts.[10] Such data comprise all criminal cases filed at the

[7]By itself, charge bargaining gives to prosecutors considerable control over the sentences assigned in plea bargained cases. In a survey conducted by the North Carolina Sentencing and Policy Advisory Commission, a defense attorney states that "the prosecutor looks at the desired sentence and calculates backward to determine how many charges are needed to reach the sentence." See http://www.nccourts.org/courts/crs/councils/spac/documents/disparityreportforwebr_060209.pdf.
[8]In practice, most defendants waive the probable cause hearing, and their cases proceed immediately from the the first appearance to the Superior Court.
[9]Other states that grant broad discovery rights to defendants include California, Florida, Illinois, Michigan, New Jersey and Pennsylvania (Roberts, 2003).
[10]Dyke (2007) uses the same data to analyze prosecutors' political incentives.

Superior Courts from 1996 to 2009 and include detailed information on case disposition, arrest charges and characteristics of the defendants. The data also identify the judge hearing each case.

The structural analysis to be presented in Section 5 requires a degree of homogeneity across cases in the sample. Such homogeneity cannot be achieved with samples that comprise offenses too different from each other. It is likely, for example, that the prosecution of a traffic offense is far different than that of a drug-related crime. Therefore, I restrict my attention to cases in which the main offense is a non-homicide violent crime—a category that consists of assault, sexual assault and robbery. Non-homicide violent crimes are, at the same time, homogeneous and numerous enough for estimation purposes. Unless otherwise specified, the tables and results presented throughout the paper are based on this sub-sample.

For each case, I observe the following defendant's characteristics: Gender, race, ethnicity, age and previous criminal record. The latter variable is reported in terms of points, which the North Carolina justice system assigns for the purposes of setting sentencing guidelines. In my sample, these points range from zero to 98. I also observe the type of defense counsel employed in the case. Counsel is provided by a public defender or, alternatively, by a privately-retained or court-appointed attorney.[11]

Table 1 contains descriptive statistics on the defendants. Over half of the defendants in the data are African-American, and the vast majority of the remaining ones are non-Hispanic white. Over 90 percent of the defendants are male, and their average age is roughly 29 years. Public defenders and court-appointed attorneys represent about 23 percent and 50 percent of the defendants, respectively. The online Appendix shows histograms of the defendant's age and criminal record points.

Table 2 presents descriptive statistics on case outcomes. 88.69 percent of the cases are solved by plea agreements—the majority of which result in an alternative sentence. The share of cases that settle for an incarceration sentence is 32.83 percent. Of the 9.91 percent of cases that reach trial, slightly more than half result in an acquittal. 3.91 percent of all cases result in an incarceration conviction at trial. In 83.00 percent of the cases, the main charge is a felony. In the other cases, the charge is either a misdemeanor or not known. Trial sentences tend to be longer than bargained

[11]Public defenders are full-time state employees who represent indigent defendants accused of crimes that may result in incarceration. Public defender offices are only present in 16 judicial districts in North Carolina. In the absence of such an office, indigent defendants are entitled to be represented by a court-appointed private attorney. These counsel services are provided at a subsidized rate, but not free of charge to the defendant. In the event of a conviction, the defendant is ordered to reimburse the state for the value of the counsel services.

Table 1. Descriptive statistics—Defendant's characteristics

Race, ethnicity, gender and counsel type		
Characteristic	Observations	Frequency
Race/ethnicity		
African-American	66755	56.12%
Non-Hispanic White	45618	38.35%
Hispanic	4126	3.47%
Other	2449	2.06%
Gender		
Male	110495	92.89%
Female	8453	7.11%
Counsel type		
Court-appointed	62325	52.40%
Public defender	27331	22.98%
Privately-held	25769	21.66%
Other	3523	2.96%
Defendant's age and previous criminal record		
Characteristic	Mean	Variance
Age (years)	28.88	10.64
Criminal record points	3.93	5.03

Notes: The table is based on all 118,948 cases in which the main charge is a non-homicide violent crime. The exception is the defendant's criminal record, which is only reported in the event of a conviction.

ones. The average length of the former is 100.73 months, while that for the latter is 38.98 months. Sentence dispersion is high. The standard deviations of trial and settlement sentences are 106.41 and 47.92, respectively. The focus of this study is on incarceration sentences. As shown in table 2, many of the cases in the data result in alternative punishments, such as probation and community service. In my main empirical analysis, I treat any non-incarceration sentence as no sentence whatsoever. In the online Appendix I present additional results incorporating alternative sentences.

3.3. **Lenient and harsh judges.** The identification of the structural model proposed in Section 5 requires me to observe a random variable that affects the sentences assigned, but not other aspects of the cases. In my empirical application, I use cross-judge heterogeneity in sentencing patterns as a source of such exogenous variation in sentencing. The judge rotation system in North Carolina ensures that, throughout a long enough time period, judges within the same division decide cases on the same

TABLE 2. Descriptive statistics—Case characteristics and outcomes

Distribution of cases by method of resolution		
Method of resolution	Observations	Frequency
Settlement		
Incarceration	39048	32.83%
Alternative sentence	66443	55.86%
Total	105491	88.69%
Trial conviction		
Incarceration	4654	3.91%
Alternative sentence	779	0.65%
Total	5433	4.56%
Trial acquittal / dismissed		
Absolved by jury	5049	4.24%
Dismissed by judge	1319	1.11%
Total	6368	5.35%
Dismissed by prosecutor	1656	1.39%
Total	118948	100.00%
Distribution of cases by severity of the charge		
Main arrest offense	Observations	Frequency
Felony	98726	83.00%
Other	20222	17.00%
Total	118948	100.00%
Length of incarceration sentences by method of resolution†		
Method of resolution	Mean	Standard deviation
Trial convictions	100.73	106.41
Settlement convictions	38.98	47.92

Notes: The table is based on all 118948 cases in which the main charge is a non-homicide violent crime.
† Measured in months.

locations. These judges are thus exposed to a similar pool of cases, and it is plausible to interpret systematic cross-judge heterogeneity in observed sentences as reflecting the sentencing behavior of the judges themselves. The determinants of judicial sentencing behavior are not relevant for the identification strategy, but notice that they can include the preferences of the judges and characteristics of their electorate.

To explore the observed variation in sentencing patterns, I divide the judges in the sample in two groups—lenient and harsh. Precisely, I obtain OLS estimates of the following specification:

$$sentence_i = \vartheta_{\mathbf{1}} \mathbf{X_{1i}} + \zeta_{\mathbf{1}} \; \mathbf{Judge_i} + \epsilon_{1i}, \tag{3.1}$$

where $sentence_i$ is the length of the incarceration sentence assigned in case i, $\mathbf{Judge_i}$ is a vector of dummies identifying the judge responsible for case i, $\mathbf{X_{1i}}$ is a vector of control variables and ϵ_{1i} is an error term. The control variables include the defendant's gender, racial / ethnic group, previous criminal record, age and squared age, as well as dummies indicating whether the case is settled and the type of attorney representing the defendant.[12] I also include dummies indicating the county where the case is prosecuted, the year of disposition and the severity of the main arrest offense, as defined by the North Carolina sentencing guidelines.

Column (1) of table 3 presents the results.[13] Although not the focus of the analysis, the point estimates suggest that female defendants receive sentences that are, on average, roughly four months shorter than the ones received by males. The results indicate a reduction of similar magnitude for defendants represented by private attorneys or public defenders, relative to those represented by court-assigned attorneys. The coefficients for African-American and Hispanic defendants are not statistically significant, and neither are those for age and squared age.

Table 4 contains information on the distribution of estimated judge fixed effects from specification (3.1). The mean fixed effect is 31.58 months, and the median is 31.51 months. The heterogeneity across judges is substantial. The standard deviation and interquartile range of the fixed effects are, respectively, 6.26 and 8.06 months. I divide the 169 judges in the data into two groups. Those whose estimated fixed effects are strictly smaller than the median are classified as lenient, while the others are classified as harsh. Averaging across judges within each group, the estimated fixed effects for lenient and harsh judges are 25.97 and 36.21, respectively.

To provide empirical support for the assumption that the cross-judge variation in sentencing patterns is independent from other aspects of the cases, I verify whether the distribution of observable case characteristics varies substantially across lenient

[12]The defendant's previous criminal record is incorporated as follows: Based on the criminal record points reported for each defendant and on the criteria employed by the North Carolina justice system for the determination of sentencing guidelines, I divide the cases in the data into six criminal record levels. I then add to the specifications dummies indicating the level to which the case belongs.

[13]The table presents coefficient estimates for selected variables. The complete set of estimates is available upon request

Table 3. Sentence length and judge assignment

	(1)	(2)
	Dependent Variable	
	sentence†	*harsh*
settled	-39.6722	–
	(1.3054)	
age	-0.0492	0.0002
	(0.1989)	(0.0009)
age^2	0.0012	0.0000
	(0.0029)	(0.0000)
female	-4.3054	-0.0138
	(0.7551)	(0.0073)
black	0.9320	-0.0067
	(0.6792)	(0.0109)
hispanic	-1.4155	-0.0092
	(1.2558)	(0.0144)
private attorney	-3.7823	0.0014
	(0.7492)	(0.0118)
public defender	-3.5356	-0.0344
	(0.7401)	(0.0290)
Judge dummies	Yes	No
County dummies	Yes	No
Superior Court division dummies	No	Yes
Observations	36644	97942
R^2	0.5481	0.2789

Notes: Both specifications estimated by OLS. The dummy *harsh* indicates whether the case is decided by a harsh judge, as explained in the text.
Other controls: Disposition year, offense severity and defendant's criminal record.
Standard errors (robust to clustering at the judge level) in parenthesis.
† Measured in months.

Table 4. Variation of sentencing behavior across judges

Estimated judge fixed effects†						
mean	std. dev.	min.	1st quart.	median	3rd quart.	max.
31.58	6.26	16.28	27.56	31.51	35.62	64.63
Number of judges: 169						

Notes: Fixed effects based on OLS estimates of specification (3.1).
† Measured in months.

and harsh judges. I consider the following specification:

$$harsh_i = \vartheta_{\mathbf{2}} \mathbf{X_{2i}} + \epsilon_{2i}, \tag{3.2}$$

where $harsh_i$ is a dummy indicating whether the judge responsible for case i is a harsh one, $\mathbf{X_{2,i}}$ is a vector of controls and ϵ_{2i} is an error term. The control variables are similar to the ones in specification (3.1)—except that they lack a dummy indicating settlement and, instead of county-specific dummies, include dummies for the Superior Court division where the case is prosecuted.[14] All these variables are arguably out of the judge's control. Column (2) in table 3 contains OLS estimates of this specification. All of the point estimates are very close to zero and only the coefficient associated with the defendant's gender is significant at the ten precent level. These results suggest that the cases in the data do not differ considerably across the two judge categories.[15]

4. The Model

4.1. **Setup.** In this section, I briefly describe Bebchuk's (1984) settlement model, which will guide my structural analysis. Two agents—a prosecutor and a defendant—bargain over the outcome of a case. The agents' utility functions are linear in the sentence assigned to the defendant. Specifically, the defendant wants to minimize the sentence, whereas the prosecutor wants to maximize it.[16]

If bargaining fails, the case is brought to trial, where the defendant is found guilty with probability Θ. This probability is drawn at the beginning of the game from a distribution F. I make the following technical assumptions:[17] F has support $(\underline{\theta}, \bar{\theta}) \subseteq (0, 1)$ and is twice differentiable. The associated density function f is strictly positive

[14] I control for the division as this is the level at which judges are comparable, due to rotation.

[15] An F-test rejects the null hypothesis that all variables are jointly insignificant. But this result is due solely to a dummy indicating whether the case is a class three misdemeanor, the least severe class of misdemeanors in North Carolina. Curiously, the coefficient for this variable is positive and significant—indicating that, if anything, light cases tend be assigned to harsh judges. Once I exclude this dummy from the F-test, I fail to reject the null that all other variables are jointly insignificant. In the online Appendix, I show that I obtain similar results if, instead of the harsh judge indicator, I employ the estimated judge fixed effects from (3.1) as the dependent variable in (3.2).

[16] Alschuter (1968) provides several pieces of anecdotal evidence that prosecutors mostly play the role of advocates in the process of plea bargaining. The author explains that "...in this role, the prosecutor must estimate the sentence that seems likely after a conviction at trial, discount this sentence by the possibility of an acquittal, and balance the "discounted trial sentence" against the sentence he can ensure through a plea agreement." The assumption that the prosecutor's utility is linear in the assigned sentence captures such behavior in a simple way. In the online Appendix, I consider an extension of the model, which allows for non-linearities in the defendant's utility function.

[17] These assumptions ensure that, without conditioning on Θ, the probability of settlement is strictly between zero and one.

on $(\underline{\theta}, \bar{\theta})$ and is non-increasing in a neighborhood of $\bar{\theta}$. Moreover, to rule out multiple equilibria, I assume that the hazard rate $f/[1-F]$ is strictly increasing in θ.

Only the defendant knows the realization θ of Θ. The assumption that the defendant has private information on the probability of being convicted at trial is motivated by noticing that most defendants know whether they are guilty. As Scott and Stuntz (1992) point out, such knowledge may assume several forms. A defendant may, for example, not be involved in the crime at all. Alternatively, the defendant may be involved, but may have lacked the requisite criminal intent. Regardless of guilt, the defendant may have more information than the prosecutor on the quality of the available evidence (e.g., what, exactly, a given witness knows).[18]

In the event of a conviction at trial, a common knowledge sentence t is assigned. The common knowledge of t is interpreted as follows: During bargaining, the prosecutor and the defendant know enough about the case to determine exactly what sentence the judge will assign if the defendant is found guilty at trial. Such information comes from the arrest offenses, the defendant's criminal record, sentencing guidelines, the judge's previous sentencing decisions, etc. The prosecutor and the defendant, respectively, pay costs c_p and c_d for reaching the trial stage. For the prosecutor, these are the opportunity costs of taking a case to trial. For the defendant, the costs represent attorney and court fees. c_p and c_d are common knowledge to both players.

The bargaining protocol is take-it-or-leave-it, so the model constitutes a screening game. The prosecutor offers the defendant to settle for a sentence s.[19] If the defendant accepts it, the game ends. Payoffs are then s for the prosecutor and $-s$ for the defendant. If the defendant rejects the offer, the case reaches trial. The trial payoffs for the prosecutor are $t - c_p$ if the defendant is found guilty, and $-c_p$ otherwise.

[18]The assumption that only the defendant has private information is supported by the discovery rules in North Carolina, as discussed in Section 3. Despite these rules, the prosecutor is likely to have more experience with the justice system than the defendant, which would motivate the consideration of private information on the prosecutor's side as well. My modeling assumption is arguably more plausible for defendants who have had previous contact with the criminal justice system. The empirical results that I present in the main text are for defendants with relatively short criminal records, as I formally define in Section 6, but the online Appendix contains results for defendants with long records. Reassuringly, the empirical results do not vary systematically across these groups of defendants, suggesting that my assumption of one-sided asymmetric information is a good approximation of reality for all defendants. Still, a more realistic model would allow for private information on the prosecutor's side. Such an extension, however, would make the empirical analysis of the model impractical. See Friedman and Wittman (2007) and the references therein for models of settlement with two-sided asymmetric information.

[19]I assume that the prosecutor can employ either sentence bargaining or careful charge bargaining to set an offer s. As described in Section 3, plea bargaining practices in North Carolina give to prosecutors substantial discretion over the sentences assigned in settled cases.

For the defendant, the trial payoffs are $-t - c_d$ in the case of a conviction, and $-c_d$ otherwise. Whereas imposing a take-it-or-leave-it bargaining protocol is not without loss of generality, notice that, under fairly general conditions, that procedure is optimal from the prosecutor's perspective (Spier, 1992).

I assume that the prosecutor can commit to take a case to trial even if her expected utility of doing so is negative. Such an assumption is justified in the criminal law context, as prosecutors are recurrent players in the plea-bargaining game. Reputational effects thus make the commitment to take cases to trial credible.[20] Later, in my empirical analysis, I clarify how I deal with dropped cases observed in the data.

4.2. **Equilibrium.** The relevant equilibrium concept is subgame perfection, and the game is solved by backward induction. The defendant accepts a prosecutor's offer s if and only if $s \leq \theta t + c_d$. Therefore, for every value s chosen by the prosecutor, the defendant's strategy is characterized by a cutoff

$$\theta(s) = \frac{s - c_d}{t} \tag{4.1}$$

such that the defendant accepts the prosecutor's offer if and only if the probability of conviction at trial is greater than or equal to $\theta(s)$.

The prosecutor then solves the following problem:

$$\max_{s} \quad \{1 - F[\theta(s)]\} s + F[\theta(s)] \left\{ -c_p + t \frac{\int_{\underline{\theta}}^{\theta(s)} x f(x) dx}{F[\theta(s)]} \right\}.$$

Bebchuk (1984) shows that the optimal prosecutor's offer s^* satisfies $\theta(s^*) \in (\underline{\theta}, \bar{\theta})$, so that the prosecutor's first-order condition, presented below, holds with equality

$$\frac{t}{c_p + c_d} = \frac{f[\theta(s^*)]}{\{1 - F[\theta(s^*)]\}}. \tag{4.2}$$

Equations (4.1) and (4.2) characterize the equilibrium. Without conditioning on Θ, the equilibrium probability that the prosecutor's offer is rejected is given by $F[\theta(s^*)]$ and is strictly between zero and one. The probability of conviction is the sum of the probabilities of settlement and conviction at trial, and is given by

$$1 - F[\theta(s^*)] + \int_{\underline{\theta}}^{\theta(s^*)} x f(x) dx \quad = \quad 1 - \int_{\underline{\theta}}^{\theta(s^*)} (1 - x) f(x) dx. \tag{4.3}$$

[20] Notice that the model only implicitly incorporates these reputational effects by assuming that the prosecutor can credibly commit to a trial. Explicitly modeling the prosecutor's reputation formation is not within the scope of the present paper, but it would be an interesting avenue for further research.

4.3. **Empirical implications.** Keeping the trial costs c_p and c_d constant, I perform a comparative static analysis by letting the trial sentence t vary. Define the functions $\tilde{s}(\cdot)$ and $\tilde{\theta}(\cdot)$, which, for any trial sentence t, respectively return the equilibrium prosecutor's offer and the equilibrium cutoff point for the defendant. The domain of both functions is $\Re_{++}$. Bebchuk (1984) shows that both $\tilde{s}(\cdot)$ and $\tilde{\theta}(\cdot)$ are strictly increasing, and $\tilde{s}(\cdot)$ is strictly convex.[21]

Since $\tilde{\theta}(\cdot)$ is strictly increasing and the density f is strictly positive over the whole support of Θ, the probability that the defendant rejects the prosecutor's offer decreases strictly with the trial sentence t. Furthermore, since $\tilde{\theta}(\cdot)$ and $\tilde{s}(\cdot)$ are strictly increasing and the defendant's type is always lesser than one, the right-hand side of (4.3) implies that the probability of conviction also decreases strictly with t.

In the online Appendix, I present reduced-form results indicating that the data are consistent with these empirical implications of the model. In particular, I show evidence that the settlement of a case becomes less likely as the sentence to be assigned in the event of a trial conviction increases.

5. Structural analysis

This section contains the structural analysis of the data, based on the model outlined in Section 4. I first describe the data-generating process, which adapts the model to the institutional setting presented in Section 3. Then, I discuss the identification of the model and propose an estimator for it.

5.1. **Data-generating process.**

5.1.1. *Primitives.* For every case i, let the random variable Z_i represent the judge responsible for the case. Assume that, for all i, the support of Z_i consists of only two elements—lenient ($Z = l$) and harsh ($Z = h$) judges, as described in section 3.[22] Notice that all the primitives defined below are allowed to depend on observable case characteristics, such as the defendant's race, ethnicity and criminal background. To make the notation as clean as possible, I do not explicitly condition the model

[21] That $\tilde{\theta}(\cdot)$ is strictly increasing follows from (4.2) and the assumption that the hazard rate for the defendants' type is strictly increasing. (4.1) then implies the monotonicity and convexity of $\tilde{s}(\cdot)$.

[22] I can trivially extend the identification strategy below to consider a finer classification of judges. The advantages of employing a binary classification here are twofold: First, it simplifies the notation. Second, it is consistent with the empirical application in Section 6. The non-parametric nature of my estimation procedure and the limited sample size make it important to maximize the number of judges within each group—which I achieve by dividing the judges in only two categories.

primitives on these characteristics in the present section. However I do account for observable case heterogeneity in the estimation procedure.

To each case i corresponds a potential incarceration sentence, which is assigned if the defendant is convicted at trial. This sentence is described by the random variable T_i, which, conditional on Z_i, is i.i.d. across cases. T_i follows a conditional mixture distribution. With probability $\nu(Z_i)$, it is equal to zero, which means that, in the event of conviction at trial, a non-incarceration sentence is assigned. With probability $1-\nu(Z_i)$, T_i is distributed according to the CDF $G(\cdot|Z_i)$, defined on the interval $[\underline{t}, \bar{t}]$, where $\underline{t} > 0$. Assume that $G(\cdot|Z_i)$ is continuous, and denote by $g(\cdot|Z_i)$ the associated conditional pdf. Moreover, assume that $g(t|Z_i) > 0$ for all $t \in [\underline{t}, \bar{t}]$. The realization of T_i is common knowledge to the defendant and the prosecutor.

The defendant's probability of conviction at trial also varies across cases. For case i, such a probability is described by the random variable Θ_i, with support $(\underline{\theta}, \bar{\theta}) \subseteq (0, 1)$. Assume that Θ_i and T_i are independent. Assume, also, that Θ_i is i.i.d. across cases, and let $F(\cdot)$ be its distribution. Importantly, the latter assumption implies that $F(\cdot)$ is constant on Z_i.[23] Make the following technical assumptions: $F(\cdot)$ is twice differentiable, and the associated density, denoted by $f(\cdot)$, is strictly positive over the whole support and non-increasing on a neighborhood of $\bar{\theta}$. Moreover, assume that the hazard rate $f(\cdot)/\left[1 - F(\cdot)\right]$ is strictly increasing.

Let c_d and c_p be the trial costs for the defendant and the prosecutor, respectively. I assume that these costs are deterministic, conditional on T_i. In particular, they do not depend on Z_i. In my empirical application, I allow c_d and c_p to vary with T_i. Here, to simplify the exposition, I assume that the costs are constant. Notice that these are the costs expected by the agents during negotiations, so it is possible that the actual trial costs vary across cases.

Throughout the main text, I maintain the assumption that the utility functions of the defendant and the prosecutor are linear in the assigned sentence.[24] The primitives of the structural model are then: (i) the distribution of potential trial sentences, characterized by $\nu(Z_i)$ and $G(\cdot|Z_i)$; (ii) the distribution of defendants' types, given by the $F(\cdot)$; (iii) the trial costs c_d and c_p; and (iv) the distribution of Z_i.

---

[23]In the online Appendix, I show that I can relax the assumption of independence between Θ_i and T_i and still partially identify the model. In particular, I am able to address the case in which Θ_i and T_i are positively correlated. Since the distribution of T_i depends on Z_i, this more general version of the model allows $F(\cdot)$ to vary across judges.

[24]In the online Appendix, I estimate an extension of the model, which relaxes the non-linearity assumption for the defendant's utility. The estimation results obtained using the extended model are qualitatively similar to the ones presented in the main text.

5.1.2. *Observables.* For each case i, given realizations t_i of T_i and θ_i of Θ_i, as well as $F(\cdot)$, c_d and c_p, the prosecutor and defendant play the game described in Section 4. Equations (4.1) and (4.2) define the functions $\tilde{\theta}(\cdot)$ and $\tilde{s}(\cdot)$. Such functions, respectively, return the defendant's equilibrium cutoff point and the prosecutor's equilibrium offer, given a trial sentence t. Assume that $\tilde{\theta}(0) = 0$ and $\tilde{s}(0) = 0$. Hence, if a non-incarceration sentence is to be assigned at trial, then the prosecutor offers to settle for a non-incarceration sentence, and the case settles with certainty.[25]

The prosecutor's offer for case i is thus described by the random variable $S_i = \tilde{s}(T_i)$. Conditional on Z_i, S_i is i.i.d. across cases, and follows a mixed distribution. With probability $\nu(Z_i)$, it equals zero. With probability $1 - \nu(Z_i)$, S_i follows the CDF $B(\cdot|Z_i)$, which has support $[\tilde{s}(\underline{t}), \tilde{s}(\bar{t})]$. Notice that, as $G(\cdot|Z_i)$ and $\tilde{s}(t)$ are continuous, $B(\cdot|Z_i)$ is also continuous. Let $b(\cdot|Z_i)$ be its associated density.

A selection process complicates the identification of the model. The realization t_i of T_i is observable to the econometrician only in the event of a conviction at trial. Similarly, the realization s_i of S_i is available only for settled cases. Formally, let Ψ_i be a discrete random variable describing the way case i is resolved. Define

$$\Psi_i = \begin{cases} 0 & \text{if the case is dropped or settled for an alternative sentence} \\ 1 & \text{if the case is settled for an incarceration sentence} \\ 2 & \text{if the case results in an incarceration conviction at trial} \\ 3 & \text{if the defendant is found not guilty at trial.} \end{cases}$$

The observables for each case are (i) the realization ψ_i of Ψ_i; (ii) the realization z_i of Z_i; (iii) the realization t_i of T_i, if and only if $\psi_i = 2$; and (iv) the realization s_i of S_i, if and only if $\psi_i = 1$. To identify the model's primitives, I must account for such a selection problem. In the remainder of this section, I refer to a settlement or a successful plea bargain only if $\Psi_i = 1$. Similarly, by a trial conviction, I mean $\Psi_i = 2$. Moreover, for ease of notation, I omit the index i when I refer to a specific case.

5.2. **Identification.** Since Θ is distributed continuously, and the defendants' action space consists of merely accepting or rejecting the prosecutor's offer, there is no bijection between the defendants' types and actions. Thus, techniques often employed by the auctions literature for the identification of private types cannot be used here.[26]

[25] In my main analysis, I make no distinction between cases that settle for an alternative sentence and cases that the prosecutor drops. See the online Appendix for an alternative treatment of non-incarceration sentences. Notice that the model does not allow for the assignment of alternative sentences at trial. As shown in table 2, that is a rare event.

[26] See Athey and Haile (2007) for an introduction to the identification of auction models.

But the prosecutor's equilibrium offer conveys information on the distribution $F(\cdot)$. As shown below, I explore such information to recover the model's primitives.

The strategy described here consists of two main parts. First, I identify the prosecutor's optimal offer function $\tilde{s}(\cdot)$. Then, using $\tilde{s}(\cdot)$, I show how to recover $G(\cdot|Z=z)$, $F(\cdot)$, c_d and c_p. Since Z is always observed, the identification of its distribution is trivial. The same comment holds for $\nu(z)$, as $P\left[\Psi=0|Z=z\right]$ is observable.

Consider the function

$$\phi(t) \equiv \frac{g\left(t|Z=h\right)}{g\left(t|Z=l\right)}.$$

Assume that $\phi(\cdot)$ is piecewise strictly monotone. Similar to the conditions for an instrumental variable in the context of regression analysis, the assumptions that $F(\cdot)$, c_d and c_p do not depend on Z (up to observable case characteristics) and that $\phi(\cdot)$ is piecewise strictly monotone can be regarded as exclusion and inclusion restrictions, respectively. That is, the assumptions require that the variation of Z affects the distribution of potential trial sentences, but not the other primitives in the model. Like the exclusion condition for an instrumental variable, the assumption that $F(\cdot)$, c_d and c_p do not depend on Z cannot be directly tested. But, as discussed in Section 3, the judge rotation system in North Carolina ensures that different judges in the same Superior Court division decide on similar cases. Therefore, it is valid to employ differences in the sentencing patterns between lenient and harsh judges within the same division as a source of variation in the distribution of trial sentences.

5.2.1. *Prosecutor's settlement offer function.* My first main result states that, under the conditions defined above, the prosecutor's optimal offer function $\tilde{s}(\cdot)$ is identified. I present the proof of the proposition in the main text since it is important for the understanding of the estimation procedure proposed later in the paper.

PROPOSITION 1. *If the function $\phi(\cdot)$ is piecewise strictly monotone and the primitives $F(\cdot)$, c_d and c_p do not depend on the judge Z, the prosecutor's optimal offer function $\tilde{s}(\cdot)$ is identified over the whole interval $[\underline{t},\bar{t}]$.*

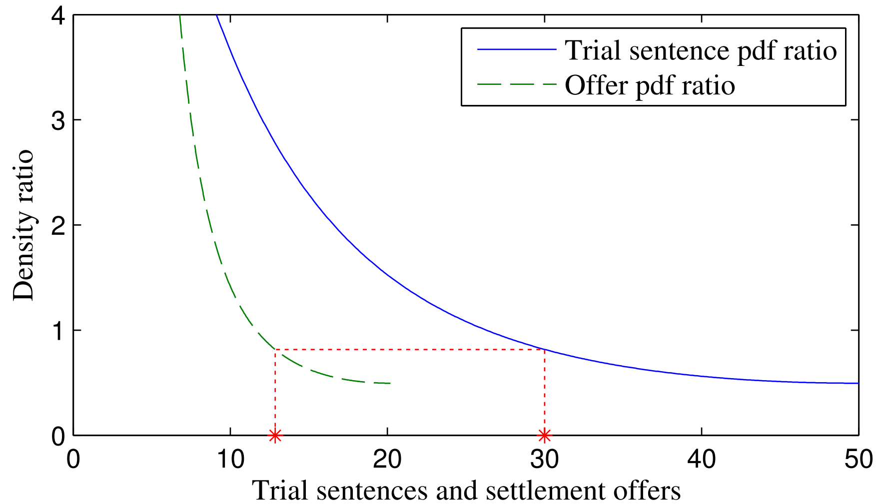
Proof. Since $\tilde{s}(\cdot)$ is strictly increasing in t, I can write

$$b(s|Z=z) = g\left(\tilde{s}^{-1}(s)|Z=z\right)\frac{d\tilde{s}^{-1}(s)}{ds} \tag{5.1}$$

for all $s \in [\tilde{s}(\underline{t}), \tilde{s}(\bar{t})]$ and all $z \in \{l,h\}$. Therefore, for all $t \in [\underline{t},\bar{t}]$, I have that

$$\frac{b\left[\tilde{s}(t)|Z=h\right]}{b\left[\tilde{s}(t)|Z=l\right]} = \frac{g\left(t|Z=h\right)}{g\left(t|Z=l\right)}. \tag{5.2}$$

Figure 1. Settlement offer function identification: Example



Note: This figure employs simulated data to illustrate the identification of the settlement offer function from (5.2). The blue and green lines represent the density ratios of trial sentences and settlement offers, respectively. As shown by the red dotted line, the offer associated with a trial sentence of 30 is 12.85.

Notice that, on the right-hand side of (5.2), I have $\phi(t)$. This equation is key to the identification of $\tilde{s}(t)$. To understand the argument, it is useful to first consider the case in which $\phi(\cdot)$ is (non-piecewise) strictly monotone. Under the strict monotonicity of $\phi(\cdot)$ and $\tilde{s}(\cdot)$, the function

$$\phi_S(s) \equiv \frac{b\,(s|Z=h)}{b\,(s|Z=l)} \tag{5.3}$$

is also strictly monotone. Thus, if $g\,(\cdot|Z=z)$ and $b\,(\cdot|Z=z)$ were observed for $z \in \{l, h\}$, I would be able to apply $\phi_S^{-1}(\cdot)$ to both sides of (5.2) in order to recover $\tilde{s}(t)$. To illustrate the process, figure 1 depicts density ratios of trial sentences and settlement offers calculated from simulated data. Using (5.2), I match any given point in the support of trial sentences to the point in the support of settlement offers that has the same density ratio, as shown by the red dotted line. By repeating the procedure for every point in the trial sentences' support, I identify $\tilde{s}(\cdot)$.

To verify that the argument holds if $\phi(\cdot)$ is piecewise strictly monotone, consider a partition of $[\underline{t}, \bar{t}]$ into N sub-intervals of the form $I_1^T = [\underline{t}, t_1)$, $I_2^T = [t_1, t_2)$, $\cdots$, $I_{N-1}^T = [t_{N-2}, t_{N-1})$, $I_N^T = [t_{N-1}, \bar{t}]$, and assume that, within each such sub-interval, $\phi(\cdot)$ is strictly monotone. Then, from (5.2) and the strict monotonicity of $\tilde{s}(\cdot)$, there is also a partition of $[\tilde{s}(\underline{t}), \tilde{s}(\bar{t})]$ into N sub-intervals $I_1^S, \cdots, I_N^S$ such that:

(i) $I_1^S = [\tilde{s}(\underline{t}), \tilde{s}(t_1))$, $I_2^S = [\tilde{s}(t_1), \tilde{s}(t_2))$, $\cdots$,$I_{N-1}^S = [\tilde{s}(t_{N-2}), \tilde{s}(t_{N-1}))$ and $I_N^S = [\tilde{s}(t_{N-1}), \tilde{s}(\bar{t})]$; and (ii) for all $n = 1, \cdots, N$, as $\phi(\cdot)$ is strictly increasing (decreasing) on I_n^T, then $\phi_S(\cdot)$ is strictly increasing (decreasing) on I_n^S. Therefore, I can apply the inversion of $\phi_S(\cdot)$ to both sides of (5.2) and recover $\tilde{s}(\cdot)$ over $[\underline{t}, \bar{t}]$.

As shown above, $\tilde{s}(\cdot)$ would be identified under the piecewise monotonicity of $\phi(\cdot)$ if $g(\cdot|Z = z)$ and $b(\cdot|Z = z)$ were known for $z \in \{l, h\}$. However, due to selection, none of these densities is observed directly. Still, I can employ the observed, censored versions of the distributions to obtain the density ratios in (5.2). Consider, first, the settlement offers. I observe $b(\cdot|\Psi = 1, Z = z)$, the distribution of accepted settlement offers. Given any z and any value s, the equilibrium conditional probability of a successful plea bargain is

$$P[\Psi = 1|S = s, Z = z] = 1 - F\left[\tilde{\theta}\left(\tilde{s}^{-1}(s)\right)\right]. \tag{5.4}$$

From Bayes' rule, I have that

$$b(s|\Psi = 1, Z = z) = \frac{P[\Psi = 1|S = s, Z = z]b(s|Z = z)}{P[\Psi = 1|Z = z]}. \tag{5.5}$$

Notice that the support of $b(s|\Psi = 1, Z = z)$ is still $[\tilde{s}(\underline{t}), \tilde{s}(\bar{t})]$.[27] Notice also that $P[\Psi = 1|S = s, Z = h] = P[\Psi = 1|S = s, Z = l]$ for all s, so I can write

$$\frac{b(s|\Psi = 1, Z = h)}{b(s|\Psi = 1, Z = l)} \frac{P[\Psi = 1|Z = h]}{P[\Psi = 1|Z = l]} = \frac{b(s|Z = h)}{b(s|Z = l)} \tag{5.6}$$

for all $s \in [\tilde{s}(\underline{t}), \tilde{s}(\bar{t})]$. Therefore, I recover the density ratio on the left-hand side of (5.2), using the (observable) censored distribution of accepted settlement offers.

Turning to the potential trial sentences, $g(\cdot|\Psi = 2, Z = z)$ is available for all z. For any t and $z \in \{l, h\}$, the probability of conviction at trial is

$$P[\Psi = 2|T = t, Z = z] = \int_{\underline{\theta}}^{\tilde{\theta}(t)} x\ f(x)\, dx. \tag{5.7}$$

Using Bayes' rule once more, I have that

$$g(t|\Psi = 2, Z = z) = \frac{P[\Psi = 2|T = t, Z = z]g(t|Z = z)}{P[\Psi = 2|Z = z]}. \tag{5.8}$$

[27]The defendant's equilibrium cutoff $\tilde{\theta}\left(\tilde{s}^{-1}(s)\right)$ belongs to the interval $(0, 1)$ for all values of s, so $P[\Psi = 1|S = s, Z = z]$ is always strictly positive.

The support of $g(t|\Psi = 2, Z = z)$ is still $[\underline{t}, \bar{t}]$.[28] Since $P[\Psi = 2|T = t, Z = l] = P[\Psi = 2|T = t, Z = h]$, I can write

$$\frac{g(t|\Psi = 2, Z = h)}{g(t|\Psi = 2, Z = l)} \frac{P[\Psi = 2|Z = h]}{P[\Psi = 2|Z = l]} = \frac{g(t|Z = h)}{g(t|Z = l)} \tag{5.9}$$

for all $t \in [\underline{t}, \bar{t}]$. I thus use the (observable) censored distribution of potential trial sentences for cases where the defendant is convicted at trial in order to identify the density ratio on the right-hand side of (5.2). From (5.2), (5.6) and (5.9), I can then use the piecewise strict monotonicity of $\phi(\cdot)$ to recover $\tilde{s}(t)$. □

5.2.2. *Completing the identification of the model.* I now show that, once $\tilde{s}(\cdot)$ is identified, I can recover the model's primitives. Intuitively, the prosecutor's optimal offer function conveys information about the distribution of defendants' private types, which allows the identification of the whole model. Again, I present the proof in the main text as it helps understanding the estimation procedure.

PROPOSITION 2. *Assume that the prosecutor's optimal offer function $\tilde{s}(\cdot)$ is identified over the whole interval $[\underline{t}, \bar{t}]$. Then the following objects are identified: (i) the distribution function $G(\cdot|Z = z)$, over the whole interval $[\underline{t}, \bar{t}]$; (ii) the distribution of defendants' types $F(\cdot)$, over the interval $[\tilde{\theta}(\underline{t}), \tilde{\theta}(\bar{t})]$; and (iii) the trial costs c_d and c_p.*

Proof. Using the defendant's cutoff point in equation (4.1), I have that

$$\tilde{\theta}(t) = \frac{\tilde{s}(t) - c_d}{t} \tag{5.10}$$

for all $t \in [\underline{t}, \bar{t}]$, so $\tilde{\theta}(\cdot)$ is identified, up to the constant c_d.

I can now rewrite the prosecutor's first-order condition in (4.2) as

$$\frac{t}{c_p + c_d} = \frac{f\left[\tilde{\theta}(t)\right]}{1 - F\left[\tilde{\theta}(t)\right]}.$$

Denote by $\lambda(\theta)$ the hazard function of $F(\cdot)$ evaluated at θ. The right-hand side of the expression above is $\lambda(\tilde{\theta}(t))$. Since $\tilde{\theta}(\cdot)$ is strictly increasing, I have that

$$\lambda(\theta) = \frac{\tilde{\theta}^{-1}(\theta)}{c_p + c_d} \tag{5.11}$$

[28]Indeed, $\tilde{\theta}(t) \in (0, 1)$ for all t, and $f(\theta) > 0$ for all $\theta \in (\underline{\theta}, \bar{\theta})$, so that $P[\Psi = 2|T = t, Z = z]$ is always strictly positive.

for every θ in the interval $[\tilde{\theta}(\underline{t}), \tilde{\theta}(\bar{t})]$. Using the relationship between the hazard and the distribution functions of Θ, I can then write

$$F(\theta) = 1 - \mu \exp\left(-\int_{\tilde{\theta}(\underline{t})}^{\theta} \lambda(x)\, dx\right) \tag{5.12}$$

for all $\theta \in [\tilde{\theta}(\underline{t}), \tilde{\theta}(\bar{t})]$, where

$$\mu = \exp\left(-\int_{\underline{\theta}}^{\tilde{\theta}(\underline{t})} \lambda(x)\, dx\right).$$

Together, (5.10), (5.11) and (5.12) identify $F(\cdot)$ for the whole image of $\tilde{\theta}(\cdot)$, up to the constants c_p, c_d and μ.[29] Using (5.1), (5.4), (5.5) and (5.12), I have that

$$g\left(t|Z=z\right) = \frac{b\left(\tilde{s}(t)|\Psi=1, Z=z\right) P[\Psi=1|Z=z]}{\mu \exp\left(-\int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} \lambda(x)\, dx\right)} \frac{d\tilde{s}(t)}{dt} \tag{5.13}$$

for $z \in \{l, h\}$. Thus the density of potential trial sentences can be recovered, up to c_p, c_d and μ. By integrating both sides of (5.13) over $t \in [\underline{t}, \bar{t}]$, I solve for μ in terms of c_p and c_d. Therefore, $F(\cdot)$ and $g(\cdot|Z=z)$ are identified, up to c_p and c_d. The following lemma concludes the proof of the proposition:

Lemma 1. *The trial costs c_d and c_p are uniquely identified.*

See the Appendix for the proof. (5.10) provides intuition for the separate identification of the trial costs for the defendant. For any trial sentence t, c_d is equal to the difference between $\tilde{s}(t)$ and $\tilde{\theta}(t)t$, where the latter term represents the trial outcome expected by the defendant that is just indifferent between accepting and rejecting the prosecutor's offer. Taking $\tilde{s}(t)$ as given, greater c_d thus implies lower $\tilde{\theta}(t)$—which in turn is associated with a lower probability of conviction, conditional on a trial. Therefore, the defendant's trial win rate helps separately identifying c_d. The prosecutor's costs c_p are then recovered by the settlement probability.

$\square$

5.3. **Estimation.** My estimation procedure closely follows the identification strategy. First I estimate the function $\tilde{s}(\cdot)$. Then, I use the estimated $\tilde{s}(\cdot)$ to obtain estimators for the model's primitives.

[29] The term μ depends on the behavior of $F(\cdot)$ for values of θ lower than $\tilde{\theta}(\underline{t})$. It can be rewritten as $\mu = 1 - F\left[\tilde{\theta}(\underline{t})\right]$.

5.3.1. *Estimation of $\tilde{s}(\cdot)$.* For every $z \in \{l, h\}$, I obtain the estimates

$$\begin{aligned} & \hat{P}[\Psi=\psi|Z=z] && \text{for} \quad P[\Psi=\psi|Z=z] \quad ,\psi \in \{0;1;2;3\} \\ & \hat{g}(t|\Psi=2,Z=z) && \text{for} \quad g(t|\Psi=2,Z=z) \\ \text{and} \qquad & \hat{b}(s|\Psi=1,Z=z) && \text{for} \quad b(s|\Psi=1,Z=z). \end{aligned}$$

The estimates for the distribution of Ψ are trivially computed. I estimate $g(t|\Psi = 2, Z = z)$ and $b(s|\Psi = 1, Z = z)$ by kernel smoothing, which poses two challenges. First, I must deal with the boundedness of $[\underline{t}, \bar{t}]$. To do so, I employ a boundary-correction method recently developed by Karunamuni and Zhang (2008). Broadly speaking, the method generates artificial data beyond the boundaries of the original distribution's support. A kernel estimator applied to this enlarged data is uniformly consistent on the entire support of the original distribution.[30] The second challenge consists of incorporating observed case-level heterogeneity into the estimators. I deal with this problem by, first, assuming that all observed heterogeneity variables are discrete and, then, employing smoothing techniques proposed by Li and Racine (2007) for the estimation of conditional densities. This procedure divides the space of case-level characteristics into a finite number of subsets and estimates the densities for each one of them. By smoothing across the subsets, the estimator employs all the available data. See Appendix A.2 for more details.

Equation (5.2) implies that

$$\int_{\underline{t}}^{\bar{t}} \left\{ \frac{b\left[\tilde{s}(t)|\Psi=1,Z=h\right]}{b\left[\tilde{s}(t)|\Psi=1,Z=l\right]} \frac{P[\Psi=1|Z=h]}{P[\Psi=1|Z=l]} \right. \\ \left. - \frac{g\left(t|\Psi=2,Z=h\right)}{g\left(t|\Psi=2,Z=l\right)} \frac{P[\Psi=2|Z=h]}{P[\Psi=2|Z=l]} \right\}^2 dH(t) = 0$$

for any measure H over $[\underline{t}, \bar{t}]$. Denote the shortest and longest observed trial sentences by $\hat{\underline{t}}$ and $\hat{\bar{t}}$, respectively, and notice that such values are consistent estimates for $\underline{t}$ and $\bar{t}$. Similarly, denote by $\hat{\underline{s}}$ and $\hat{\bar{s}}$, respectively, the shortest and longest observed plea-bargained sentences, and notice that these are consistent estimates for $\tilde{s}(\underline{t})$ and $\tilde{s}(\bar{t})$. Remember that $\tilde{s}(\cdot)$ is increasing and convex, and let Υ be the space of increasing and convex functions $s(\cdot)$ over $[\underline{t}, \bar{t}]$ such that $s(\hat{\underline{t}}) = \hat{\underline{s}}$ and $s(\hat{\bar{t}}) = \hat{\bar{s}}$. The estimator

[30]The bandwidth for most of the support is chosen by Silverman's "rule-of-thumb" (Silverman, 1986). Near the boundaries, a modified bandwidth is employed. See Karunamuni and Zhang (2008) for details. I use the Epanechnikov kernel function. The same correction is applied to the non-parametric estimation of first-price auction models by Hickman and Hubbard (2012).

$\hat{s}(\cdot)$ for the function $\tilde{s}(\cdot)$ solves

$$\min_{s\in\Upsilon} \int_{\underline{t}}^{\bar{t}} \left\{ \frac{\hat{b}\left[s(t)|\Psi=1,Z=h\right]}{\hat{b}\left[s(t)|\Psi=1,Z=l\right]} \frac{\hat{P}[\Psi=1|Z=h]}{\hat{P}[\Psi=1|Z=l]} - \frac{\hat{g}\left(t|\Psi=2,Z=h\right)}{\hat{g}\left(t|\Psi=2,Z=l\right)} \frac{\hat{P}[\Psi=2|Z=h]}{\hat{P}[\Psi=2|Z=l]} \right\}^2 dH(t), \tag{5.14}$$

where I set $H(\cdot)$ to be $\hat{g}\left(t|\Psi=2,Z=l\right)$. In order to solve this infinite-dimensional optimization problem, I approximate the space Υ using splines—functions that are piecewise polynomial and have a high degree of smoothness at the points where the polynomials meet. Splines can be represented as the linear combination of a finite set of basis functions, which allow me to treat problem (5.14) as a finite-dimension non-linear regression problem. A popular family of basis functions is B-splines. Here, instead, I use C-splines, a set of basis functions recently proposed by Meyer (2008). C-splines differ from B-splines in that the basis functions in the former are strictly increasing and strictly convex. Because of such properties, monotonicity and convexity can be easily imposed to the approximate solution of (5.14), by restricting the regression coefficients to be non-negative. In the online Appendix, I show that the estimator proposed above is uniformly consistent under standard assumptions.

5.3.2. *Estimation of the model's primitives.* The proof of Proposition 2 shows that the model is overidentified. For estimation purposes, I can easily generalize it by allowing the trial costs to vary deterministically with the trial sentence. Specifically, I assume that the trial costs for the defendant and the prosecutor are given by $c_d(t) \equiv \alpha_d + \beta_d t$ and $c_p(t) \equiv \alpha_p + \beta_p t$, respectively. The primitives to be estimated are then the cost parameters α_d, β_d, α_p and β_p, as well as the distributions of defendants' types and trial sentences.[31]

[31] Notice that the model imposes constraints on α_d and β_d. Indeed, since the support of Θ is $(\underline{\theta},\bar{\theta}) \subseteq (0,1)$, it must be the case that $\tilde{\theta}(t) \in (0,1)$ for all t. From (5.10), I conclude that

$$\begin{aligned} &\alpha_d + \beta_d t < \tilde{s}(t) \\ \text{and} \qquad &\alpha_d + \beta_d t > \tilde{s}(t) - t \end{aligned}$$

for all $[\underline{t},\bar{t}]$. Also, as shown in Section 4, the function $\tilde{\theta}(\cdot)$ is strictly increasing in t,which implies

$$\alpha_d > \tilde{s}(t) - \tilde{s}'(t)\,t.$$

Some algebra shows that the following conditions are sufficient for the three inequalities above to hold: (i) $\tilde{s}(t)$ is convex in t; (ii) $\tilde{s}(t) < t$ for all $t \in [\underline{t},\bar{t}]$; (iii) $\alpha_d \in [\mathrm{Max}\left\{0, \tilde{s}(\underline{t}) - \tilde{s}'(\underline{t})\underline{t}\right\}, \tilde{s}(\underline{t})]$; and (iv) $\beta_d \in [0, \tilde{s}'(\underline{t})]$. As argued in Section 4, condition (i) is true. Condition (ii) is strongly supported by the estimates of $\tilde{s}(\cdot)$ reported later in the paper. Conditions (iii) and (iv) are then enough to guarantee that the function $\tilde{\theta}(\cdot)$ behaves as predicted by the theory. I impose these conditions in the estimation of the model.

For simplicity, I refer to the vector of parameters $[\alpha_d \; \beta_d \; \alpha_p \; \beta_p \; \mu]$ as ω.[32] Equations (5.12)-(5.13) show how, using an estimate $\hat{s}(\cdot)$ of the settlement offer function, I can obtain estimates for the distribution of defendant's types $F(\cdot)$ and the distribution of trial sentences $g(\cdot|Z = z)$, up to ω.[33] To complete the estimation of the model, I must estimate ω, which I do by maximum likelihood. See Appendix A.2 for details.

6. Empirical Results

I restrict the analysis to cases in which the main arrest offense is a non-homicide violent crime. Also, to control for observed heterogeneity across cases, I separate the observations in the sample into covariate groups, according to the following variables: The Superior Court division where the case is prosecuted, the type of defense counsel employed and the defendant's gender, race and criminal record. I then estimate the model separately across groups.

The motivation for separating cases by the Superior Court division is twofold: First, doing so helps controlling for any case-specific characteristic that may be correlated with the place of prosecution. Importantly, these characteristics include the district of the prosecutor in charge of the case.[34] Second, the judge rotation mandated by the state constitution takes place at the division level. Thus it is plausible to assume that cases decided by different judges within a division are drawn from the same pool.

I further separate the cases according to the type of defense attorney. I consider the following three categories of counsel: privately-retained, public defender and court-appointed. Lastly, I divide cases by defendant's gender, race (African-Americans vs. others) and criminal record. To avoid dividing the data into too many sparsely populated groups, I consider two categories of criminal record: short (four points or less) and long (more than four points).[35] According to this definition, approximately one-fourth of the defendants in the data have a long criminal record.

Separating the observations in the data according to the eight Superior Court divisions, three defense attorney types, two gender groups, two race groups and two categories of criminal records results in 192 covariate bins. Many such bins contain

[32]For the reasons outlined in Appendix A.2, I treat μ as a parameter to be estimated.

[33]The distribution of trial sentences is characterized by $g(\cdot|Z = z)$ and $\nu(z)$. The latter can be trivially estimated using the empirical probability that $\Psi = 0$, conditional on $Z = z$.

[34]North Carolina is divided into prosecutorial districts that roughly correspond to the Superior Court districts described in Section 3.

[35]To put this classification into perspective, notice that four points are equivalent to having a previous conviction for setting fire to an unoccupied building (second-degree arson) or breaking and entering an unoccupied residency with the intent to commit another crime (second-degree burglary).

relatively few observations. In particular, female defendants are only represented in 7.11% of all cases. The estimation of the model based on such few observations would be inappropriate, due to the non-parametric nature of my estimator. Therefore, I focus on covariate groups in which the defendant is male. Still, there are 96 such groups. Estimating the model separately for each one of them would be computationally too costly. I thus restrict my analysis to cases prosecuted at the 5th Superior Court division, which mostly comprises counties located in the Piedmont Triad region of the state and is the division with the largest sample size in my data set.[36] These restrictions leave me with 12 covariate groups, which are summarized in table 5.[37]

In the interest of space, I only report in the main text estimation results for covariates groups one and two. Both groups comprise defendants with short criminal history and who are represented by court-appointed attorneys. Defendants in group two are African-American, while those in group one are not. The online Appendix contains the results for covariate groups three to 12. The differences in the estimation results across groups one and two, presented below, suggest a non-trivial variation across races in the outcome of criminal cases. The results for groups three to 12 largely point in the same direction.

Figure 2 depicts estimates of the trial sentence densities, conditional on a conviction at trial, for lenient and harsh judges in covariate groups one and two.[38] All distributions show signs of multi-modality—with most of their mass concentrated at sentences shorter than 100 months. The densities for group two vary as expected across lenient and harsh judges. Indeed, the density associated with lenient judges has more mass at short sentences, relative to that associated with harsh judges. For group one the pattern is less clear.

To estimate the prosecutor's settlement offer function, I implement the spline regression procedure described previously. I do so independently for each covariate

[36]The 5th division has 22,658 cases overall. The 1st, 2nd, 3rd, 4th, 6th, 7th and 8th divisions have, respectively, 11,084, 10,110, 19,906, 13,743, 13,047, 19,033 and 9,364 cases. A map of the divisions can be found at http://www.nccourts.org/Courts/Trial/District/Documents/SuperiorCourtmap.pdf.

[37]Previous criminal records are only available for defendants who are convicted—either at trial or by plea bargain. Therefore, I do not observe directly the number of cases by covariate group. To compute the numbers reported in table 5, I first calculate the proportion of cases matching each covariate group among all cases that result in a conviction. Then I multiply this proportion by the number of cases in the data matching the following reduced list of covariates: Race/ethnicity, gender, counsel type and Superior Court division. This procedure assumes that, conditional on the other case characteristics, the distribution of criminal history points is independent from a conviction. Considering, from table 2, that only 5.63 percent of the cases in my sample result in an acquittal or a dismissal, such an assumption has a negligible impact on my empirical results.

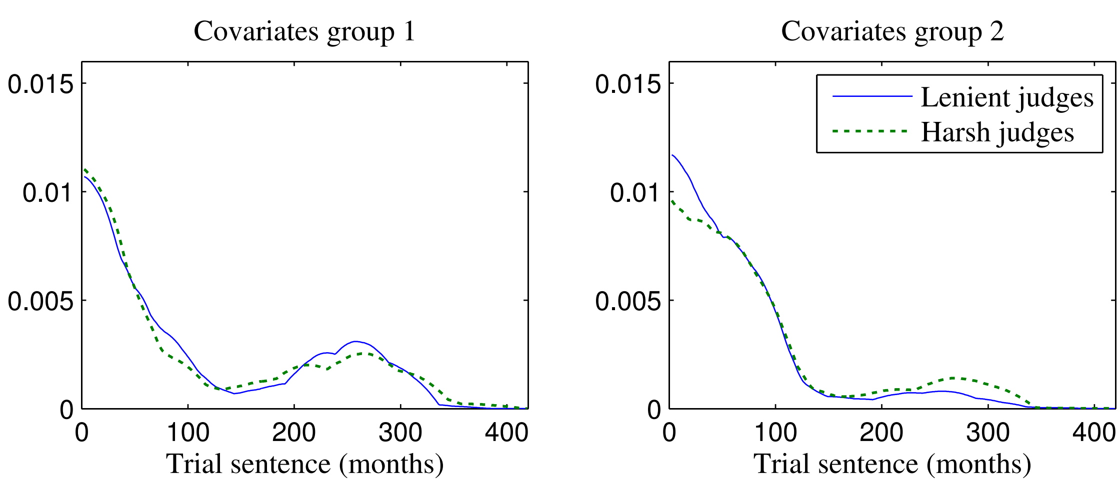
[38]Bandwidths are reported in Appendix A.2, together with other estimation details.

Table 5. Covariate groups

Group	Covariates					Obs.
	Defendant's race	Defendant's gender	Defense counsel	Defendant's record[a]	Sup. Court division	
1	other	Male	appointed	short	5	2141
2	Afr. American	Male	appointed	short	5	2727
3	other	Male	appointed	long	5	1989
4	Afr. American	Male	appointed	long	5	2868
5	other	Male	public	short	5	956
6	Afr. American	Male	public	short	5	2049
7	other	Male	public	long	5	915
8	Afr. American	Male	public	long	5	2547
9	other	Male	private	short	5	1922
10	Afr. American	Male	private	short	5	1308
11	other	Male	private	long	5	882
12	Afr. American	Male	private	long	5	707

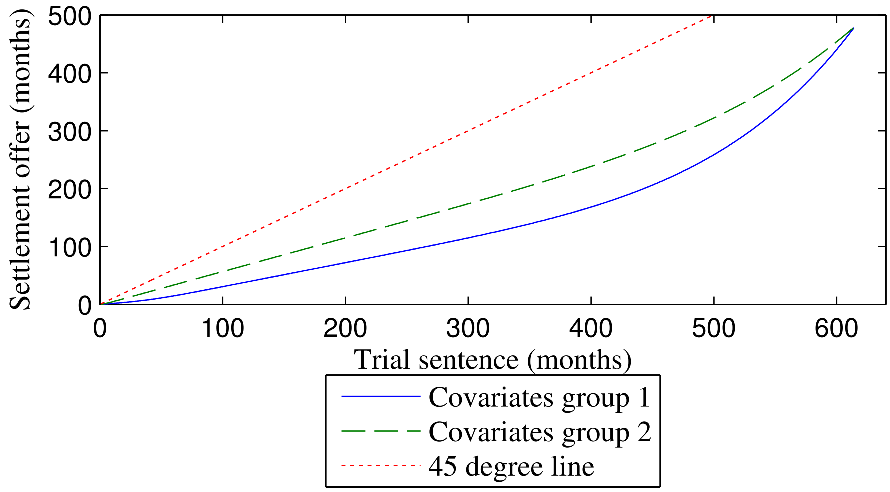
Notes: I estimate the structural model separately for each one of these groups. I only report the estimation results for groups one and two in the main text. See the Online Appendix for the other groups' results.
[a] North Carolina assigns points to each defendant, for the purposes of sentencing guidelines. I classify a defendant's record as long if it has more than four points.

Figure 2. Conditional trial sentence density estimates—Lenient and harsh judges



Note: Kernel density estimates of trial sentences assigned by lenient and harsh judges, conditional on covariates. See table 5 for a description of the covariate groups and Section 3 for details on the classification of judges.

FIGURE 3. Settlement offer function estimates



Note: Spline regression estimates of the prosecutor's settlement offer function for covariate groups one and two. See table 5 for a description of the covariate groups.

group. I set the knots—the points in the domain of the function where the polynomial pieces that constitute the spline connect— at the 25th, 50th and 75th percentiles of the entire sample of trial sentences, irrespective of the covariate group. The chosen knots are 21, 70 and 125 months. Notice that, because I constrain the estimated offer function to be increasing and convex, the selection of knots is not as critical in my analysis as it is in other spline regression applications (Meyer, 2008).

Figure 3 presents the estimated settlement offer functions.[39] The variation of the estimates across groups one and two is substantial. Specifically, the estimated offer function is more convex for group one than for group two. Remember that the only difference between these groups is that defendants in group two are African-American. The result thus indicates that prosecutors tend to offer longer sentences to African-American defendants than to their non-African-American counterparts, conditional on the length of the trial sentence. This is not a rigorous result, however, since I do not formally test whether one offer function is more convex than the other.

Once offer function estimates are obtained, I proceed with the estimation of the model's primitives. For each group, that consists of estimating five scalar parameters: α_p, β_p, α_d and β_d, which characterize the trial costs; and μ, which captures the behavior of the distribution of defendants' types for values of θ smaller than $\tilde{\theta}(\underline{t})$. Let $\hat{\alpha}_p$, $\hat{\beta}_p$, $\hat{\alpha}_d$, $\hat{\beta}_d$ and $\hat{\mu}$ be the respective estimates.

[39]The online Appendix contains a discussion of the fit, precision and robustness of these estimates.

Table 6. Parameter estimates by covariate group

Group	Parameters				
	$\hat{\alpha}_d$	$\hat{\beta}_d$	$\hat{\alpha}_p$	$\hat{\beta}_p$	$\hat{\mu}$
1	0.02 (0.01)	0.00 (0.00)	0.00 (0.00)	0.97 (0.41)	1.00 (0.00)
2	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	1.06 (0.26)	1.00 (0.00)

Notes: MLE estimates of the model parameters, conditional on covariates. See table 5 for a description of the covariate groups. Bootstrap standard errors in parenthesis.

Table 6 reports the estimation results, together with bootstrap standard errors.[40] Trial costs are measured in terms of months. The defendant's trial costs, for example, represent that agent's disutility of having her case decided at trial, relative to a month of incarceration. Accordingly, the trial cost intercepts $\hat{\alpha}_d$ and $\hat{\alpha}_p$ are expressed in months, while $\hat{\beta}_d$ and $\hat{\beta}_p$ are dimensionless. For both covariate groups, $\hat{\alpha}_d$ and $\hat{\beta}_d$ are very small. My estimates thus suggest that defendants barely take the trial costs into consideration when deciding whether to accept the prosecutor's offer. The discussion in Section 5.2.2 is useful for understanding what patterns in the data drive such a result. From (5.10), for any trial sentence t, the separate identification of the defendant's trial costs is based on the difference between $\tilde{s}(t)$ and $\tilde{\theta}(t)t$. This equation thus clarifies that the low defendant's trial cost estimates are due to the relatively short settlement offers observed in the data. Given these offers, higher trial costs would imply lower $\tilde{\theta}(t)$, which would drive the defendant's trial win rate away from that observed in the data. Indeed, in the online Appendix, I present results obtained from modified versions of the data. There, I show that, by artificially increasing either the settlement offers or the probability of acquittal, conditional on a trial, I obtain substantially higher estimates for the defendant's trial costs.

For both groups, $\hat{\alpha}_p$ is equal to zero and $\hat{\beta}_p$ is roughly one, suggesting that trial costs for the prosecutor increase rapidly as the trial sentence gets longer. At first glance, it may seem surprising that, according to the estimates, the prosecutor has negative expected utility of taking a case to trial. But, as argued in Section 4, the prosecutor is a recurrent player in the plea bargaining game. Because of reputation effects, it is reasonable to assume that she can commit to go to trial, even if her expected utility of

[40]See Appendix A.2 for details on the computation of the standard errors.

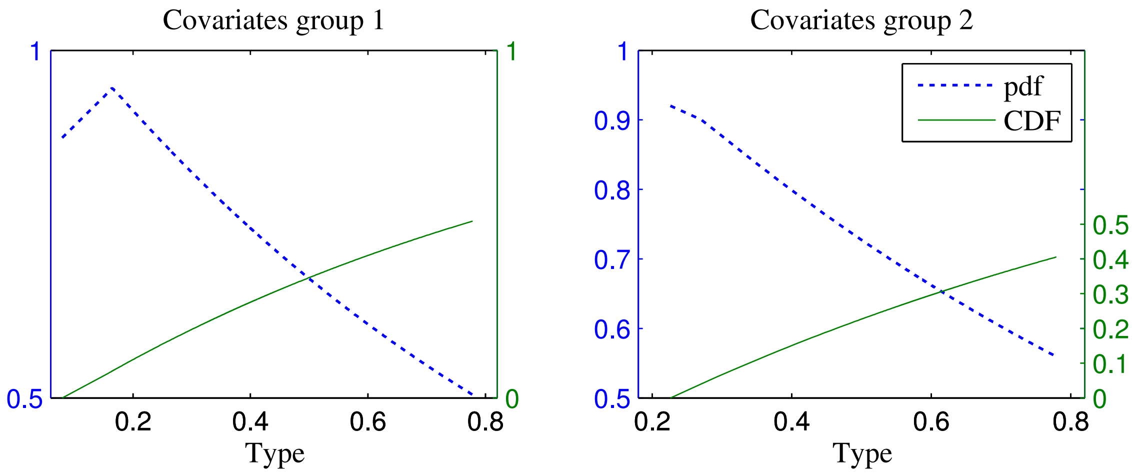
doing so is negative. Also, due to career-concerns, the costs of dropping a case could be substantial—meaning that the prosecutor's outside option would correspond to a negative payoff. In the online Appendix, I compute the prosecutor's expected payoffs, unconditional on the success of plea bargaining. For both groups, these expected payoffs are positive—although, as discussed above, it is unclear whether the opposite result would be inconsistent with rational behavior by the prosecutor. Estimates of μ are one for both groups. Notice that $F(\tilde{\theta}(\underline{t})) = 1 - \mu$ (see footnote 29). My results thus indicate that defendants whose types belong to the interval $[0, \tilde{\theta}(\underline{t})]$ are very rare.

Having estimated all of the remaining parameters, I compute estimates for the distribution of defendants' types. Figure 4 depicts these distributions. For each covariate group, the figure shows both the estimated density and cumulative distribution functions. All distributions have a mode at a type lower than 0.3. The densities then sharply decrease, giving the impression of being unimodal. However, as shown in Section 5, these distributions are only identified over the interval $[\tilde{\theta}(\underline{t}), \tilde{\theta}(\bar{t})]$. For both covariate groups, this interval comprises a large portion of the unit line. The lower bound varies from zero to just above 0.2, depending on the group, and the upper bound is roughly 0.8 for both groups. But both distributions have substantial mass outside of this range. For group one, the estimated $F(\tilde{\theta}(\bar{t}))$ is approximately 0.5, while for group two this value is roughly 0.4. Such numbers suggest that the distributions of defendants' types have at least one more mode, located at a relatively high type. The results are thus consistent with distributions that concentrate mass at types located near the boundaries of the unit line. Interestingly, these bimodal distributions of types are precisely the ones that would arise if trials were generally, but not always, successful at convicting the guilty defendants and acquitting the innocent ones.

The distributions shown in figure 4 help rationalizing the differences between the estimated settlement offer functions of African-American and non-African-American defendants, which were pointed-out in the discussion of figure 3. The estimate of $F(\tilde{\theta}(\bar{t}))$ is greater for group one than for group two, indicating that the distribution of types for African-American defendants concentrates more mass at high types than the distribution for other defendants. Thus, controlling for the other covariates considered in my analysis, the results suggest that African-Americans are more likely than others to to be convicted by the jury in the event of a trial.[41] According to the model, these differences in the distributions of types are considered by the prosecutor in

[41]These results are consistent with the raw data. Conditional on a trial, the African-American defendants in my sample are 9.16 p.p. more likely to be convicted than their non-African-American counterparts. I report detailed statistics on trial outcomes by race in the online Appendix.

Figure 4. Defendants' types distribution estimates



Note: Estimated pdf and CDF of the distribution of defendants' types (probabilities of conviction at trial), conditional on covariates. See table 5 for a description of the covariate groups. The distributions are only identified over part of their support.

the process of making settlement offers. Specifically, given an African-American and a non-African-American defendant who face the same trial sentence, the prosecutor offers to the former a longer sentence. The remaining model primitive is the full distribution of potential trial sentences—i.e., the distribution without conditioning on a trial conviction. In the interest of space, I report such distribution for groups one and two in the online Appendix.

Table 7 presents information on the fit of the model. It separately shows the probabilities of conviction to incarceration by plea bargain and at trial for each covariate group. The model fits well the probability of a plea bargain, while it slightly under-estimates the probability of a conviction at trial (by approximately 2 to 3 p.p., depending on the group). As a result, the model under-estimates the total probability of a conviction to incarceration by roughly the same amount. The table also shows three versions of the average assigned sentences: (i) The overall average—i.e., the average sentence, conditional on either a trial conviction or a plea bargain; (ii) the average plea bargain sentence; and (iii) the average trial sentence. The model fits the overall average well. However, it substantially over-estimates the average trial sentence for group one. It is useful to put such over-estimation into perspective by comparing it to the standard deviation of the observed trial sentences reported in table 2. The fitted model predicts the average trial sentence of group one to be 0.59 standard deviations longer than observed. An explanation for that is the small

Table 7. Fitted values versus data

Group		Conviction probability		
		Any ($\Psi \in \{1,2\}$)	Settlement ($\Psi = 1$)	Trial ($\Psi = 2$)
1	Data	38.35%	34.60%	3.75%
	Model	36.59%	34.86%	1.73%
2	Data	44.11%	36.66%	7.45%
	Model	42.22%	37.50%	4.71%

Group		Average sentence, conditional on method of resolution†		
		All ($\Psi \in \{1,2\}$)	Settlement ($\Psi = 1$)	Trial ($\Psi = 2$)
1	Data	63.13	54.93	138.73
	Model	57.12	49.97	201.18
2	Data	48.14	41.51	80.73
	Model	45.08	40.63	80.50

Notes: $\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
See table 5 for a description of the covariate groups.
† Measured in months.

number of trial conviction observations in my sample, which leads the maximum likelihood procedure to prioritize reproducing other features of the data. For the exact same reason, the overestimation of the average trial sentences has a very weak effect on the general fit of the model—as shown by the other moments on table 7.

7. Policy Experiments

Sentencing reform has been at the center of the public policy discussion, thanks to the growing consensus that the United States incarcerates too many people for too long. About 0.7 percent of Americans were in prison or jail at the end of 2010 (Glaze, 2011)—to some accounts, the highest incarceration rate in the world (Walmsley, 2009). Besides drastically affecting the lives of millions of inmates and their families, the correctional system constitutes a major component of public spending. Nationwide direct expenditures on corrections surpassed \$70 billion in every year between 2000 and 2007 (Kyckelhahn, 2011). That has led policy makers to consider reforms intended to reduce the number of prisoners. In particular, there have been an increasing number of proposals for reversing some of the "tough on crime" laws of the 1980s and 1990s. See Porter (2011) for examples of policy reforms recently proposed and implemented by state legislatures. At the federal level, a bipartisan effort introduced in the United States Senate the Sentencing Reform and Corrections Act of 2015.[42]

Two of the proposals that have recently gained political currency are the reduction of mandatory minimum sentence lengths, and the broader use of alternative punishments for less serious offenses. In this section, I use the estimated model to conduct counterfactual policy experiments on such reforms. Specifically, I consider a decrease in trial sentences, capturing the effect of an across-the-board reduction in the mandatory minimum sentence lengths. Also, I analyze the scenario in which relatively short trial sentences are set to zero. Such a scenario simulates the effect of abolishing incarceration sentences for the mildest offenses.

The focus of my counterfactual analysis is on the number of defendants who receive incarceration sentences, as well as on the total time of incarceration assigned. The outcomes of interest are: (i) the probability that a case results in an incarceration conviction at trial or by plea bargain; and (ii) the expected sentence length, unconditional on an incarceration conviction. The total number of months of incarceration assigned in any given period is determined by the number of defendants prosecuted over that period times the expected sentence length. Outcome (ii) is, thus, the main variable of interest if, for example, one is mostly concerned about the total incarceration costs of current sentencing decisions. However, a decrease in the expected sentence length that is solely due to shorter sentences may have no immediate impact on the incarceration rate. Changes in the probability of conviction are more likely to capture short-term effects of sentencing reforms.

My counterfactual analysis ignores important implications of sentencing reform, such as crime deterrence and recidivism. A more complete exercise would need to simultaneously account for these effects and plea bargaining. My analysis is one of partial equilibrium, and can be thought of as a step towards the better understanding of policy interventions in the criminal justice system. Endogenizing crime rates in a model similar to the one considered here is an exciting avenue for further research.

In a third policy experiment, I consider the scenario in which plea bargaining is not allowed, so that every case is decided at trial. Admittedly, given that roughly 90 percent of all criminal cases are currently settled, completely eliminating plea bargaining is likely to be too radical an intervention. But considering this extreme scenario allows me to assess the loss of information that arises due to settlements.

[42] See https://www.congress.gov/bill/114th-congress/senate-bill/2123 for details.

Specifically, I compute the proportion of defendants who are currently convicted to incarceration sentences, but who would be acquitted if their cases were tried. This policy experiment also allows me to measure the defendants' informational rents by comparing the average length of the incarceration sentences assigned in the scenarios with and without plea negotiations.

7.1. **Reducing mandatory minimum sentences.** Table 8 shows the effects of a twenty percent reduction in the length of potential trial sentences for all cases in the sample.[43] The top half of the table reports the effect on the probability of an incarceration conviction, either by plea bargain or at trial. The results indicate that shorter potential trial sentences increase the probability of a settlement, which raises the total probability of conviction. The latter probability increases by 1.89 percent for group one and 0.64 percent for group two.

The bottom half of the table shows the impact on the expected length of the assigned sentences. For both covariate groups, the elasticity of the expected sentence with respect to the potential trial sentence is roughly one, indicating that this intervention may be highly effective in reducing the total incarceration time assigned by the courts. This magnitude is not surprising, considering the results in figure 3, which show that the prosecutor's offer functions are quite steep. Since most cases are resolved by plea bargain, the main effect of a reduction in the potential trial sentences is a decrease in the settlement sentences. Thus, a reduction in the length of potential trial sentences is likely to increase incarceration rates slightly in the short run. But in the long run, the same intervention may lead to a major decrease in these rates.

7.2. **Assigning alternative punishments for mild cases.** Now I replace the ten percent of cases with the lowest positive potential trial sentence with cases in which the potential trial sentence is zero.[44] Table 8 shows the results. The impact of the intervention on the probability of conviction is high. A ten percent decrease in the number of cases with positive trial sentence reduces the total probability of conviction by roughly ten percent for both groups. This effect is due to the types of cases affected

[43]This policy experiment assumes that an across-the-board reduction in mandatory minimum sentences would lower the sentences to be assigned in the event of a trial conviction for all cases. While it is true that a case's potential trial sentence is determined by a conjunction of factors, mandatory minimums are likely to serve as a reference for the determination of sentences by the judges, and even be a binding constraint in a considerable number of cases. The results presented here can be interpreted as the effects of changes in mandatory minimum sentences insofar as such changes actually affect the punishments to be assigned at trial.

[44]The tenth percentiles of the distributions of potential trial sentences for groups one and two are 48.23 and 21.75 months, respectively.

Table 8. Counterfactual results—Sentencing reform

Group		Conviction probability		
		Any	Settlement	Trial
1	Current	36.59%	34.86%	1.73%
	-20% trial sentence length	37.28%	35.91%	1.37%
	-10% incarceration cases	32.74%	31.04%	1.70%
2	Current	42.22%	37.50%	4.71%
	-20% trial sentence length	42.49%	37.94%	4.55%
	-10% incarceration cases	37.94%	33.38%	4.56%

Group		Expected sentence†	
		$\Psi \in \{1, 2\}$	Unconditional on Ψ
1	Current	57.12	20.90
	-20% trial sentence length	41.23	15.37
	-10% incarceration cases	63.64	20.84
2	Current	45.08	19.03
	-20% trial sentence length	35.63	15.14
	-10% incarceration cases	49.95	18.95

Notes: This table reports the results of two counterfactual exercises. In the first one I reduce the length of the trial sentences of every case by 20 percent. Such a scenario intends to simulate the lowering of mandatory minimum sentences for all types of cases. In the second exercise I set the incarceration sentences of all cases below the tenth percentile to zero, with the objective of capturing the broader assignment of alternative sentences. The current values are the ones fitted by the estimated model.
$\Psi = 1$ and $\Psi = 2$ indicate incarceration convictions by plea bargaining and at trial, respectively.
See table 5 for a description of the covariate groups.
† Measured in months.

by the intervention. Cases with short potential trial sentences are very likely to be settled before trial, which results in a guaranteed conviction. The elimination of such cases, therefore, has a direct impact on the probability of conviction.

Nevertheless, the impact of the intervention on the expected sentence is relatively low. A ten percent decrease in the number of cases with positive potential trial sentence reduces the expected sentence by less than one percent for both covariate groups. The policy intervention considered here is rather extreme, and it is surprising that its effect on the expected sentence is so modest. An explanation for such a low magnitude is that the current expected sentences are largely influenced by the length of the most severe sentences. Even the complete elimination of the mildest cases has a small effect on the average outcome.

The results of the two policy simulations undertaken so far suggest that the broader use of alternative sentences and the reduction of mandatory minimum sentences complement each other. Assigning alternative sentences reduces the total probability of conviction, which immediately affects incarceration rates. Lowering the mandatory minimum sentences, particularly for severe offenses, greatly reduces the total incarceration time assigned but does not decrease incarceration rates in the short run.

7.3. **Eliminating plea bargaining.** If no plea bargaining is allowed and every case reaches the trial stage, each defendant is convicted with a probability equal to her type. Given a trial sentence t and a defendant of type θ, the expected sentence is θt. A challenge here is that I only identify the distribution of defendant's types for values within the interval $[\tilde{\theta}(\underline{t}), \tilde{\theta}(\bar{t})]$. Recovering the probability of conviction—that is, the mean of Θ—and the full distribution of settlement offers in the counterfactual scenario would require me to know the distribution of types outside of this range. Nevertheless, I am able to compute bounds for the probability of conviction and the mean settlement offer. To do so, I consider two extreme cases. In the first one, which leads to lower bounds for the probability and the mean, I assume that all defendants with type θ lesser than $\tilde{\theta}(\underline{t})$ have type equal to zero, and all defendants with type greater or equal to $\tilde{\theta}(\bar{t})$ have type exactly equal to $\tilde{\theta}(\bar{t})$. In the second extreme case, I obtain an upper bound for the probability and the mean by assuming that all defendants with type lesser or equal to $\tilde{\theta}(\underline{t})$ have type exactly equal to $\tilde{\theta}(\underline{t})$, and all defendants with type above $\tilde{\theta}(\bar{t})$ have type equal to one.

Table 9 presents the results. Relative to current levels, the probability of incarceration conviction falls substantially for both covariate groups. The differences between the current probabilities and the estimated lower bounds are 11.28 p.p. for group one and 9.13 p.p. for group two. When, instead, the upper bounds are considered, the differences are 6.38 p.p. and 2.53 p.p., respectively. These magnitudes are large, especially considering that the current rates of incarceration convictions are roughly 40 percent. For group one, eliminating plea bargaining leads to a decrease of 17.44 to 30.83 percent in the incarceration conviction rate. For group two, the decrease is between 5.99 and 21.62 percent.

The expected sentences increase considerably in the scenario without plea bargaining, relative to the current one. The lower bounds for this increase are 76.17 and 28.01 percent for groups one and two, respectively. The upper bounds are 109.86

Table 9. Counterfactual results—No plea bargaining

Group		Outcome	
		Probability of conviction	Expected sentence†
1	Current	36.59%	20.90
	No plea bargaining	[25.31% , 30.21%]	[36.81 , 43.86]
2	Current	42.22%	19.03
	No plea bargaining	[33.09% , 39.69%]	[24.36 , 29.22]

Notes: This table reports the results of forcing all cases to go to trial. Because the distribution of defendants' types is not identified over its entire support, I can only calculate bounds for the probability of conviction and the expected sentence. The current values are the ones fitted by the estimated model.
See table 5 for a description of the covariate groups.
† Measured in months.

and 53.55 percent. These results suggest that, on average, defendants greatly benefit from informational rents in the process of plea bargaining. Therefore, although a large proportion of the defendants would be acquitted at trial after a ban on plea bargains, the average defendant would be considerably worse off in such a scenario.

8. Conclusion

I develop a framework for the empirical analysis of plea bargaining, which allows me to evaluate the impact of several policy interventions on the outcome of prosecuted cases. My results suggest that lowering mandatory minimum sentences may greatly reduce the total amount of incarceration time assigned by the courts. However, the same intervention would slightly increase the proportion of cases resulting in incarceration convictions, which, in the short run, may have unintended effects on the incarceration statistics. A different reform, the wider use of alternative sentences in less serious cases, leads to a decrease in conviction rates but has little effect on the total incarceration time assigned. Hence, these two interventions have the potential to complement each other. I also evaluate the impact of prohibiting cases from being settled. My findings suggest that a large proportion of defendants who presently receive incarceration convictions by plea bargaining would be acquitted if their cases reached the trial stage.

Appendix A.

A.1. **Proof of Lemma 1.** Notice that (5.7), (5.8) and (5.13) imply

$$\int_{\underline{\theta}}^{\tilde{\theta}(t)} x\, f(x)\, dx$$
$$= \mu \exp\left(- \int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} \lambda(x)\, dx \right) \frac{P[\Psi = 2|Z = z]}{P[\Psi = 1|Z = z]} \frac{g\left(t|\Psi = 2, Z = z\right)}{b\left(\tilde{s}(t)|\Psi = 1, Z = z\right)} \frac{1}{\frac{d\tilde{s}(t)}{dt}}. \tag{A.1}$$

The equation above holds for all $t \in [\underline{t}, \bar{t}]$. Evaluating the equation for $\underline{t}$ and any other value t and subtracting on both sides, I obtain

$$\frac{P[\Psi = 1|Z = z]}{P[\Psi = 2|Z = z]} \int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} x\, \lambda(x) \exp\left(- \int_{\tilde{\theta}(\underline{t})}^{x} \lambda(u)\, du \right) dx$$
$$= \frac{g\left(t|\Psi = 2, Z = z\right)}{b\left(\tilde{s}(t)|\Psi = 1, Z = z\right)} \frac{1}{\frac{d\tilde{s}(t)}{dt}} \exp\left(- \int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} \lambda(x)\, dx \right) - \frac{g\left(\underline{t}|\Psi = 2, Z = z\right)}{b\left(\tilde{s}(\underline{t})|\Psi = 1, Z = z\right)} \frac{1}{\frac{d\tilde{s}(\underline{t})}{dt}}$$

for $z \in \{l, h\}$, where I use the relationship between the hazard and density functions: $f(\theta) = \lambda(\theta) \exp\left(- \int_{\underline{\theta}}^{\theta} \lambda(x)\, dx \right) = \lambda(\theta)\, \mu \exp\left(- \int_{\tilde{\theta}(\underline{t})}^{\theta} \lambda(x)\, dx \right)$. Rearranging gives

$$g\left(t|\Psi = 2, Z = z\right) = \Xi(c_d, c_p, t, z)\, b\left(\tilde{s}(t)|\Psi = 1, Z = z\right) \tag{A.2}$$

for all $t \in [\underline{t}, \bar{t}]$, where

$$\Xi(c_d, c_p, t, z) = \frac{d\tilde{s}(t)}{dt} \exp\left(\int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} \lambda(x)\, dx \right)$$
$$\times \left[\frac{P[\Psi = 1|Z = z]}{P[\Psi = 2|Z = z]} \int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} x \lambda(x) \exp\left(- \int_{\tilde{\theta}(\underline{t})}^{x} \lambda(u) du \right) dx + \frac{g\left(\underline{t}|\Psi = 2, Z = z\right)}{b\left(\tilde{s}(\underline{t})|\Psi = 1, Z = z\right)} \frac{1}{\frac{d\tilde{s}(\underline{t})}{dt}} \right]$$

for $z \in \{l, h\}$. Since (A.2) holds for all t in the support, it implies a system of infinitely many equations. This system relates two observed distributions—that of trial sentences, conditional on a conviction at trial, and that of settlement offers, conditional on a successful bargain. Given any t and z, the link between these two distributions, $\Xi(c_d, c_p, t, z)$, depends on c_d and c_p as follows: Both c_d and c_p identically affect the hazard function $\lambda(\theta)$ for any value of θ, but only c_d affects the equilibrium cutoff points $\tilde{\theta}(\underline{t})$ and $\tilde{\theta}(t)$. Thus c_d and c_p are independent in the system and separately identified. Indeed, since (A.2) holds both for $z = l$ and $z = h$, I can write

$$\begin{aligned}
&\frac{P[\Psi=2|Z=l]}{P[\Psi=1|Z=l]} \\
&\times \left[\frac{g\left(t|\Psi=2,Z=l\right)}{b\left(\tilde{s}(t)|\Psi=1,Z=l\right)} \frac{1}{\frac{d\tilde{s}(t)}{dt}} \exp\left(-\int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} \lambda(x)\,dx \right) - \frac{g\left(\underline{t}|\Psi=2,Z=l\right)}{b\left(\tilde{s}(\underline{t})|\Psi=1,Z=l\right)} \frac{1}{\frac{d\tilde{s}(\underline{t})}{dt}} \right] \\
&= \frac{P[\Psi=2|Z=h]}{P[\Psi=1|Z=h]} \\
&\times \left[\frac{g\left(t|\Psi=2,Z=h\right)}{b\left(\tilde{s}(t)|\Psi=1,Z=h\right)} \frac{1}{\frac{d\tilde{s}(t)}{dt}} \exp\left(-\int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} \lambda(x)\,dx \right) - \frac{g\left(\underline{t}|\Psi=2,Z=h\right)}{b\left(\tilde{s}(\underline{t})|\Psi=1,Z=h\right)} \frac{1}{\frac{d\tilde{s}(\underline{t})}{dt}} \right]
\end{aligned}$$

for all $t \in [\underline{t}, \bar{t}]$. Rearranging, I obtain

$$\int_{\tilde{\theta}(\underline{t})}^{\tilde{\theta}(t)} \lambda(x)\,dx = \Gamma(t), \tag{A.3}$$

where the function

$$\begin{aligned}
\Gamma(t) &\equiv \log\left(\frac{d\tilde{s}(\underline{t})}{dt}\right) - \log\left(\frac{d\tilde{s}(t)}{dt}\right) \\
&+ \log\left[\frac{P[\Psi=1|Z=h]}{P[\Psi=2|Z=h]} \frac{g\left(t|\Psi=2,Z=h\right)}{b\left(\tilde{s}(t)|\Psi=1,Z=h\right)} - \frac{P[\Psi=1|Z=l]}{P[\Psi=2|Z=l]} \frac{g\left(t|\Psi=2,Z=l\right)}{b\left(\tilde{s}(t)|\Psi=1,Z=l\right)} \right] \\
&+ \log\left[\frac{P[\Psi=1|Z=h]}{P[\Psi=2|Z=h]} \frac{g\left(\underline{t}|\Psi=2,Z=h\right)}{b\left(\tilde{s}(\underline{t})|\Psi=1,Z=h\right)} - \frac{P[\Psi=1|Z=l]}{P[\Psi=2|Z=l]} \frac{g\left(\underline{t}|\Psi=2,Z=l\right)}{b\left(\tilde{s}(\underline{t})|\Psi=1,Z=l\right)} \right]
\end{aligned}$$

consists only of observables and $\tilde{s}(\cdot)$, which was previously identified. Differentiating both sides of (A.3), I have that

$$\lambda\left[\tilde{\theta}(t)\right] \frac{d\tilde{\theta}(t)}{dt} = \frac{d\Gamma(t)}{dt},$$

which, from (5.10) and (5.11), can be rewritten as

$$\frac{\frac{d\tilde{s}(t)}{dt} t - \tilde{s}(t) + c_d}{c_d + c_p} = \frac{d\Gamma(t)}{dt} t \tag{A.4}$$

for all $t \in [\underline{t}, \bar{t}]$. The system of equations implied by (A.4) can be trivially solved for c_d and c_p. $\square$

A.2. **Estimation appendix.**

A.2.1. *Estimation of the model primitives.* The primitives to be estimated are the cost parameters α_d, β_d, α_p and β_p; the distribution of defendant's types $F(\cdot)$; and the distribution of trial sentences, characterized by $\nu(z)$ and $g(\cdot|Z = z)$.

For $z \in \{l, h\}$, I can trivially estimate $\nu(z)$ by using the empirical probabilities that $\Psi = 0$, conditional on $Z = z$. To recover the other primitives, I follow the steps outlined in the proof of proposition 2, which shows how to recover $F(\cdot)$ and $g(\cdot|Z = z)$, based on the parameters $\alpha_d, \beta_d, \alpha_p$ and β_p. (A.1) holds for all $t \in [\underline{t}, \bar{t}]$, and thus defines a system of infinitely many equations. Notice that, besides $\alpha_d, \beta_d, \alpha_p$ and β_p, the system contains the parameter μ, which captures the behavior of $F(\cdot)$ for values of θ lower than $\tilde{\theta}(\underline{t})$.[45] Let $\omega \equiv [\alpha_d \; \beta_d \; \alpha_p \; \beta_p \; \mu]$ be the vector of all unknown variables in (A.1). The model is overidentified, so I estimate ω by maximum likelihood.

Specifically, let $\Omega \equiv \Re^4_{++} \times [0, 1]$ be the space of possible values for ω, and consider $\ddot{\omega} \in \Omega$. From $\hat{s}(\cdot)$ and (5.10), I numerically obtain $\ddot{\theta}(\cdot; \ddot{\omega})$, the function $\tilde{\theta}(\cdot)$ consistent with $\ddot{\omega}$. Using (5.11) and (5.12), I then obtain $\ddot{f}(\cdot|\ddot{\omega})$, the density function $f(\cdot)$ consistent with $\ddot{\omega}$. Similarly, from (5.13) and the estimated density $\hat{b}(\cdot|\Psi = 1, Z = z)$, I numerically compute $\ddot{g}(\cdot|Z = z; \ddot{\omega})$, the density $g(\cdot|Z = z)$ consistent with $\ddot{\omega}$. Using $\ddot{f}(\cdot|\ddot{\omega})$ and $\ddot{g}(\cdot|Z = z; \ddot{\omega})$, I obtain the likelihood that $\Psi = 3$, given Z, and consistently with $\ddot{\omega}$. Such likelihood is

$$\ddot{P}[\Psi = 3|Z = z; \ddot{\omega}] = \int_{[\hat{\underline{t}}, \hat{\bar{t}}]} \int_{\underline{\theta}}^{\ddot{\theta}(t;\ddot{\omega})} (1 - x) \; \ddot{f}(x|\ddot{\omega}) \ddot{g}(t|Z = z; \ddot{\omega}) \, dx \, dt.$$

From (5.4), I can compute the likelihood that $\Psi = 1$, given Z, and consistently with $\ddot{\omega}$. This likelihood is given by

$$\ddot{P}[\Psi = 1|Z = z; \ddot{\omega}] = \int_{[\hat{\underline{t}}, \hat{\bar{t}}]} 1 - \ddot{F}\left[\ddot{\theta}(t; \ddot{\omega})|\ddot{\omega}\right] \ddot{g}(t|Z = z; \ddot{\omega}) \, dt,$$

where $\ddot{F}[\cdot|\ddot{\omega}]$ is a CDF obtained from $\ddot{f}(\cdot|\ddot{\omega})$. From (5.7) and (5.8), the likelihood that $T = t$ and $\Psi = 2$, given Z, and consistently with $\ddot{\omega}$, is

$$\ddot{P}[\Psi = 2|Z = z; \ddot{\omega}] \; \ddot{g}(t|\Psi = 2, Z = z; \ddot{\omega}) = \int_{\underline{\theta}}^{\ddot{\theta}(t;\ddot{\omega})} x \; \ddot{f}(x|\ddot{\omega}) \, dx \, \ddot{g}(t|Z = z; \ddot{\omega}).$$

I am ready to define an observation's likelihood contribution. I consider the likelihood, conditional on $\Psi \neq 0$.[46] Let W_i be the data corresponding to observation i.[47]

---

[45]The system also implicitly contains $\pi \equiv \int_{\underline{\theta}}^{\tilde{\theta}(\underline{t})} x \; f(x) \, dx$. Like μ, the parameter π depends on the behavior of $F(\cdot)$ outside of the range of the support over which this function is identified. In estimating the model, I set $\pi = 0$. My empirical results suggest that μ is very close to one. Since $\pi \leq 1 - \mu$, setting $\pi = 0$ is unlikely to make any practical difference in the estimation.

[46]Notice that the empirical probability that $\Psi = 0$ is useful only for identifying $\nu(z)$.

[47]That is, if $\Psi_i \in \{1, 3\}$, W_i consists of z_i and ψ_i, the realizations of Z_i and Ψ_i. If $\Psi_i = 2$, W_i also includes t_i, the realization of T_i. Notice that I do not take into account the realization s_i of S_i, which is observed when $\Psi_i = 1$. That is because the likelihood of $S = s$, given $\Psi = 1$ and $Z = z$ is simply $\hat{b}(z|\Psi = 1, Z = z)$, which does not depend on $\ddot{\omega}$.

Given $\ddot{\omega}$, the likelihood contribution of an observation i, conditional on $\Psi_i \neq 0$, is

$$l(\ddot{\omega}, W_i) = \ddot{P}[\Psi = 1|Z = z; \ddot{\omega}]^{\mathbb{1}\{\Psi_i=1\}} \times \left\{\ddot{P}[\Psi = 2|Z = z; \ddot{\omega}]\, \ddot{g}(t_i|\Psi = 2, Z = z; \ddot{\omega})\right\}^{\mathbb{1}\{\Psi_i=2\}} \times \ddot{P}[\Psi = 3|Z = z; \ddot{\omega}]^{\mathbb{1}\{\Psi_i=3\}}. \tag{A.5}$$

I obtain an estimate $\hat{\omega}$ for ω by performing a numerical search to find the parameters that maximize the sum of the logarithms of $l(\ddot{\omega}, W_i)$ over all observations for which $\psi \neq 0$.[48] Finally, estimates for $g(\cdot|Z = z)$ and $f(\cdot)$ are defined by $\ddot{g}(\cdot|Z = z; \hat{\omega})$ and $\ddot{f}(\cdot|Z = z; \hat{\omega})$, respectively.

A.2.2. *Observed heterogeneity.* I divide the observations in the data into a finite number of covariate groups and implement the estimator described in Section 5 separately for each one of them. The first step of the estimator consists of computing two types of conditional densities: that of trial sentences, conditional on a conviction at trial, and that of settlement offers, conditional on a plea bargain. These conditional densities must be estimated for cases under the responsibility of both lenient and harsh judges. Therefore, for each one of the covariate groups under consideration in my analysis, I must estimate four conditional densities. I use the smoothing method by Li and Racine (2007), which I briefly describe below. Notice that the notation employed in this part of the Appendix differs from that of the rest of the paper.

Let Y be an univariate continuous random variable and X an r-dimensional discrete random variable. Denote by $f(\cdot)$, $g(\cdot)$ and $\mu(\cdot)$ the joint density of (X, Y) and the marginal densities of Y and X, respectively. For each dimension s of X, let c_s be the number of values in the support of X_s and λ_s be a real number between zero and $(c_s - 1)/c_s$. Define the vector $\lambda = (\lambda_1, \ldots, \lambda_r)$ and consider the following estimators of $f(\cdot)$ and $\mu(\cdot)$:

$$\hat{f}(x, y) = n^{-1} \sum_{i=1}^{n} L(x, X_i, \lambda) k_{h_0}(y - Y_i)$$

$$\text{and} \qquad \hat{\mu}(x) = n^{-1} \sum_{i=1}^{n} L(x, X_i \lambda),$$

where n is the sample size, $k_{h_0}(\cdot)$ is a kernel function with bandwidth h_0 and

$$L(x, X_i \lambda) = \prod_{s=1}^{r} [\lambda_s/(c_s - 1)]^{\mathbb{1}(X_{is} \neq x_s)} (1 - \lambda_s)^{\mathbb{1}(X_{is} = x_s)}.$$

[48] I constrain $\ddot{\alpha}_d$ and $\ddot{\beta}_d$ to satisfy the conditions in footnote 31.

Finally, define the the estimator of the conditional density $g(y|x)$ as

$$\hat{g}(y|x) = \hat{f}(x, y)/\hat{\mu}(x).$$

Notice that $\hat{g}(y|x)$ is obtained using all observations in the data—even those in which $X \neq x$. These observations are weighted down, relative to the ones satisfying $X = x$. The weights are given by the vector $\lambda = (\lambda_1, \ldots, \lambda_r)$. In one extreme case, λ_s is zero for all s, and $\hat{g}(y|x)$ is calculated employing only observations such that the realization of X is x. In the other extreme case, $\lambda_s = (c_s - 1)/c_s$ for all s, and $\hat{g}(y|x)$ becomes the estimate of $g(\cdot)$, the unconditional density of Y. The vector λ can be regarded as a collection of smoothing parameters—one for each dimension of X. Together, λ and h_0 determine the extent to which points away from (y, x) affect $\hat{g}(y|x)$. As argued by Li and Racine (2007), positive values of λ increase the finite sample bias of $\hat{g}(y|x)$ but also reduce its variance, with an ambiguous effect on the mean squared error.

The greatest challenge in implementing this estimator, therefore, is the choice of the smoothing parameters λ and h_0. In my application, I follow Li and Racine (2007) and select λ by maximum likelihood cross-validation. For any given sample size and any covariate dimension c, this method aims to select relatively large values of λ_c if the distribution of Y is not largely affected by variations in X_c, and small values of λ_c if the distribution of Y varies considerably with X_c. Moreover, the selected values of λ_c tend to decrease as the sample size increases.

For each covariate group, I estimate four conditional densities. Using the notation of Li and Racine's estimator presented above, Y may represent four random variables: Trial sentences assigned by lenient judges, trial sentences assigned by harsh judges, settlement offers made under lenient judges and settlement offers made under harsh judges. The discrete random variable X refers to the covariates used to divide the data into groups.[49] This random variable has the following five dimensions: (i) defendant's gender (male or female), (ii) defendant's race (African American or non-African American), (iii) the type of defense counsel (public defender, court-assigned attorney or privately-held attorney), (iv) the length of the defendant's criminal record (short or long, as defined in Section 6) and (v) Superior Court division (numbers one to eight). The function $k_{h_0}(\cdot)$ is the Epanechnikov kernel.

[49]To be sure, I estimate four conditional densities. The densities of trial sentences are conditional on a conviction at trial, and those of settlement offers are conditional on a plea bargain. Besides conditioning on the case outcome, I estimate these densities conditioning on five covariates. In the notation of this Appendix, X refers only to these covariates.

TABLE 10. Conditional density estimators—Covariates' smoothing parameters

Covariate	Conditional density estimator				
	Trial sentences (lenient)	Trial sentences (harsh)	Offers (lenient)	Offers (harsh)	Upper endpoint
Gender	0.39	0.11	0.03	0.03	0.50
Race	0.12	0.05	0.09	0.09	0.50
Counsel	0.23	0.18	0.20	0.25	0.67
Record	0.02	0.03	0.01	0.00	0.50
Division	0.41	0.33	0.44	0.41	0.88

Notes: Covariates' smoothing parameters selected by maximum-likelihood cross-validation. The parameters are used for smoothing across covariate groups in the kernel estimation of conditional densities of trial sentences and settlement offers.

Table 10 contains the smoothing parameters λ obtained by maximum-likelihood cross-validation for each of the four conditional densities of my analysis. Notice that, for every covariate c, lambda must belong to the interval $[0, (c_s - 1)/c_s]$, where c_s is the covariate's support. The upper endpoints of this interval are shown in the last column of the table. All the selected smoothing parameters are far away from these endpoints, suggesting that the covariates under consideration are important in explaining the distributions of trial sentences and settlement offers. In particular, the smoothing parameters associated with the defendant's previous criminal record are very close to zero. The parameters associated with race are also relatively low—ranging from 0.05 to 0.12. The gender parameters are larger for the densities of trial sentences than for those of settlement offers, which can be explained by the larger sample sizes used to compute the latter.

As explained in Section 5, the supports of trial sentences and settlement offers are bounded, which complicates the estimation of the conditional densities described above. I use a boundary correction proposed by Karunamuni and Zhang (2008). Using the notation of this Appendix, the approach consists of reflecting a transformation of the data near the boundary of Y. The reflected data points have the same x as the corresponding observations in the original data set, but y is modified. The estimator uses separate bandwidths h_0 for points near the boundary and away from it. Differently from the naive reflection of the untransformed data, this method allows the partial derivative of $g(y|x)$ with respect to y to be different from zero at the boundary of the support. See Karunamuni and Zhang (2008) for details.

TABLE 11. Conditional density estimators—Trial sentences and settlement offers' bandwidths†

	Trial sentences (lenient)	Trial sentences (harsh)	Settlement offers (lenient)	Settlement offers (harsh)
Bandwidth	21.83	25.31	6.59	7.37

Notes: Bandwidths selected by Silverman's "rule-of-thumb" (Silverman, 1986).
† Measured in months.

Table 11 reports the bandwidths h_0 for points away from the boundary, which are computed using Silverman's "rule-of-thumb" (Silverman, 1986). The bandwidths for trial sentences are 21.83 months (lenient judges) and 25.31 months (harsh judges). Those for settlement offers are 6.59 and 7.37 months (lenient judges) and 7.37 months (harsh judges). The larger bandwidths for trial sentences reflect the relative scarcity of cases that result in an incarceration conviction at trial.

A.2.3. *Standard errors.* I use 1200 bootstrap samples for each group to compute standard errors for the parameters reported in table 6. For each such sample, I estimate the densities of trial sentences and settlement offers using the same bandwidths and smoothing parameters employed in the main data. There are two main issues with this procedure. First, I do not offer a proof of the validity of the bootstrap for my estimator. Subsampling methods (Politis, Romano and Wolf, 1999) are more robust than the bootstrap, but, to apply these methods, the convergence rate of the estimator must be known. The second issue is that, for part of the bootstrap samples, the last step of the estimation procedure—i.e., obtaining maximum likelihood estimates for α_d, β_d, α_p, β_p and μ—becomes computationally too costly. This is the case whenever the estimated settlement offer function is too convex. I do not implement the last estimation step for these samples.[50] Thus the standard deviations reported in table 6 may slightly overstate the actual precision of my estimator.

REFERENCES

Alschuter, Albert W., "The Prosecutor's Role in Plea Bargaining," *University of Chicago Law Review*, 1968, *36*, 50–112.

[50]More precisely, I drop from my analysis every bootstrap sample in which the coefficient associated with the last C-spline basis is greater than four. As a reference, using the main data, I estimate this coefficient to be 2.43 for covariate group one and 1.27 for group two. This procedure eliminates 13.83% and 32.50% of the 1200 bootstrap samples for groups one and two, respectively.

Athey, Susan. and Philip.A. Haile, "Nonparametric Approaches to Auctions," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6A, New York: Elsevier, 2007, pp. 3847–3965.

Bagwell, Kyle, Robert W. Staiger, and Ali Yurukoglu, "Multilateral Trade Bargaining: A First Look at the GATT Bargaining Records," *NBER Working Paper Series*, 2015, *21488.*

Baker, Scott and Claudio Mezzetti, "Prosecutorial Resources, Plea Bargaining and the Decision to Go to Trial," *Journal of Law, Economics and Orgainzation*, 2001, *17*, 149–167.

Bebchuk, Lucian Arye, "Litigation and Settlement under Imperfect Information," *RAND Journal of Economics*, 1984, *15*, 404–415.

Bjerk, David, "Guilt Shall Not Escape or Innocence Suffer: the Limits of Plea Bargaining When Defendant Guilt is Uncertain," *American Law and Economics Review*, 2007, *9*, 305–329.

Boylan, Richard T., "The Effect of Punishment Severity on Plea Bargaining," *Journal of Law and Economics*, 2012, *55* (3), pp. 565–591.

Buescher, Paul A. and Ziya Gizlice, "Healthy Life Expectancy in North Carolina, 1996-2000," SCHS Studies 129, State Center for Health Statistics, 1908 Mail Service Center, Raleigh, N.C. January 2002.

Chen, Xiaohong, "Large Sample Sieve Estimation of Semi-nonparametric Models," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6B, New York: Elsevier, 2007, pp. 5549–5632.

Daughety, Andrew F. and Jennifer F. Reinganum, "Settlement," in Chris W. Sanchirico, ed., *Encyclopedia of Law and Economics*, second ed., Vol. 8 - Procedural Law and Economics, Cheltenham, UK: Edward Elgar Publishing Co, 2012.

Durose, Matthew R. and Patrick A. Langan, "Felony Sentences in State Courts 2000," Bulletin NCJ 198821, U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC: U.S. Department of Justice June 2003.

Dyke, Andrew, "Electoral Cycles in the Administration of Criminal Justice," *Public Choice*, 2007, *133*, 417–437.

Eckstein, Zvi and Kenneth I Wolpin, "Duration to first job and the return to schooling: Estimates from a search-matching model," *The Review of Economic Studies*, 1995, *62* (2), 263–286.

Elder, Harold W., "Trials and Settlement in the Criminal Courts: an Empirical

Analysis of Dispositions and Sentencing," *Journal of Legal Studies*, 1989, *18*, 191–208.

Fischer, George, *Plea Bargain's Triumph: A History of Plea Bargaining in America*, Palo Alto: Stanford University Press, 2003.

Friedman, Daniel and Donald Wittman, "Litigation with Symmetric Bargaining and Two-Sided Incomplete Information," *Journal of Law, Economics and Organization*, 2007, *23*, 98–126.

Glaze, Lauren E., "Correctional Population in the United States, 2010," Bulletin NCJ 236319, U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC: U.S. Department of Justice December 2011.

Gordon, Sanford C. and Gregory A. Huber, "The Effect of Electoral Competitiveness on Incumbent Behavior," *Quarterly Journal of Political Science*, 2007, *2*, 107–138.

Hickman, Brent R. and Timothy P. Hubbard, "Replacing Sample Trimming with Boundary Corrections in Nonparametric Estimation of First-price Auctions," 2012. Mimeo, Northwestern University.

Huber, Gregory A. and Sanford C. Gordon, "Accountability and Coercion: Is Justice Blind when it Runs for Office?," *American Journal of Political Science*, 2004, *48*, 247–263.

Karunamuni, Rohana J. and Shunpu Zhang, "Some Improvements on a Boundary Corrected Kernel Density Estimator," *Statistics and Probability Letters*, 2008, *78*, 499–507.

Keniston, Daniel E., "Bargaining and Welfare: a Dynamic Structural Analysis," 2011. Mimeo, Massachusetts Institute of Technology.

Kennan, John and Robert Wilson, "Bargaining with Private Information," *Journal of Economic Literature*, 1993, *31*, 45–104.

Kuziemko, Ilyana, "Does the Threat of the Death Penalty Affect Plea Bargaining in Murder Cases? Evidence from New York's 1995 Reinstatement of Capital Punishment," *American Law and Economics Review*, 2006, *8* (1), 116–142.

Kyckelhahn, Tracey E., "Justice Expenditures and Employment, FY 1982-2007 - Statistical Tables," Bulletin NCJ 236218, U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC: U.S. Department of Justice December 2011.

LaCasse, Chantele L. and A. Abigail Payne, "Federal Sentencing Guidelines and Mandatory Minimum Sentences: Do Defendants Bargain in the Shadow of the Judge?," *Journal of Law and Economics*, 1999, *42*, 245–270.

Larsen, Bradley, "The Efficiency of Dynamic, Post-Auction Bargaining: Evidence from Wholesale Used-Auto Auctions," 2014. Mimeo, Stanford University.

Li, Qi and Jeffrey S. Racine, *Nonparametric Econometrics: Theory and Practice*, first ed., New Jersey: Princeton University Press, 2007.

Lim, Claire S. H., "Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges," *American Economic Review*, 2013, *103* (4), 1360–1397.

Lim, Claire S.H., Bernardo S. Silveira, and James M. Snyder, "Do Judges? Characteristics Matter? Ethnicity, Gender, and Partisanship in Texas State Trial Courts," 2016. forthcoming at the *American Law and Economics Review*.

Merlo, Antonio, "Bargaining over governments in a stochastic environment," *Journal of Political Economy*, 1997, pp. 101–131.

_______ **and Xun Tang**, "Bargaining with Optimism: A Structural Analysis of Medical Malpractice Litigation," 2015. Mimeo, Rice University.

_______**, François Ortalo-Magné, and John Rust**, "The Home Selling Problem: Theory and Evidence," *International Economic Review*, 2015, *56*, 457–484.

Meyer, Mary C., "Inference Using Shape-restricted Regression Splines," *Annals of Applied Statistics*, 2008, *2*, 1013–1033.

Moore, Janet, "Democracy and Criminal Discovery Reform after Connick and Garcetti," *Brooklyn Law Review*, 2012, *77* (4).

Nalebuff, Barry, "Credible Pretrial Negotiation," *RAND Journal of Economics*, 1987, *18*, 198–210.

Newey, Whitney K. and Daniel L. McFadden, "Large Sample Estimation and Hypothesis Testing," in Robert F. Engle and Daniel L. McFadden, eds., *Handbook of Econometrics*, Vol. 4, New York: Elsevier, 1994, pp. 2113–2245.

Politis, Dimitris N., Joseph P. Romano, and Michael Wolf, *Subsampling*, first ed., New York: Springer, 1999.

Porter, Nicole D., "The State of Sentencing 2010 – Developments in Policy and Practice," Technical Report, The Sentencing Project February 2011.

Priest, George and Benjamin Klein, "The Selection of Disputes for Litigation," *Journal of Legal Studies*, 1984, *13*, 1–56.

Reinganum, Jennifer F., "Plea Bargaining and Prosecutorial Discretion," *American Economic Review*, 1988, *78*, 713–728.

_______ **and Louis L. Wilde**, "Settlement, Litigation, and the Allocation of Litigation Costs," *RAND Journal of Economics*, 1986, *17*, 557–566.

Roberts, Jenny, "Too Little, Too Late: Ineffective Assistance of Counsel, the Duty to Investigate, and Pretrial Discovery in Criminal Cases," *Fordham Urban Law Review*, 2003, *31* (4).

Rubbinfeld, Daniel L. and David E. M. Sappington, "Efficient Awards and Standards of Proof in Judicial Proceedings," *RAND Journal of Economics*, 1987, *18*, 308–315.

Scott, Robert E. and William J. Stuntz, "Plea Bargaining as Contract," *The Yale Law Journal*, 1992, *101* (8), pp. 1909–1968.

Sieg, Holger, "Estimating a Bargaining Model with Asymmetric Information: Evidence from Medical Malpractice Disputes," *Journal of Political Economy*, 2000, *108*, 1006–1021.

Silverman, Bernard W., *Density Estimation for Statistics and Data Analysis*, London, UK: Chapman and Hall, 1986.

Spier, Kathryn E., "The dynamics of pretrial negotiation," *The Review of Economic Studies*, 1992, *59* (1), 93–108.

Waldfogel, Joel, "The Selection Hypothesis and the Relationship between Trial and Plaintiff Victory," *Journal of Political Economy*, 1995, *103*, 229–260.

———, "Reconciling Asymmetric Information and Divergent Expectations Theories of Litigation," *Journal of Legal Studies*, 1998, *41*, 451–476.

Walmsley, Roy, "World Prison Population List – Eigth Edition," Technical Report, International Centre for Prison Studies January 2009.

Watanabe, Yasutora, "Learning and Bargaining in Dispute Resolution: Theory and Evidence from Medical Malpractice Litigation," 2009. Mimeo, Northwestern University.