The Effect of Punishment Severity on Plea Bargaining

Richard T. Boylan   *Rice University*

Abstract

This study examines whether criminal suspects facing more severe punishments are more likely to go to trial. Sample selection makes it difficult to obtain valid proxies for severity; for instance, I expect severity to be positively related to the prosecutor's decision to indict, to indict in federal court (versus state court), and to try the suspect. Theoretical and empirical findings indicate that in samples containing only indicted, convicted, or tried suspects, reasonable proxies for severity may be negatively related to actual severity. The assignment of defendants to judges randomizes the severity of punishment in a manner that is unrelated to sample selection. Thus, by examining the effect of these assignments, I find that a 10-month increase in prison sentences raises trial rates by 1 percentage point.

1. Introduction

In the United States, trials are adversarial: opposing parties acquire and present evidence, while impartial judges and juries decide the outcome of the dispute. However, most cases are settled without a trial. For civil cases, federal judges are actively encouraged to participate in negotiations among litigants and thus influence settlement outcomes (Galanter 1986). For criminal cases, federal judges are not allowed to play any role in plea negotiations or make any comments that might indirectly influence the bargaining process (Fed. R. Crim. P. 11[c][1]). However, the reputations of federal judges may indirectly affect negotiations; for instance, defendants may be more willing to plead guilty if they are assigned to judges reputed to be favorable to prosecutors.

Prior to the enactment of mandatory minimum sentences and sentencing guidelines, federal judges had almost complete discretion over sentencing. Thus, prosecutors saw their role as securing convictions (Stuntz 1997; Simons 2009). Mandatory minimum sentences and sentencing-guidelines laws shifted some of

the sentencing authority from judges to prosecutors. For instance, prosecutors can charge a defendant with possessing a gun while selling drugs, which is punished by a 5-year mandatory minimum sentence on top of the drug-trafficking sentence, while declining to file gun charges for identical defendants. The lack of judicial supervision over charging decisions thus gives prosecutors considerable power over sentencing (Vorenberg 1981).

Judges and scholars have recommended that prosecutors alter their role as adversaries with respect to sentencing and instead seek just outcomes. In particular, some scholars advocate that prosecutors ensure that similar defendants be punished similarly (Green 1999; Simons 2009).[1] In this paper, I provide evidence that prosecutors seek harsher penalties against defendants who are assigned to harsher judges. This evidence suggests that prosecutors act as adversaries in sentencing decisions rather than seek similar penalties for defendants who commit similar crimes.

To reach these conclusions, I note that if prosecutors seek to punish similar defendants similarly, they will make their plea offers independent of the judge assigned to the case. Consequently, defendants find their plea offers more attractive and are less likely to go to trial when they are assigned to harsher judges. In contrast, prosecutors who act as adversaries offer longer sentences to defendants assigned to harsher judges.

I define a judge as harsh if defendants receive longer prison sentences when assigned to that judge compared to other judges in the same courthouse and year. I find that defendants are 1 percentage point more likely to go to trial when they are assigned to a judge who is 10 months harsher. This evidence contradicts the hypothesis that prosecutors seek equitable sentences. Further, the magnitude of this effect is large since only 9 percent of defendants go to trial.

Prior work has shown that the length of prison sentences varies significantly among federal judges (Anderson, Kling, and Stith 1999; LaCasse and Payne 1999). For instance, judges differ in sentencing, in the evidence that they allow at trial over a hearsay objection, and in whether they allow particular expert testimony. Earlier studies found that judge assignments do not affect trial rates (Ashenfelter, Eisenberg, and Schwab 1995; LaCasse and Payne 1999). However, these studies lack statistical power since they are based on only two or three districts and include a different dummy variable for each judge.[2] In contrast, Waldfogel (1995) assumes that differences among judges can be summarized by a single parameter and finds that judges who have higher percentages of their cases go to trial also have higher rates of defendant wins. Waldfogel (1995) examines civil cases; thus, his results are not applicable to criminal cases.

If prosecutors do not seek uniform sentences, then my empirical estimates allow me to quantify the effect of increasing prison sentences on the adminis-

---

[1] In contrast, Easterbrook (1983) remarks that if deterrence is the only goal of criminal law, then prosecutors should not seek similar penalties for similar defendants.

[2] Anderson, Kling, and Stith (1999) look at more districts but do not examine the decision to go to trial.

tration of criminal justice. This is important because scholars and practitioners have expressed concerns that the severe sentences required by the U.S. Federal Sentencing Guidelines discourage defendants from accepting plea bargains and lead to more trials than the criminal justice system can handle. Conversely, scholars have argued that death penalty statutes may save resources by making it more likely for defendants to plea-bargain rather than go to trial and face the threat of execution.[3] Given that trials take priority over other activities, higher trial rates reduce the amount of time that prosecutors spend on other valuable activities.[4]

Prior empirical work provides contradictory evidence on the effect of punishment severity on the likelihood that a case goes to trial.[5] Some of these empirical findings can be criticized because they use the prison sentence at trial as a proxy for severity. However, cases that go to trial are not representative of the set of litigated cases, and thus prison sentences at trial are endogenous (Priest and Klein 1984). Other empirical studies use characteristics of the offense for indicted defendants as proxies for punishment severity. These proxies can also lead to specious conclusions, as illustrated in the following example.

In my example, two defendants are indicted in federal court for trafficking heroin: the first defendant is accused of trafficking 1 gram of heroin, while the second is accused of trafficking 50 grams. Without any other aggravating circumstance, the first defendant faces a 15- to 21-month prison sentence, while the second faces a 33- to 41-month sentence. Federal authorities prosecute the worst offenders; thus, it is unlikely that the first defendant is suspected solely of trafficking 1 gram of heroin. However, I assume in my example that individuals accused of trafficking 50 grams are indicted even absent any aggravating circumstance.

Specifically, I assume that the first defendant carried a gun, while the second did not. Thus, the defendant with 1 gram faces a longer prison sentence than the defendant with 50 grams, since possession of a gun adds a minimum of 5 years to the prison sentence. Of course, the defendant with 50 grams of heroin could also have carried a gun. However, prosecutors' decisions regarding which

---

[3] Concerns that the U.S. Federal Sentencing Guidelines would increase the number of trials are discussed in Cohen (1991), Karle and Sager (1991), and Sisk, Heise, and Morriss (1998), while the expectation that death penalty statutes decrease the number of plea bargains is discussed in Kuziemko (2006) and Cook (2009).

[4] For instance, between October 1994 and September 1995, the District Court for the Southern District of New York disposed of 1,048 criminal cases. Of these cases, 93 were disposed by trial. Over the same period, the Southern District of New York spent 1,252 attorney working hours on district court activity and 1,023 attorney working hours on witness preparation (U.S. Department of Justice, response to author's Freedom of Information Act request for 1993–98 Department of Justice personnel data, September 20, 1999). Thus, a 1-percentage-point increase in the number of cases disposed by trial increases the number of trials in that district by 11 percent and has a significant negative effect on the activities of that district.

[5] Elder (1989), Snyder (1990), and Boylan and Long (2005) provide evidence that more severe cases are more likely to go to trial. Bjerk (2005) and Kuziemko (2006) find that severity does not affect the likelihood that a case goes to trial. Landes (1971) reaches different conclusions in the two samples that he examines.

defendants to indict make it more likely that the defendant with 1 gram of heroin possessed a gun. Thus, it is conceivable that the average prison sentence of indicted defendants with 1 gram of heroin is longer than the average prison sentence of defendants with 50 grams of heroin. In Section 2, I examine all federal drug-trafficking defendants between 1994 and 2008 for whom I have information on the weight of the heroin involved in the case. Consistent with my example, I find that the weight of the heroin is negatively correlated with the prison sentence received by defendants. Thus, the weight of the heroin is a very poor proxy for the expected prison sentence faced by a heroin trafficker when I do not control for other characteristics of the offense (for example, possession of a gun).

When the researcher does not observe all aspects of the crime, he or she may obtain incorrect estimates of the effect of punishment severity on the likelihood that a case goes to trial. In my example, the defendant with 1 gram of heroin pleads guilty to trafficking, and in exchange the prosecutor does not file gun charges. Thus, the researcher does not know that the defendant with 1 gram of heroin possessed a gun. This example is not far-fetched, since gun charges are often dropped as part of a plea agreement (Nagel and Schulhofer 1992).[6] Suppose further that the defendant with 50 grams of heroin is convicted at trial. Then the researcher incorrectly concludes that the defendant with 1 gram committed the less severe crime and thus that the defendant who committed the less severe crime agreed to plea-bargain.

Note in my example that the defendant with 1 gram of heroin is sentenced to a shorter prison sentence than the defendant with 50 grams. However, I can easily modify this example to ensure that the weight of the heroin is negatively correlated with the prison sentence. For instance, I can include several additional defendants with 1 gram of heroin who possessed a gun. Unlike the first defendant, these additional defendants are convicted at trial of trafficking heroin and possessing a gun. In this more complicated example, the average prison sentence for individuals with 1 gram of heroin is higher than the average prison sentence for individuals with 50 grams. Further, individuals with 1 gram of heroin are still more likely to plea-bargain than individuals with 50 grams. Finally, presence of a gun is a very good predictor of whether a case goes to trial. However, the relation is not causal since in my example gun charges are observed only in cases that go to trial.

In Section 3 I describe a theoretical model for the selection of cases that prosecutors decline, try, and plea-bargain. The expected prison sentence at trial (severity) is the sum of two terms: the characteristics of the crime that are always observed by the researcher (observed severity) and the characteristics of the crime that are unobserved for an endogenous subset of observations (unobserved

---

[6] Thus, the researcher observes whether the prosecutor included the gun charges in the indictment but does not observe whether the prosecutor could have included the charges but decided not to do so in exchange for the defendant pleading guilty of drug trafficking.

severity). In my example, the weight of the heroin was always observed, while the presence of a gun was not observed for the first defendant as part of the plea agreement. I show in my theoretical model that case selection can lead to a negative correlation between observed severity and the expected prison sentence at trial. This theoretical result is consistent with my empirical evidence that the weight of the heroin is negatively correlated with the prison sentence. Thus, my empirical evidence and my theoretical model indicate that the selection of cases to prosecute can make proxies for severity invalid. Prior empirical work provides contradictory evidence on the effect of punishment severity on the likelihood that a case goes to trial. One explanation for these contradictory findings is that case selection makes the proxies for severity unreliable.

There are other explanations for the contradictory empirical evidence on the effect of punishment severity on the expected prison sentence at trial. For instance, a defendant may be able to trade information about other cases in exchange for a reduced sentence. Thus, a defendant who traffics a large amount of heroin may be able to obtain a more lenient prison sentence because he or she has more valuable information to trade. This could explain my empirical finding that the weight of the heroin is negatively correlated with the prison sentence. For some crimes, the individuals charged with more serious offenses may have more information to trade and are more likely to plea-bargain, while for other crimes, the relation could be reversed. Consequently, to estimate whether trials are more frequent when penalties are more severe, I need an instrument for severity—namely, a variable that is correlated with the expected prison sentence for a defendant but uncorrelated with the cost of going to trial, the likelihood that the defendant is guilty, the strength of the evidence, the ability of the defendant to provide information in other cases, and sample selection. Judge assignments are a valid instrument for severity, since defendants are randomly assigned to judges after prosecutors decide to indict a suspect and since judges vary in their severity. Section 5 provides empirical evidence that when defendants are assigned harsher judges, they are more likely to go to trial and to plead to longer prison sentences.

2. Example of the Effect of Case Selection

I present two empirical examples of how case selection affects the validity of proxies for severity. In the first example, I have data for all arrests, including arrests that did not lead to an indictment. This example illustrates how the severity of the crime may not be a good predictor for the expected prison sentence received by a defendant. In the second example, I do not have data on declinations. However, I show that the natural proxy for case severity is negatively related to the average prison sentence received by defendants. One explanation for this finding is that excluding cases that are declined corrupts the proxy for severity.

2.1. State Cases

I examine data on 450 suspects arrested for assault in Portland, Oregon, in 1982 and 1986 (Abt Associates 1987; Boland 2004) because this data set includes all felony arrests, including all cases declined for prosecution. I divide assaults into first-, second-, and third-degree assault. First-degree assault requires the physical injury to have been serious, intentional, and caused by a dangerous weapon, while second- and third-degree assault require, respectively, two or one of these conditions to be present.

Figure 1 shows the average prison sentences (in months) and trial rates for first-, second-, and third-degree assault. If more severe cases go to trial, then first-degree assault would be most likely to be tried, while in our data the reverse is true. One explanation for this finding is that there are other factors that affect the likelihood of a trial and that these factors are correlated with the severity of the offense and the selection of cases to prosecute. Consistent with this explanation, when I examine all suspects, individuals arrested for third-degree assault receive the highest average prison sentences, while when I examine suspects in cases that are not declined, individuals arrested for first-degree assault receive the highest average prison sentence.[7]

2.2. Federal Cases

Next, I examine federal drug trafficking cases for the years 1994–2008 (U.S. Department of Justice 2009). I restrict my sample to observations for 8,587 indicted defendants for which I have information about the weight of the heroin involved in the case. The U.S. Federal Sentencing Guidelines made the weight of the drug one of the most important factors in determining the prison sentence. For this reason, in their study of the determinants of trial rates, Boylan and Long (2005) take the weight of the drug as the main proxy for the severity in the case. Figure 2 displays actual prison sentences as a function of the weight of the heroin. For better readability, the graph was smoothed using locally weighted regression. Prison sentences are shorter when the weight of the heroin is greater.[8] This finding can be explained by the theoretical results that I present next, namely, that average severity in truncated samples of indicted defendants can be negatively correlated with average severity in nontruncated samples of suspects.

3. Model

I model two important sample selection problems. First, cases that go to trial are not representative of the set of litigated cases. In 2004, 96 percent of defen-

---

[7] The differences in prison sentences for first-, second-, and third-degree manslaughter are statistically insignificant. However, in both of my samples, the differences in the trial rates between second-degree and first-degree assault cases are statistically significant at the 10 percent confidence level.

[8] This is not the case for all drugs. For marijuana and crack cocaine, the prison sentences are strongly increasing in the weight of the drug.
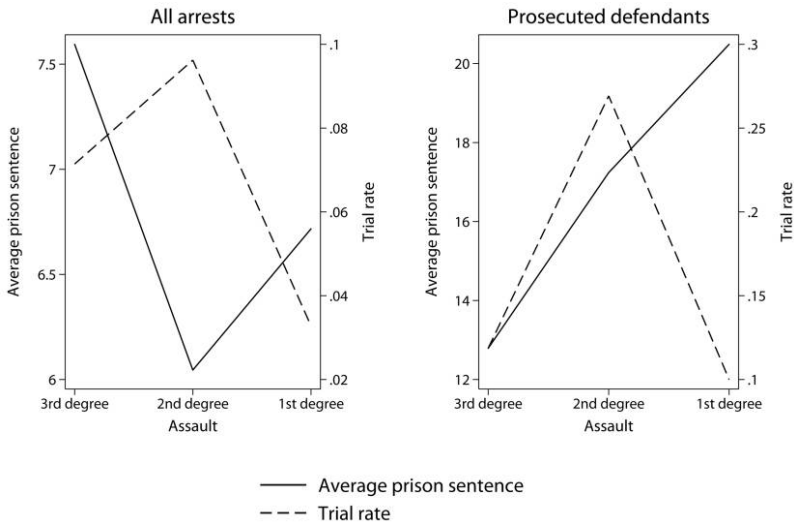
Figure 1. Average prison sentences and average trial rates for all suspects (*left*) and excluding suspects for whom charges were declined (*right*).

dants convicted of felony offenses pleaded guilty to their charges (U.S. Department of Justice 2006). Thus, prison sentences at trial are endogenous and cannot proxy for offense severity. Second, indicted defendants are not representative of the set of suspects referred to prosecutors. In 2004, prosecutors declined to prosecute 21.5 percent of suspects (U.S. Department of Justice 2006). Thus, the observed characteristics of the defendant are endogenous and cannot proxy for offense severity.

I present theoretical results that characterize the relation between sample selection, observed characteristics of the suspect, and outcomes. In the Appendix, the criminal process is described as a game between a defendant and a prosecutor in a manner similar to Reingenaum (1988). The following notation is used: $\tilde{S}$ is the severity of the punishment before indictment, $S = \tilde{S} + w$ is the severity after indictment, δ is the probability that a suspect is guilty, and aS is the expected prison sentence for an innocent defendant who goes to trial. Proposition A1 states that in equilibrium there are scalars $\underline{S}$ and $\overline{S}$ such that

i) if $\tilde{S} < \underline{S}$, the prosecutor declines to indict the suspect,

ii) if $\tilde{S} > \underline{S}$ and $S < \overline{S}$, the prosecutor indicts the suspect, the expected prison sentence for the suspect is aS, and the probability of a trial is 0, and

iii) if $\tilde{S} > \underline{S}$ and $S > \overline{S}$, the prosecutor indicts the suspect, the expected prison sentence for the suspect is $\delta S + (1 - \delta)aS$, and the probability of a trial is $1 - \delta$.

The first consequence of proposition A1 is that exclusion restrictions are unlikely to be valid if I simultaneously estimate the prison sentence received by
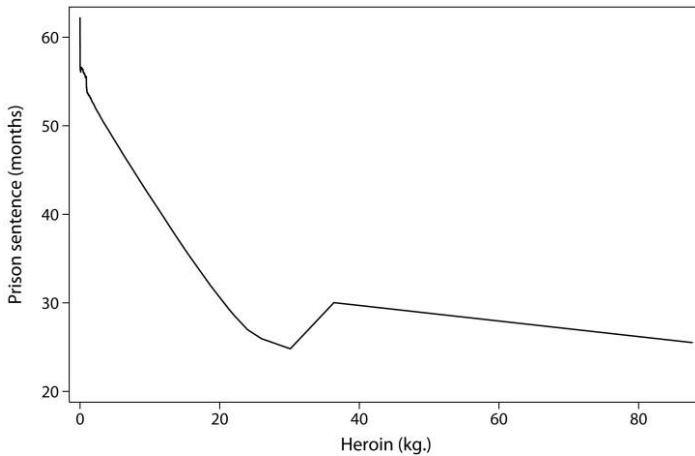
Figure 2. Prison sentences and heroin weight

a suspect and whether the suspect is tried. Specifically, any proxy for case severity affects both the likelihood of trial and the prison sentence.[9]

The second consequence of proposition A1 is that individuals accused of more severe offenses are more likely to be indicted. Thus, the observed severity among indicted defendants is endogenous. Specifically, suppose the severity of the punishment before indictment, $\tilde{S}$, can be decomposed into two independent components: an observed component o and an unobserved component u. Then, individuals are indicted if $\tilde{S} = o + u > \underline{S}$ or, alternatively, if $u > \underline{S} - o$. Therefore, unobserved severity is decreasing in observed severity.

In Boylan and Long's (2005) study of the determinants of trial rates, observed severity is the weight and type of the drug. However, Section 2.2 demonstrates that in federal heroin cases, the weight of the drug is negatively correlated with the expected prison sentence. Thus, among indicted defendants, severity $\tilde{S} = o + u$ is negatively correlated with observed severity o. Thus, an additional distributional assumption is needed to be able to find the relation between the severity of the offense and trial rates.

Proposition 1 characterizes conditions under which observed severity is positively correlated with trial rates and prison sentences.

Proposition 1
i) Suppose that severity before indictment is $\tilde{S} = o + u$, severity after indictment is $S = \tilde{S} + w$, and u and w are independent. Then, in a subgame perfect

---

[9] The same is true for trial costs that affect $\overline{S}$ and thus affect both the likelihood of trial and the prison sentence.

equilibrium, cases with a higher observed severity (o) are more likely to go to trial and have longer average prison sentences.

ii) Suppose that severity before indictment is $\tilde{S} = o + u$, severity after indictment is $S = \tilde{S} + w$, and u and w are independent. Then, in a subgame perfect equilibrium, if the distribution of unobserved severity (u) satisfies the monotone likelihood ratio property, indicted defendants with a higher observed severity (o) have longer average prison sentences and higher trial rates.

iii) Suppose that severity before indictment is $\tilde{S} = o + u$, severity after indictment is $S = \tilde{S} + w$, and u and w are independent. Then, in a subgame perfect equilibrium, for some distributions of unobserved severity (u), indicted defendants with higher observed severity (o) have shorter average prison sentences.

iv) Suppose that severity before indictment is $\tilde{S} = o + u$, severity after indictment is $S = \tilde{S} + w$, and u and w are independent. Then, in a subgame perfect equilibrium, indicted defendants with higher postindictment severity (w) have longer average prison sentences and higher trial rates, but they need not have longer average prison sentences at trial.

Proof

i) First, I fix w. Note that $\tilde{S}$ satisfies condition a in lemma A1. Thus, the probability of trial conditional on w and o is increasing in o, and consequently the probability of trial conditional on o is also increasing in o. Let $\phi(\tilde{S}, w)$ be the expected prison sentence; namely,

$$P(\tilde{S}, w) = \begin{cases} 0 & \text{if } \tilde{S} \leq \underline{S} \\ aS + T^D - A^D & \text{if } \tilde{S} > \underline{S} \quad \text{and} \quad S > \overline{S} \\ \delta(S + T^D - A^D) + (1 - \delta)aS & \text{if } \tilde{S} > \underline{S} \quad \text{and} \quad S > \overline{S}. \end{cases}$$

Since P is increasing in $\tilde{S}$, by lemma A1, the expected prison sentence conditional on o and w is increasing in o. Therefore, the expected prison sentence conditional on o is increasing in o.

ii) By lemmas A3 and A4, the probability density of $\tilde{S}$ conditional on the case not being dismissed satisfies condition a in lemma A1. Thus, I can prove the result the same way that I did in part i.

iii) Suppose that u is distributed according to the Student's t distribution with 2 degrees of freedom; namely, u has a density

$$g(u) = \left(\frac{1}{2 + u^2} \right)^{3/2}.$$

Then,

$$E[\tilde{S} | \tilde{S} > \underline{S}] = o + E[u | u > \underline{S} - o]$$
$$= o + (\underline{S} - o) + \sqrt{2 + \underline{S} - o}$$
$$= \underline{S} + \sqrt{2 + \underline{S} - o},$$

and thus the expected prison sentence conditional on the case not being declined is decreasing in observed severity o.

iv) By lemma A1, the trial rates and the expected prison sentence are increasing in w. To simplify the notation, assume that $o = 0$, $\overline{S} - \underline{S} > w$, and u has a density

$$g(u) = \left(\frac{1}{2 + u^2}\right)^{3/2}.$$

Then, the expected prison sentence at trial for indicted cases

$$E[\tilde{S} + w \,|\, \tilde{S} > \underline{S} \quad \text{and} \quad \tilde{S} > \overline{S}] = E[u + w \,|\, u > \overline{S} - w]$$

$$= w + (\overline{S} - w) + \sqrt{2 + \underline{S} - w}$$

is decreasing in w. Q.E.D.

Proposition 1 provides three methods by which to examine whether criminal suspects who face more severe punishment are more likely to go to trial. First, I can assume that severity is independently distributed from other characteristics that affect the likelihood of a trial and examine the set of all suspects (proposition 1.i). Second, I can assume that severity is independently distributed from other characteristics that affect the likelihood of a trial and that the distribution of unobserved severity satisfies the monotone likelihood ratio property (proposition 1.ii).[10] Third, I can use an instrument for severity w that is uncorrelated with sample selection (proposition 1.iv). I follow the third method and use the random assignment of defendants to judges as my instrument w.

4. Empirical Framework

I estimate two sets of regressions. In the first set of regressions, I seek to determine whether defendants are more likely to go to trial when they are assigned to harsher judges. Thus, the unit of observation is the defendant, and I proxy the harshness of judge j in year t by the average prison sentence for all defendants assigned to judge j in year t. When I regress the likelihood of a plea on the proxy for severity, I expect a downward bias in my estimates since the severity of the judge is observed with error. I can, however, obtain consistent estimates by two-stage least squares regression, where the instrument for severity is the average prison sentence for defendants assigned to judge j in other years. I provide below conditions for which my procedure gives consistent estimates; namely, within a particular courthouse and year, defendants must be randomly assigned to judges.

In the second set of regressions, I examine whether defendants plead to longer prison sentences when they are assigned to judges who are harsher at trial. Let

---

[10] Roosen and Hennessy (2004) provide a statistical test for whether a distribution satisfies the monotone likelihood ratio property.

Y_i^1 be the prison sentence that defendant i receives if the case goes to trial, and let Y_i^0 be the prison sentence that defendant i receives in a plea. Further, let $j(i)$ be the judge assigned to defendant i. I seek to estimate

$$Y_i^0 = \gamma Y_i^1 + u_i.$$

Since both trial and plea sentences cannot be observed for a particular defendant, the average prison sentence for cases that go to trial assigned to judge $j = j(i)$ in year t, $\overline{Y}_{jt}^1$ is my proxy for Y_i^1. Similarly, the average prison sentence for plea cases assigned to judge $j = j(i)$ in year t, $\overline{Y}_{jt}^0$, is my proxy for Y_i^0. Let $\hat{\gamma}$ be the coefficient obtained when regressing $\overline{Y}_{jt}^0$ on $\overline{Y}_{jt}^1$. The coefficient $\hat{\gamma}$ does not provide consistent estimates for γ because the trial sentences affect which cases are concluded by trial. However, I am able to bound γ; specifically, I show that $\hat{\gamma} < \gamma$.

4.1. Likelihood of a Plea

For simplicity, I consider a linearized version of my model, examine only one courthouse, and assume that each judge is assigned to n defendants per year. For a defendant i, $t(i)$ is the year, $j(i)$ is the assigned judge, $w_{j(i)}$ is the harshness of the judge, and S_i is severity, where for some vector γ,

$$S_i = w_{j(i)} + \gamma o_i + u_i.$$

Let $\eta_{j(i)}$ be nonharshness judge-specific characteristics that affect the likelihood of a trial. Then, the probability of trial and the prison sentence are

$$P_i = \alpha S_i + \eta_{j(i)} + \varepsilon_i$$

and

$$Y_i = \beta S_i + \delta_i,$$

where ε_i and δ_i are independently and identically distributed zero mean variables.

By construction, $w_{j(i)}$ is orthogonal to $\eta_{j(i)}$. I assume that defendants are randomly assigned to judges; specifically, $[w_{j(i)}, \eta_{j(i)}]$ is orthogonal to $(o_i, u_i, \varepsilon_i, \delta_i)$, and $(o_i, u_i, \varepsilon_i, \delta_i)$ is orthogonal to $(o_l, u_l, \varepsilon_l, \delta_l)$ when $j(i) = j(l)$.

Since I do not observe $w_{j(i)}$, I proxy $w_{j(i)}$ with the average prison sentence for judge $j(i)$ in year $t(i)$. For any variable X_i, let $\overline{X}_i$ be the average value of X for the other defendants assigned to judge $j(i)$ in year $t(i)$; namely,

$$\overline{X}_i = \frac{\sum_{k \in \mathcal{I}_i} X_k}{\sum_{k \in \mathcal{I}_i} 1},$$

where

$$\mathcal{I}_i = \{k \neq i : j(k) = j(i) \quad \text{and} \quad t(k)t(i)\}.$$

Then the average prison sentence for the other cases assigned to judge $j(i)$ is $\overline{Y}_i$, where

$$\overline{Y}_i = \beta(w_{j(i)} + \boldsymbol{\gamma}\overline{o}_i + \overline{u}_i) + \boldsymbol{\gamma}\overline{\delta}_i,$$

or, alternatively,

$$S_i = \frac{1}{\beta}\overline{Y}_i - \boldsymbol{\gamma}\overline{o}_i - \overline{u}_i - \frac{1}{\beta}\overline{\delta}_i + \boldsymbol{\gamma}o_i + u_i.$$

Thus, when I use $\overline{Y}_i$ as a proxy for S_i, I obtain

$$P_i = \frac{\alpha}{\beta}\overline{Y}_i + v_i, \tag{1}$$

where

$$v_i = -\alpha\boldsymbol{\gamma}\overline{o}_i - \alpha\overline{u}_i - \frac{\alpha}{\beta}\overline{\delta}_i + \alpha\boldsymbol{\gamma}o_i + \alpha u_i + \varepsilon_i + \eta_{j(i)}.$$

Denote by Σ_o the covariance matrix for o, σ_u^2 and σ_δ^2 the variance of u and δ, and n the number of defendants assigned to judge $j(i)$. Then,

$$\text{cov}(\overline{Y}_i, v_i) = \frac{\alpha}{n}\boldsymbol{\gamma}\Sigma_o\boldsymbol{\gamma}' + \frac{\alpha}{n}\sigma_u^2 + \frac{\alpha}{\beta(n-1)}\sigma_\delta^2.$$

In general, I expect great variation in unobserved severity across defendants; namely, I expect σ_u^2 to be very large. So, even for relatively large values of n, $\text{cov}(\overline{Y}_i, v_i) \neq 0$, and thus I expect the ordinary least squares estimate of equation (1) to be inconsistent.

Instrumental variables estimation can be used to remedy the inconsistency arising from the error in variables, where the instrument for the average prison sentence for defendants assigned to a particular judge is its value in the previous year. Specifically, for any variable X_i, $\overline{X}_{i,t(i)-1}$ is the average value of X for cases assigned to judge $j(i)$ in year $t(i) - 1$; namely,

$$\overline{X}_{i,t(i)-1} = \frac{\sum_{k \in \mathcal{I}_{i,t(i)-1}} X_k}{\sum_{k \in \mathcal{I}_{i,t(i)-1}} 1},$$

where

$$\mathcal{I} = \{k : j(k) = j(i) \quad \text{and} \quad t(k) = t(i) - 1\}.$$

The instrument for judge severity is $\overline{Y}_{i,t(i)-1}$, where

$$\overline{Y}_{i,t(i)-1} = (\beta w_{j(i),t-1} + \boldsymbol{\gamma}\overline{o}_{i,t(i)-1} + \overline{u}_{i,t(i)-1}) + \overline{\delta}_{i,t(i)-1}$$

and

$$\text{cov}(Y_{i,t(i)-1}, v_i) = 0.$$

4.2. Plea Sentences

For simplicity, I consider the case in which there is only one courthouse, 1 year, and K judges, and each judge is assigned to n defendants per year. This

includes the case in which there are multiple years and courthouses if I interpret each variable as the deviation from the courthouse-year mean. So, for instance, with multiple courthouses and years, the prison sentence of defendant i is interpreted as the difference between the actual sentence and the mean prison sentence in the same courthouse and year.

I write severity as

$$S_i = w_{j(i)} + u_i,$$

where $E[w_j] = 0$, $E[u_i|w] = 0$, and $E[u_i^2|w] = \sigma^2$. Let $T_i = 1$ if defendant i is tried, and let $T_i = 0$ if defendant i pleads. The probability that defendant i is tried is

$$P_i = \alpha S_i$$

and $\alpha > 0$. Thus,

$$E[u_i P_i | j(i) = j] = \alpha\sigma^2 > 0,$$

$$E[u_i(1 - P_i)|j(i) = j] = -\alpha\sigma^2 < 0,$$

$$\begin{aligned} E[u_i w_j P_i | j(i) = j] &= E[E[u_i P_i|w, j(i) = j]w_j] \\ &= E[E[u_i P_i|w, j(i) = j]w_j] \\ &= \alpha\sigma^2 E[w_j] = 0, \end{aligned}$$

and

$$\begin{aligned} E[u_i w_j(1 - P_i)|j(i) = j] &= E[E[u_i(1 - P_i)|w, j(i) = j]w_j] \\ &= -\alpha\sigma^2 E[w_j] = 0. \end{aligned}$$

For any variable X_i, let $\overline{X}_j^T$ be the average value of X for the cases assigned to judge j with $T(i) = T$; namely,

$$\overline{X}_j^T = \frac{\sum_{i \in \mathcal{I}_j^T} X_i}{\sum_{i \in \mathcal{I}_j^T} 1},$$

where

$$\mathcal{I}_j^T = \{i : j(i) = j \quad \text{and} \quad T(i) = T\}.$$

Then,

$$E[u_j^1] = E[u_i | T_i = 1 \text{ and } j(i) = j]$$

$$= \frac{E[u_i P_i | j(i) = j]}{E[P_i | j(i) = j]} > 0,$$

$$E[u_j^0] = E[u_i | T_i = 0 \quad \text{and} \quad j(i) = j]$$

$$= \frac{E[u_i(1 - P_i) | j(i) = j]}{E[(1 - P_i) | j(i) = j]} < 0,$$

$$E[w_j u_j^1] = E[u_i w_j | T_i = 1 \quad \text{and} \quad j(i) = j]$$

$$= \frac{E[u_i w_j P_i | j(i) = j]}{E[P_i | j(i) = j]} = 0,$$

and

$$E[w_j u_j^0] = E[u_i w_j | T_i = 0 \quad \text{and} \quad j(i) = j]$$

$$= \frac{E[u_i w_j(1 - P_i) | j(i) = j]}{E[1 - P_i | j(i) = j]} = 0.$$

I assume that the prison sentence received by defendant i is

$$Y_i = \begin{cases} S_i & \text{if } T_i = 1 \\ \gamma S_i & \text{if } T_i = 0. \end{cases}$$

Note that

$$E[Y_j^1]^2 = E[(w_j)^2 + 2w_j u_j^1 + (u_j^1)^2]$$

and

$$E[Y_j^0 Y_j^1] = \gamma E[(w_j)^2 + w_j(u_j^0 + u_j^1) + u_{jt}^0 u_j^1] < \gamma E[Y_j^1]^2,$$

since $E[w_j u_j^1] = E[w_j u_j^0] = 0$ and $E[u_j^0 u_j^1] = E[u_j^0]E[u_j^1] < E[(u_{jt}^1)^2]$. Thus, for large values of n,

$$\hat{\gamma} = \frac{\sum_j Y_j^0 Y_j^1}{\sum_{jt} (Y_j^1)^2} < \gamma.$$

5. Judge Assignments

5.1. Sample

All felony cases in U.S. district courts are assigned to judges appointed for life by the president. For each case, the Transactional Records Access Clearing-

house provides the name of the judge, investigative agency, offense, conviction, and prison sentence.[11] I examined defendants for the fiscal years 1992–2002 in 32 federal judicial districts.[12] I restrict my analysis to drug and violent crimes because they present a large and relatively homogenous type of case.[13]

In most federal courts, a court clerk assigns cases randomly to judges. Random assignment is implemented to prevent judge shopping and to ensure fairness. According to the U.S. House Committee on the Judiciary (U.S. Congress 1998, p. 67), "all federal districts have rules that require the random assignment of cases." I identified several examples of districts with multiple courthouses in which local rules specify random assignment within the courthouse rather than at the district level.[14] For this reason, I assume that cases are randomly assigned to judges within a courthouse instead of within a district (the same assumption was made by Hofer, Blackwell, and Ruback [1999]).

Because the random assignment is important for my analysis, for each courthouse I estimate a multinomial logit model in which the dependent variables are characteristics of the case (for example, involves multiple defendants or was investigated by the Drug Enforcement Administration), while the explanatory variables are the judge assigned to the case and year fixed effects. I exclude from my analysis courthouses for which the likelihood ratio test rejects the hypothesis that the characteristics of the case have no effect on the identity of the judge.[15] To ensure convergence of the maximum likelihood algorithm, I do not include courthouses with fewer than 300 drug defendants and judges with fewer than 25 drug defendants. Even though the overall sample is large, there are few observations for a particular judge-year-offense. For this reason, I use the small-sample distribution of the likelihood ratio test computed with a permutation

---

[11] See Transactional Records Access Clearinghouse, TRACfed: A Unique Source of Authoritative Information about the Federal Government's Enforcement (http://tracfed.syr.edu/).

[12] I examined the following districts: middle Alabama, northern Alabama, southern Alabama, Alaska, Arizona, central California, southern California, eastern California, northern California, Colorado, Connecticut, District of Columbia, Delaware, northern Florida, southern Florida, middle Georgia, Hawaii, Idaho, southern Illinois, northern Indiana, southern Indiana, Kansas, eastern Kentucky, western Louisiana, northern Mississippi, southern Mississippi, eastern New York, southern New York, Utah, northern West Virginia, western Wisconsin, and Wyoming.

[13] Violent crimes include bank robberies, civil rights and racial violence, domestic violence, and weapons violations. The other main types of criminal cases in federal courts are immigration and white-collar crimes.

[14] The U.S. district courts include the U.S. District Court for the District of Arizona (2008), U.S. District Court for the Central District of California (2008), U.S. District Court for the Northern District of Florida (2008), U.S. District Court for the Northern District of Illinois (2008), U.S. District Court for the Eastern District of Kentucky (2008), U.S. District Court for the Eastern District of Michigan (2008), U.S. District Court for the Eastern District of New York (2008), U.S. District Court for the Eastern District of Pennsylvania (2008), U.S. District Court for the Western District of Tennessee (2008), U.S. District Court for the Northern District of Texas (1978), U.S. District Court for the Western District of Texas (2008), and U.S. District Court for the Eastern District of Wisconsin (2008).

[15] Often, in cases with multiple defendants, all defendants are assigned to the same judge. Thus, defendants are not randomly assigned, even though cases may be randomly assigned to judges. I avoid this complication by treating defendants in a multiple-defendant case as one defendant.

sampling method based on selecting 1,000 random samples (the method used in Abrams, Bertrand, and Mullainathan [2011]).

At the 10 percent confidence level, I reject the random-assignment hypothesis in the following courthouses: Birmingham (Alabama), East St. Louis (Illinois), Indianapolis (Indiana), Montgomery (Alabama), Salt Lake City (Utah), Tallahassee (Florida), and Tucson (Arizona). Thus, the remainder of the analysis excludes these courthouses and courthouses with fewer than 300 defendants.

5.2. Summary Statistics and Results

My sample consists of drug and violent offenders in the courthouses that randomly assign defendants to judges, namely, the courthouses listed in Table 1 for which $p > .1$. Table 2 provides summary statistics for the variables used in the regressions. Of the 73,155 cases, 83 percent are concluded through a plea agreement, 9 percent are concluded through trial, and 9 percent are dismissed. To reduce the influence of outliers, I top coded prison sentences at 360 months. Consistent with the hypothesis that more severe cases go to trial, the average prison sentence for cases that go to trial is much higher than for cases concluded by plea (141 versus 55 months). There is also great variability in defendants' sentences (SD = 72) and in average prison sentences for a particular judge and year (SD = 28).

Except for the inclusion of courthouse-year fixed effects, I consider the same model as in Section 4.1; namely, the unit of observation is a defendant i assigned to judge j in courthouse c and year t, and the probability of trial (in percentages) is as follows:

$$\text{Probability of Trial}_{ict} = \alpha \text{Judge Severity}_{jt} + \psi_{ct} + \nu_{it},$$

where the primary explanatory variable, Judge Severity, is the average prison sentence (in months) for the other cases assigned to a judge in a particular year, ψ_{ct} are courthouse-year dummy variables, and ν_{it} is the error term.

I expect downward bias in the ordinary least squares estimates of $\hat{\alpha}$ because my proxy for severity is measured with error. For this reason, I also estimate the regressions by the two-stage least squares method with the following instruments: the average prison sentence for cases assigned to judge j in other years and an indicator for whether judge j has prosecutorial experience (see the discussion in Section 4.1). I expect the observations to be correlated if there are other judge-specific factors that affect the probability of trial but are not accounted for by severity. To account for this possibility, I cluster standard errors at the judge level.

To determine whether my model is properly specified, I regress the prison sentence received by a defendant on judge severity. If the model is correctly specified, the regression results should imply that a defendant receives a sentence with an additional month in prison when assigned to a judge whose sentences are 1 month longer. The results are summarized in the top of Table 3. The ordinary least squares coefficient for judge severity is statistically insignificant.

Table 1
Random Assignment of Cases to Judges (*p*-Values)

Courthouse	Large Sample	Simulated	Sample Size	Judges
Anchorage	.26	.28	214	3
Biloxi	.36	.37	276	2
Birmingham	.03	.05	982	8
Brooklyn	.33	.35	4,026	15
Cheyenne	.65	.66	168	2
Denver	.55	.55	1,162	7
East St. Louis	.03	.04	572	6
Fort Lauderdale	.18	.17	1,110	5
Fresno	.29	.31	248	3
Hammond	.69	.71	666	2
Honolulu	.55	.57	846	5
Indianapolis	.03	.03	612	5
Jackson	.38	.41	439	3
Kansas City	.85	.86	138	4
Lexington	.20	.21	282	2
Los Angeles	.52	.58	2,307	28
Macon	.12	.13	542	3
Miami	.12	.14	4,952	17
Mobile	.86	.88	1,067	3
Montgomery	.02	.03	128	3
New York	.21	.28	2,201	30
Pensacola	.46	.46	704	2
Phoenix	.75	.80	1,180	8
Sacramento	.57	.59	473	5
Salt Lake City	.00	.00	742	8
San Diego	.51	.53	8,065	12
San Francisco	.75	.85	175	8
Santa Ana	.61	.62	276	3
South Bend	.29	.36	60	2
Tallahassee	.03	.03	464	3
Tucson	.10	.10	3,564	5
Washington	.55	.68	1,522	17
West Palm Beach	.57	.63	692	3
Wichita	.55	.57	398	3

Note. For each courthouse, a multinomial logit regression estimates the probability that a drug case involves multiple defendants, as a function of judge indicators. The *p*-values shown are for a test that the judge indicators are zero under the large-sample distribution and a small-sample distribution obtained according to 1,000 random samples from the data. Similar values are obtained when the dependent variable is the probability that the case is investigated by the Drug Enforcement Agency.

Thus, it is important to control for the error-in-variables problem. When I reestimate the model by two-stage least squares regression, I find that defendants receive prison sentences that are 1.04 months longer when they are assigned to judges who give sentences that are harsher by 1 month. I cannot reject the hypothesis that the two-stage least squares coefficient is one, and thus I cannot reject the hypothesis that my model is correctly specified.[16]

[16] The *F*-test has a value of 1.1 and a *p*-value of .3.

Table 2
Summary Statistics

Variable	Mean	SD	Min	Max	N
Prison sentence (months)	58.20	72.1	0	360	71,170
If plea (months)	54.88	59.4	0	360	58,741
If trial (months)	140.99	123.6	0	360	6,359
Iff trial conviction (months)	169.74	116.2	0	360	5,282
Judge-year average prison sentence (Judge Severity)	57.55	29.7	0	360	73,382
Judge-(other years) average prison sentence	58.42	25.5	19	136	73,334
Judge with prosecutorial experience	.53	.5	0	1	73,408
Plea (100 if plea, 0 otherwise)	83.26	37.3	0	100	73,155
Trial (100 if trial, 0 otherwise)	8.78	28.3	0	100	73,155
Judges' average trial rate (%)	9.33	5.7	0	43	73,408

Note. Statistics are for drug and violent offenders for 1991–2002 for courthouses that randomly assigned defendants to judges.

The coefficients of the first-stage regression provide further evidence that the empirical model is correctly specified. Judge severity is persistent: judges whose sentences are harsher by 1 month in other years give sentences that are harsher by .2 month in the current year. Further, judges with prosecutorial experience give sentences that are harsher by 3.3 months. For specifications with one endogenous regressor, the rule of thumb is that the *F*-statistic should be at least 10 for weak identification not to be considered a problem (Staiger and Stock 1997). The *F*-statistics here are just above this threshold (see Table 3). I also estimate all regressions by limited-information maximum-likelihood instrumental variables regression, since Stock and Yogo (2005) find this estimator to be far superior to two-stage least squares regression with weak instruments. I obtain the same results using both methods, and thus my results do not seem to be subject to a weak-identification bias.

The dependent variable for the regressions at the bottom of Table 3 is the probability that a defendant goes to trial. Perhaps because of measurement error, the ordinary least squares coefficient of judge severity is close to zero. The two-stage least squares estimates imply that a 1-month increase in sentence severity increases the probability of a trial by .106 percentage point. I obtain similar but more precise estimates when I rerun the regression with additional controls for judge and offense.[17]

Prosecutorial experience could affect the likelihood of a trial independent of its effect on average prison sentences. If this was the case, I would obtain different coefficient estimates depending on whether I include prosecutorial experience as an instrument. A robust test of overidentifying restrictions allows me to determine whether judge severity is a valid instrument (Wooldridge 2002, p. 138). In all regressions, I fail to reject the null hypothesis of valid overidentifying

---

[17] Biographical information for the judges is from Federal Judicial Center, History of the Federal Judiciary: Biographical Directory of Federal Judges, 1789–Present (http://www..fjc.gov/history/home.nsf/judges.html).

Table 3
Estimates of the Effect of Judge Assignments on Trials

	Judge Severity	R^2	N	First-Stage F-Statistic	Overridentifying χ^2
Effect of judge severity on prison sentences (months):					
OLS	−.044	.136	71,147		
	(.051)				
2SLS	1.039**	.092	71,074	10.7**	.126
	(.037)				
Effect of judge severity on trial rates (%):					
OLS	−.024*	.052	73,129		
	(.012)				
2SLS	.106[+]	.048	73,065	10.9**	.004
	(.058)				
2SLS with judge controls	.138*	.046	73,065	11.2**	.099
	(.056)				
2SLS with offense controls	.116*	.061	73,065	11.2**	.005
	(.055)				
Logit: marginal effect	.122**		70,689		
	(.040)				

Note. The sample consists of drug and violent offenders who were randomly assigned to judges. The instruments for the two-stage least squares (2SLS) estimates are the average prison sentence for cases assigned to the same judge in other years and an indicator variable for judge prosecutorial experience. Additional judge controls are black, female, Baptist, and appointed by a Republican president. Offense controls are offense-courthouse dummies and whether the case involves multiple defendants. In the logit regression, judge severity is the average prison sentence for other cases assigned to the same judge in all years. All regressions include courthouse-year fixed effects, and standard errors are clustered at the judge level. OLS = ordinary least squares.
  [+] Significant at the 10% level.
  * Significant at the 5% level.
  ** Significant at the 1% level.

restrictions (see Table 3). This supports the assumption that prosecutorial ex-perience is a valid instrument for judge severity.

I estimate the probability of trial with a linear probability model because linear models are more robust to misspecification. For instance, ordinary least squares estimates are consistent in the presence of omitted variables that are uncorrelated with the endogenous variables, while probit estimates are inconsistent (Yatchew and Griliches 1985). Further, when facing measurement error in the regressor, linear two-stage least squares estimates are consistent, while nonlinear two-stage least squares estimates are inconsistent (Hausman 2001). Nonetheless, to provide evidence of the robustness of my results, I estimate the probability of a trial using a conditional logit regression.[18] To reduce the bias due to measurement error, I define judge severity as the average prison sentence for all other cases assigned to the judge in all years. With this proxy for severity, I find that a 1-

---

[18] Because of the courthouse-year fixed effects, I use conditional logit regression rather than logit or probit regression.

Table 4

Effect of Average Trial Sentences on Average Plea Sentences

Estimation	Average Trial Sentences	R^2	N
OLS, all observations	.028**	.661	1,270
	(.008)		
OLS, more than five trials	.059**	.862	377
	(.019)		

Note. The sample consists of drug and violent offenders who were randomly assigned to judges. In each regression, the unit of observation is the judge-year, the regressor is the average prison sentence for cases that go to trial for a particular judge and year, and the dependent variable is the average prison sentence for cases that are plead. All regressions include courthouse-year fixed effects, and standard errors are clustered at the judge level. OLS = ordinary least squares.

** Significant at the 1% level.

month increase in severity raises the probability of a trial by .122 percentage point.

My measure of severity is the average prison sentence for all cases. This definition is justified, since I expect longer plea sentences for defendants assigned to judges with longer expected trial sentences. Specifically, suppose a defendant assigned to a harsh judge expects a sentence with 1 additional month in prison if the case goes to trial. Then, the defendant agrees to a plea bargain with $1 > \gamma > 0$ additional months in prison. Table 4 provides estimates of γ ($\hat{\gamma}$) obtained by regressing the average prison sentence in a plea on the average prison sentence in a trial, where the averages are computed for each judge and year.

These estimates of γ are inconsistent because defendants are more likely to go to trial when they are assigned to harsher judges. That is, harsher judges try criminals who commit less serious offenses than defendants tried by more lenient judges. As a result, the average prison sentence at trial underestimates the severity of harsher judges. However, I showed in Section 4.2 that $\hat{\gamma}$ provides a lower bound for the causal effect γ. Thus, although I cannot infer the magnitude of γ, I can conclude from my estimates in Table 4 that $\gamma > 0$. Consequently, the average prison sentence over all cases is a valid measure for the severity of judges at trial.

6. Conclusion

Economic models provide explanations for bargaining failures, such as labor negotiations that lead to strikes or litigation that goes to trial. Scholars have been interested in providing empirical support for these models because they are based on strong rationality and informational assumptions. However, Priest and Klein (1984) pointed out that if parties are strategic, we should be cautious in drawing inferences from subsets of negotiations, such as negotiations that lead to bargaining failures or, in the context of litigation, cases that go to trial.

I have provided evidence consistent with strategic behavior by defendants and

prosecutors. Specifically, I find that prison sentences agreed upon in plea bargains depend on the assignment of the judge despite the fact that, in federal criminal cases, judges are not allowed to be involved in settlement negotiation. Further, I provide evidence that trial rates are increasing in the expected prison sentence or, more generally, that the size of the stakes affects the likelihood of a bargaining failure. While this is not the first study to provide this type of evidence, it is the first to be able to instrument for the stakes in a way that is uncorrelated with case selection.

A large literature has discussed the legitimacy of adjudicating criminal cases through plea bargaining rather than trial. Plea bargaining gains legitimacy if settlements largely reflect trial outcomes (Bibas 2004). However, there are several obstacles in providing evidence that settlements reflect trial outcomes. First, for defendants who plea-bargain, we do not know the sentence they would have received if they had gone to trial. Second, we expect cases that are settled to differ from cases that go to trial. Third, the relation between trial outcomes and plea outcomes could occur through reverse causality; specifically, plea outcomes affect the composition of cases that go to trial, which in turn affects the trial prison sentences.

I have provided empirical evidence that the harshness of the judge at trial explains the prison sentence in plea bargains. Thus, my empirical evidence supports the legitimacy of plea bargaining, since plea bargains reflect trial outcomes. One may object that it is unjust that prison sentences agreed to in plea bargains depend on judge assignments. However, as pointed out by Easterbrook (1983), the inequities of plea-bargaining outcomes point to a deficiency in the trial process rather than defects in the plea-bargaining process.

Appendix

Formal Model and Proofs

Bargaining Model

Denote by S the expected prison sentence at trial for a guilty suspect and by aS the expected prison sentence at trial of an innocent suspect, where $a < 1$. Suspects know whether they are guilty, while the prosecutor believes that a defendant is guilty with probability δ. Part of the information on severity is available before the indictment ($\tilde{S}$), while the other part ($w = S - \tilde{S}$) is revealed after the indictment.

Figure A1 describes the sequencing of the game. The prosecutor first decides whether to indict the suspect. If the prosecutor indicts the suspect, both learn additional information about the severity of the punishment (w), and the prosecutor makes a plea offer. If the plea offer is accepted, the case is settled by plea bargain; if it is rejected, the case goes to trial.

The utility function of the suspect is decreasing in the length of the prison

sentence and the costs incurred in a trial or plea agreement. Specifically, if I denote the prison sentence received at trial, the prison sentence received in a plea agreement, the trial costs, and the plea agreement costs by x, y, T^D, and A^D, respectively, the utility function of the suspect is

$$U^D = \begin{cases} -x - T^D & \text{if the case is settled through trial} \\ -y - A^D & \text{if the case is settled through plea agreement} \\ 0 & \text{if the suspect is not indicted.} \end{cases}$$

The utility function of the prosecutor is increasing in the prison sentence that a guilty defendant receives, decreasing in the prison sentence that an innocent defendant receives, and decreasing in the costs incurred in a trial and in a plea agreement. Specifically, if I denote trial and plea costs by $T^P \geq 0$ and $A^P \geq 0$, the utility function for the prosecutor is

$$U^P = \begin{cases} \gamma x - T^P & \text{if the case is settled through trial and the defendant is guilty} \\ \gamma y - A^P & \text{if the case is settled through plea and the defendant is guilty} \\ -\lambda x - T^P & \text{if the case is settled through trial and the defendant is innocent} \\ -\lambda y - A^P & \text{if the case is settled through plea and the defendant is innocent} \\ 0 & \text{if the suspect is not indicted.} \end{cases}$$

To simplify the analysis, I make the following restrictions on the parameters.

Assumption A1. Let $\underline{S} = -(T^D - A^D) + A^P/a[\delta\gamma - (1 - \delta)\lambda]$ and $\overline{S} = (1 - \delta)[T^P - (T^D - A^D)\lambda - A^P]/\delta(1 - a)\gamma$. Then,
i) $\delta > \lambda/(\gamma + \lambda)$,
ii) $T^D \geq A^D \geq 0$, $T^P > A^P > 0$, and
iii) $Ew = 0$ and $|v| < \max\{\overline{S} - \underline{S}, A^P\}$.

Proposition A1. In the subgame perfect Nash equilibrium of the game,
i) if $\tilde{S} < \underline{S}$, the prosecutor declines to indict the suspect,
ii) if $\tilde{S} > \underline{S}$ and $S < \overline{S}$, the expected prison sentence is aS, and the probability of a trial is 0, and
iii) if $\tilde{S} > \underline{S}$ and $S > \overline{S}$, the expected prison sentence is $\delta S + (1 - \delta)aS$, and the probability of a trial is $1 - \delta$.

Proof. I first consider the optimal plea offer. Because of the trial costs, the prosecutor never makes a settlement offer that all defendants reject. Thus, the prosecutor can dismiss a case, make a plea offer that all defendants accept (pooling equilibrium), or make an offer that only guilty defendants accept (separating equilibrium). In a pooling equilibrium, the prosecutor offers a plea of $aS + T^D - A^D$, obtains an expected prison sentence of $aS + (T^D - A^D) + \delta(T^D - A^D)$, and has a payoff of

$$(aS + T^D - A^D)[\delta\gamma - (1 - \delta)\lambda] - A^P.$$

In a separating equilibrium, the prosecutor offers a plea of $S + T^D - A^D$, obtains an expected prison sentence of $\delta S + (1 - \delta)aS$, and has a payoff of
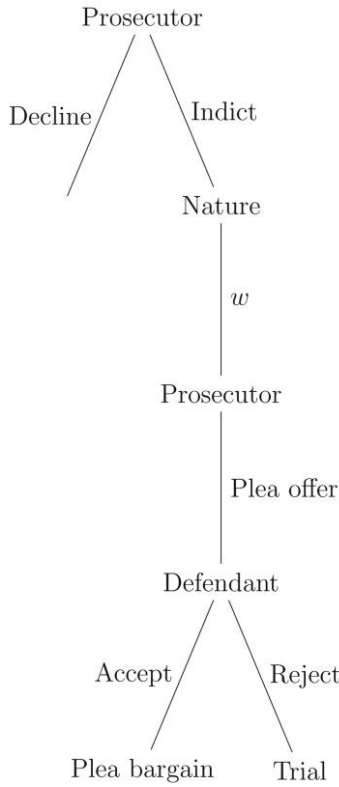
Figure A1. Disposition of criminal suspects

$$\delta[(S + T^{\mathrm{D}} - A^{\mathrm{D}})\gamma - A^{\mathrm{P}}] + (1 - \delta)(-aS\lambda - T^{\mathrm{P}}).$$

Therefore, the separating equilibrium gives the prosecutor a higher payoff when

$$S > \frac{(1 - \delta)[T^{\mathrm{P}} - (T^{\mathrm{D}} - A^{\mathrm{D}})\lambda - A^{\mathrm{P}}]}{\delta(1 - a)\gamma} = \overline{S}.$$

Finally, note that the expected payoff in a separating equilibrium is nonnegative when $S \geq \underline{S}$.

Next, I consider the decision of whether or not to decline prosecution. First, consider the case in which $\tilde{S} \geq S$. Note that the prosecutor can guarantee a nonnegative payoff by indicting and then selecting the pooling equilibrium. Therefore, when $\tilde{S} \geq S$, the prosecutor decides to indict. Second, consider the case in which $\tilde{S} < S$. By assumption A1.iii, if the prosecutor indicts, he or she selects a pooling equilibrium that gives a negative payoff. Therefore, the prosecutor decides not to indict. Q.E.D.

Characterizations of Ordered Distributions

Lemma A1.   Let X be a random variable with density $g(X; o)$. Consider the following conditions:

a)   for all $o' > o$, there is a probability space (Ω, A, P) and random variables $Z'(\omega)$ and $Z(\omega)$ with densities $g(Z; o)$ and $g(Z; o')$ such that $Z' \geq Z$.

b)   For all a, $o' > o$, $\Pr(X \geq a; o') \geq \Pr(X \geq a; o)$.

c)   For all increasing functions ϕ, $o' > o$, $E_{o'}\phi(X) \geq E_{o}\phi(X)$.

Then conditions *a*, *b*, and *c* above are equivalent. Further, conditions *a*, *b*, and *c* hold if g satisfies the monotone likelihood ratio property.

Proof.   See Lehmann (1955). Q.E.D.

Example A1.   In my plea-bargaining model, $Z = X = \tilde{S} = u + o$.

Lemma A2.   Let X be a random variable with density $g(X; o)$. Suppose for all intervals (a, b), $E[X; o', X \in (a, b)] \geq E[X; o, X \in (a, b)]$. Then, for all increasing functions ϕ, $E_{o'}\phi(X) \geq E_{o}\phi(X)$.

Proof.   This follows from the definition of an integral as the limit of step functions. Q.E.D.

Lemma A3.   Let X be a random variable with density $g(X; o)$. The density g satisfies the monotone likelihood ratio property if and only if for all intervals (a, b), $E[X; o', X \in (a, b)] \geq E[X; o, X \in (a, b)]$.

Proof.   See Whitt (1980). Q.E.D.

Definition A1.   A random variable X depending on a real parameter o with density $g(X; o)$ satisfies the monotone likelihood ratio property if for $o' > o$, $X' > X$,

$$\frac{g(X'; o')}{g(X; o)} \geq \frac{g(X; o')}{g(X; o)}.$$

Note that when g is differentiable, the definition of the monotone likelihood ratio property is equivalent to

$$\frac{\partial^2 \log g(X; o)}{\partial X \partial o} \geq 0.$$

Example A2.   Suppose that u has a normal density with mean μ and variance σ^2. Then $\tilde{f}$ satisfies the monotone likelihood ratio property since

$$\frac{\partial^2 \log \tilde{f}(\tilde{S}; o)}{\partial \tilde{S} \partial o} = 2\frac{\mu + o}{\sigma^4} > 0.$$

Lemma A4.   Let $\tilde{X}$ and $X = \tilde{X} + w$ be random variables with density $\tilde{g}$ and g, and assume that X and w are independent. The density h satisfies the monotone likelihood ratio property if $\tilde{g}$ satisfies the monotone likelihood ratio property.

Proof. Let $o' > o$ and $S' > S$. Then, by lemma A1, for any w,

$$E[X, o, w, X \in (a, b)] = E[\tilde{X}, o, \tilde{X} \in (a - w, b - w)] + w$$

$$\geq E[\tilde{X}, o', \tilde{X} \in (a - w, b - w)] + w$$

$$\geq E[X, o', w, X \in (a, b)].$$

Therefore, $E[X, o, X \in (a, b)] \geq E[X, o', X \in (a, b)]$, and by lemma A1, f satisfies the monotone likelihood ratio property. Q.E.D.

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2011. Do Judges Vary in Their Treatment of Race? Working paper. University of Pennsylvania Law School, Philadelphia.

Abt Associates. 1987. Prosecution of Felony Arrests, 1982: Portland, Pregon and Washington, D.C. (computer file). Ann Arbor, Mich.: Inter-university Consortium for Political and Social Research.

Anderson, James M., Jeffrey R. Kling, and Kate Stith. 1999. Measuring Interjudge Sentencing Disparity: Before and after the Federal Sentencing Guidelines. *Journal of Law and Economics* 42(1):271–307.

Ashenfelter, Orley, Theodore Eisenberg, and Stewart J. Schwab. 1995. Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes. *Journal of Legal Studies* 24(2):257–81.

Bibas, Stephanos. 2004. Plea Bargaining Outside the Shadow of Trial. *Harvard Law Review* 117(8):2463–2547.

Bjerk, David. 2005. Making the Crime Fit the Penalty: The Role of Prosecutorial Discretion under Mandatory Minimum Sentencing. *Journal of Law and Economics* 48(2):591–625.

Boland, Barbara. 2004. Prosecution of Felony Arrests, 1986: Indianapolis, Los Angeles, New Orleans, Portland, St. Louis, and Washington, DC (computer file). Ann Arbor, Mich.: Inter-university Consortium for Political and Social Research.

Boylan, Richard T., and Cheryl X. Long. 2005. Salaries, Plea Rates, and the Career Objectives of Federal Prosecutors. *Journal of Law and Economics* 48(2):627–51.

Cohen, Mark A. 1991. Explaining Judicial Behavior or What's "Unconstitutional" about the Sentencing Commission? *Journal of Law, Economics, and Organization* 7:183–99.

Cook, Philip J. 2009. Potential Saving from Abolition of the Death Penalty in North Carolina. *American Law and Economic Review* 11(2):498–529.

Easterbrook, Frank H. 1983. Criminal Procedure as a Market System. *Journal of Legal Studies* 12(2):289–332.

Elder, Harold W. 1989. Trials and Settlements in the Criminal Courts: An Empirical Analysis of Dispositions and Sentencing. *Journal of Legal Studies* 18(1):191–208.

Galanter, Marc. 1986. The Emergence of the Judge as a Mediator in Civil Cases. *Judicature* 69(5):2231–36.

Green, Bruce A. 1999. Why Should Prosecutors Seek Justice? *Fordham Urban Law Journal* 26:607–43.

Hausman, Jerry. 2001. Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives* 15(4):57–67.

Hofer, Paul J., Kevin R. Blackwell, and R. Barry Ruback. 1999. The Effect of the Federal Sentencing Guidelines on Inter-judge Sentencing Disparity. *Journal of Criminal Law and Criminology* 90(1):239–321.

Karle, Theresa Walker, and Thomas Sager. 1991. Are the Federal Sentencing Guidelines Meeting Congressional Goals? Empirical and Case Law Analysis. *Emory Law Journal* 40:393–444.

Kuziemko, Ilyana. 2006. 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* 8(1):116–42.

LaCasse, Chantale, and A. Abigail Payne. 1999. Federal Sentencing Guidelines and Mandatory Minimum Sentences: Do Defendants Bargain in the Shadow of the Judge? *Journal of Law and Economics* 42:245–69.

Landes, William M. 1971. An Economic Analysis of the Courts. *Journal of Law and Economics* 14:61–107.

Lehmann, E. L. 1955. Ordered Families of Distributions. *Annals of Mathematical Statistics* 26(3):399–419.

Nagel, Ilene H., and Stephen J. Schulhofer. 1992. A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices under the Federal Sentencing Guidelines. *Southern California Law Review* 66:501–65.

Priest, George L., and Benjamin Klein. 1984. The Selection of Disputes for Litigation. *Journal of Legal Studies* 13(1):1–55.

Reinganum, Jennifer F. 1988. Plea Bargaining and Prosecutorial Discretion. *American Economic Review* 78(4):713–28.

Roosen, Jutta, and David A. Hennessy. 2004. Testing for the Monotone Likelihood Ratio Assumption. *Journal of Business and Economic Statistics* 22:358–66.

Simons, Michael A. 2009. Prosecutors as Punishment Theorists: Seeking Sentencing Justice. *George Mason Law Review* 16(2):303–56.

Sisk, Gregory C., Michael Heise, and Andrew P. Morriss. 1998. Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning. *New York University Law Review* 73(5):1377–1500.

Snyder, Edward A. 1990. The Effect of Higher Criminal Penalties on Antitrust Enforcement. *Journal of Law and Economics* 33(2):439–62.

Staiger, Douglas, and James H. Stock. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica* 65(3):557–86.

Stock, James H., and Motohiro Yogo. 2005. Testing for Weak Instruments in Linear IV Regression. Pp. 80–108 in *Identification and Inference for Econometric Models—Essays in Honor of Thomas Rothenberg,* edited by Donald W. K. Andrews and James H. Stock. New York: Cambridge University Press.

Stuntz, William J. 1997. The Uneasy Relationship between Criminal Procedure and Criminal Justice. *Yale Law Journal* 107(1):1–76.

U.S. Congress. House. 1998. *Judicial Reform Act of 1998.* 105th Cong., 2d Sess. H. Rep. 478. http://www.gpo.gov/fdsys/pkg/CRPT-105hrpt478/pdf/CRPT-105hrpt478.pdf.

U.S. Department of Justice. Bureau of Justice Statistics. 2006. *Compendium of Federal Justice Statistics, 2004.* Washington, D.C.: Office of Justice Programs.

———. 2009. Defendants in Criminal Cases Terminated during Fiscal Year 1994–2008 (computer file). Washington D.C.: Washington, D.C.: Office of Justice Programs.

U.S. District Court for the Central District of California. 2008. Method of Allocating Criminal and Complex Cases among the Divisions of the Central District Court of California. http://court.cacd.uscourts.gov/Cacd/GenOrders.nsf/0 /62415B130E6F7D8D88257371006F660C/$file/ATTH2YP6.

U.S. District Court for the District of Arizona. 2008. Rules of Practice of the United States District Court for the District of Arizona. http://www.azd.uscourts.gov/azd/courtinfo.nsf/ 40047025CF342B460725750F005BB5EA/$file/2008+Local+Rules.pdf?openelement.

U.S. District Court for the Eastern District of Kentucky. 2012. Case Assignment and Recusal Order. http://www.kyed.uscourts.gov/general-orders/gen12-8.pdf.

U.S. District Court for the Eastern District of Michigan. 2012. Local Rules. http://www .mied.uscourts.gov/rules/localrules/printer/localrulespackage.pdf.

U.S. District Court for the Eastern District of New York. 2012. Local Rules of the United States for the Southern and Eastern Districts of New York. http://www.nyed.uscourts .gov/sites/default/files/local_rules/localrules.pdf.

U.S. District Court for the Eastern District of Pennsylvania. 2010. Local Criminal Rules of the United States District Court for the Eastern District of Pennsylvania. http:// www.paed.uscourts.gov/documents/locrules/criminal/crrules.pdf.

U.S. District Court for the Eastern District of Wisconsin. 2008. General Order Regarding Assignment of Cases to the United States District Court Judge Designated to Hold Court in Green Bay, Wisconsin. http://www.wied.uscourts.gov/dmdocuments/GBCaseAssign.pdf.

U.S. District Court for the Northern District of Florida. 2008. Assignment of Civil and Criminal Cases: Amended Case Assignment Order. http://www.flnd.uscourts.gov/forms/ Administrative%20Orders/20070121_Amended_Case_Assignments.pdf.

U.S. District Court for the Northern District of Illinois. 2008. Local Rule 40.1. https:// www.ilnd.uscourts.gov/LocalRules.aspx.

U.S. District Court for the Northern District of Texas. 1978. Miscellaneous Order No. 22. http://www.txnd.uscourts.gov/pdf/sp_order3/10-27-78.pdf.

U.S. District Court for the Western District of Tennessee. 2008. Local Rules. http:// www.tnwd.uscourts.gov/pdf/content/LocalRules.pdf.

U.S. District Court for the Western District of Texas. 2008. Place of Trial within District. http://www.txwd.uscourts.gov/rules/online/cr/cr-18.pdf.

Vorenberg, James. 1981. Decent Restraint of Prosecutorial Power. *Harvard Law Review* 94(7):1521–73.

Waldfogel, Joel. 1995. The Selection Hypothesis and the Relationship between Trial and Plaintiff Victory. *Journal of Political Economy* 103(2):229–60.

Whitt, Ward. 1980. Uniform Conditional Stochastic Order. *Journal of Applied Probability* 17(1):112–23.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, Mass.: MIT Press.

Yatchew, Adonis, and Zvi Griliches. 1985. Specification Error in Probit Models. *Review of Economics and Statistics* 67(1):134–39.