

BARGAINING WITH ASYMMETRIC INFORMATION: AN EMPIRICAL STUDY OF PLEA NEGOTIATIONS

BERNARDO S. DA SILVEIRA

OLIN BUSINESS SCHOOL
WASHINGTON UNIVERSITY IN ST. LOUIS

SEPTEMBER, 2014

ABSTRACT. This paper empirically investigates how sentences to be assigned at trial impact plea bargaining. The analysis is based on a variation of a bargaining model with asymmetric information due to 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 lowering mandatory minimum sentences has the potential to greatly reduce the total amount of incarceration time assigned by the justice system, 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 negligible effect over the total assigned incarceration time. I also consider the effects of a ban of plea bargains. Depending on the characteristics of the case, more than 20 percent of the defendants who currently receive incarceration sentences would be acquitted at trial if settlements were forbidden. Nevertheless, in expectation, defendants would be substantially worse off in a scenario without plea bargaining.

1. INTRODUCTION

In this paper, I analyze how the harshness of sentences to be assigned at trial affects plea bargaining. I build upon a model due to Bebchuk (1984) and establish sufficient conditions under which the model is non-parametrically identified. I then propose a non-parametric estimator and implement it using data on criminal cases from North Carolina state courts. The estimates allow me to evaluate the impact of commonly proposed sentencing reforms on the final result of cases. In particular, I consider a reduction in minimum mandatory sentences, and the wider use of non-incarceration sentences for mild offenses. I am also able to assess the proportion of cases that are

I am grateful to Alessandro Lizzeri for invaluable support, encouragement and advice. I also benefited from guidance by and insightful conversations with John Asker and Quang Vuong. I would like to thank Andrew Daughety, Sandy Gordon, Jakub Kastl, Fernanda de Leon, Konrad Menzel, Isabelle Perrigne, Nicola Persico, Luigi Pistaferri, Jennifer Reinganum and seminar and conference participants at Caltech, Carnegie Mellon, Cornell, FGV-EESP, ITAM, Microsoft Research, Princeton, UCLA, Maryland, Minnesota, NYU, PUC-Rio, Stanford, Toronto, USC, Wisconsin, Vanderbilt, WUSTL, 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 .

currently settled for incarceration sentences but that would result in an acquittal at trial if plea bargains were banned.

The results presented here are relevant for the recent discussion on sentencing reform. Due mainly to the high incarceration rates in the United States, which have strong welfare and fiscal implications, a debate has developed over whether the current sentencing guidelines should be replaced by more lenient ones. Prior to any consideration of what the optimal guidelines should be, however, a better understanding is required of the possible effects of different sentencing reforms on the outcomes of prosecuted criminal cases. Such a task is complicated by the prevalence of plea bargaining in the American justice system. Indeed, the vast majority of criminal cases in the United States are settled without a trial.¹ Because of that, it is likely that the total impact of any sentencing policy depends largely on how it affects the plea-bargaining process, rather than on the outcome of cases actually brought to trial. Thus, in order to contemplate the effects of changes in the sentencing guidelines on the final outcome of criminal cases, it is necessary to take plea bargaining into account.

In the model I analyze, 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 that bargaining takes place under asymmetric information. The sentence that will be assigned in the event of a trial conviction is common knowledge.

Although the defendant's private information is distributed on a continuum, her action space consists simply of accepting or rejecting the prosecutor's settlement 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 the information on the distribution of defendants' types conveyed by the prosecutor's equilibrium offers. Specifically, I show that there is a one-to-one mapping between the the potential trial sentence and the prosecutor's optimal offer, and that, if such a mapping is known by the econometrician, the distribution of defendants' types can be recovered. The identification of the prosecutor's optimal offer as a function of the potential trial sentence, however, is complicated by a selection problem. Indeed, the prosecutor's offer is observed only in cases in which plea bargaining is successful, while the trial sentence is available only in cases resulting in a conviction at trial. Nevertheless, I show that the prosecutor's optimal offer function satisfies an equation relating the (observed) distributions of accepted settlement offers and assigned trial sentences. I then present sufficient conditions under which that functional equation can be solved. These conditions amount to the existence of a source of variation in the distribution of assigned sentences that does not affect the other primitives of the model. In my application, cross-judge heterogeneity in sentencing patterns serves as such a source of variation.

The estimation of the model closely follows the identification strategy. First, I compute kernel density estimates for the observed distributions of trial and settlement sentences. I then use sieve methods to solve an empirical analogue to the functional equation defined above and to obtain

¹In 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).

an estimator for the prosecutor's optimal offer function, which I show to be uniformly consistent. Employing such a function, I can recover all of the model's primitives..

I implement the estimation procedure using data on cases filed in the North Carolina Superior Courts, the most important general-purpose trial courts in that state, 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 incarceration sentences. I find that the asymmetric information between prosecutors and defendants is quantitatively relevant and that the prosecutors incur high costs for bringing a case to trial. Defendants, however, behave as if their costs of going to trial were negligible. Such differences in trial costs, together with the informational asymmetry, help to explain why sentences assigned in settled cases are shorter than those assigned in trial convictions.

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 an across-the-board 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 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 policy 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 consider the scenario of no asymmetric information between prosecutors and defendants. Besides serving as a reference for evaluating policies such as the enhancement of discovery rules, the results of this counterfactual analysis allow me to quantify the informational rents obtained by the defendants in the process of plea bargaining. I estimate such rents to be substantial. Without asymmetric information, I find that more defendants would be convicted and, conditional on a conviction, most of the sentences assigned would be longer than in the current scenario. In my last 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 well above 20 percent. Insofar as trials accurately convict the guilty and acquit the innocent, these numbers suggest that plea bargaining leads a disturbing number of defendants to be unjustly incarcerated. Nevertheless, my results indicate that, in expectation, the defendants would be considerably worse off if settlements were not allowed, relative to the present situation.

The identification and estimation strategies proposed here are adaptable to contexts that extend beyond the resolution of legal disputes. Many important situations in economics can be modeled as bargaining games in which one party is privately informed on the consequences of bargaining failure. An example is the acquisition of public companies, where the board of the target firm must decide whether to accept the acquirer's proposal for a friendly takeover. If negotiations fail, the acquirer may proceed with a hostile takeover, but the target's board has private information on the quality of the company's takeover defenses or on the likelihood of winning a proxy fight. Another example is debt renegotiation, where the debt holder is privately informed on the likelihood of going into bankruptcy if the renegotiation fails. To widen the applicability of my empirical strategy, I

generalize the identification result to settings in which more than two outcomes are possible, after bargaining failure.

The remainder of the paper is organized as follows: Section 2 discusses the contribution of this study to the literature. Section 3 describes relevant features of the North Carolina Superior Courts, as well as the data used in the paper. In Section 4, I present the theoretical model on which the empirical analysis is based. Section 5 provides preliminary evidence supporting the model and the identifying assumptions employed throughout the paper. The structural model is then presented in Section 6, followed by the empirical results, in Section 7. Section 8 contains the results of the policy experiments. Section 9 offers a deeper discussion on how my study relates to the existing literature, and Section 10 concludes. In the Appendix, I present the proofs omitted from the main text, details concerning the manipulation of the data, and extensions of my identification strategy to more general settings.

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 sentences to be assigned at trial affects plea bargaining outcomes. For example, Elder (1989) uses data on cases prosecuted in a group of U.S. cities and, consistent with theoretical predictions, provides evidence that factors that may aggravate a trial punishment (e.g., the possession of a firearm by the defendant at the moment of the alleged offense) reduce the probability of settlement. LaCasse and Payne (1999), examining cases from federal courts in New York, show a positive correlation between plea and trial sentences of cases under the responsibility of the same judge. Boylan (2012) finds similar results using data on a large set of federal courts. Kuziemko (2006) analyzes homicide cases in the state of New York and finds that defendants subject to the death penalty tend to accept harsher settlements. Taken together these results serve as strong evidence that plea bargaining takes place in the shadow of a trial.

My investigation differs from the existing plea-bargaining literature in that I develop and implement a framework for the structural analysis of data on judicial cases. This approach has several advantages over those of previous studies. First, I am able to quantify the relationship between potential trial sentences and plea-bargaining results. Also, I can recover objects that are not directly observable, such as the full distribution of sentences to be assigned at trial. Most importantly, my results allow me to conduct a series of policy experiments in order to evaluate the impact of different interventions on the justice system.

²The theoretical papers in the Law and Economics literature differ in how they explain why parties fail to reach a settlement agreement and, instead, proceed to a costly trial. One theory 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 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 two branches of literature. See Daughety and Reinganum (2012) for a detailed review.

³Theoretical papers that apply settlement models to the criminal context include Rubinfeld and Sappington (1987), Reinganum (1988), Kobayashi (1992), Miceli (1996), Baker and Mezzetti (2001) and Bjerk (2007).

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) and Watanabe (2009). My study differs from those in three important aspects: First, my focus is on criminal, rather than civil, cases. Second, I conduct a careful identification analysis, which shows what can be learned from the data using only restrictions derived from the theoretical model, and without imposing parametric assumptions on the model’s primitives. The non-parametric approach is particularly valuable in the empirical investigation of legal disputes since, due to the endogenous selection of cases for trial, the raw data provide little guidance about the adequacy of different parametric specifications for the model. Accordingly, my estimation procedure is totally non-parametric. Third, and due mainly to my focus on plea bargaining, the model estimated here departs from the ones used in previous papers. In particular, 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) is interested primarily in dynamic aspects of the bargaining process, such as the costs of bargaining delays, which are more relevant in the context of tort cases than in that of criminal ones. His model, for example, emphasizes 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. However, some features of Sieg’s model make it inappropriate for the analysis of criminal cases. Specifically, that model does not allow changes in the distribution of potential trial sentences to affect conviction rates, which are of great interest in the discussion of sentencing reform. As Section 4 makes clear, the model estimated in my paper does not have such a limitation. A detailed comparison between Sieg’s paper and mine is postponed to Section 9, after my model and results are presented.

More generally, my paper proposes a new strategy for the non-parametric identification and estimation of bargaining games with asymmetric information, which can, in principle, be applied to settings other than the settlement of judicial cases. Models of bargaining with asymmetric information are used in several fields of Economics, and recent years have seen a renewed interest in the empirical analysis of such models. Besides the papers on settlement mentioned above, recent studies by Merlo, Ortalo-Magné and Rust (2008), Keniston (2011) and Larssen (2014) estimate models of bargaining over the sale of a good in which agents have private information about their valuations. Using rich data on offers and counteroffers, these papers analyze interesting dynamic aspects of bargaining. However, data on many relevant topics are rarely available at such a level of detail. The techniques developed here are useful in the far more common scenario, where only the final outcome of bargaining is observable.

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) and Lim (2013) 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. Lim (2013) treats trial and settlement sentences equally, which may lead to overestimation of the differences in the sentencing patterns across different judges. Indeed, according to most settlement

models, including the one I consider, harsher trial sentences make it more difficult to settle cases. Moreover, settled sentences are normally considerably shorter than the ones assigned at trial. Therefore, if one specific judge tends to be harsher than another, the difference between the observed sentencing patterns of these two judges may be explained not only because the former assigns longer sentences than the latter, but also because cases under the responsibility of the former judge are more likely to be resolved at trial. Huber and Gordon (2004) and Gordon and Huber (2007) use indicators of whether a case is resolved by plea bargain as a control variable in their estimation procedures, but that does not account for the endogeneity of settlement. The empirical framework presented here allows me to recover the full distribution of sentences to be assigned at trial, which I can use to obtain appropriate measures of the sentencing behavior of different judges.

3. DATA AND INSTITUTIONAL DETAILS

3.1. North Carolina Superior Courts. To empirically evaluate how potential trial sentences affect plea bargaining, I use data on criminal cases prosecuted in the North Carolina Superior Courts—the highest of the general trial courts in North Carolina’s justice system. The Superior Courts have exclusive jurisdiction over all felony cases prosecuted in the state, 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.⁴ The state is divided into eight Superior Court divisions, and each division is further divided into districts for electoral and administrative purposes.⁵ There are 46 such districts statewide. Figure 1 contains a map that shows the exact apportionment of the state into Superior Court divisions and districts.

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.⁶ 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 decides whether the defendant is guilty or not guilty of the charged offenses.⁷ If the defendant is convicted, the judge decides on a sentence. The judge must follow a set of structured sentencing guidelines—i.e., a sentence is selected from a predetermined range of options, which depends on the severity of the

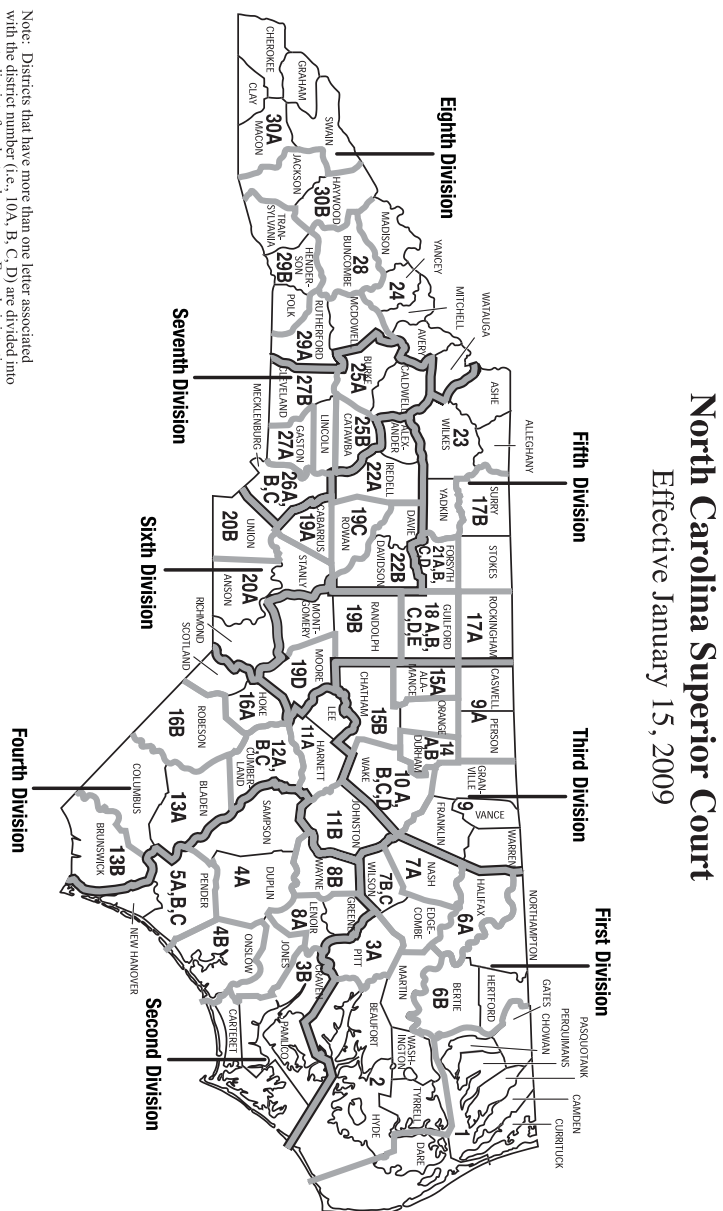
⁴Felonies are the most serious offenses in the criminal code, whereas misdemeanors are less serious crimes. Whether the main charge in a case is a felony or a misdemeanor often depends on several circumstances. In assault cases, for example, a felony may be characterized by the use of a lethal weapon, the intention to seriously injure, or the actual injuries suffered by the victim.

⁵Until 1999, there were only four Superior Court Divisions statewide. In that year, each division was split in two.

⁶A few judges may be appointed by the Governor for five-year terms.

⁷Jury trials take place even in cases appealed from the District Courts. In such cases, the defendant receives a whole new trial (*trial de novo*).

FIGURE 1. Superior Court Divisions and Districts



Note: Districts that have more than one letter associated with the district number (i.e., 10A, B, C, D) are divided into separate districts for electoral purposes. For administrative purposes, they are combined into a single district.

Copyright © 2009
School of Government
The University of North Carolina at Chapel Hill

crime and on the defendant’s previous criminal record.⁸ The sentence may include incarceration time and alternative punishments, such as fines, community service and probation.

⁸The judge sets only the minimum sentence length. The maximum length is determined according to a formula. There is no parole under the structured sentencing system in North Carolina. Offenders must serve at least their entire minimum sentence in incarceration and may serve up to the maximum length. Whether or not the offender serves more than the minimum sentence is determined by the Department of Corrections, not by the judge. For convenience, in the remainder of the paper, I refer to the assigned minimum sentence simply as the assigned sentence.

3.2. Data. I use case-level data provided by the North Carolina Administrative Office of the Courts.⁹ Such data comprise all criminal cases filed at the Superior Courts from 1996 to 2009 and include detailed information on case disposition, charged offenses and characteristics of the defendants. The data also identify the judge hearing each case. I match judges with their respective judicial district using the annual editions of the North Carolina Manual.¹⁰

The structural analysis to be presented in Section 6 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, the analysis must be undertaken using a reduced sample of cases, comprising offenses that do not vary much in nature. But the chosen sub-sample cannot be too small, in practice, since my non-parametric estimation procedure, described in Section 6, demands a relatively large number of observations to be implemented. I consider a sample of cases in which the main offense is a non-homicide violent crime—a category that consists of assault, sexual assault and robbery.¹¹ Drug-related offenses are not included. Non-homicide violent crimes are, at the same time, homogeneous and numerous enough to allow the estimation of the model.¹² 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 are assigned by the North Carolina justice system for the purposes of setting sentencing guidelines. In my data, these criminal record points range from zero to 98. I also observe the type of defense counsel employed in the case. Counsel services are provided by a public defender or, alternatively, by a privately-retained or court-appointed attorney.¹³

Table 1 contains descriptive statistics of the defendant’s characteristics. Over half of the defendants in the data are African-American, and the vast majority of the remaining defendants are non-Hispanic white. Less than four percent of the defendants are Hispanic and around two percent are from other race or ethnicity. More than 90 percent of the defendants are male, and the mean defendant’s age is approximately 29 years. Public defenders and court-appointed attorneys represent about 23 percent and 50 percent of the defendants, respectively. In roughly 22 percent

⁹Dyke (2007) uses the same data to analyze prosecutors’ political incentives.

¹⁰See the Appendix for details.

¹¹Robbery is defined as the taking or attempted taking of personal property, either by force, or by threat of force. It is, therefore, both a property crime and a crime against the person.

¹²The larger category of violent crimes, which in addition to the aforementioned crimes includes homicides, is employed by the Federal Bureau of Investigation in its Uniform Crime Reporting Program. Homicides are excluded from the sample because of the prevalence of death and life sentences assigned to defendants convicted of such crimes. There would be no clear way of incorporating death sentences in my analysis. See the Appendix for details on the classification of the offenses in the data.

¹³Public 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. Notice that 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 provided either by public defenders or court-appointed attorneys.

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%
Total	118948	100.00%
Gender		
Male	110495	92.89%
Female	8453	7.11%
Total	118948	100.00%
Counsel type		
Court-appointed	62325	52.40%
Public defender	27331	22.98%
Privately-held	25769	21.66%
Other	3523	2.96%
Total	118948	100.00%
Defendant’s age and previous criminal record		
Characteristic	Mean	Variance
Age [†]	28.88	10.64
Criminal record points	3.93	5.03
† : measured in years		

of the cases, the defendants are represented by a privately-retained attorney. Less than 3 percent of the defendants either employ others types of counsel services or waive counsel.

The focus of this study is on incarceration sentences assigned at trial or by plea bargain. In many of the cases in the data, the defendant is sentenced to an alternative punishment, such as probation, community service or the payment of a fine. Unfortunately, such punishments cannot be incorporated into my analysis since every criterion for representing incarceration and non-incarceration sentences on the same space would be arbitrary.¹⁴ Thus, in my empirical analysis, unless otherwise specified, I treat any alternative sentence as no sentence whatsoever.

Descriptive statistics of the case outcomes are presented in table 2. A total of 88.69 percent of the cases in the sample are solved by plea agreements—the majority of which result in an alternative sentence. The share of cases that settle for a 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 a conviction with incarceration time at trial. In 83.00 percent of the cases in the

¹⁴Moreover, in many cases where an alternative sentence is assigned, the data do not report the exact nature or the severity of such a sentence. Thus, in order to take alternative sentences into account in my analysis, I would either have to treat all such sentences as equivalent to each other, or I would have to deal with a severe missing-variables problem.

TABLE 2. Descriptive statistics – non-homicide violent crimes

Distribution of cases by outcome		
Case outcome	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 charged offense	Observations	Frequency
Felony	98726	83.00%
Other	20222	17.00%
Total	118948	100.00%
Sentences' length		
Conviction method	Mean	Standard deviation
Trial convictions [†]	100.73	106.41
Settlement convictions [†]	38.98	47.92
† : Measured in months		

sample, the main charge is a felony. In the other cases, the charge is either a misdemeanor or not known. Sentences assigned at trial tend to be much longer than those of settled cases. The average length of the former is 100.73 months, while that for the latter is 38.98 months. The sentence dispersion is very high. The standard deviations of trial and settlement sentences are 106.41 and 47.92, respectively.

4. THE MODEL

4.1. **Setup.** In this section, I briefly describe Bebchuk’s (1984) settlement model, which will guide my analysis in subsequent sections.¹⁵ Two agents—the prosecutor and the defendant—bargain over the outcome of a case. The agents’ utility functions are linear in the sentence assigned to

¹⁵Bebchuk’s game was originally proposed for the study of civil cases. Since I use the model for analyzing plea bargaining, the names of the agents and certain variables are changed in order to better suit the criminal-law context. For example, the plaintiff in the original model is named the prosecutor, the judgment to which a winning plaintiff is entitled is named a trial sentence, etc.

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 represents the strength of the case against the defendant and is drawn at the beginning of the game from a distribution F . The following technical assumptions are made:¹⁷ F is distributed over an interval $(\underline{\theta}, \bar{\theta}) \subseteq (0, 1)$ and is twice differentiable. The associated density function f is strictly positive on $(\underline{\theta}, \bar{\theta})$ and is non-increasing in a neighborhood of $\bar{\theta}$. Moreover, in order to rule out multiple equilibria, I assume that the hazard rate $f/[1 - F]$ is strictly increasing in θ .

I assume that bargaining takes place under asymmetric information. More precisely, only the defendant knows the realization θ of Θ . The assumption that the defendant has private information on the probability of being found guilty at trial may be motivated by noticing that most defendants know whether or not 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 committed the offense without the requisite criminal intent. Regardless of guilt, the defendant may have more information than the prosecutor on the quality of the evidence that may be used at trial (e.g., what, exactly, a given witness knows).¹⁸

If the defendant is convicted at trial, a common knowledge sentence t is assigned. The assumption that t is common knowledge may be interpreted as follows: At the bargaining stage, 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, for example, from (common knowledge) facts about the offense with which the defendant is charged, the defendant’s criminal record, sentencing guidelines, the judge’s previous sentencing patterns, etc. The prosecutor and the defendant, respectively, pay costs c_p and c_d for reaching the trial stage. The costs for the prosecutor are interpreted as the opportunity costs of taking a case to trial. For the defendant, the costs are associated with attorney fees and court fees in general. Similarly to t , the trial costs c_p and c_d are assumed to be common knowledge to both players.

The bargaining protocol is take-it-or-leave-it, so that the model constitutes a standard screening game. The prosecutor offers the defendant to settle for a sentence s . 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 the trial stage. The trial payoffs for the prosecutor are $t - c_p$ if the defendant is found guilty, and $-c_p$ otherwise. For the defendant, the trial payoffs are $-t - c_d$

¹⁶Alschuter (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. A similar utility function for the prosecutor is used in the classical paper by Landes (1971).

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

¹⁸Ideally, I should allow for private information on the prosecutor’s side, as well. Such an extension, however, would make the empirical analysis of the model impractical. See Schweizer (1989), Sobel (1989), Daughety and Reinganum (1994) and Friedman and Wittman (2007) for models of settlement with two-sided asymmetric information.

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). Since the prosecutor is a recurrent player in the plea-bargaining game, reputational incentives make the commitment with a take-it-or-leave-it offer plausible.

For now, I assume that the prosecutor cannot withdraw the case. Therefore, the case may be brought to trial even if the expected value of doing so is negative for the prosecutor. Such an assumption is justified in the criminal-law context since prosecutors often have incentives to stay in a case—either due to career concerns or for the sake of their reputation.¹⁹ Later, in my empirical analysis, I will clarify how dropped cases are dealt with.

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

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

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:

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

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

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

The equilibrium is characterized by equations (4.1) and (4.2). 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

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

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 and verifying how that affects the equilibrium outcome. 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

¹⁹Bebchuk (1984) assumes that $\theta t \geq c_p$ holds, so that it is always better for the prosecutor to go to trial than to drop the case. See Nalebuff (1987) for a game in which the prosecutor cannot commit to bringing a case to trial if, by doing that, the expected payoff becomes negative.

both functions is \mathfrak{R}_{++} . Bebchuk (1984) shows that both $\tilde{s}(\cdot)$ and $\tilde{\theta}(\cdot)$ are strictly increasing, and $\tilde{s}(\cdot)$ is strictly convex.²⁰

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, it is clear from the right-hand side of (4.3) that the probability of conviction also decreases strictly with t .

5. REDUCED-FORM ANALYSIS

In the following sections, I develop a structural analysis of the data based on the model of judicial bargaining under asymmetric information described in Section 4. Before that, however, I present the results of a reduced-form analysis that has two main objectives: The first one is to verify whether the data are consistent with basic predictions of the model. I do so by presenting evidence that the settlement of a case becomes less likely as the sentence to be assigned in the event of a trial conviction increases. The second objective is to classify the judges in the data according to their sentencing behavior. Such a classification, which is based on the results of a regression of assigned sentence lengths on judge-specific dummies, is employed in subsequent sections for the estimation of the structural model.

5.1. Reduced-form evidence for the model. An important prediction of the bargaining model discussed in Section 4 is that the probability of a successful settlement is decreasing in the length of the trial sentence. It is useful to present evidence supporting this prediction. A simple way of doing so is by comparing the likelihoods of settlement across cases in which the main offenses differ in nature. Table 3 presents settlement rates and average incarceration sentences for several categories of crimes in the data. The table separately reports the average sentence lengths for cases resolved at trial and by plea bargain. The offense categories considered are homicide, non-homicide violent crimes, property crimes, drug-related crimes and other offenses.²¹ Here, I consider as settled all cases decided by plea bargain—independent of whether the sentence includes incarceration time. That is because, otherwise, the settlement rates would largely capture differences in the probabilities of incarceration across crime categories. The table shows that, as expected, sentences assigned to defendants convicted of homicide are very long.²² Among the categories displayed in the table, non-homicide violent crimes have the second longest average sentences, followed by drug-related crimes and property crimes. More interestingly, the table suggests a negative relationship between average sentences and the likelihood of settlement. The settlement ratios for homicides, non-homicide violent crimes, drug-related crimes and property crimes are, respectively, 81.83 percent, 89.94 percent, 96.95 percent and 98.26 percent. It is also worth noticing that cases in which the average trial sentences are long tend to settle for relatively long sentences. This observation provides further evidence that settlement negotiations take place in the shadow of the trial.

²⁰That $\tilde{\theta}(\cdot)$ is strictly increasing follows from (4.2) and the assumption that the hazard rate for the defendants’ type is strictly increasing. The strict monotonicity and convexity of $\tilde{s}(\cdot)$ are then obtained from (4.1).

²¹Property crimes include burglary, larceny and arson. Drug-related crimes comprise both trafficking and possession.

²²The numbers in table 3 underestimate the average homicide sentence since death sentences are not accounted for in the computation.

TABLE 3. Settlement rates and sentences by type of offense

Distribution of cases by outcome				
Nature of the offense	Observations	% Settled	Average sentence (settled cases) [†]	Average sentence (trial cases) [†]
Violent (non-homicide)	118948	89.94%	38.98	100.73
Homicide	6611	81.83%	134.71	236.76
Property	215505	98.26%	13.36	49.68
Drugs	199692	96.95%	18.63	51.44
Other	313622	96.09%	18.47	37.03
Total	854378	95.86%	25.05	81.30

† : Measured in months

Another method of testing whether cases with long potential trial sentences are settled less often is to explore the correlation between settlement ratios and the sentencing patterns of different judges. There are several reasons why some judges can be harsher than others. North Carolina Superior Court judges are elected, and the electorate’s preferences can vary across districts. Also, judges in different stages in their careers may have distinct incentives to pander to voters. For example, it is possible that young judges want to establish a reputation as tough on crime and, thus, tend to assign harsher punishments than their senior counterparts. Moreover, since electoral accountability is unlikely to be perfect, judges’ personal preferences probably explain much of the heterogeneity in their sentencing behavior.

As previously mentioned, an important feature of the North Carolina justice system is the rotation of judges across districts within the same Superior Court divisions. The rotation schedule is centrally determined by the state administration, which makes it plausible to assume that cases are randomly assigned to judges.²³ That allows me to treat the caseloads of different judges as identical and to attribute any variation in the sentencing patterns to judges’ characteristics. I can then verify whether the settlement ratios of cases decided by harsher judges are smaller than those of cases decided by more lenient ones.

More precisely, consider the following regression model:²⁴

$$sentence_i = \vartheta_1 X_i + \zeta_1 \mathbf{Judge}_i + \epsilon_{1i}, \quad (5.1)$$

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 , X_i is a vector of control variables and ϵ_{1i} is an error term. Consider, also, the following model

$$settled_i = \vartheta_2 X_i + \zeta_2 \mathbf{Judge}_i + \epsilon_{2i}, \quad (5.2)$$

²³Later in this section I present empirical evidence supporting the random assignment hypothesis.

²⁴For an analogous exercise using data on federal criminal cases, see Boylan (2012). Waldfoegel (1998) undertakes a similar analysis using data on civil cases. The results of both papers are similar to the ones presented in this section.

TABLE 4. Regression Outcomes

	(1)	(2)	(3)
	Dependent Variable		
	<i>sentence</i> [†]	<i>settled</i>	<i>harsh</i>
<i>settled</i>	-39.67 (1.31)	-	-
<i>age</i>	-0.05 (0.20)	-0.01 (0.00)	0.00 (0.00)
<i>age</i> ²	0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
<i>female</i>	-4.31 (0.76)	- 0.06 (0.00)	-0.01 (0.01)
<i>black</i>	0.93 (0.68)	0.00 (0.00)	-0.01 (0.01)
<i>hispanic</i>	-1.42 (1.26)	0.19 (0.01)	-0.01 (0.01)
<i>private attorney</i>	-3.78 (0.75)	-0.09 (0.00)	0.00 (0.01)
<i>public defender</i>	-3.54 (0.74)	-0.02 (0.01)	-0.03 (0.03)
Judge dummies	Yes	Yes	No
County dummies	Yes	Yes	No
Superior Court division dummies	No	No	Yes
Observations	36644	97942	97942
<i>R</i> ²	0.51	0.27	0.28
Other controls: Year of disposition, offense severity and defendant's criminal record.			
Reference category for counsel type: Court-assigned attorney.			
Standard errors (robust to clustering at the judge level) in parenthesis.			
†: Measured in months			

where $settled_i$ is a dummy indicating whether case i is resolved by plea bargain, \mathbf{Judge}_i and X_i are defined as before, and ϵ_{2i} is another error term.²⁵ I estimate the two equations above by OLS, using only cases in which the main offense is a non-homicide violent crime. In both specifications, the vector of control variables X_i includes the defendant's gender, racial / ethnic group, previous criminal record, age and squared age, as well as dummies indicating the type of attorney representing the defendant.²⁶ I also include as controls dummies indicating the county where the case is prosecuted, the year of disposition and the severity of the main charged offense,

²⁵Here, and in all the empirical analysis in the remainder of this paper, I define a case as settled only if the prosecutor and the defendant agree on a incarceration sentence.

²⁶The defendant's previous criminal record is incorporated as follows in specifications (5.1) and (5.2): 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.

as defined by the North Carolina sentencing guidelines. Finally, I add the dummy *settled* as a control variable in specification 5.1. The results of each regression are presented in columns (1) and (2) of table 4. Here, I am directly interested in the relation between $\hat{\zeta}_1$ and $\hat{\zeta}_2$, the vectors of estimated coefficients for the judge-specific dummies in equations (5.1) and (5.2), respectively. A negative correlation between such two vectors suggests that cases decided by harsh judges are less likely to be resolved by a plea bargain, as predicted by the model in Section 4. I find the Pearson's correlation coefficient to be -0.23 and significant at the ten percent level.²⁷

There are evident problems with the OLS estimation of equation (5.1). Indeed, the regression includes only cases in which the defendant receives a incarceration sentence, which may generate sample selection. Also, it is not plausible to assume that $E(\epsilon_1\epsilon_2) = 0$, so that the variable *settled*_{*i*} is endogenous in (5.1). Unfortunately, there is no clear instrument available for that variable. Therefore, one should be careful in interpreting the above results. Still, together with the variation in settlement ratios from table 3, the negative correlation between $\hat{\zeta}_1$ and $\hat{\zeta}_2$ provides support for the model discussed in Section 4. Starting in Section 6, I use that model as a base for the structural analysis of the data.

5.2. Classifying judges as lenient or harsh. The identification of the structural model to be presented in Section 6 requires the econometrician to observe variation in the distribution of potential trial sentences. Such a variation must be unrelated to other aspects of the the case, such as the distribution of trial costs and the probability that the jury finds the defendant guilty in the event of a trial.²⁸ In this paper I employ a particular source on variation in trial sentences: differences in the observed sentencing patterns across judges. Precisely, I consider the estimated judge fixed effects from specification (5.1). Table 5 contains information on the distribution of such estimates. The mean fixed effect is 31.58 months, and the median is 31.51 months. The results in the table suggest that 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, depending on whether their estimated fixed effects are smaller or larger the median judge fixed effect reported in table 5. Judges whose 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. Such a classification of judges is useful for the estimation of the structural model of Section 6. The estimation procedure proposed in that section is based on the comparison of the distributions of sentences assigned by two or more groups of judges. In principle, I could implement the estimator employing a finer classification of the judges in the data. The advantage of a binary classification is that it maximizes the number of judges within each group. Therefore, given the non-parametric nature of the estimator to be proposed and the limited number of sentences observed in the data, the use of a binary classification of judges, such as the one considered here, is justified.

²⁷The standard deviation of the correlation coefficient, calculated from 1000 bootstrap samples, is 0.1242. The associated p-value is 0.067. Replicating the exercise with the entire data set (i.e., not only non-homicide violent crimes), I find a correlation coefficient of -0.18 that is significant at all conventional levels.

²⁸These conditions are formally defined in Section 6.

TABLE 5. Variation of sentencing behavior across judges

Estimated judge fixed effects [†]						
mean	standard deviation	minimum	1st quartile	median	3rd quartile	maximum
31.58	6.26	16.28	27.56	31.51	35.62	64.63
Number of judges: 169						
†: Fixed effects measured in months and estimated using specification (5.1).						

As mentioned above, for the structural model to be identified, the variation in the distribution of trial sentences arising from differences in sentencing behavior across judges must be uncorrelated with other variables that may affect the outcome of the case. This is an assumption that, unfortunately, cannot be tested. Notice, however, that the assumption holds if the cases in the data are indeed randomly assigned to the judges. It is possible to provide some support for the random assignment hypothesis by verifying whether the distribution of observable case characteristics varies substantially across lenient and harsh judges. I consider the following specification:

$$harsh_i = \vartheta_3 X_i + \epsilon_{3i}, \quad (5.3)$$

where $harsh_i$ is a dummy indicating whether the judge responsible for case i is a harsh one, X_i is a vector of case characteristics and ϵ_{3i} is an error term. The variables included in X_i are the defense attorney type and the defendant’s gender, race/ethnicity, previous criminal record, age and age squared. I also include, as controls, dummies indicating the year of disposition of the case, the severity of the offense and the Superior Court division where the case is prosecuted. All these variables are arguably out of the judge’s control. Column (3) in table 4 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 percent level. These results suggest that the cases in the data do not differ considerably across the two judge categories.

The empirical evidence presented in this section provides support for the bargaining model developed in Section 4. Moreover, the results above document substantial cross-judge heterogeneity in sentencing behavior, and indicate that such a heterogeneity is unrelated to other variables that may affect the outcome of the cases in the data. These differences in sentencing behavior across judges play a fundamental role in the identification of the structural model proposed in the next section.

6. STRUCTURAL MODEL

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 show sufficient conditions for the model to be identified and propose an estimator for it.

6.1. Data-generating process.

6.1.1. *Primitives.* For every case i , let the (possibly vector valued) random variable Z_i , defined on the space Δ , describe case-level characteristics that are observable to the econometrician. Such characteristics include variables related to the charged offense, the defendant, the judge responsible for the case, the time and the location where prosecution takes place, etc. The other primitives of the structural model, which are presented below, are defined conditional on Z_i .

To each case i corresponds a potential incarceration sentence, which is assigned if the defendant is convicted at trial. Note that such a sentence exists even for cases settled by a plea bargain or for cases in which there is a trial, but the defendant is found not guilty. The potential trial sentence is described by the random variable T_i , which, conditional on Z_i , is assumed to be 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 $\underline{t} > 0$.²⁹ 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 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, conditional on Z_i , Θ_i is i.i.d. across cases, and let $F(\cdot|Z_i)$ be its conditional distribution. Assume, also, that Θ_i and T_i are independent, conditional on Z_i .³⁰ Make the following technical assumptions: $F(\cdot|Z_i)$ is twice differentiable, and the associated conditional density, denoted by $f(\cdot|Z_i)$, is strictly positive over the whole support and non-increasing on a neighborhood of $\bar{\theta}$. Moreover, assume that the conditional hazard rate $f(\cdot|Z_i)/[1 - F(\cdot|Z_i)]$ is strictly increasing.

Let $c_d(Z_i)$ and $c_p(Z_i)$ be the trial costs for the defendant and the prosecutor, respectively. Notice that, conditional on Z_i , such variables are deterministic. At this point, two observations are in order. First, $c_d(Z_i)$ and $c_p(Z_i)$ are just the costs expected by the agents at the moment of plea bargaining. It is possible that the actual trial costs vary across cases. Second, as shown below, the model is actually overidentified under the assumption that, conditional on Z_i , the trial costs are constant. Indeed, I am still able to identify the model after allowing the trial costs for the defendant and the prosecutor to vary with T_i in a simple (i.e., depending on few parameters) and deterministic way. In the estimation of the model, I assume that $c_d(Z_i)$ and $c_p(Z_i)$ are linear functions of the trial costs.

The primitives of the structural model are then: (i) the conditional distribution of potential trial sentences, characterized by the functions $\nu(Z_i)$ and $G(\cdot|Z_i)$; (ii) the conditional distribution of defendants' types, given by the $F(\cdot|Z_i)$; (iii) the trial costs $c_d(Z_i)$ and $c_p(Z_i)$; and (iv) the distribution of Z_i .

6.1.2. *Observables.* For every case i , given realizations z_i of Z_i , t_i of T_i and θ_i of Θ_i , as well as $F(\cdot|Z_i = z_i)$, $c_d(z_i)$ and $c_p(z_i)$, the prosecutor and defendant play the game described in Section 4. A prosecutor's settlement offer to the defendant corresponds to each case. As in Section 4, use

²⁹I could, in principle, allow the support $[\underline{t}, \bar{t}]$ to vary with Z_i . In the estimation of the model, which is discussed below, I chose not to do so due to sample size limitations. Accordingly, here I assume a constant support, in order to keep the notation as clear as possible.

³⁰I can relax the independence assumption and still obtain partial identification of the model. See the Appendix for details.

equations (4.1) and (4.2) to define the functions $\tilde{\theta}(t, z)$ and $\tilde{s}(t, z)$. Such functions, respectively, return the defendant's equilibrium cutoff point and the prosecutor's equilibrium offer, given a potential trial sentence t and case characteristics z ³¹. Assume that $\tilde{\theta}(0, z) = 0$ and $\tilde{s}(0, z) = 0$ for every z . Hence, if a non-incarceration sentence is to be assigned following a conviction at trial, then the prosecutor offers to settle for a non-incarceration sentence. In equilibrium, the defendant always takes such offer, so that the case settles with certainty.³²

The prosecutor's offer for case i , therefore, is described by the random variable $S_i = \tilde{s}(T_i, Z_i)$. Conditional on Z_i , S_i is i.i.d. across cases, and, similarly to T_i , follows a mixed distribution. With probability $\nu(Z_i)$, it equals zero. With probability $1 - \nu(Z_i)$, S_i is conditionally distributed according to the CDF $B(\cdot|Z_i)$, which has support $[\tilde{s}(\underline{t}, z), \tilde{s}(\bar{t}, z)]$. Notice that $B(\cdot|Z_i)$ is continuous, and let $b(\cdot|Z_i)$ be the associated density.³³

The model implies a selection process that complicates identification. The realization t_i of T_i is observable to the econometrician only in cases in which the defendant is convicted at trial. Similarly, the realization s_i of S_i is available only in cases settled by plea bargain. Formally, let Ψ_i be a discrete random variable describing the way case i is resolved. Define

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

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 $\psi_i = 1$. In order 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.

6.2. Identification. Since Θ is distributed continuously on $(\underline{\theta}, \bar{\theta})$, and the defendants' action space consists of merely accepting or rejecting the prosecutor's offer, there is no one-to-one mapping between the defendants' types and actions. As a consequence, techniques often employed by the auctions literature for the identification of private types cannot be used here.³⁴ However, the prosecutor's equilibrium offer conveys information on the distribution F . As shown below, such information can be explored to recover all the model's primitives.

The strategy described here provides the identification of the model's primitives, conditional on a realization z of Z . It consists of two main parts. First, I identify the prosecutor's optimal offer

³¹Notice that $\tilde{\theta}(t, z)$ and $\tilde{s}(t, z)$ depend on z through $F(\cdot|Z_i = z)$, $c_d(z)$ and $c_p(z)$.

³²In my analysis, I consider all non-incarceration sentences as no sentence whatsoever. Thus, I make no distinction between cases that settle for an alternative sentence and cases that the prosecutor drops. Notice that the model does not allow for the assignment of alternative sentences at trial. As shown in table 2, which describes the data employed in the structural analysis, that is a rare event. Accommodating it would make the identification of the model considerably more difficult, so I chose not to do that.

³³To see that $B(\cdot|Z_i)$ is, indeed, continuous, notice that both $G(\cdot|Z_i)$ and $\tilde{s}(t, z)$ are continuous.

³⁴See Athey and Haile (2007) for an introduction to identification strategies for the empirical analysis of auctions.

function $\tilde{s}(\cdot, z)$. Then, using $\tilde{s}(\cdot, z)$, I show how to recover $G(\cdot|Z = z)$, $F(\cdot|Z = z)$, $c_d(z)$ and $c_p(z)$. Remember that the case characteristics, represented by Z , are distributed over the space Δ . Since Z is always observable, the identification of its distribution is straightforward, and I do not discuss it here. The same observation holds for $\nu(z)$, as $P[\Psi = 0|Z = z]$ is observable.

6.2.1. *Conditions A and B.* I begin by defining the following two conditions, which pose restrictions on subsets of the space of case characteristics Δ . Together with the technical assumptions stated in the definition of the primitives, such conditions are sufficient for the identification of the model.

DEFINITION 1. *A subset Δ_ρ of Δ satisfies **condition A** if, for every pair z', z'' in Δ_ρ , the following equalities hold*

$$\begin{aligned} c_d(z') &= c_d(z'') \\ c_p(z') &= c_p(z'') \\ \text{and} \quad F(\theta|Z = z') &= F(\theta|Z = z'') \end{aligned}$$

for all $\theta \in (\underline{\theta}, \bar{\theta})$.

The trial costs and the distribution of defendants' types are, thus, constant across all z within a subset Δ_ρ that satisfies condition A. Notice that Δ_ρ satisfying condition A implies that $\tilde{s}(\cdot, z') = \tilde{s}(\cdot, z'')$ and $\tilde{\theta}(\cdot, z') = \tilde{\theta}(\cdot, z'')$ for all z', z'' in Δ_ρ . Notice, also, that the condition does not restrict the distribution of potential trial sentences within Δ_ρ . Condition A is not vacuous since it is trivially valid for every singleton subset of Δ . It is also plausible to assume that the condition holds for subsets of Δ that present little variation in the nature of the main offense, as well as in the characteristics of potential jury members, prosecutors and defendants—the main determinants of trial costs and the probability of convictions at trial.

Before I state the next definition, notice that a partition of an interval $[a, b]$ is a finite sequence of the form

$$a = x_0 < x_1 < \dots < x_N = b$$

I now define the second identification condition.

DEFINITION 2. *Consider the function*

$$\phi(t, z', z'') \equiv \frac{g(t|Z = z')}{g(t|Z = z'')}.$$

*A subset Δ_ρ of Δ satisfies **condition B** on a bounded interval $X \subseteq [t, \bar{t}]$ if there is a partition $\{x_0, x_1, \dots, x_N\}$ of X such that, for all $n \in \{1, \dots, N\}$, there are elements $z', z'' \in \Delta_\rho$ such that the function $\phi(\cdot, z', z'')$ is strictly monotonic in t on the interval $[x_{n-1}, x_n]$.*

The function $\phi(\cdot, z', z'')$ consists of the ratio of the trial sentence densities $g(\cdot|Z = z')$ and $g(\cdot|Z = z'')$. If Δ_ρ contains only the two elements z' and z'' , then it satisfies condition B on an interval X if and only if $\phi(\cdot, z', z'')$ is piecewise monotone on X . If Δ_ρ has more than two elements, then condition B is easier to satisfy. It is enough that X can be partitioned in a finite number of intervals, and, on each of such intervals, $\phi(\cdot, z', z'')$ is piecewise monotone for some pair z', z'' of elements of Δ_ρ .

Conditions A and B can be regarded as exclusion and inclusion restrictions, respectively, similar to the conditions for an instrumental variable. They require that variation in some dimension of Z affects the conditional distribution of potential trial sentences, but not the other primitives in the model. However, for Δ_ρ to satisfy condition B , notice that there must be enough variation within Δ_ρ , so that the distribution of trial sentences is affected along its whole support.³⁵

The selection of a subset Δ_ρ that simultaneously satisfies conditions A and B involves a clear trade-off. On the one hand, the fewer elements Δ_ρ has, the more plausible condition A is. Condition B , on the other hand, becomes harder to satisfy as Δ_ρ gets small.³⁶ I propose the following choice of Δ_ρ : In order to satisfy condition A , I must maintain a degree of homogeneity across cases. With that intent, I restrict Δ_ρ according to the nature of the charged offense and to characteristics of the defendant such as gender, race and previous criminal background. Moreover, I restrict Δ_ρ by the place where the case is prosecuted, which helps controlling for characteristics of the prosecutor and the jury members. The variation in the conditional distributions of trial sentences, which is necessary for condition B to be satisfied, is then obtained by allowing judge-specific characteristics to vary within Δ_ρ . Specifically, in my application, I explore variation in sentencing behavior across judges classified as lenient and harsh, as defined in section 5. Each Δ_ρ considered has, thus, two elements—one for each of these groups of judges. The rotation system imposed by the North Carolina constitution ensures that, throughout a long enough time period, different judges decide on similar cases in the same location. The evidence presented in section 5 indicates that the distributions of cases under the responsibility of lenient and harsh judges are, indeed, indistinguishable from one another—making it valid, for the purposes of my analysis, to employ differences in the sentencing patterns between these two groups as a source of variation in the distribution of trial sentences.

6.2.2. *Prosecutor's settlement offer function.* Remember that $g(\cdot|Z)$ is the pdf of potential trial sentences, and $b(\cdot|Z)$ is the pdf of prosecutors' settlement offers, conditional on Z . The following lemma is useful for the identification of $\tilde{s}(\cdot, z)$:

LEMMA 1. *Assume that $(b[s|Z = z'] / b[s|Z = z''])$ and $(g(t|Z = z') / g(t|Z = z''))$ are known for every $z', z'' \in \Delta_\rho$, $s \in [\tilde{s}(\underline{t}, z), \tilde{s}(\bar{t}, z)]$ and $t \in [\underline{t}, \bar{t}]$. Assume, also, that condition B holds on $[\underline{t}, \bar{t}]$, and that there exists $\check{z} \in \Delta_\rho$ such that*

$$\frac{b[\tilde{s}(t, \check{z})|Z = z']}{b[\tilde{s}(t, \check{z})|Z = z'']} = \frac{g(t|Z = z')}{g(t|Z = z'')}$$

for all $z', z'' \in \Delta_\rho$ and $t \in [\underline{t}, \bar{t}]$. Then, the function $\tilde{s}(\cdot, \check{z})$ is identified over $[\underline{t}, \bar{t}]$.

See the appendix for the proof.

The functional equation in the statement of Lemma 1 relates ratios of densities for two different elements of Δ_ρ . On the right-hand side is a ratio of trial sentence densities, evaluated at t . On the left-hand side is a ratio of settlement offer densities, evaluated at $\tilde{s}(t, \check{z})$. The lemma shows that, if

³⁵Like the exclusion condition for an instrumental variable, condition A cannot be tested. The validity of condition B , however, can be verified empirically. See footnote 35 below.

³⁶As discussed below, the estimation of the model for a small Δ_ρ may also be impractical, given a limited sample size.

such a functional equation is valid for all t , then the left-hand side can be inverted to recover the function $\tilde{s}(\cdot, \check{z})$. Condition B ensures that the left-hand side of the functional equation is piecewise invertible.

My first main result states that, under conditions A and B , the prosecutor's optimal offer function is identified.

PROPOSITION 1. *Assume that the set $\Delta_\rho \subseteq \Delta$ satisfies conditions A and B on $[\underline{t}, \bar{t}]$. Then, for all $z \in \Delta_\rho$, the prosecutor's optimal offer function $\tilde{s}(\cdot, z)$ is identified over the whole interval $[\underline{t}, \bar{t}]$.*

Condition A allows me to write a functional equation for $\tilde{s}(t, z)$, equivalent to the one that appears in the statement of Lemma 1. The same condition is also used to recover the two sides of the equation from the observables. Using condition B , I can then apply Lemma 1 and identify $\tilde{s}(t, z)$ for all $z \in \Delta_\rho$. I present the proof of the proposition in the main text since it is important for understanding the estimation procedure proposed later in the paper.

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

$$b(s|Z = z) = g(\tilde{s}^{-1}(s, z)|Z = z) \frac{d\tilde{s}^{-1}(s, z)}{ds} \quad (6.1)$$

for all $s \in [\tilde{s}(\underline{t}, z), \tilde{s}(\bar{t}, z)]$ and all $z \in \Delta$

Therefore, for every pair $z', z'' \in \Delta_\rho$ and all $t \in [\underline{t}, \bar{t}]$, I have that

$$\begin{aligned} \frac{b[\tilde{s}(t, z')|Z = z']}{b[\tilde{s}(t, z'')|Z = z'']} &= \frac{g(\tilde{s}^{-1}[\tilde{s}(t, z'), z']|Z = z') \frac{d\tilde{s}^{-1}(\tilde{s}(t, z), z')}{ds}}{g(\tilde{s}^{-1}[\tilde{s}(t, z''), z'']|Z = z'') \frac{d\tilde{s}^{-1}(\tilde{s}(t, z), z'')}{ds}} \\ &= \frac{g(t|Z = z')}{g(t|Z = z'')}, \end{aligned}$$

where the derivatives of the right-hand side of the first equality cancel out due to condition A . Now consider any $\check{z} \in \Delta_\rho$, and notice that, again by condition A , $\tilde{s}(\cdot, \check{z}) = \tilde{s}(\cdot, z') = \tilde{s}(\cdot, z'')$. The equation above then implies

$$\frac{b[\tilde{s}(t, \check{z})|Z = z']}{b[\tilde{s}(t, \check{z})|Z = z'']} = \frac{g(t|Z = z')}{g(t|Z = z'')} \quad (6.2)$$

for all $z', z'' \in \Delta_\rho$ and $t \in [\underline{t}, \bar{t}]$.

If the full distribution of offers and potential trial sentences were observed, I would be able to apply Lemma 1 to equation (6.2) in order to recover $\tilde{s}(t, z)$ for all $z \in \Delta_\rho$. However, due to selection, I do not observe such distributions, so that (6.2) cannot be used directly. Still, I can employ the censored versions of the distributions of settlement offers and trial sentences to obtain the density ratios in (6.2). Consider, first, the settlement offers. Although $b(\cdot|Z = z)$ is not available, 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

$$P[\Psi = 1|S = s, Z = z] = 1 - F\left[\tilde{\theta}(\tilde{s}^{-1}(s, z), z)|Z = z\right]. \quad (6.3)$$

From Bayes' rule, I have that

$$b(s|\Psi = 1, Z = z) = \frac{P[\Psi = 1|S = s, Z = z]b(s|Z = z)}{P[\Psi = 1|Z = z]}. \quad (6.4)$$

Notice that the support of $b(s|\Psi = 1, Z = z)$ is still $[\tilde{s}(\underline{t}, z'), \tilde{s}(\bar{t}, z')]$.³⁷ Since Δ_ρ satisfies condition A, $P[\Psi = 1|S = s, Z = z'] = P[\Psi = 1|S = s, Z = z'']$ for all s and every z' and z'' in Δ_ρ . Hence, I can write

$$\frac{b(s|\Psi = 1, Z = z')}{b(s|\Psi = 1, Z = z'')} \frac{P[\Psi = 1|Z = z']}{P[\Psi = 1|Z = z'']} = \frac{b(s|Z = z')}{b(s|Z = z'')} \quad (6.5)$$

for every $s \in [\tilde{s}(\underline{t}, z'), \tilde{s}(\bar{t}, z')]$ and all $z', z'' \in \Delta_\rho$. Therefore, I recover the density ratio on the left-hand side of equation (6.2), using the (observable) censored distribution of accepted settlement offers.

Now, turning to the potential trial sentences, I do not observe the full conditional density $g(\cdot|Z = z)$. However, its censored counterpart, $g(\cdot|\Psi = 2, Z = z)$, is available for all z . Given any sentence value t and any $z \in \Delta$, the probability of conviction at trial is

$$P[\Psi = 2|T = t, Z = z] = \int_{\underline{\theta}}^{\tilde{\theta}(t, z)} x f(x|Z = z) dx. \quad (6.6)$$

Using Bayes' rule once more, I have that

$$g(t|\Psi = 2, Z = z) = \frac{P[\Psi = 2|T = t, Z = z]g(t|Z = z)}{P[\Psi = 2|Z = z]}. \quad (6.7)$$

Notice that the support of $g(t|\Psi = 2, Z = z)$ is still $[\underline{t}, \bar{t}]$.³⁸ Condition A implies that, for all t and every z' and z'' in Δ_ρ , $P[\Psi = 2|T = t, Z = z'] = P[\Psi = 2|T = t, Z = z'']$. Hence,

$$\frac{g(t|\Psi = 2, Z = z')}{g(t|\Psi = 2, Z = z'')} \frac{P[\Psi = 2|Z = z']}{P[\Psi = 2|Z = z'']} = \frac{g(t|Z = z')}{g(t|Z = z'')} \quad (6.8)$$

for all $t \in [\underline{t}, \bar{t}]$ and all z', z'' in Δ_ρ . 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 equation (6.2).³⁹ The proof of the proposition then follows from Lemma 1. \square

6.2.3. Completing the identification of the model. I now show that, once $\tilde{s}(\cdot, z)$ is identified for a given value z , the model's primitives can be recovered, conditional on $Z = z$. Intuitively, the prosecutor's optimal offer function conveys information about the distribution of defendants' private types, which allows the identification of the whole model. The next proposition formalizes this idea.

PROPOSITION 2. *Assume that, for all $z \in \Delta_\rho$, the prosecutor's optimal offer function $\tilde{s}(\cdot, z)$ is identified over the whole interval $[\underline{t}, \bar{t}]$. Then, for all $z \in \Delta_\rho$, 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 $F(\cdot|Z = z)$, over the interval $[\tilde{\theta}(\underline{t}, z), \tilde{\theta}(\bar{t}, z)]$; and (iii) the trial costs $c_d(z)$ and $c_p(z)$.*

The proposition is based on the insight that, knowing $\tilde{s}(\cdot, z)$, I can employ the first-order condition of the prosecutor's optimization problem to recover the hazard function for the distribution

³⁷That is because the defendant's equilibrium cutoff $\tilde{\theta}(\tilde{s}^{-1}(s, z), z)$ belongs to the interval $(0, 1)$ for all values of s , so that $P[\Psi = 1|S = s, Z = z]$ is always strictly positive.

³⁸Indeed, $\tilde{\theta}(t, z) \in (0, 1)$ for all t , and $f(\theta|Z = z) > 0$ for all $\theta \in (\underline{\theta}, \bar{\theta})$, so that $P[\Psi = 2|T = t, Z = z]$ is always strictly positive.

³⁹Equation (6.8) points to how condition B can be verified using the data, although a formal test would require the computation of confidence bands for the density ratios.

of defendants' types. Using the hazard function, I am able to identify the distribution itself, followed by all the other primitives of the model. Notice that the distribution of defendants' types $F(\cdot|Z = z)$ is identified only over the interval $[\tilde{\theta}(\underline{t}, z), \tilde{\theta}(\bar{t}, z)]$. As in Proposition 1, the proof of Proposition 2 facilitates the understanding of the estimation procedure, so that I present it in the main text.

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

$$\tilde{\theta}(t, z) = \frac{\tilde{s}(t, z) - c_d(z)}{t} \quad (6.9)$$

for all $t \in [\underline{t}, \bar{t}]$, so that $\tilde{\theta}(\cdot, z)$ is identified up to the constant $c_d(z)$. I later discuss how to recover such a constant.⁴⁰

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

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

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

$$\lambda(\theta, z) = \frac{\tilde{\theta}^{-1}(\theta, z)}{c_p(z) + c_d(z)} \quad (6.10)$$

for every θ in the interval $[\tilde{\theta}(\underline{t}, z), \tilde{\theta}(\bar{t}, z)]$. I can then write

$$F(\theta|Z = z) = 1 - \mu(z) \exp\left(-\int_{\tilde{\theta}(\underline{t}, z)}^{\theta} \lambda(x, z) dx\right) \quad (6.11)$$

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

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

⁴⁰The proof of the proposition shows that the model is overidentified under the assumption that $c_d(z)$ and $c_p(z)$ are constant. It is trivial to extend the analysis to the case in which the trial costs are parametric functions of the trial sentence T . In the estimation of the model, I allow $c_d(z)$ and $c_p(z)$ to vary linearly with T . More precisely, I assume that the trial costs for the defendant and the prosecutor are given, respectively, by $c_d(t, z) \equiv \alpha_d(z) + \beta_d(z)t$ and $c_p(t, z) \equiv \alpha_p(z) + \beta_p(z)t$, where $\alpha_d(z)$, $\beta_d(z)$, $\alpha_p(z)$ and $\beta_p(z)$ are constants and t is the realization of T . Notice that the model imposes bounds for $\alpha_d(z)$ and $\beta_d(z)$. Indeed, since the support of Θ is $(\underline{\theta}, \bar{\theta}) \subseteq (0, 1)$, it must be the case that $\tilde{\theta}(t, z) \in (0, 1)$ for all t . From equation 6.9, I conclude that, for all $[\underline{t}, \bar{t}]$,

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

Also, as shown in Section 4, the function $\tilde{\theta}(\cdot, z)$ is strictly increasing in t . Therefore, the following inequality must hold:

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

where $\tilde{s}_t(\cdot, z)$ is the partial derivative of $\tilde{s}(\cdot, z)$ w.r.t. t . Some algebra shows that the following conditions are sufficient for the three inequalities above to hold: (i) $\tilde{s}(t, z)$ is convex in t , (ii) $\tilde{s}(t, z) < t$ for all $t \in [\underline{t}, \bar{t}]$, (iii) $\alpha_d(z) \in [\text{Max}\{0, \tilde{s}(\underline{t}, z) - \tilde{s}'(\underline{t}, z)\underline{t}\}, \tilde{s}(\underline{t}, z)]$ and (iv) $\beta_d(z) \in [0, \tilde{s}'(\underline{t}, z)]$. As argued in Section 4, condition (i) is true. Condition (ii) is strongly supported by the estimates of $\tilde{s}(\cdot, z)$ reported later in the paper. Conditions (iii) and (iv) are then enough to guarantee that the function $\tilde{\theta}(\cdot, z)$ behaves as predicted by the theory. These conditions will be useful for the estimation of the model.

Together, (6.9), (6.10), (6.10) and (6.11) identify $F(\cdot|Z = z)$ for the whole image of $\tilde{\theta}(\cdot, z)$, up to the constants $c_p(z)$, $c_d(z)$ and $\mu(z)$.⁴¹

The full density of potential trial sentences, $g(t|Z = z)$, can then be recovered, up to $c_p(z)$, $c_d(z)$ and $\mu(z)$, using the functions $\tilde{s}(\cdot, z)$ and $F(\cdot|Z = z)$ —as well as equations (6.1), (6.3) and (6.4). To complete the identification of the model's primitives, I must show how to recover the scalars $c_p(z)$, $c_d(z)$ and $\mu(z)$. But first I need to introduce a new parameter. Given z and a potential trial sentence t , the equilibrium probability that a defendant is convicted at trial is

$$\int_{\underline{\theta}}^{\tilde{\theta}(t,z)} x f(x|Z = z) dx.$$

Thus,

$$\begin{aligned} P[\Psi = 2|Z = z] &= \int_{[\underline{t}, \bar{t}]} \int_{\underline{\theta}}^{\tilde{\theta}(t,z)} x f(x|Z = z) g(t|Z = z) dx dt \\ &= \pi(z) + \int_{[\underline{t}, \bar{t}]} \int_{\tilde{\theta}(t,z)}^{\tilde{\theta}(t,z)} x f(x|Z = z) g(t|Z = z) dx dt. \end{aligned} \quad (6.12)$$

where $\pi(z) = \int_{\underline{\theta}}^{\tilde{\theta}(\underline{t}, z)} x f(x|Z = z) dx$.⁴² Like $\mu(z)$, the parameter $\pi(z)$ captures the behavior of the distribution $F(\cdot|Z = z)$ for θ lower than $\tilde{\theta}(\underline{t}, z)$.

The identification of $c_p(z)$, $c_d(z)$, $\mu(z)$ and $\pi(z)$ may proceed as follows. Using equations (6.3), (6.4) and (6.9) to (6.11), I write the density $b(\cdot|Z = z)$ in terms of $c_p(z)$, $c_d(z)$, $\mu(z)$ and the conditional density $b(\cdot|\Psi = 1, Z = z)$. Integrating both sides over $[\tilde{s}(\underline{t}), \tilde{s}(\bar{t})]$, I recover $\mu(z)$ as a function of $c_p(z)$ and $c_d(z)$. As a consequence, I can recover the densities $b(\cdot|Z = z)$ and $f(\cdot|Z = z)$, up to $c_p(z)$ and $c_d(z)$ only. Using equation (6.12), I am then able to recover $\pi(z)$, up to $c_p(z)$ and $c_d(z)$. Finally, using equations (6.4) and (6.7), I have that

$$\frac{b(\tilde{s}(t, z)|\Psi = 1, Z = z)}{g(t|\Psi = 2, Z = z)} = \left(\frac{1}{\frac{d\tilde{s}(t, z)}{dt}} \right) \frac{P[\Psi = 1|S = \tilde{s}(t, z), Z = z]}{P[\Psi = 1|Z = z]} \frac{P[\Psi = 2|Z = z]}{P[\Psi = 2|T = t, Z = z]} \quad (6.13)$$

for all $t \in [\underline{t}, \bar{t}]$. Notice that the left-hand side of the equation above depends only on $c_d(z)$ and $c_p(z)$, since $\mu(z)$ and $\pi(z)$ are known, up to the trial costs. The right-hand side is observable. Since the equation holds for all t in the support, it implies a system of infinitely many equations. In such a system, the variables $c_d(z)$ and $c_p(z)$ are independent from each other, since only the former affects $\tilde{\theta}(\cdot, z)$. Intuitively, $c_d(z)$ and $c_p(z)$ jointly determine the distribution of the defendant's type—and, thus, the probability of bargaining failure. Given this probability, $c_d(z)$ is separately identified by the probability of a conviction at trial. Hence, evaluating equation (6.13) at two different values of t (say \underline{t} and \bar{t}), I can solve for $c_d(z)$ and $c_p(z)$.⁴³ \square

This result concludes the identification of the model.

⁴¹The term $\mu(z)$ depends on the behavior of $F(\cdot|Z = z)$ for values of θ lower than $\tilde{\theta}(\underline{t}, z)$. It can be rewritten as $\mu(z) = 1 - F(\tilde{\theta}(\underline{t}, z)|Z = z)$.

⁴²Notice that $\pi(z)$ must belong to the interval $[0, (1 - \mu(z))\tilde{\theta}(\underline{t}, z)]$.

⁴³As mentioned above, I estimate an extended version of the model, in which I allow the trial costs to vary linearly with the trial sentence. Recovering the trial costs in that case requires the identification of two intercepts and two slope coefficients. The identification of these four parameters follows the argument above—except that equation (6.13) must be evaluated at four different values of t , in order to form a system of four equations and four unknowns.

6.3. Estimation. I propose an estimation procedure that closely follows my identification strategy. It consists of two parts. First I present an estimator of $\tilde{s}(\cdot, z)$. Then, I use the estimated function $\tilde{s}(\cdot, z)$ in order to obtain estimators for the model’s primitives.

6.3.1. *Estimation of $\tilde{s}(\cdot, z)$.* Let Δ contain case-specific variables that affect $F(\cdot)$, $c_d(\cdot)$ and $c_p(\cdot)$, as well as judge-specific variables that affect sentencing. I choose a subset $\Delta_\rho \subseteq \Delta$ in which the case-specific variables are constant (so that condition *A* holds), but where the judge-specific characteristics vary.

For every $z \in \Delta_\rho$, I obtain the estimates

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

The estimates for the distribution of Ψ are trivially computed. The estimation of the densities $g(t|\Psi = 2, Z = z)$ and $b(s|\Psi = 1, Z = z)$ can be conducted by a kernel density estimator (KDE). There are two challenges to the estimation of these densities. First, I must deal with the boundedness of $[t, \bar{t}]$. When the standard KDE is applied to distributions with bounded support, it tends to underestimate the density near the boundaries. The resulting estimates are inconsistent. To overcome this problem, I employ a boundary-correction method recently developed by Karunamuni and Zhang (2008).⁴⁴ Broadly speaking, the method consists of generating 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.⁴⁵ The rate of convergence is the same as the one obtained away from the boundaries using the standard KDE. The second challenge arises because the dimensionality of Z may be large. The curse of dimensionality can thus prevent the estimation of $g(t|\Psi = 2, Z = z)$ and $b(s|\Psi = 1, Z = z)$ using conventional KDE methods, even for large sample sizes. I deal with this problem by, first, assuming that all variables in Δ are discrete and, then, employing smoothing techniques proposed by Li and Racine (2007) for the estimation of conditional densities with mixtures of discrete and continuous variables. The idea of this procedure is to divide the space of case-level characteristics into a finite number of subsets and estimate the densities for each one of these subsets. As opposed to a naive implementation of this idea, which would employ only observation satisfying $Z = z$ for the estimation of $g(t|\Psi = 2, Z = z)$ and $b(s|\Psi = 1, Z = z)$, Li and Racine’s conditional density estimator uses all the available data, assigning to each observation i a weight that depends on the realization of Z_i . See the appendix for a more detailed discussion about this procedure and how it is implemented in the present study.

⁴⁴The Karunamuni and Zhang correction is an improvement on the correction proposed by Zhang, Karunamuni and Jones (1999). It is applied to the non-parametric estimation of first-price auction models by Hickman and Hubbard (2012).

⁴⁵The 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 kernel function.

Equation (6.2) implies that

$$\int_{\underline{t}}^{\bar{t}} \left\{ \left\{ \frac{b[\tilde{s}(t, z)|\Psi = 1, Z = z']}{b[\tilde{s}(t, z)|\Psi = 1, Z = z'']} \frac{P[\Psi = 1|Z = z']}{P[\Psi = 1|Z = z'']} \right. \right. \\ \left. \left. - \frac{g(t|\Psi = 2, Z = z')}{g(t|\Psi = 2, Z = z'')} \frac{P[\Psi = 2|Z = z']}{P[\Psi = 2|Z = z'']} \right\}^2 \right\} dH(t) = 0$$

for any measure H over $[\underline{t}, \bar{t}]$. Denote the lowest and highest observed trial sentences for $z \in \Delta_\rho$ by \hat{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 lowest and highest observed plea-bargained sentences for $z \in \Delta_\rho$, and notice that these are consistent estimates for $\tilde{s}(\underline{t}, z)$ and $\tilde{s}(\bar{t}, z)$. Remember that $\tilde{s}(\cdot, z)$ is increasing and convex, and let Υ be the space of increasing and convex functions $s(\cdot)$ over $[\underline{t}, \bar{t}]$ such that $s(\hat{t}) = s(\hat{\underline{s}})$ and $s(\hat{\bar{t}}) = s(\hat{\bar{s}})$. Also, let \mathcal{C} be the set of all pairwise combinations of points in Δ_ρ . The estimator $\hat{s}(\cdot, z)$ for the function $\tilde{s}(\cdot, z)$ can be found by solving

$$\min_{s \in \Upsilon} \sum_{(z', z'') \in \mathcal{C}} \int_{\underline{t}}^{\bar{t}} \left\{ \frac{\hat{b}[s(t, z')|\Psi = 1, Z = z']}{\hat{b}[s(t, z'')|\Psi = 1, Z = z'']} \frac{\hat{P}[\Psi = 1|Z = z']}{\hat{P}[\Psi = 1|Z = z'']} \right. \\ \left. - \frac{\hat{g}(t|\Psi = 2, Z = z')}{\hat{g}(t|\Psi = 2, Z = z'')} \frac{\hat{P}[\Psi = 2|Z = z']}{\hat{P}[\Psi = 2|Z = z'']} \right\}^2 dH(t), \quad (6.14)$$

where H is an arbitrary measure over $[\underline{t}, \bar{t}]$.⁴⁶ 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.⁴⁷ Splines can be conveniently represented as the linear combination of a finite set of basis functions, which allow me to treat problem (6.14) 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 of the basis functions, monotonicity and convexity can be easily imposed to the approximate solution of (6.14), by restricting the regression coefficients to be non-negative.

To state the result showing the consistency of the estimator described above, define the functions

$$\tilde{b}_r(s, z', z'') \equiv \frac{b[s|\Psi = 1, Z = z']}{b[s|\Psi = 1, Z = z'']} \frac{P[\Psi = 1|Z = z']}{P[\Psi = 1|Z = z'']} \\ \text{and} \quad \tilde{g}_r(t, z', z'') \equiv \frac{g(t|\Psi = 2, Z = z')}{g(t|\Psi = 2, Z = z'')} \frac{P[\Psi = 2|Z = z']}{P[\Psi = 2|Z = z'']}$$

for $(z', z'') \in \mathcal{C}$. I can now state the following proposition:

PROPOSITION 3. *Assume that, for every pair $(z', z'') \in \mathcal{C}$, the following conditions hold: (i) the function $\tilde{b}_r(\cdot, z', z'')$ is differentiable and the absolute value of its derivative is bounded by $\hat{b}_r(z', z'')$; and (ii) the functions $\tilde{b}_r(\cdot, z', z'')$ and $\tilde{g}_r(\cdot, z', z'')$ are positive and bounded by $\bar{B}_r(z', z'')$ and $\bar{G}_r(z', z'')$,*

⁴⁶In the application described below, I set such a measure to be $\hat{g}(t|\Psi = 2, Z = z'')$.

⁴⁷See Eubank (1999) for an excellent introduction to spline approximation.

respectively. Then, the estimator of the prosecutor's settlement offer function $\tilde{s}(\cdot, z)$ for $z \in \Delta_\rho$ is uniformly consistent.

See the appendix for the proof. The estimation of $\tilde{s}(\cdot, z)$ has two steps. First, I obtain $\hat{P}[\Psi = \psi|Z = z]$, $\hat{g}(t|\Psi = 2, Z = z)$ and $\hat{b}(s|\Psi = 1, Z = z)$, and then these estimates are used in the construction of the objective function shown in (6.14). The proof of Proposition 3 is facilitated by noticing that, given the first-step estimates, the objective function is non-random. The restrictions on the density ratios $\bar{b}_r(\cdot, z', z'')$ and $\bar{g}_r(\cdot, z', z'')$ ensure the continuity of the objective function. Such continuity allows me to apply standard techniques for achieving the consistency of two-step estimators.

6.3.2. Estimation of the model's primitives. Once the estimate $\hat{s}(\cdot, z)$ is obtained, I am able to recover the model's primitives. I extend the basic model by allowing the trial costs to vary linearly with the potential trial sentence T . Formally, I assume that the trial costs for the defendant and the prosecutor are given, respectively, by $c_d(t, z) \equiv \alpha_d(z) + \beta_d(z)t$ and $c_p(t, z) \equiv \alpha_p(z) + \beta_p(z)t$. The primitives to be estimated are then the cost parameters $\alpha_d(z)$, $\beta_d(z)$, $\alpha_p(z)$ and $\beta_p(z)$, the distribution of defendant's types $F(\cdot|Z = z)$ and the distribution of trial sentences, characterized by $\nu(z)$ and $g(\cdot|Z = z)$. Notice that, for all $z \in \Delta_\rho$, 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 Subsection 6.2.3, which show that $F(\cdot|Z = z)$ and $g(\cdot|Z = z)$ can be written in terms of the cost parameters $\alpha_d(z)$, $\beta_d(z)$, $\alpha_p(z)$ and $\beta_p(z)$, as well as the auxiliary parameters $\mu(z)$ and $\pi(z)$. Since equation (6.13) holds for all $t \in [\underline{t}, \bar{t}]$, the model is overidentified. In order to recover the primitives, therefore, I estimate $\alpha_d(z)$, $\beta_d(z)$, $\alpha_p(z)$, $\beta_p(z)$ and $\mu(z)$ by maximum likelihood.⁴⁸

More precisely, consider guesses $\ddot{\alpha}_p$, $\ddot{\beta}_p$, $\ddot{\alpha}_d$, $\ddot{\beta}_d$ and $\ddot{\mu}$ for $\alpha_p(z)$, $\beta_p(z)$, $\alpha_d(z)$, $\beta_d(z)$ and $\mu(z)$, respectively. From $\hat{s}(\cdot, z)$ and equation (6.9), I numerically obtain $\ddot{\theta}(\cdot, z, \ddot{\alpha}_d, \ddot{\beta}_d)$, the function $\ddot{\theta}(\cdot, z)$ consistent with the guesses $\ddot{\alpha}_d$ and $\ddot{\beta}_d$. Using equations (6.10) and (6.11), I then obtain $\ddot{f}(\cdot|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu})$, the density function $\ddot{f}(\cdot|Z = z)$ consistent with $\ddot{\alpha}_d$, $\ddot{\beta}_d$, $\ddot{\alpha}_p$, $\ddot{\beta}_p$ and $\ddot{\mu}$. Employing $\hat{s}(\cdot, z)$, the estimated conditional density $\hat{b}(\cdot|\Psi = 1, Z = z)$, and equations (6.3), (6.4) and (6.9) to (6.11), I can numerically compute $\ddot{g}(\cdot|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu})$, the density $\ddot{g}(\cdot|Z = z)$ consistent with $\ddot{\alpha}_d$, $\ddot{\beta}_d$, $\ddot{\alpha}_p$, $\ddot{\beta}_p$ and $\ddot{\mu}$. From equation (6.12), I calculate $\ddot{\pi}(z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_d, \ddot{\mu})$, the value of $\pi(z)$ consistent with $\ddot{\alpha}_d$, $\ddot{\beta}_d$, $\ddot{\alpha}_p$, $\ddot{\beta}_p$ and $\ddot{\mu}$. Using $\ddot{g}(\cdot|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu})$, $\ddot{f}(\cdot|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu})$ and $\ddot{\pi}(z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu})$, I can numerically compute the likelihood that $\Psi = 3$, given Z , and consistently with $\ddot{\alpha}_d$, $\ddot{\beta}_d$, $\ddot{\alpha}_p$, $\ddot{\beta}_p$ and $\ddot{\mu}$. Such likelihood is

$$\begin{aligned} & \hat{P}[\Psi = 3|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}] \\ &= \int_{[\underline{t}, \bar{t}]} \int_{\underline{\theta}}^{\ddot{\theta}(t, z, \ddot{\alpha}_d, \ddot{\beta}_d)} (1-x) \ddot{f}(x|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}) \ddot{g}(t|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}) dx dt. \end{aligned}$$

⁴⁸In principle, I could also allow $\pi(z)$ to be a free parameter. But doing so, in practice, would have a negligible effect on the fit of the model if the bounds on $\pi(z)$ discussed in footnote 42 were to be respected.

From equation (6.3), I can numerically compute the likelihood that $\Psi = 1$, given Z , and consistently with $\ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p$ and $\ddot{\mu}$. This likelihood is given by

$$\ddot{P}[\Psi = 1|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}] = \int_{[\hat{t}, \hat{t}]} 1 - \ddot{F} \left[\ddot{\theta}(t, z, \ddot{\alpha}_d, \ddot{\beta}_d) | Z = z \right] \ddot{g}(t|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}) dt,$$

where $\ddot{F}[\cdot|Z = z]$ is a CDF obtained from $\ddot{f}(\cdot|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu})$. From equations (6.6) and (6.7), the likelihood that $T = t$ and $\Psi = 2$, given Z , and consistently with $\ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p$ and $\ddot{\mu}$, is

$$\begin{aligned} \ddot{P}[\Psi = 2|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}] \ddot{g}(t|\Psi = 2, Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}) \\ = \int_0^{\ddot{\theta}(t, z, \ddot{\alpha}_d, \ddot{\beta}_d)} x \ddot{f}(x|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}) dx \ddot{g}(\cdot|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}). \end{aligned}$$

I am now ready to define the likelihood contribution of an observation i . I consider the likelihood conditional on $\Psi \neq 0$.⁴⁹ Let W_i be the data corresponding to observation i .⁵⁰ Given $\ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p$ and $\ddot{\mu}$, the likelihood contribution of an observation i , conditional on $\Psi_i \neq 0$, is

$$\begin{aligned} l(\ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}, W_i) &= \ddot{P}[\Psi = 1|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}]^{\mathbb{1}\{\Psi_i=1\}} \\ &\times \left\{ \ddot{P}[\Psi = 2|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}] \ddot{g}(t_i|\Psi = 2, Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}) \right\}^{\mathbb{1}\{\Psi_i=2\}} \\ &\times \ddot{P}[\Psi = 3|Z = z, \ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}]^{\mathbb{1}\{\Psi_i=3\}}. \end{aligned}$$

Estimates $\hat{\alpha}_d(z), \hat{\beta}_d(z), \hat{\alpha}_p(z), \hat{\beta}_p(z)$ and $\hat{\mu}(z)$ for $\alpha_d(z), \beta_d(z), \alpha_p(z), \beta_p(z)$ and $\mu(z)$ can then be obtained by performing a numerical search to find the parameters that maximize the sum of the logarithms of $l(\ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p, \ddot{\beta}_p, \ddot{\mu}, W_i)$ over all observations for which $\psi \neq 0$. I constrain $\ddot{\alpha}_d, \ddot{\beta}_d, \ddot{\alpha}_p$ and $\ddot{\beta}_p$ to be positive, and $\ddot{\mu}$ to belong to the unit interval.⁵¹ Finally, estimates for $g(\cdot|Z = z)$ and $f(\cdot|Z = z)$ are defined by $\ddot{g}(\cdot|Z = z, \hat{\alpha}_d(z), \hat{\beta}_d(z), \hat{\alpha}_p(z), \hat{\beta}_p(z), \hat{\mu}(z))$ and $\ddot{f}(\cdot|Z = z, \hat{\alpha}_d(z), \hat{\beta}_d(z), \hat{\alpha}_p(z), \hat{\beta}_p(z), \hat{\mu}(z))$, respectively.

7. EMPIRICAL RESULTS

To implement the estimator described above, I must focus my analysis on cases that are relatively homogeneous, so that condition A is plausible. To achieve this homogeneity, I restrict the analysis to cases in which the main charged offense is a non-homicide violent crime. This category of offenses includes robberies, assaults and sexual assaults. Also, to control for observed heterogeneity across cases in my estimation, 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 attorney employed in the case and the gender, race and previous criminal record of

⁴⁹I do so because the empirical probability that $\Psi = 0$ is useful only for identifying $\nu(z)$.

⁵⁰That 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 consists of 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{c}_p, \ddot{c}_d or $\ddot{\mu}$.

⁵¹I further constrain $\ddot{\alpha}_d$ and $\ddot{\beta}_d$ to satisfy conditions (iii) and (iv) described in footnote 40, where $\tilde{s}(\cdot, z)$ is replaced by the estimate $\hat{s}(\cdot, z)$. Restricting the trial cost coefficients to be positive is consistent with the notion that it is more costly to bring relatively serious cases to trial. Notice that assuming $\alpha_d(z) > 0$ and $\alpha_p(z) > 0$ also ensures that the hazard rate of the distribution of defendant's type is strictly increasing, as can be shown by applying the implicit function theorem on equation (4.2).

the defendant. I then estimate the primitives of the model separately across these groups. The variation in the distribution of trial sentences within such groups, which is necessary for condition B to hold, is due to differences in the sentencing patterns across judges.

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.⁵² Second, the judge rotation mandated by the state constitution takes place at the division level. Thus, over the period of 14 years covered by my data set, it is reasonable 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. The distinction between these categories in my analysis is motivated by the empirical literature linking the type of defense counsel and the outcome of criminal cases.⁵³ Lastly, I divide cases by defendant's gender, race (African-Americans vs. others) and previous 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).⁵⁴ 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 previous criminal records results in 192 covariate bins. Some of these bins contain 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 estimation procedure. Therefore, I restrict the implementation of my estimator to 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 focus the analysis on cases prosecuted at the 5th Superior Court division, which comprises counties located in the northwestern part of the state and is the division with the largest sample size in my data set.⁵⁵ These restrictions leave me with a total of 12 covariate groups, which are summarized in table 6. The table also shows how many cases are observed in the data for each one of the groups.⁵⁶ Notice that, even though I

⁵²North Carolina is divided into prosecutorial districts that roughly correspond to the Superior Court districts described in Section 3. Sample size limitations prevent me from undertaking the analysis at the Superior Court district level. Dividing the cases in the data by the Superior Court division is a feasible alternative.

⁵³See, for example, Iyengar (2007) and Anderson and Heaton (2012). The former paper examines Federal District Court cases, while the latter analyzes homicide cases in Philadelphia. Both studies find evidence that cases assigned to public defenders tend to result in more lenient sentences than those assigned to court-appointed attorneys.

⁵⁴To put this classification into perspective, notice that four points are equivalent to having a previous conviction for an offense like 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).

⁵⁵The 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.

⁵⁶Previous 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 6, 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 and gender of the defendant, defense counsel type and the Superior Court division of prosecution.

TABLE 6. Covariate groups

Group	Covariates					Observations
	Defendant's race	Defendant's gender	Defense counsel	Defendant's record	Sup. Court division	
1	other	Male	assigned	short	5	2141
2	Afr. American	Male	assigned	short	5	2727
3	other	Male	assigned	long	5	1989
4	Afr. American	Male	assigned	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

do not recover the model primitives for female defendants and divisions other than the 5th, I still employ observations corresponding to these groups for the estimation of conditional densities of trial sentences and settlement offers. See Section 6 and the Appendix for details.

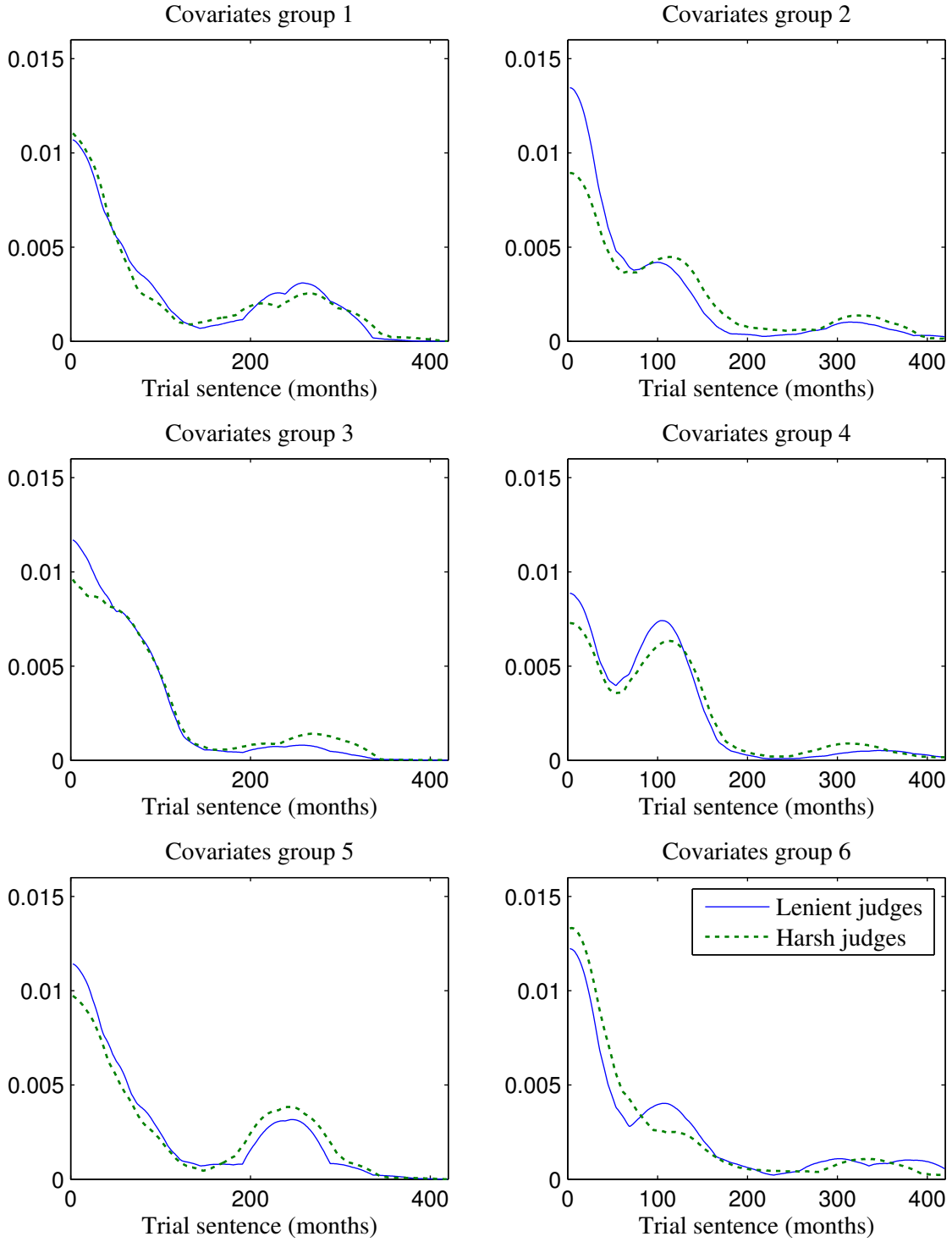
Within each covariate group, I explore variation in the sentencing patterns across judges to obtain condition B . Specifically, I compare sentencing decisions taken by lenient and harsh judges, as defined on Section 5. Figures 2 and 3 depict estimates of the trial sentence densities, conditional on a conviction at trial, for these two types of judge in each one the 12 covariate groups of my analysis.⁵⁷ All of the distributions show signs of multi-modality—with most of their mass concentrated at sentences shorter than 100 months. Importantly, notice that, for each covariate group, the densities vary as expected across lenient and harsh judges. Indeed, the densities associated with lenient judges have more mass at short sentences, relative to those associated with harsh judges. Although a formal test of condition B is beyond the scope of the present study, the figures suggest that there is enough variation in the sentencing behavior across judges to implement the estimator described in Section 6.

In order to estimate the prosecutor's settlement offer function, I implement the spline regression procedure described previously. I do so independently for each covariate group. An important step in any spline regression is the selection of knots—the points in the domain of the function to be estimated where the polynomial pieces that constitute the spline connect. I set such knots 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).

Figures 4 to 6 present the estimated settlement offer functions for each covariate group. The variation of the estimates across groups is substantial. Specifically, the estimated offer functions

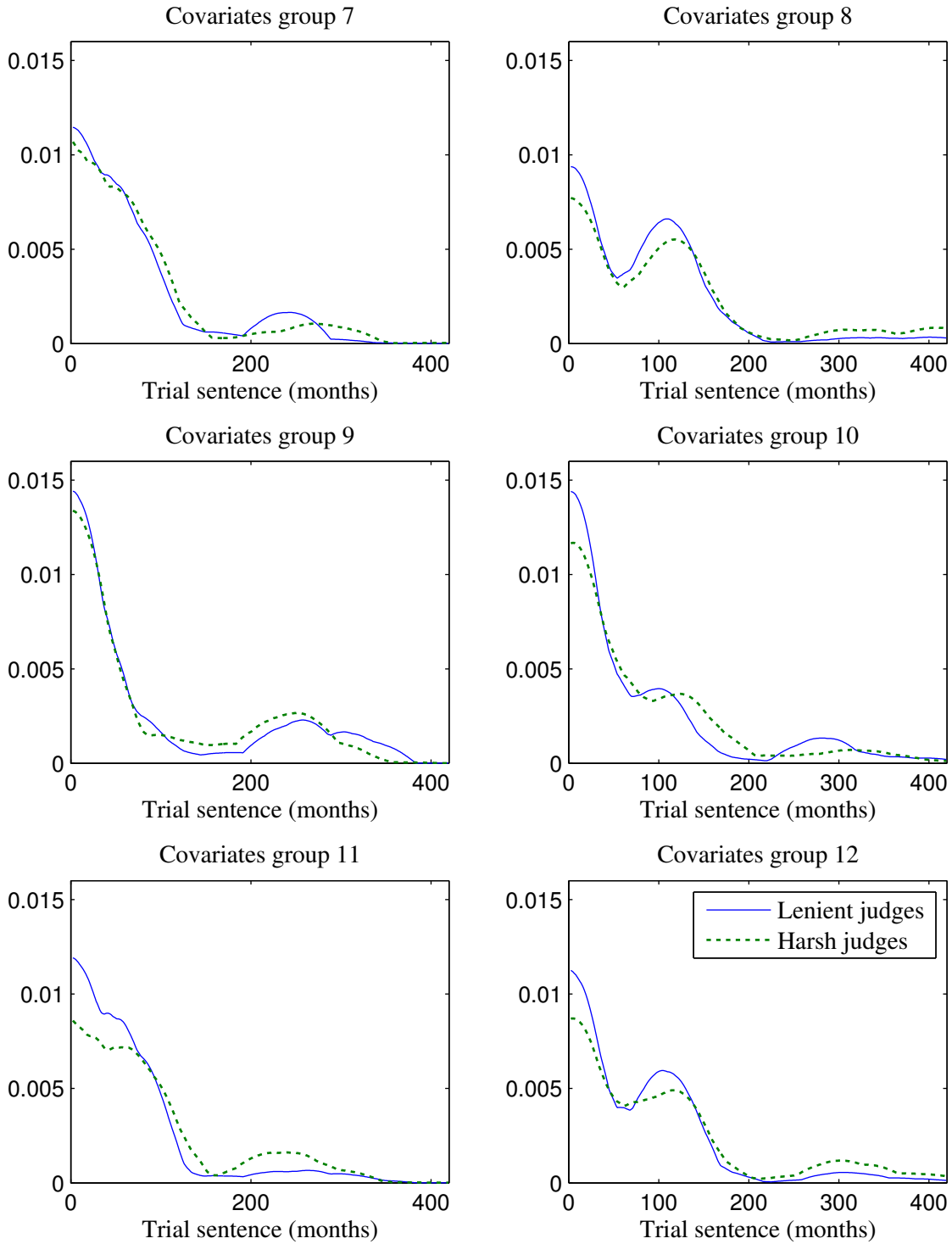
⁵⁷Bandwidths are reported in the Appendix, together with other estimation details.

FIGURE 2. Conditional trial sentence densities by covariate group – lenient and harsh judges



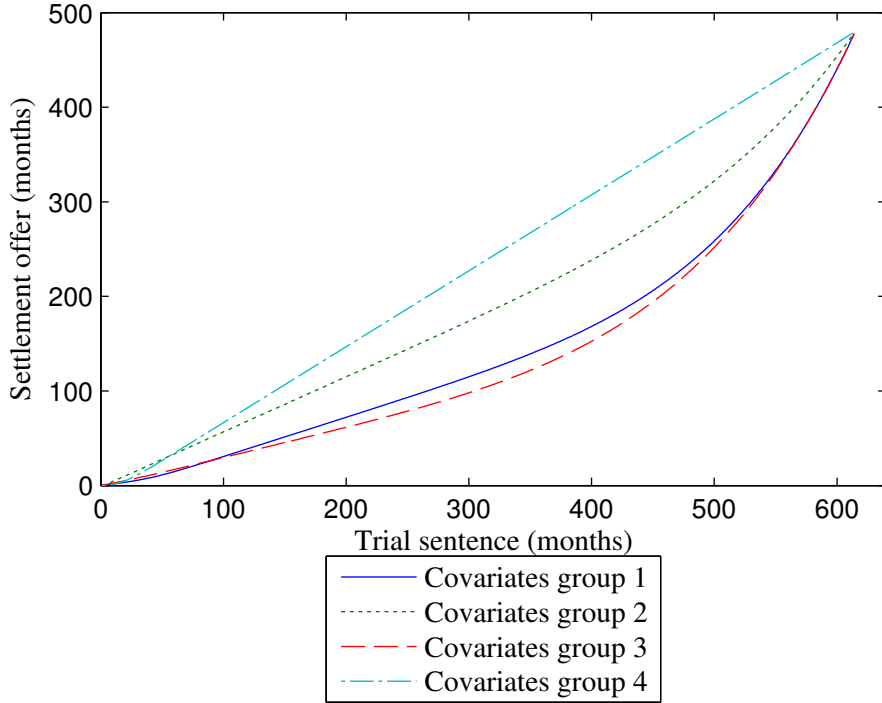
See table 6 for a description of the covariate groups

FIGURE 3. Conditional trial sentence densities by covariate group – lenient and harsh judges (cont.)



See table 6 for a description of the covariate groups

FIGURE 4. Settlement offer – covariate groups 1 to 4



for groups two, four, ten and 12 are almost linear, while the functions for other groups are much more convex. Notice that the covariate group classification described in table 6 attributes even numbers to groups in which the defendant is African-American and odd numbers to groups in which the defendant is not. In figures 4 to 6, a comparison of the estimates for subsequent odd and even-numbered groups shows that, in most cases, the settlement offer function associated with African-American defendants is less convex than the one associated with other defendants, controlling for all other covariates in the analysis. The only exception occurs in the comparison of the estimates for groups seven and eight. That the settlement offer functions associated with African-Americans are relatively linear suggests that prosecutors tend to offer longer sentences to these 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 the offer functions of even-numbered groups are more convex than those of odd-numbered ones.

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

Table 7 reports the estimation results, together with bootstrap standard errors.⁵⁸ Notice that the trial costs are measured in terms of months. Accordingly, the trial cost intercepts $\hat{\alpha}_d$ and $\hat{\alpha}_p$ are

⁵⁸See the Appendix for details on the computation of the standard errors.

FIGURE 5. Settlement offer – covariate groups 5 to 8

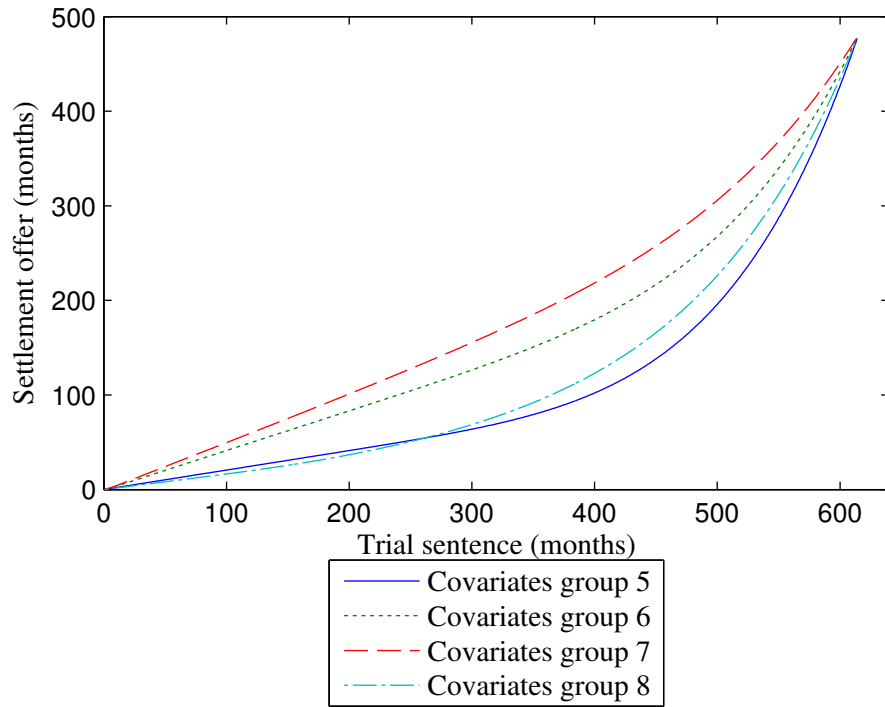
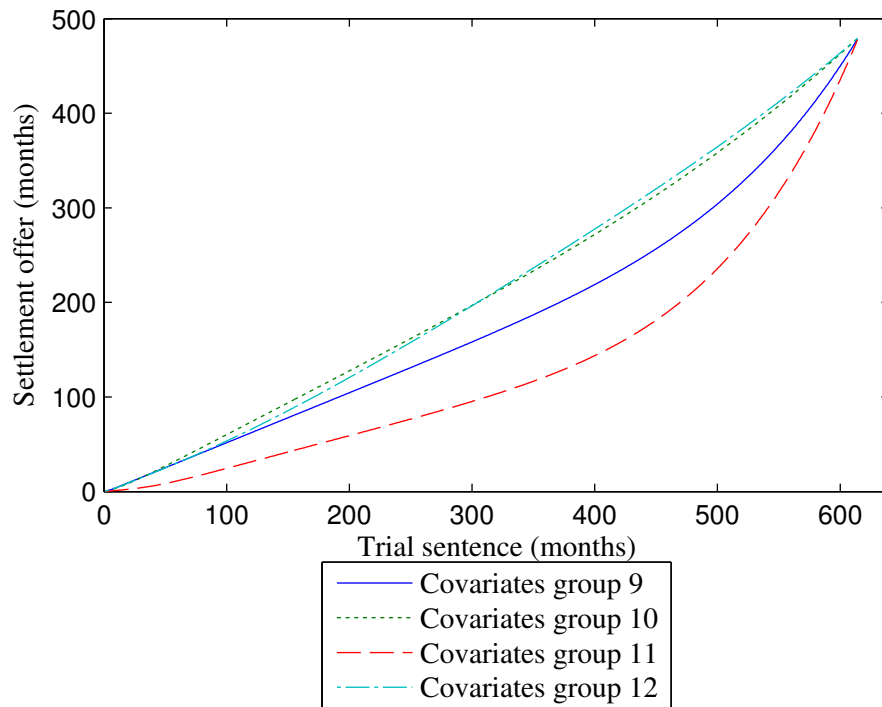


FIGURE 6. Settlement offer – covariate groups 9 to 12



See table 6 for a description of the covariate groups

TABLE 7. 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)
3	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.64 (0.22)	1.00 (0.00)
4	0.03 (0.00)	0.00 (0.00)	0.00 (8.77)	1.43 (0.73)	1.00 (0.00)
5	0.02 (0.00)	0.00 (0.00)	0.00 (0.00)	0.67 (0.11)	1.00 (0.00)
6	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	1.06 (0.31)	1.00 (0.00)
7	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	2.43 (0.83)	1.00 (0.00)
8	0.02 (0.01)	0.00 (0.00)	0.00 (0.00)	0.71 (0.25)	1.00 (0.00)
9	0.03 (0.01)	0.00 (0.00)	0.00 (0.00)	1.25 (0.28)	0.92 (0.04)
10	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	1.16 (0.38)	1.00 (0.00)
11	0.03 (0.00)	0.00 (0.00)	0.00 (0.00)	0.48 (0.08)	1.00 (0.00)
12	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	0.91 (0.39)	1.00 (0.00)

Bootstrap standard errors in parenthesis
 See table 6 for a description of the covariate groups

expressed in months, while $\hat{\beta}_d$ and $\hat{\beta}_p$ are coefficients. For all covariate groups, both $\hat{\alpha}_d$ and $\hat{\beta}_d$ are very small. The former parameter varies from 0.00 to 0.03, depending on the group, and the latter is always 0.00. These numbers suggest that defendants, even those with privately-retained counsel, barely take the costs of going to trial into consideration when deciding whether or not to accept the prosecutor's offer. One possible interpretation of this result is that, for the defendant, the costs of serving time in prison are so high that every expenditure associated with the trial procedure is of second order.

For every group, $\hat{\alpha}_p$ is equal to zero and $\hat{\alpha}_d$ is large, ranging from 0.48 to 2.43. These estimates suggest that trial costs increase rapidly as the trial sentence gets longer. In particular, the prosecutor has no costs of bringing to trial a case in which the trial sentence is zero. Given a trial sentence of 100.73 months, the average observed in the data, the prosecutor’s trial costs vary from 48.35 to 244.77 months. The differences between the estimated trial costs for the defendant and the prosecutor may, at first glance, seem striking. Notice, however, that the model does not allow for the comparison of cardinal utility between the prosecutor and the defendant—which implies that the trial cost estimates for the two parties are not directly comparable either.⁵⁹ Estimates of μ are zero for all covariate groups other than group nine. For the latter group I find $\hat{\mu}$ equal to 0.91. Remember that $F(\tilde{\theta}(\underline{t})) = 1 - \mu$. My results thus indicate that, for almost all groups, defendants whose type belongs to the interval $[0, \tilde{\theta}(\underline{t})]$ are very rare.

After the estimation of the remaining parameters, I am able to compute estimates for the distribution of defendants’ types. Figures 7 and 8 depict these distributions. For each covariate group, the figures show both the estimated density function and the estimated cumulative distribution function. All distributions have a mode at a type lower than 0.3. In fact, for about half of the covariate groups, this mode is located at zero. The densities then sharply decrease, giving the impression that the distributions of defendants’ types are unimodal. However, from Section 6, we know that these distributions are only identified over part of their support—specifically, the interval $[\tilde{\theta}(\underline{t}), \tilde{\theta}(\bar{t})]$. For all covariate groups, this interval comprises a large portion of the unit line. The lower bound of the identified range of the support varies from zero to just above 0.2, depending on the covariate group, and the upper bound is roughly 0.8 for all groups. But all distributions have substantial mass outside of this range. The most extreme case is that of covariate group seven, for which the estimate of $F(\tilde{\theta}(\bar{t}))$, the cumulative distribution function evaluated at the upper bound of the identified range of the support, is approximately 0.2. For the other groups, such a value varies from about 0.4 to 0.6. Therefore, these numbers suggest that the distributions of defendants’ types have at least one more mode, located at a relatively high type. In particular, my results are 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 figures 7 and 8 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 figures 4 to 6. In most cases, the estimate of $F(\tilde{\theta}(\bar{t}))$ for an odd-numbered group is greater than the estimate of the same variable for the subsequent even-numbered group, indicating that the distributions of types for African-Americans concentrate more mass at high types than the distributions 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. According to the model, these differences in the distributions of types are considered by the prosecutors in the process of making settlement

⁵⁹Indeed, the equilibrium outcome of the bargaining game is invariant to any monotone linear transformation affecting both the parameters α_p and β_p and the utility derived by the prosecutor from the assignment of a given incarceration sentence.

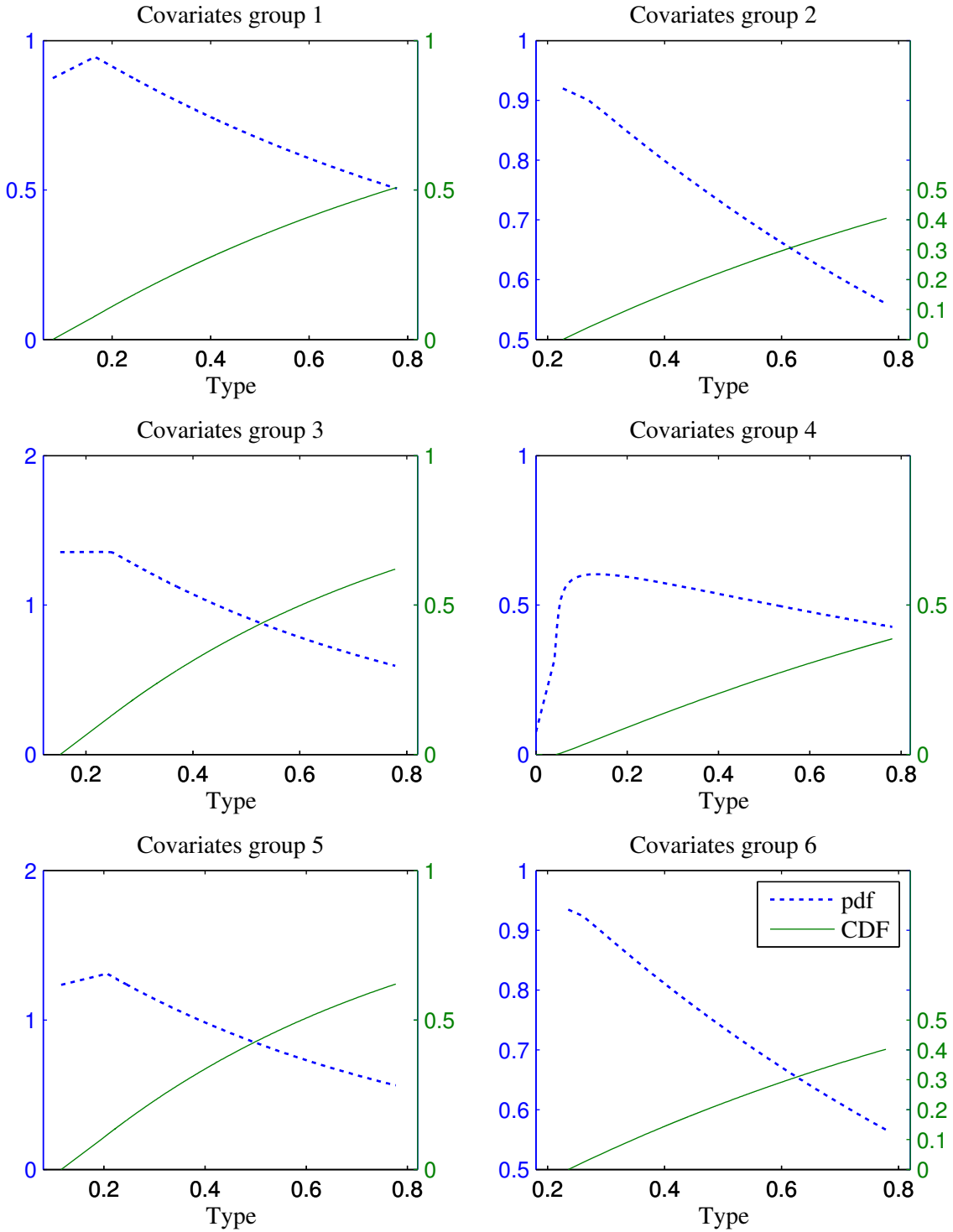
offers. In particular, given an African-American and a non-African-American defendant who face the same trial sentence, a prosecutor offers to the former a longer sentence. Relatively generous offers made to non-African-American defendants result in more convex settlement offer functions for these groups, which is precisely the finding in figures 4 to 6.⁶⁰ Two comments are in order here: First, as in the discussion of the offer function estimates, a rigorous comparison between the distribution of types of African-American and non-African-American defendants would require formal tests that I do not carry out in this paper. Second, even assuming that the likelihood of trial convictions is indeed higher for African-Americans than for other defendants, it is not clear whether this difference is due to the discretion of the jury or to variation in other case characteristics across groups of defendants. These are interesting topics for further research.

As a final step in the estimation of the model primitives, I obtain the full distribution of potential trial sentences for each covariate group—i.e., the distribution without conditioning on a trial conviction. In the interest of space, such distributions are not reported in the paper. They are available from the author upon request.

Tables 8 and 9 present information on the fit of the model. Table 8 separately shows the probabilities of conviction to incarceration by plea bargain and at trial for each covariate group. The model offers a very good fit to the probability of a plea bargain, while it slightly under-estimates the probability of a conviction at trial (by approximately 0.5 to 3.0 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. Table 9 contains the average assigned sentences for each covariate group. It reports three different versions of the average sentence: (i) The overall average, i.e., the average sentence, conditional on either a conviction at trial or on a plea bargain; (ii) the average sentence, conditional on a plea bargain; and (iii) the average sentence, conditional on a trial conviction. The model fits the the overall average sentence well. However, a further examination of the fitted average sentences reveals that, for six out of the 12 covariate groups, the model substantially over-estimates the average trial sentence. 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. For six groups, the fitted model predicts the average trial sentence to be from 0.38 (group 6) to 1.46 (group 5) standard deviations longer than observed. A possible explanation for that is the small number of trial conviction observations in my sample, which leads the maximum likelihood procedure used for completing the estimation of the model 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 9. As long as one is not particularly interested in the distribution of outcomes for the small minority of cases that reach the trial stage, the model is able to replicate the key aspects of the data very well for all covariate groups.

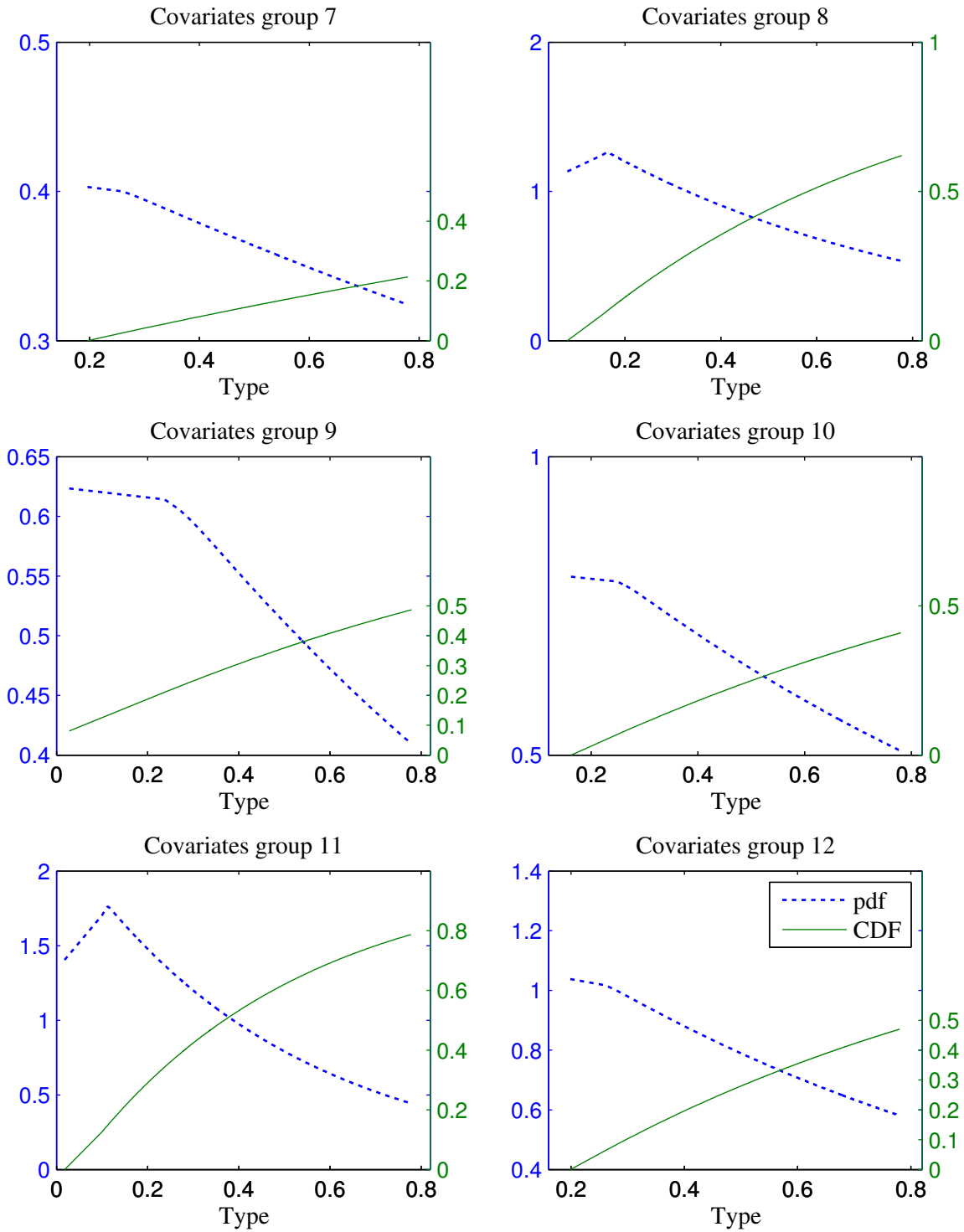
⁶⁰The only case in which the settlement offer function associated with an odd-numbered group is less convex than the function associated with the subsequent group is that of groups seven and eight. Interestingly, the estimated $F(\hat{\theta}(\bar{t}))$ is smaller for group seven than it is for group eight.

FIGURE 7. Distribution of defendants' types – density and CDF



See table 6 for a description of the covariate groups

FIGURE 8. Distribution of defendants' types – density and CDF (cont.)



See table 6 for a description of the covariate groups

TABLE 8. Fitted values versus data – Distribution of outcomes

Group		Outcome		
		Any conviction ($\Psi \in \{1, 2\}$)	Settlement ($\Psi = 1$)	Trial conviction ($\Psi = 2$)
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%
3	Data	38.35%	34.60%	3.75%
	Model	36.66%	34.59%	2.08%
4	Data	44.11%	36.66%	7.45%
	Model	40.68%	36.64%	4.04%
5	Data	38.16%	35.11%	3.05%
	Model	35.55%	34.44%	1.10%
6	Data	40.32%	37.53%	2.79%
	Model	39.68%	37.64%	2.04%
7	Data	38.16%	35.11%	3.05%
	Model	37.53%	36.10%	1.43%
8	Data	40.32%	37.53%	2.79%
	Model	38.62%	37.60%	1.01%
9	Data	21.65%	19.37%	2.28%
	Model	21.59%	19.51%	2.08%
10	Data	29.43%	25.51%	3.92%
	Model	29.09%	25.83%	3.26%
11	Data	21.65%	19.37%	2.28%
	Model	20.91%	19.33%	1.58%
12	Data	29.43%	25.51%	3.92%
	Model	28.61%	25.02%	3.59%

See table 6 for a description of the covariate groups

8. POLICY EXPERIMENTS

Sentencing reform has been at the center of the recent 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)—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, combined with the fiscal crisis in the past years, 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” sentencing reforms of the 1980s and 1990s.

TABLE 9. Fitted values versus data – Sentences

Group		Average sentence, conditional on method of resolution [†]		
		All ($\Psi \in \{1, 2\}$)	Settlement ($\Psi = 1$)	Trial conviction ($\Psi = 2$)
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
3	Data	52.43	46.79	104.45
	Model	54.29	45.64	198.33
4	Data	60.48	50.51	109.54
	Model	51.98	45.96	106.46
5	Data	53.53	47.65	121.25
	Model	52.76	45.55	277.60
6	Data	41.27	38.92	72.95
	Model	42.93	39.13	113.04
7	Data	47.35	42.96	98.01
	Model	43.43	41.21	99.55
8	Data	44.81	38.49	129.96
	Model	47.07	40.66	284.73
9	Data	47.76	41.13	104.00
	Model	46.15	42.25	82.86
10	Data	42.67	37.19	78.36
	Model	43.55	38.83	80.99
11	Data	44.68	41.35	72.95
	Model	55.21	41.45	223.86
12	Data	54.54	47.64	99.47
	Model	50.59	43.56	99.63

† : Measured in months
 See table 6 for a description of the covariate groups

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.⁶¹ In this section, I use the estimated model in order to conduct counterfactual policy experiments on such reforms. Specifically, I consider a decrease in the potential trial sentence of all cases in my sample, which captures the effect of an across-the-board reduction in the mandatory minimum sentence lengths. Also, I analyze the scenario in which cases with low, but strictly positive potential

⁶¹The following are examples of reforms adopted in 2010 alone: New Jersey reduced the mandatory minimum sentence for defendants convicted of drug-related offenses committed within 1000 feet of school property or a school bus. South Carolina equalized penalties for offenses involving crack and powder cocaine. In the same state, the minimum sentence for non-violent second-degree burglary was reduced from 15 to ten years. Tennessee enhanced alternative sentences for several non-violent property offenses, including felony theft. See Porter (2011) for other policy reforms.

trial sentences are replaced by cases with trial sentences equal 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 for each judicial division group. In all the policy experiments evaluated here, the outcomes of interest are: (i) the average probability that a case results in a incarceration conviction, either at trial or due to a settlement; (ii) the average observed incarceration sentence length; and (iii) the expected incarceration sentence length at the beginning of the prosecution process, which is simply the product of outcomes (i) and (ii). The total number of months of incarceration assigned in any given period is determined by the number of defendants prosecuted at that period times the expected sentence length. Outcome (iii) is, thus, the main variable of interest if, for example, one is mostly concerned about the total incarceration costs of current sentencing decisions.⁶² However, a decrease in the expected sentence length that is solely due to lower observed sentences may have no short-term effects on the incarceration rate. To see that, consider the immediate impact of reducing a convicted defendant's sentence from 40 to 20 years. Changes in the probability of conviction are more likely to capture the short-term effects of sentencing reforms. Indeed, if fewer defendants receive incarceration sentences, that immediately reduces the number of entering inmates and, consequently, current expenditures on corrections.

Evidently, my counterfactual analysis ignores some of the most important outcomes of sentencing reform, such as crime deterrence and recidivism. A more complete exercise would need to simultaneously account for these effects and the intricacies of plea bargaining. The analysis conducted here 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. Embedding a model similar to the one considered here into a framework in which agents decide whether to commit crimes is an exciting avenue for further research.

The estimated model also allows me to consider counterfactual experiments other than sentencing reform. An interesting scenario is that in which there are no asymmetries of information between the prosecutor and the defendant. In principle, better discovery procedures could reduce such asymmetries, although at the cost of longer and more burdensome negotiations. Therefore, this counterfactual analysis may serve as a reference for assessing the potential effects of more detailed discovery. But the most interesting aspect of the experiment is that it provides information on the extent to which defendants currently benefit from the asymmetric information. By comparing the average length of the incarceration sentences assigned in the scenarios with and without asymmetric information, it is possible to quantify the defendants' informational rents.

The last counterfactual analysis undertaken in this section addresses 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 the possibility of plea bargaining is likely to be too radical an intervention. However, considering this extreme scenario allows me to measure the loss of information that arises due to settlements. Specifically, I compute the proportion of defendants who are convicted to incarceration sentences by plea bargain but who would be acquitted if their cases were decided at trial. Insofar as trials correctly convict

⁶²Of course, that assumes the linearity of costs on incarceration time.

the guilty and acquit the innocent, such a proportion reveals the extent to which plea bargaining results in the conviction of innocent defendants. I am also able to calculate the losses incurred by defendants who are prohibited from settling their cases. In debating the merits of plea bargaining, legal scholars make a distinction between its private and social benefits. A plea bargaining is a voluntary contractual arrangement between the prosecutor and the defendant. If both parties are rational, they will only agree to settle a case if such an arrangement is mutually beneficial (Scott and Stuntz, 1992). But the conviction of innocent individuals imposes a negative externality to society—as attested, for example, by the requirement that judges find factual basis before accepting guilty pleas (Schulhofer, 1992). Although my analysis does not provide direct estimates of this externality, it allows me to assess how often innocent defendants agree to settle their cases. This result, along with my measure of the private costs of forbidding settlements, contributes to a better understanding of the welfare implications of plea bargaining.

8.1. Reducing mandatory minimum sentences. Tables 10 and 11 show the effects of a twenty percent reduction in the length of potential trial sentences for all cases in the sample.⁶³ The former table reports the effect of this reduction on the probability that a case results in a conviction to incarceration, 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 magnitude of the effect varies across covariate groups, but is always moderate. For most groups, the probability of conviction increases by roughly one to two percent.⁶⁴

Table 11 reports the impact of decreasing trial sentences on the expected length of the assigned sentences. For all 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 figures 4 to 6, which show that the prosecutor’s offer functions are quite steep for all covariate groups. 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. This decrease is captured, in table 11, by the drop in the average observed 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.

8.2. Assigning alternative punishments for mild cases. Now I simulate a policy intervention in which alternative punishments are assigned to defendants who currently face the prospect of a short incarceration sentence. With that intent, I consider the scenarios in which the ten percent of the cases with the lowest positive potential trial sentence are replaced by cases in which the

⁶³This 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.

⁶⁴For groups seven and nine, the table indicates a very small decrease in the probability of conviction, following a reduction in the trial sentence. These results are due to approximation error in the simulation used to compute the counterfactual exercises.

TABLE 10. Counterfactual results – Sentencing reform (distribution of outcomes)

Group		Outcome		
		Any conviction	Settlement	Trial conviction
1	Current	36.59%	34.86%	1.73%
	-20% trial sentence length	37.28%	35.91%	1.37%
	-10% potential 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% potential incarceration cases	37.94%	33.38%	4.56%
3	Current	36.66%	34.59%	2.08%
	-20% trial sentence length	37.08%	35.49%	1.59%
	-10% potential incarceration cases	32.86%	31.02%	1.84%
4	Current	40.68%	36.64%	4.04%
	-20% trial sentence length	41.58%	38.21%	3.37%
	-10% potential incarceration cases	36.40%	32.56%	3.84%
5	Current	35.55%	34.44%	1.10%
	-20% trial sentence length	36.08%	35.43%	0.65%
	-10% potential incarceration cases	32.02%	31.23%	0.79%
6	Current	39.68%	37.64%	2.04%
	-20% trial sentence length	40.00%	37.95%	2.05%
	-10% potential incarceration cases	35.76%	33.70%	2.06%
7	Current	37.53%	36.10%	1.43%
	-20% trial sentence length	37.44%	36.18%	1.26%
	-10% potential incarceration cases	33.58%	32.36%	1.22%
8	Current	38.62%	37.60%	1.01%
	-20% trial sentence length	39.46%	38.81%	0.65%
	-10% potential incarceration cases	34.91%	34.07%	0.84%
9	Current	21.59%	19.51%	2.08%
	-20% trial sentence length	21.57%	19.56%	2.01%
	-10% potential incarceration cases	19.24%	17.36%	1.88%
10	Current	29.09%	25.83%	3.26%
	-20% trial sentence length	29.44%	26.64%	2.80%
	-10% potential incarceration cases	26.07%	23.09%	2.98%
11	Current	20.91%	19.33%	1.58%
	-20% trial sentence length	21.52%	20.55%	0.97%
	-10% potential incarceration cases	18.23%	17.07%	1.16%
12	Current	28.61%	25.02%	3.59%
	-20% trial sentence length	28.98%	25.92%	3.06%
	-10% potential incarceration cases	25.69%	22.53%	3.16%

See table 6 for a description of the covariate groups

TABLE 11. Counterfactual results – Sentencing reform (sentences)

Group		Outcome	
		Average observed sentence [†] (given $\Psi \in \{1, 2\}$)	Expected sentence [†] (unconditional on Ψ)
1	Current	57.12	20.90
	-20% trial sentence length	41.23	15.37
	-10% potential incarceration cases	63.64	20.84
2	Current	45.08	19.03
	-20% trial sentence length	35.63	15.14
	-10% potential incarceration cases	49.95	18.95
3	Current	54.29	19.91
	-20% trial sentence length	38.83	14.40
	-10% potential incarceration cases	59.38	19.51
4	Current	51.98	21.14
	-20% trial sentence length	38.79	16.13
	-10% potential incarceration cases	57.77	21.03
5	Current	52.76	18.75
	-20% trial sentence length	34.47	12.44
	-10% potential incarceration cases	56.46	18.08
6	Current	42.93	17.03
	-20% trial sentence length	33.85	13.54
	-10% potential incarceration cases	47.69	17.05
7	Current	43.43	16.30
	-20% trial sentence length	33.71	12.62
	-10% potential incarceration cases	47.80	16.05
8	Current	47.07	18.18
	-20% trial sentence length	31.42	12.40
	-10% potential incarceration cases	51.09	17.84
9	Current	46.15	9.96
	-20% trial sentence length	36.41	7.85
	-10% potential incarceration cases	51.18	9.85
10	Current	43.55	12.67
	-20% trial sentence length	33.45	9.85
	-10% potential incarceration cases	48.06	12.53
11	Current	55.21	11.54
	-20% trial sentence length	36.73	7.90
	-10% potential incarceration cases	58.95	10.75
12	Current	50.59	14.47
	-20% trial sentence length	38.57	11.18
	-10% potential incarceration cases	55.52	14.26

[†] : Measured in months

See table 6 for a description of the covariate groups

potential trial sentence is zero.⁶⁵ The effects of such a counterfactual are presented in tables 10 and 11. The impact of the intervention on the probability of conviction is high. A ten percent decrease in the number of cases with positive potential trial sentence reduces the total probability of conviction by roughly ten percent for all covariate groups. This effect can be explained by the types of cases affected 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 assigning alternative punishments on the expected sentence at the beginning of the prosecution process 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 four out of the 12 covariate groups considered in the analysis. The greatest impact is a reduction by 6.85 percent, for group 11. 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 more often 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. However, the latter policy is not likely to decrease incarceration rates in the short run.

8.3. Eliminating asymmetric information. In this counterfactual simulation I examine the scenario in which there is no asymmetric information between the prosecutor and the defendant. With complete information, given a trial sentence t , defendant's trial costs c_d and a defendant of type θ , the prosecutor offers to settle for $\theta t + c_d$. This offer leaves the defendant just indifferent between accepting it and facing a trial. As a consequence, every case is settled. A natural outcome of eliminating informational asymmetries is thus an increase in the conviction rate. Such an increase can be easily measured using the estimated model.

A more difficult computation is that of the distribution of sentences. The 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 full distribution of settlement offers in the complete information scenario would require me to know the distribution of types outside of this range. Nevertheless, I am able to compute bounds for the mean settlement offer. To do so, I consider two extreme cases. In the first one, which leads to a lower bound for the mean settlement offer, 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

⁶⁵Exactly what sentence length constitutes the 20th percentile in the distribution of positive potential trial sentences varies across covariate groups. For groups one to 12, the values are 48.23, 21.75, 38.61, 30.92, 42.60, 26.00, 22.01, 62.47, 20.23, 25.46, 55.48 and 23.37 months, respectively. It is arguably more plausible that an actual reform would affect all cases below a specific threshold (say one year) statewide. Still, the results presented here serve as an illustration of the potential of a broader use of alternative sentences in each judicial group.

mean settlement offer 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 12 shows the results of this analysis. The probability of a conviction to incarceration increases considerably for all 12 covariate groups. Such an increase ranges from seven percent, for group seven, to 44.39 percent, for group 11. Very substantial increases, as the one computed for group 11, are due to a relatively large proportion of cases currently resulting in a trial acquittal. In the scenario without asymmetric information, all such cases would be settled, leading to an incarceration conviction. The lower bounds for the expected sentence, conditional on an incarceration conviction, are estimated to be longer than the current level for all covariate groups other than group nine. For many groups, even the lower bound is markedly longer than the current expected sentence. For group eight, for example, these values are 96.11 months and 47.07 months, respectively. My findings indicate that eliminating the asymmetric information between the prosecutor and the defendant would lead to more incarceration convictions and, conditional on such a conviction, longer assigned sentences. Therefore, it is not a surprise that the unconditional expected sentence increases considerably for all covariate groups. As shown in the table, the lower bound for this increase varies from 18.13 percent to 134.55 percent, depending on the covariate group. The upper bound for the increase ranges from 50.00 percent to 172.20 percent. These results suggest that, on average, defendants greatly benefit from informational rents in the process of plea bargaining.

8.4. Eliminating plea bargaining. I now consider the scenario in which no plea bargaining is allowed, and every case reaches the trial stage. In such a scenario, each defendant is convicted with a probability equal to her type. Thus, given a trial sentence t and a defendant of type θ , the expected sentence is θt . Here, both the computation of the overall probability of incarceration conviction and that of the expected sentence are complicated by the lack of identification of the distribution of types $F(\cdot)$ over its whole support. I can only calculate bounds for these values. To obtain the bounds, I consider the same two extreme cases as in the previous counterfactual exercise. First, I assume that all defendants with type lesser than $\tilde{\theta}(\underline{t})$ have type zero, and all defendants with type greater or equal to $\tilde{\theta}(\bar{t})$ have type $\tilde{\theta}(\bar{t})$. These assumptions generate lower bounds for the probability of conviction and the expected sentence. Then, to compute the upper bounds, I assume that all defendants with type lesser or equal to $\tilde{\theta}(\underline{t})$ have type $\tilde{\theta}(\underline{t})$, and all defendants with type above $\tilde{\theta}(\bar{t})$ have type one.

Table 13 presents the lower and upper bounds for the probability of conviction. Relative to current levels, such a probability falls substantially for all covariate groups. The difference between the current probability and the estimated lower bound varies from 1.62 p.p., for group seven, to 11.33 p.p., for group eight. When, instead, the upper bounds are considered, the difference ranges from 5.37 p.p. to 15.18 p.p.. These magnitudes are large, especially considering that the current rates of incarceration convictions are roughly between 20 to 40 percent, depending on the covariate group. For four out of the 12 groups, eliminating the possibility of plea bargains leads to decreases of more than 20 percent in the conviction rate, even taking the lower bound for such a decrease into consideration. The explanation for these large effects is that the estimated distributions of defendants' types, depicted in figures 7 and 8, place a considerable portion of their probability mass

TABLE 12. Counterfactual results – No asymmetric information

Group		Outcome		
		Probability of conviction	Expected sentence [†] (given $\Psi \in \{1, 2\}$)	Expected sentence [†] (unconditional on Ψ)
1	Current	36.59%	57.12	20.90
	No asymmetric info.	43.61%	[84.41 , 100.57]	[36.81 , 43.86]
2	Current	42.22%	45.08	19.03
	No asymmetric info.	50.26%	[48.47 , 58.15]	[24.36 , 29.22]
3	Current	36.66%	54.29	19.91
	No asymmetric info.	43.61%	[78.07 , 90.09]	[34.05 , 39.29]
4	Current	40.68%	51.98	21.14
	No asymmetric info.	50.26%	[49.70 , 60.40]	[24.98 , 30.36]
5	Current	35.55%	52.76	18.75
	No asymmetric info.	40.50%	[[99.78 , 115.60]	[40.41 , 46.82]
6	Current	39.68%	42.93	17.03
	No asymmetric info.	44.36%	[62.82 , 75.40]	[27.87 , 33.45]
7	Current	37.53%	43.43	16.30
	No asymmetric info.	40.50%	[58.37 , 72.62]	[23.64 , 29.41]
8	Current	38.62%	47.07	18.18
	No asymmetric info.	44.36%	[96.11 , 111.54]	[42.63 , 49.48]
9	Current	21.59%	46.15	9.96
	No asymmetric info.	30.19%	[45.86 , 55.44]	[13.84 , 16.74]
10	Current	29.09%	43.55	12.67
	No asymmetric info.	35.48%	[44.26 , 53.28]	[15.70 , 18.90]
11	Current	20.91%	55.21	11.54
	No asymmetric info.	30.19%	[64.39 , 71.70]	[19.44 , 21.65]
12	Current	28.61%	50.59	14.47
	No asymmetric info.	35.48%	[51.40 , 61.19]	[18.24 , 21.71]

† : Measured in months

See table 6 for a description of the covariate groups

at relatively low values. These estimates indicate that, despite facing low probabilities of being convicted at trial, many defendants currently accept to settle their cases for short incarceration sentences.

Notice that, for each defendant, the expected sentences in the scenarios without plea bargaining and without asymmetric information are given by θt and $\theta t + c_d$, respectively. That is, the difference between the sentences in each scenario consists of the defendants' trial costs. As shown in table 7, I estimate such costs to be essentially zero for all covariate groups. Thus, for each group, the expected sentence without plea bargaining is the same as the one reported in the third column of table 12, for the case of no asymmetric information. These values are considerably longer than the current expected sentences. Therefore, as discussed above, a large proportion of the defendants

TABLE 13. Counterfactual results – No plea bargaining

Group	Outcome	
	Probability of conviction	
1	Current	36.59%
	No plea bargaining	[25.31% , 30.21%]
2	Current	42.22%
	No plea bargaining	[33.09% , 39.69%]
3	Current	36.66%
	No plea bargaining	[24.03% , 27.81%]
4	Current	40.68%
	No plea bargaining	[31.53% , 38.32%]
5	Current	35.55%
	No plea bargaining	[21.62% , 25.25%]
6	Current	39.68%
	No plea bargaining	[29.35% , 35.19%]
7	Current	37.53%
	No plea bargaining	[28.68% , 35.91%]
8	Current	38.62%
	No plea bargaining	[23.44% , 27.29%]
9	Current	21.59%
	No plea bargaining	[16.22% , 19.91%]
10	Current	29.09%
	No plea bargaining	[22.39% , 27.71%]
11	Current	20.91%
	No plea bargaining	[12.04% , 13.61%]
12	Current	28.61%
	No plea bargaining	[21.84% , 26.21%]

See table 6 for a description of the covariate groups

would be acquitted at trial if plea bargains were not allowed. But the results in table 12 suggest that, in expectation, defendants would be considerably worse off in such a scenario.

9. DISCUSSION

9.1. Comparison with Sieg (2000). The model estimated by Sieg (2000), which is based on an earlier contribution due to Nalebuff (1987), is closely related to the one studied here. There are, however, important differences between Sieg’s paper and mine. First, I identify and estimate my model without assuming any parametric specification, whereas Sieg’s analysis is totally parametric. Also, I analyze data on criminal cases, whereas Sieg uses data on medical malpractice disputes. Not surprisingly, the different applications lead to distinctions between the models used in the two papers, which turn out to have substantial implications. I now discuss these differences.

In Sieg’s model, the defendant is privately informed about the merit of the case, and the damages suffered by the plaintiff are commonly known by the two parties. If the case reaches the trial stage, the plaintiff is, with probability one, awarded the product of the damages and the merit. In such a setup, the defendant is found not liable if and only if the case has no merit at all. A

direct implication is that changes in the distribution of damages across cases have no impact on the proportion of cases won by the plaintiff. That is not a shortcoming in Sieg’s application since, in the context of medical malpractice disputes, the plaintiff win rate is not an outcome of major interest to policy makers. But the same setup in the context of criminal cases would imply that changes in the distribution of potential trial sentences do not affect the probability that the defendant is found guilty. That would prevent the evaluation of how changes in sentencing guidelines influence conviction rates, a topic of great interest in the discussion on sentencing reform. In the model estimated in my paper, the defendant is privately informed—not about the merit of the case, but about the probability of being found guilty at trial. Changes in the distribution of potential trial sentences affect conviction rates by altering the types of defendants who, in equilibrium, choose to go to trial.

Another difference between the model used by Sieg and the one analyzed here concerns the prosecutor’s outside option. There, the plaintiff cannot bring a case to trial if doing so implies a negative expected payoff. In other words, the prosecutor’s outside option is set to zero, which makes perfect sense in the context of tort cases. Here, in contrast, I assume that, for cases with positive potential trial sentence, the prosecutor’s outside option is low enough that such cases are never withdrawn. This assumption is arbitrary. However, as argued in previous sections, prosecutors often face career-concern incentives to not withdraw even cases that are very difficult. Although it is theoretically appealing to consider the prosecutor’s outside option, setting it to zero in the analysis of criminal cases is at least as arbitrary as setting it to any negative number. Future research can incorporate the estimation of the prosecutor’s outside option into the empirical analysis of criminal proceedings.

10. CONCLUSION

In this paper, I develop a framework for the empirical analysis of plea bargaining, which allows me to evaluate how sentences to be assigned at trial affect the final result of criminal cases. I adapt a model of bargaining with asymmetric information due to Bebchuk (1984), and I present conditions for it to be non-parametrically identified. I then propose a consistent estimator for the model, which I apply to data on criminal cases prosecuted in the North Carolina Superior Courts. My findings suggest that, while the opportunity costs of going to trial are high for prosecutors, defendants behave as if trials were costless.

Using the estimated model, I evaluate the impact of different sentencing reforms on the outcome of prosecuted cases. My experiments suggest that lowering mandatory minimum sentences may greatly reduce the total amount of incarceration time assigned by the courts. However, the same intervention would 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 assigned incarceration time. Hence, these two interventions have the potential to complement each other. I also evaluate the impact of two other policy experiments: Eliminating the asymmetric information between the prosecutor and the defendant and prohibiting

cases from being settled. The results from these analysis 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. Nevertheless, in expectation, defendants are considerably worse off in the scenario without plea bargaining than in the current one.

REFERENCES

- Alschuter, Albert W.**, “The Prosecutor’s Role in Plea Bargaining,” *University of Chicago Law Review*, 1968, *36*, 50–112.
- Anderson, James M. and Paul Heaton**, “How Much Difference Does the Lawyer Make? The Effect of Defense Counsel on Murder Case Outcomes,” *Yale Law Journal*, 2012, *122* (1), 154–2694.
- Athey, Susan. and Philip.A. Haile**, “Nonparametric Approaches to Auctions,” in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6A, Elsevier, 2007, pp. 3847–3965.
- Baker, Scott and Claudio Mezzetti**, “Prosecutorial Resources, Plea Bargaining and the Decision to Go to Trial,” *Journal of Law, Economics and Organization*, 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, Elsevier, 2007, pp. 5549–5632.
- Daughety, Andrew F. and Jennifer F. Reinganum**, “Settlement Negotiations with Two Sided Asymmetric Information: Model Duality, Information Distribution, and Efficiency,” *International Review of Law and Economics*, 1994, *14*, 283–298.
- and ———, “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.
- 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.

- Eubank, Randall L.**, *Nonparametric Regression and Spline Smoothing*, second ed., New York: Marcel Dekker, 1999.
- 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.
- Iyengar, Radha**, "An Analysis of the Performance of Federal Indigent Defense Counsel," *NBER Working Paper Series*, 2007, 13187.
- 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.
- Kobayashi, Bruce H.**, "Deterrence with Multiple Defendants: an Explanation for 'Unfair' Plea Bargaining," *RAND Journal of Economics*, 1992, 23, 507–517.
- 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.
- Landes, William M.**, "An Economic Analysis of the Courts," *Journal of Law and Economics*, 1971, 14, 61–107.
- Larssen, 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.
- Merlo, Antonio, François Ortalo-Magné, and John Rust**, “The Home Selling Problem: Theory and Evidence,” 2008. Mimeo, University of Maryland.
- Meyer, Mary C.**, “Inference Using Shape-restricted Regression Splines,” *Annals of Applied Statistics*, 2008, 2, 1013–1033.
- Miceli, Thomas J.**, “Plea Bargaining and Deterrence: an Institutional Approach,” *European Journal of Law and Economics*, 1996, 3, 249–264.
- 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, 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.
- Rubinfeld, Daniel L. and David E. M. Sappington**, “Efficient Awards and Standards of Proof in Judicial Proceedings,” *RAND Journal of Economics*, 1987, 18, 308–315.
- Schulhofer, Stephen J.**, “Plea bargaining as disaster,” *Yale Law Journal*, 1992, 101 (8), 1979–2009.
- Schweizer, Urs**, “Litigation and Settlement under Two Sided Incomplete Information,” *Review of Economic Studies*, 1989, 56, 163–177.
- 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.
- Sobel, Joel**, “An Analysis of Discovery Rules,” *Law and Contemporary Problems*, 1989, 52, 133–160.
- 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 – Eighth 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.

Zhang, Shunpu, Rohana J. Karunamuni, and M. Chris Jones, “An Improved Estimator of the Density Function at the Boundary,” *Statistics and Probability Letters*, 1999, *94*, 1231–1241.

11. APPENDIX

11.1. Extensions.

11.1.1. *Relaxing the independence between T and Θ .* In the empirical model described in Section 6, I assume that the potential trial sentences and the defendants’ types are independently distributed, conditional on the case characteristics Z . Below, I relax this assumption and show that it is still possible to obtain partial identification of the model. In fact, the optimal settlement offer function is exactly identified and can be estimated by the procedure proposed in Section 6. Therefore, the estimates of settlement offer functions presented in this paper are robust to the dependence between defendants’ types and potential trial sentences, in the way defined below.

Let $F(\cdot|Z, T)$ be the distribution function of defendants’ types, conditional on the potential trial sentence T and on case characteristics Z . Assume that, for all $t \in [\underline{t}, \bar{t}]$ and $z \in \Delta$, the function $F(\cdot|Z, T)$ satisfies the technical assumptions outlined in section 4. Denote the density and the hazard functions associated with $F(\cdot|Z, T)$ by $f(\cdot|Z, T)$ and $\lambda(\theta, t, z)$, respectively. Assume that $\lambda(\theta, t)$ is differentiable in both its arguments, and that

$$\frac{\partial}{\partial \theta} \lambda(\theta, t, z) > 0 \quad \text{and} \quad \frac{\partial}{\partial t} \lambda(\theta, t, z) < \frac{1}{c_p(z) + c_d(z)} \quad (11.1)$$

for all $t \in [\underline{t}, \bar{t}]$, $\theta \in (\underline{\theta}, \bar{\theta})$ and $z \in \Delta$. The first inequality ensures that the hazard function associated with $F(\cdot|Z, T)$ is increasing in θ . The second one limits the amount of mass $F(\cdot|Z, T)$ can redistribute towards low defendants’ types as the potential trial sentence increases.

Given realizations t of T , z of Z , the equilibrium of the bargaining game can be found in the same way as in Section 4. It is characterized by equation (4.1) and by the first-order condition for the prosecutor, which is given by

$$\frac{t}{c_p + c_d} = \frac{f[\theta(s^*)|T = t, Z = z]}{1 - F[\theta^*|T = t, Z = z]}. \quad (11.2)$$

As before, I define the equilibrium settlement offer and defendant’s threshold type as functions of t , and denote them by $\tilde{s}(\cdot, s)$ and $\tilde{\theta}(\cdot, s)$, respectively. Using condition (11.1), and applying the implicit function theorem to equation (11.2), I have that $\tilde{s}(\cdot, s)$ and $\tilde{\theta}(\cdot, s)$ are strictly increasing in t , and $\tilde{s}(\cdot, z)$ is strictly convex. The argument in Section 6 can then be easily adapted to show that, under condition B and an analogue to condition A , the function $\tilde{s}(\cdot, z)$ is identified. Notice that the estimator of $\tilde{s}(\cdot, z)$ proposed in the same section does not make direct use of the distribution of

defendants' types. That means that it can be applied under the more general conditions described here, and the estimates of $\tilde{s}(\cdot, z)$ presented in Section 7 are still valid.

Although I am able to recover the optimal offer function after relaxing the independence assumption between T and Θ , the exact identification of the model's primitives does not hold. I now outline a strategy for the partial identification of such primitives. I begin by noticing that I can still recover $\tilde{\theta}(t, z)$ and the hazard function $\lambda(\tilde{\theta}(t), t, z)$ for all $t \in [\underline{t}, \bar{t}]$, up to the scalars $c_d(z)$ and $c_p(z)$. I now strengthen condition (11.1), by assuming that

$$\frac{\partial}{\partial \theta} \lambda(\theta, t, z) > 0 \quad \text{and} \quad \frac{\partial}{\partial t} \lambda(\theta, t, z) < 0. \quad (11.3)$$

The second inequality in this condition implies that $F(\cdot|Z, T)$ places more mass on high defendants' types as the potential trial sentence increases. A consequence of this inequality is that $\lambda(\theta, t, z) \leq \lambda(\theta, \tilde{\theta}^{-1}(\theta), z)$ for all $\tilde{\theta}(\underline{t}) \leq \theta \leq \tilde{\theta}(t)$. Therefore, I have that

$$\begin{aligned} F(\theta|Z, T) &= 1 - \exp\left(-\int_0^\theta \lambda(x, t, z) dx\right) \\ &\leq 1 - \underline{\mu}(z) \exp\left(-\int_{\tilde{\theta}(\underline{t})}^\theta \lambda(x, \tilde{\theta}^{-1}(x), z) dx\right) \end{aligned}$$

for all $\tilde{\theta}(\underline{t}) \leq \theta \leq \tilde{\theta}(t)$ and $z \in \Delta$, where $\underline{\mu}(z) = \exp\left(-\int_0^{\tilde{\theta}(\underline{t})} \lambda(x, \underline{t}, z) dx\right)$. Define the function

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

I then have that $\tilde{F}(t, z) \geq F(\tilde{\theta}(t)|Z = z, T = t)$ for all $t \in [\underline{t}, \bar{t}]$. Now consider the function

$$\ddot{g}(t, z) = \left[1 - \tilde{F}(t, z)\right]^{-1} \left[\frac{\partial}{\partial t} \tilde{s}(t, z)\right] b(\tilde{s}(t)|\Psi = 1, Z = z) P[\Psi = 1|Z = z].$$

Notice that

$$\ddot{g}(t, z) \geq [1 - F(t|Z = z)]^{-1} \left[\frac{\partial}{\partial t} \tilde{s}(t, z)\right] b(\tilde{s}(t)|\Psi = 1, Z = z) P[\Psi = 1|Z = z] = g_j(t)$$

for all $t \in [\underline{t}, \bar{t}]$ and $z \in \Delta$, where the equality comes from (6.1), (6.3) and (6.4). That allows me to write

$$1 - \int_{[\underline{t}, \bar{t}]} \tilde{F}(t, z) \ddot{g}(t, z) dt \geq 1 - \int_{[\underline{t}, \bar{t}]} F(\tilde{\theta}(t)|Z = z, T = t) g(t|Z = z) dt \quad (11.4)$$

for all $t \in [\underline{t}, \bar{t}]$ and $z \in \Delta$.

The expression on the right-hand side of (11.4) is the probability that a case is settled, conditional on $Z = z$ and $\Psi \neq 0$ (i.e., on it not being withdrawn by the prosecutor). That probability is observed by the econometrician. The expression on the left-hand side is known only up to the scalars $c_d(z)$, $c_p(z)$ and $\underline{\mu}(z)$. Thus, equation (11.4) establishes a non-linear bound for such scalars. If a subset $\Delta_\rho \subset \Delta$ satisfies condition A, then $c_d(z') = c_d(z'')$, $c_p(z') = c_p(z'')$ and $\underline{\mu}(z') = \underline{\mu}(z'')$ for all z' and z'' in Δ_ρ . Therefore, (11.4) establishes multiple bounds on the model's parameters—one for each $z \in \Delta_\rho$. Those can be combined with the bounds for c_d described in footnote 36, in order to partially identify the model's primitives. Implementing such a strategy is an interesting topic for future research.

11.1.2. *Allowing for multiple outcomes following bargaining failure.* My identification and estimation strategies can be adapted to the analysis of settings other than the resolution of criminal cases. However, in the model presented above, a disagreement between the bargaining parties may lead to only two possible outcomes: A trial conviction or an acquittal. That assumption may be unrealistic in some applications. I now extend my identification strategy to allow for multiple possible outcomes in the event of a disagreement. Unfortunately, to do that, I must impose a parametric relationship between the private information received by one of the bargaining parties and the distribution of outcomes, conditional on a bargaining failure.

Assume that there are two agents, which I denote by p and d . The game has two stages. Bargaining takes place in the first stage, and, if both parties reach an agreement, the game ends. Otherwise, the game proceeds to a second stage. Payoffs in the second stage depend on a random variable O , distributed over the finite support $\{1, \dots, N\}$. Specifically, the second stage payoffs are given by

$$\begin{aligned} & -c_p + \sum_{n \in \{1, \dots, N\}} \mathbb{1}\{o = n\} \kappa_{p,n} t \quad , \text{ for agent } p \\ \text{and} \quad & -c_d + \sum_{n \in \{1, \dots, N\}} \mathbb{1}\{o = n\} \kappa_{d,n} t \quad , \text{ for agent } p, \end{aligned}$$

where $o \in \{1, \dots, N\}$ is the realization of O , and c_d and c_p are strictly positive real numbers. The value t , which is assumed to be strictly positive, measures the stake of the game and is common knowledge to both agents at the moment of bargaining.

Before bargaining begins, agent d receives a signal about O . That signal is represented by a random variable Θ , which is distributed according to the CDF F over the support $[\underline{\theta}, \bar{\theta}]$. The distribution of O , conditional on the signal Θ , is characterized by the functions $\{\mathcal{Y}_n(\cdot)\}_{i \in \{1, \dots, N\}}$, so that

$$P [O = n | \Theta = \theta] = \mathcal{Y}_n(\theta)$$

for all $n \in \{1, \dots, N\}$ and $\theta \in [\underline{\theta}, \bar{\theta}]$. My identification strategy assumes that the functions $\{\mathcal{Y}_n(\cdot)\}_{i \in \{1, \dots, N\}}$ are known by the econometrician. For example, such functions may be implied by assuming that Θ follows a parametric distribution over $\{1, \dots, N\}$, such as a binomial or a truncated Poisson distribution with parameter θ .

Assume that $\mathcal{Y}_n(\theta)$ is twice differentiable for every $n \in \{1, \dots, N\}$. Assume, also, that the technical conditions on F listed in Section 4 also hold here—i.e., F is twice differentiable and has an associated density function f ; the density f is strictly positive on an interval $(\underline{\theta}, \bar{\theta})$, and is zero outside of such interval; f is non-increasing in a neighborhood of $\bar{\theta}$; and the hazard rate $f/[1 - F]$ is strictly increasing in θ .

Bargaining follows a take-it-or-leave-it protocol. At stage one, agent p makes a proposal s to agent d . If the proposal is rejected, the game reaches the second stage described above. Conversely, if the proposal is accepted, the game ends, and the payoffs are

$$\begin{aligned} & \kappa_{p,1} t + (\kappa_{p,N} - \kappa_{p,1}) s \quad , \text{ for agent } p \\ \text{and} \quad & \kappa_{d,1} t + (\kappa_{d,N} - \kappa_{d,1}) s \quad , \text{ for agent } d. \end{aligned}$$

Assume that the payoffs are symmetric, in the following sense: Let $\kappa_{p,1} < \dots < \kappa_{p,N}$, and $\kappa_{d,1} > \dots > \kappa_{d,N}$, so that, if bargaining fails, high realizations of O benefit agent p at the expense of agent d . Assume, also, that

$$\tilde{\kappa}_n \equiv \frac{\kappa_{p,n} - \kappa_{p,1}}{\kappa_{p,N} - \kappa_{p,1}} = \frac{\kappa_{d,1} - \kappa_{d,n}}{\kappa_{d,1} - \kappa_{d,N}}$$

for all $n \in \{1, \dots, N\}$. Lastly, assume that

$$\begin{aligned} & \sum_{n \in \{1, \dots, N\}} \tilde{\kappa}_n \mathcal{Y}'_n(\theta) > 0 \\ \text{and} \quad & \sum_{n \in \{1, \dots, N\}} \tilde{\kappa}_n \mathcal{Y}''_n(\theta) < \frac{\tilde{c}_p + \tilde{c}_d}{t} \lambda'(\theta) \end{aligned}$$

for all $\theta \in [(\underline{\theta}, \bar{\theta})]$, where $\lambda(\cdot)$ is the hazard function associated with the distribution F . The two last conditions ensure the uniqueness of the equilibrium. They hold, for example, if $(\underline{\theta}, \bar{\theta}) \in [0, 1]$; $\mathcal{Y}_1(\theta) = 1 - \sum_{n \in \{2, \dots, N\}} a_n \theta$; and $\mathcal{Y}_n(\theta) = a_n \theta$ for $n \in \{2, \dots, N\}$; where $\sum_{n \in \{2, \dots, N\}} a_n \leq 1$, and $a_n \geq 0$ for all $n \in \{2, \dots, N\}$.

The payoffs may be conveniently normalized, so that the second-stage payoffs are now given by

$$\begin{aligned} & -\tilde{c}_p + \sum_{n \in \{1, \dots, N\}} \mathbb{1}\{o = n\} \tilde{\kappa}_n t \quad , \text{ for agent } p \\ \text{and} \quad & -\tilde{c}_d - \sum_{n \in \{1, \dots, N\}} \mathbb{1}\{o = n\} \tilde{\kappa}_n t \quad , \text{ for agent } d, \end{aligned}$$

and the payoffs in the event of an agreement in the first stage are

$$\begin{aligned} & s \quad , \text{ for agent } p \\ \text{and} \quad & -s \quad , \text{ for agent } d, \end{aligned}$$

where $\tilde{c}_p = c_p / (\kappa_{p,N} - \kappa_{p,1})$, and $\tilde{c}_d = c_d / (\kappa_{d,1} - \kappa_{d,N})$.

I solve for the subgame perfect equilibria of this game. Agent d , after receiving the signal θ , accepts an offer s made by agent p if and only if

$$s \leq \tilde{c}_d + \sum_{n \in \{1, \dots, N\}} \tilde{\kappa}_n \mathcal{Y}_n(\theta) t.$$

Thus, an offer s is accepted if and only if the signal θ received by agent d is greater than or equal to the cutoff

$$\theta(s) = T^{-1} \left(\frac{s - \tilde{c}_d}{t} \right), \quad (11.5)$$

where $T(x) \equiv \sum_{n \in \{1, \dots, N\}} \tilde{\kappa}_n \mathcal{Y}'_n(x)$.

Agent p then chooses s in order to solve

$$\max_s \left\{ [1 - F[\theta(s)]] s + F[\theta(s)] \left\{ -\tilde{c}_p + t \frac{\int_{\underline{\theta}}^{\theta(s)} \sum_{n \in \{1, \dots, N\}} \tilde{\kappa}_n \mathcal{Y}_n(\theta) f(x) dx}{F[\theta(s)]} \right\} \right\}.$$

Under the assumptions above, the optimal offer s^* satisfies $\theta(s^*) \in (\underline{\theta}, \bar{\theta})$. The first-order condition for agent p 's problem is then given by

$$\frac{t \sum_{n \in \{1, \dots, N\}} \tilde{\kappa}_n \mathcal{Y}'_n(\theta)}{\tilde{c}_p + \tilde{c}_d} = \frac{f[\theta(s^*)]}{\{1 - F[\theta(s^*)]\}}. \quad (11.6)$$

The equilibrium is characterized by equations 11.5 and 11.6. From the assumptions above, it follows that the equilibrium is unique. Holding \tilde{c}_p and \tilde{c}_d constant, I can define the equilibrium offer s^* and cutoff point $\theta(s^*)$ as functions of t , and denote such functions by $\tilde{s}(\cdot)$ and $\tilde{\theta}(\cdot)$, respectively.

Given a data-generating process similar to the one in Section 6, the identification of the model's primitives would proceed exactly as described in that section. As before, it is necessary to observe s whenever bargaining is successful. Also, notice that, to achieve identification, it is necessary only to observe t in the event of a single realization o of O .

11.2. Proofs.

11.2.1. *Proof of Lemma 1:* By condition B , I can divide $[\underline{t}, \bar{t}]$ into N sub-intervals of the form $I_1^T = [\underline{t}, t_1)$, $I_2^T = [t_1, t_2)$, \dots , $I_{N-1}^T = [t_{N-2}, t_{N-1})$, $I_N^T = [t_{N-1}, \bar{t}]$, and, for each $n = 1, \dots, N$, choose elements z'_n, z''_n of Δ_ρ such that the function $\phi(t, z'_n, z''_n) \equiv g(t|Z = z'_n)/g(t|Z = z''_n)$ is strictly monotonic on I_n^T . Without loss of generality, I assume that, for all n , if $\phi(t, z'_n, z''_n)$ is strictly increasing on I_n^T , then $\phi(t, z'_{n+1}, z''_{n+1})$ is strictly decreasing on I_{n+1}^T ; and, if $\phi(t, z'_n, z''_n)$ is strictly decreasing on I_n^T , then $\phi(t, z'_{n+1}, z''_{n+1})$ is strictly increasing on I_{n+1}^T .

That, together with equation (6.2) and the strict monotonicity of $\tilde{s}(\cdot, \tilde{z})$, implies that the interval $[\tilde{s}(t, \tilde{z}), \tilde{s}(\bar{t}, \tilde{z})]$ can be divided into sub-intervals I_1^S, \dots, I_N^S with the following property: for all $n = 1, \dots, N$, I have that (i) if $\phi(t, z'_n, z''_n)$ is strictly increasing on I_n^T , then the function $\phi_b(s, z'_n, z''_n) \equiv b(s|Z = z'_n)/b(s|Z = z''_n)$ is strictly increasing on I_n^S ; and (ii) if $\phi(t, z'_n, z''_n)$ is strictly decreasing on I_n^T , then $\phi_b(s, z'_n, z''_n)$ is strictly decreasing on I_n^S . Notice that $I_1^S = [\tilde{s}(\underline{t}, \tilde{z}), \tilde{s}(t_1, \tilde{z})]$, $I_2^S = [\tilde{s}(t_1, \tilde{z}), \tilde{s}(t_2, \tilde{z})]$, \dots , $I_{N-1}^S = [\tilde{s}(t_{N-2}, \tilde{z}), \tilde{s}(t_{N-1}, \tilde{z})]$ and $I_N^S = [\tilde{s}(t_{N-1}, \tilde{z}), \tilde{s}(\bar{t}, \tilde{z})]$.

Finally, for all $n = 1, \dots, N$, I identify the function $\tilde{s}(\cdot, \tilde{z})$ on the interval I_n^T by the following equation

$$\tilde{s}(t, \tilde{z}) = \phi_b^{-1}(\phi(t, z'_n, z''_n), z'_n, z''_n)$$

for all $t \in I_n^T$. \square

11.2.2. *Proof of Proposition 3.* The estimator of $\tilde{s}(\cdot, z)$ proposed in Section 6 has two stages. First, I estimate conditional densities of trial sentences and settlement offers. Then I use such sentences for constructing a functional equation whose solution is the estimate for $\tilde{s}(\cdot, z)$. In order to show the consistency of this estimator, I now state and prove a lemma, offering sufficient conditions for the consistency of two-steps estimators in which the second step employs sieve methods. The lemma is adapted from the well known results on consistency of two-steps estimators by Newey and McFadden (1994), and from the results on the consistency of sieve estimators by Chen (2007). After that, I show that its conditions are satisfied by the estimator of $\tilde{s}(\cdot, z)$ from Section 6.

LEMMA 2. *Let Ξ and Γ be two (possibly infinite-dimensional) parameter spaces endowed with metrics d_ξ and d_γ , respectively. Consider a data-generating process that can be described by the true*

parameters $\xi_0 \in \Xi$ and $\gamma_0 \in \Gamma$. An estimate $\hat{\gamma}_n$ of γ_0 is available from a previous estimation procedure. Let $\hat{Q}_n : \Xi \times \Gamma \rightarrow \mathfrak{R}$ be an empirical criterion, and Ξ_k be a sequence of approximating spaces to Ξ . Also, let

$$\hat{\xi}_n = \underset{\xi \in \Xi_k}{\operatorname{argmax}} Q_n(\xi, \hat{\gamma}_n).$$

Assume that the following conditions are true:

(a) (i) Under the metric d_ξ : Ξ is compact; and $Q(\xi, \gamma_0)$ is continuous on ξ_0 and upper semi-continuous on Ξ

$$(ii) \xi_0 = \underset{\xi \in \Xi}{\operatorname{argmax}} Q(\xi, \gamma_0) \text{ and } Q(\xi_0, \gamma_0) > -\infty$$

(b) (i) For any $\xi \in \Xi$ there exists $\pi_k \xi \in \Xi_k$ such that $d_\xi(\xi, \pi_k \xi) \rightarrow 0$ as $k \rightarrow \infty$

(ii) Under d_ξ , and for all $k \geq 1$: Ξ_k is compact and $Q_n(\xi, \gamma_0)$ is upper semi-continuous on Ξ_k

(c) For all $k \geq 1$, $\operatorname{plim}_{n \rightarrow \infty} \sup_{(\xi, \gamma) \in \Xi_k \times \Gamma} |Q_n(\xi, \gamma) - Q(\xi, \gamma)| = 0$

(d) (i) $\hat{\gamma}_n \xrightarrow{P} \gamma_0$ under d_γ

$$(ii) \sup_{\xi \in \Xi} |Q(\xi, \gamma) - Q(\xi, \gamma')| \rightarrow 0 \text{ as } \gamma \rightarrow \gamma' \text{ under } d_\gamma$$

Then $\hat{\xi}_n \xrightarrow{P} \xi_0$ under d_ξ .

Proof. Consider any $\epsilon > 0$, and notice that, by assumption (a), $d_\xi(\hat{\xi}_n, \xi_0) > \epsilon$ implies $Q(\hat{\xi}_n, \gamma_0) - Q(\xi_0, \gamma_0) < -2\eta$ for some $\eta > 0$. By assumptions (a.i) and (b), for k high enough, there is $\pi_k \xi_0 \in \Xi_k$ such that $Q(\xi_0, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) < \eta$.

$$\begin{aligned} d_\xi(\hat{\xi}_n, \xi_0) > \epsilon &\Rightarrow Q(\hat{\xi}_n, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) + Q(\pi_k \xi_0, \gamma_0) - Q(\xi_0, \gamma_0) < -2\eta \\ &\Leftrightarrow Q(\hat{\xi}_n, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) < -2\eta + Q(\xi_0, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) \\ &\Rightarrow Q(\hat{\xi}_n, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) < -\eta. \end{aligned}$$

Define $A_n \equiv \left\{ Q(\hat{\xi}_n, \gamma_0) - Q(\pi_k \xi_0, \gamma_0) < -\eta \right\}$. Clearly, $P \left[d_\xi(\hat{\xi}_n, \xi_0) > \epsilon \right] \leq P[A_n]$. To complete the proof, I just need to show that $A_n = o_p(1)$. Define

$$B_n \equiv \{ |Q(\pi_k \xi_0, \hat{\gamma}_n) - Q_n(\pi_k \xi_0, \hat{\gamma}_n)| > \eta/5 \}$$

$$C_n \equiv \{ |Q(\pi_k \xi_0, \gamma_0) - Q(\pi_k \xi_0, \hat{\gamma}_n)| > \eta/5 \}$$

$$D_n \equiv \left\{ |Q(\hat{\xi}_n, \hat{\gamma}_n) - Q_n(\hat{\xi}_n, \hat{\gamma}_n)| > \eta/5 \right\}$$

$$E_n \equiv \left\{ |Q(\hat{\xi}_n, \gamma_0) - Q(\hat{\xi}_n, \hat{\gamma}_n)| > \eta/5 \right\}.$$

Notice that $B_n = o_p(1)$ and $D_n = o_p(1)$ (by assumption (c)). Similarly, $C_n = o_p(1)$ and $E_n = o_p(1)$ (by assumption (d.ii)).

Now I argue that $A_n \cap B_n^C \cap C_n^C \cap D_n^C \cap E_n^C = \emptyset$. Indeed,

$$\begin{aligned}
D_n^C &\Rightarrow Q_n(\hat{\xi}_n, \hat{\gamma}_n) \leq Q(\hat{\xi}_n, \hat{\gamma}_n) + \eta/5 \\
E_n^C &\Rightarrow Q(\hat{\xi}_n, \hat{\gamma}_n) \leq Q(\hat{\xi}_n, \gamma_0) + \eta/5 \\
A_n &\Rightarrow Q(\hat{\xi}_n, \gamma_0) < Q(\pi_k \xi_0, \gamma_0) - \eta \\
C_n^C &\Rightarrow Q(\pi_k \xi_0, \gamma_0) \leq Q(\pi_k \xi_0, \hat{\gamma}_n) + \eta/5 \\
B_n^C &\Rightarrow Q(\pi_k \xi_0, \hat{\gamma}_n) \leq Q_n(\pi_k \xi_0, \hat{\gamma}_n) + \eta/5.
\end{aligned}$$

Hence, $A_n \cap B_n^C \cap C_n^C \cap D_n^C \cap E_n^C$ implies

$$\begin{aligned}
Q_n(\hat{\xi}_n, \hat{\gamma}_n) &\leq Q(\hat{\xi}_n, \gamma_0) + 2(\eta/5) < Q(\pi_k \xi_0, \gamma_0) + 2(\eta/5) - \eta \\
&\leq Q(\pi_k \xi_0, \hat{\gamma}_n) + 3(\eta/5) - \eta \leq Q_n(\pi_k \xi_0, \hat{\gamma}_n) + 4(\eta/5) - \eta \\
&= Q_n(\pi_k \xi_0, \hat{\gamma}_n) - \eta/5.
\end{aligned}$$

That contradicts $\hat{\xi}_n = \operatorname{argmax}_{\xi \in \Xi_k} Q_n(\xi, \hat{\gamma}_n)$.

Finally, notice that

$$\begin{aligned}
P[A_n] &\leq P[A_n \cup (B_n \cup C_n \cup D_n \cup E_n)] \\
&= P\left[A_n \cap (B_n \cup C_n \cup D_n \cup E_n)^C\right] + P[B_n \cup C_n \cup D_n \cup E_n] \\
&= P[\emptyset] + P[B_n] + P[C_n] + P[D_n] + P[E_n].
\end{aligned}$$

Therefore, $A_n = o_p(1)$. □

I can now show that the estimator for $\tilde{s}(\cdot, z)$ proposed in Section 6 is consistent. A feature of the second estimation stage is that, given the first-stage estimates for the conditional distribution of Ψ and the censored densities of trial sentences and settlement offers, the objective function is not random. In the notation of Proposition 3, $Q_n(\cdot, \cdot) = Q(\cdot, \cdot)$. As a consequence, in order to apply Lemma 2, I do not need to verify condition (c).

The estimator is defined for a set $\Delta_\rho \subseteq \Delta$. Remember that \mathcal{C} is the set of all pairwise combinations of elements in Δ_ρ . In the notation of Lemma 2, the true parameter ξ_0 is the function $\tilde{s}(\cdot, z)$, for $z \in \Delta_\rho$. The true parameter γ_0 consists of the two finite sequences of functions $\left\{\tilde{b}_r(\cdot, z', z'')\right\}_{(z', z'') \in \mathcal{C}}$ and $\left\{\tilde{g}_r(\cdot, z', z'')\right\}_{(z', z'') \in \mathcal{C}}$. In order to avoid burdensome notation, I'll present the proof for the case where Δ_ρ has only two elements, so that $\mathcal{C} = (z', z'')$. I denote $\tilde{s}(\cdot, z')$ and $\tilde{s}(\cdot, z'')$ by $\tilde{s}(\cdot)$; $\tilde{b}_r(\cdot, z', z'')$ by $\tilde{b}_r(\cdot)$; and $\tilde{g}_r(\cdot, z', z'')$ by $\tilde{g}_r(\cdot)$. The generalization for any finite Δ_ρ is conceptually trivial.

Remember that, by the assumptions in the statement of Proposition 2: (i) the function $b_r(\cdot)$ is differentiable and the absolute value of its derivative is bounded by \mathring{b}_r ; and (ii) the functions $b_r(\cdot)$ and $g_r(\cdot)$ are positive and bounded by \bar{B}_r and \bar{G}_r , respectively. Also, from equation 4.1 and the boundedness of $[\underline{t}, \bar{t}]$, I can assume without loss of generality that the function $\tilde{s}(\cdot)$ is bounded from above by \bar{s} , and its derivative is bounded from above by a constant \mathring{s} .

Using the boundary correction discussed in Section 6, the kernel density estimators of $b[s|\Psi = 1, Z]$ and $g[t|\Psi = 2, Z]$ are uniformly consistent. The estimators of $P[\Psi = 1|Z]$ and $P[\Psi = 2|Z]$ are

consistent, as well. From Slutsky's theorem, therefore, the first stage of the estimation procedure returns uniformly consistent estimates of $\tilde{b}_r(s)$ and $\tilde{g}_r(t)$.

The objective function in the second stage of estimation is given by

$$Q(s(\cdot), b_r(\cdot), g_r(\cdot)) = E \left\{ [b_r(s(t)) - g_r(t)]^2 \right\}$$

where $s(\cdot)$ is an element from the space of increasing and convex functions on $[\underline{t}, \bar{t}]$, and $g_r(\cdot)$ and $b_r(\cdot)$ are positive valued functions.

In order to show that the estimator in the second step is consistent, I need to prove that $Q(s(\cdot), b_r(\cdot), g_r(\cdot))$ satisfies conditions (a) and (d) from Lemma 2. That is what I do in the following lemmas.

LEMMA 3. *Under the assumptions in the statement of Proposition 3, the estimator described in Section 6 satisfies condition (a) from Lemma 2, where d_ξ is the sup norm*

Proof. Using equation (6.2), it is easy to verify that

$$Q(\tilde{s}(\cdot), \tilde{b}_r(s), \tilde{g}_r(t)) = 0$$

and that the objective function is strictly positive for any other continuous function $s(\cdot)$. Condition (a.ii) is then trivially verified.

Let Ξ be the space of functions defined on $[\underline{t}, \bar{t}]$ that are increasing and convex, uniformly bounded by zero and \bar{s} , and whose derivative is uniformly bounded by zero and \bar{b}_r . To verify condition (a.i), notice first that, by the Arzela-Ascoli theorem, the space of differentiable functions on $[\underline{t}, \bar{t}]$ that are uniformly bounded and have a uniformly bounded derivative is compact under the sup norm. The space Ξ is the intersection of that space and the space of increasing and convex functions on $[\underline{t}, \bar{t}]$, which is closed. Hence, Ξ is compact.

It remains to verify the upper semi-continuity of the objective function on Ξ and the continuity at $\tilde{s}(\cdot, z)$. Here, I show that the objective function is continuous on Ξ under the sup norm. Let $s(\cdot)$ and $\check{s}(\cdot)$ be two functions in Ξ such that

$$\sup_{t \in [\underline{t}, \bar{t}]} |s(t) - \check{s}(t)| \leq \eta.$$

I can write

$$\begin{aligned} & \left| E \left\{ \left[\tilde{b}_r(s(t)) - \tilde{g}_r(t) \right]^2 \right\} - E \left\{ \left[\tilde{b}_r(\check{s}(t)) - \tilde{g}_r(t) \right]^2 \right\} \right| \\ &= \left| E \left[b_r(s(t))^2 - b_r(\check{s}(t))^2 \right] + 2E \left\{ g_r(t) [b_r(s(t)) - b_r(\check{s}(t))] \right\} \right| \\ &\leq \left| E \left\{ [b_r(s(t)) - b_r(\check{s}(t))] [2g_r(t) + b_r(s(t)) + b_r(\check{s}(t))] \right\} \right|. \end{aligned}$$

Some algebra shows that the right-hand side of the last inequality is lower than $2 \bar{b}_r [\bar{B}_r + \bar{G}_r] \eta$, so that the objective function is continuous on the whole space Ξ . \square

LEMMA 4. *Under the assumptions in the statement of Proposition 3, the estimator described in Section 6 satisfies condition (d) from Lemma 2, where d_γ is the sup norm.*

Proof. Since I have uniformly consistent estimates from the first stage, condition (d.i) holds under the sup norm. To verify condition (d.ii), assume that $b_r(\cdot)$, $\check{b}_r(\cdot)$, $g_r(\cdot)$ and $\check{g}_r(\cdot)$ are such that

$$\begin{aligned} \sup_s \left| b_r(s) - \check{b}_r(s) \right| &\leq \eta \\ \text{and} \quad \sup_t \left| g_r(t) - \check{g}_r(t) \right| &\leq \eta \end{aligned} \tag{11.7}$$

for some $\eta > 0$. I can write

$$\begin{aligned} &\left| E \left\{ [b_r(s(t)) - g_r(t)]^2 \right\} - E \left\{ [\check{b}_r(s(t)) - \check{g}_r(t)]^2 \right\} \right| \\ &= \left| E \left[b_r(s(t))^2 - \check{b}_r(s(t))^2 \right] + E \left[g_r(t)^2 - \check{g}_r(t)^2 \right] + 2E \left[\check{b}(s(t))\check{g}(t) - \underline{b}(s(t))g(t) \right] \right| \\ &\leq E_1(t) + E_2(t) + E_3(t), \end{aligned}$$

where

$$\begin{aligned} E_1(t) &\equiv \left| E \left[b_r(s(t))^2 - \check{b}_r(s(t))^2 \right] \right| \\ E_2(t) &\equiv \left| E \left[g_r(t)^2 - \check{g}_r(t)^2 \right] \right| \\ \text{and} \quad E_3(t) &\equiv \left| 2E \left[\check{b}(s(t))\check{g}(t) - b(s(t))g(t) \right] \right|. \end{aligned}$$

Some algebra shows that, for all functions $s(\cdot)$, the following inequalities hold

$$\begin{aligned} E_1(t) &\leq \max \left\{ \sup_{t \in t \in [\underline{t}, \bar{t}]} 2 \eta b_r(s(t)) + \eta^2 ; \sup_{t \in t \in [\underline{t}, \bar{t}]} 2 \eta \check{b}_r(s(t)) + \eta^2 \right\} \\ &\leq \eta^2 + 2 \eta \bar{B}_r, \\ E_2(t) &\leq \max \left\{ \sup_{t \in t \in [\underline{t}, \bar{t}]} 2 \eta g_r(t) + \eta^2 ; \sup_{t \in t \in [\underline{t}, \bar{t}]} 2 \eta \check{g}_r(t) + \eta^2 \right\} \\ &\leq \eta^2 + 2 \eta \bar{G}_r, \\ E_3(t) &\leq 2 \max \left\{ \sup_{t \in t \in [\underline{t}, \bar{t}]} \eta [b_r(s(t)) + g_r(t)] + \eta^2 ; \sup_{t \in t \in [\underline{t}, \bar{t}]} \eta [\check{b}_r(s(t)) + \check{g}_r(t)] + \eta^2 \right\} \\ &\leq 2 \eta^2 + 2 \eta [\bar{B}_r + \bar{G}_r]. \end{aligned}$$

Since neither \bar{B}_r nor \bar{G}_r depends on $s(\cdot)$, condition (d) is satisfied. □

That concludes the proof of Proposition 3. □

11.3. Data appendix.

11.3.1. *Reducing multiple-counts cases to a single count.* My unit of analysis is a case. Some cases in the data are associated with multiple counts. To reduce such cases to a single count, I employ the following procedure: If sentences are assigned to more than one count of the same case, I consider only the count with the longest sentence.⁶⁶ If no sentence is assigned to any count in a case, I

⁶⁶I observe in the data set whether a sentence consists of incarceration time, intermediate punishment (such as probation) or community service. In order to classify the sentences, I use a lexicographic order, so that any time in incarceration is considered a higher sentence than any intermediate punishment. Similarly, any intermediate

classify the charged offenses for each count according to their severity (using the same classification adopted by the structured sentencing guidelines in North Carolina) and consider only the count with the most severe charged offense.

11.3.2. *Classification of offenses.* Each count in the data is associated with an offense code assigned by the North Carolina Justice System. The offense codes employed in North Carolina are based on the Uniform Offense Classifications, organized by the National Crime Information Center (NCIC). Like in the NCIC code system, the first two digits of the North Carolina codes classify the offenses into relatively broad categories (e.g., robbery, fraud, vehicle theft, etc). Based on these two digits, I determine whether each count in my data qualifies as a robbery, an assault or a sexual assault—the offense categories that I use in my structural analysis.

11.3.3. *Identification of judges.* In the main data, judges are identified only by their initials. In most cases, three initials are used. I match the initials to the full names of the judges as reported annually in the North Carolina Manual. In the period comprised by the sentencing data, only two pairs of judges have the same three initials. Cases decided by these judges were excluded from the data. I also excluded all the cases in which the judge was either identified by fewer than three initials or not identified at all.

11.3.4. *Life sentences.* To convert life sentences into a length of incarceration time, I consider the life expectancy in North Carolina for individuals of age 29.04, which is the average defendant’s age in my sample. This life expectancy is 77.14 years (Buescher and Gizlice, 2002). I want to make sure that any life sentence is at least as long as the longest non-life sentence, which, in North Carolina, is forty years. I then define the length of a life sentence as $\text{Max}\{\text{defendant's age} - 77.14; 40\}$. Cases resulting in a death sentence are excluded from the analysis.⁶⁷ Cases whose sentence length is missing in the data are treated as cases in which only an alternative sentence is assigned.

11.4. Estimation appendix.

11.4.1. *Observed heterogeneity.* I incorporate observed heterogeneity across cases to my estimation procedure by dividing the observations in the data into a finite number of covariate groups and implementing the estimator described in Section 6 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 successful plea bargain. These two types of conditional densities must be estimated both for cases under the responsibility of lenient judges and for those under the responsibility of harsh ones. Therefore, for each one of the 12 covariate groups under consideration in my analysis, I must estimate four conditional densities. I estimate these conditional densities using the smoothing method developed 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.

sentence is considered higher than any type of community service. Notice that I only take intermediate or community punishments into consideration in order to organize the data set by case. As discussed in the main text, the empirical analysis in this paper only accounts for incarceration sentences, and treats cases where only alternative punishments are assigned as cases with no sentence whatsoever.

⁶⁷Cases of non-homicide crimes against the person, which constitute the subsample considered in the structural analysis, never result in a capital sentence.

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 $\lambda = (\lambda_1, \dots, \lambda_r)$ and consider the following estimates of $f(\cdot)$ and $\mu(\cdot)$:

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

and

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

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

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

Finally, define the estimate 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—i.e., even those in which $X \neq x$. These observations, however, are weighted down, relative to the ones satisfying $X = x$. The weights are given by the vector $\lambda = (\lambda_1, \dots, \lambda_r)$. 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. Depending on the magnitudes of these two effects, the net result of increasing λ can be a decrease in the mean squared error associated with $\hat{g}(y|x)$.

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 cross-validation.⁶⁸ 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 in my analysis, 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

⁶⁸Li and Racine (2007) propose two basic approaches for the selection of smoothing parameters: least squares cross-validation and maximum likelihood cross-validation. The former method is too computationally costly for me to employ it, given the sample sizes that I deal with in my application. I therefore use maximum likelihood cross-validation in the present paper.

TABLE 14. Smoothing parameters

Covariate	Conditional density estimator [†]				
	Trial sentences (lenient judges)	Trial sentences (harsh judges)	Settlement offers (lenient judges)	Settlement offers (harsh judges)	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

[†] : Trial sentence densities are conditional on a conviction at trial and on the covariates. Settlement offer densities are conditional on a plea bargain and on the covariates.

variable X refers to the covariates used to divide the data into groups.⁶⁹ 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.

Table 14 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 6, the supports of trial sentences and settlement offers are bounded, which complicates the estimation of the conditional densities described above. To deal with this problem, 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.

⁶⁹To 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 15. Kernel bandwidths for trial sentences and settlement offers[†]

	Trial sentences (lenient judges)	Trial sentences (harsh judges)	Settlement offers (lenient judges)	Settlement offers (harsh judges)
Bandwidth	21.83	25.31	6.59	7.37
†: Measured in months				

The bandwidths h_0 for points away from the boundary are reported in table 15. The bandwidths for trial sentences are 21.83 months (lenient judges) and 25.31 months (harsh judges). Those for settlement offers are 6.59 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.

11.4.2. *Standard errors.* I use bootstrap methods to compute standard errors for the parameters reported in table 7. Specifically, I consider 1200 bootstrap samples for each covariate group. For each of these samples, I estimate the densities of trial sentences and settlement offers using the same bandwidths and smoothing parameters employed in the main data. I then estimate the offer function and, finally, the model parameters.

There are two main issues with this procedure. The first one is that 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 development of suitable inference methods in the context of my estimator is an interesting topic for further research. 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.⁷⁰ As a consequence, the standard deviations reported in table 7 may underestimate the actual standard deviation of the model parameters.

⁷⁰More 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, 1.27 for group two, 2.04 for group three and 0.00 for group four. This procedure eliminates 13.83% of the 1200 bootstrap samples for group one, 32.50% for group two, 12.58% for group three, 37.00% for group four, 3.33% for group five, 4.83% for group six, 20.25% for group seven, 14.50% for group eight, 11.25% for group nine, 10.83% for group ten, 36.08% for group 11 and 14.75% for group 12.