Summer 2013

Putting the Trial Penalty on Trial

David S. Abrams
University of Pennsylvania Carey Law School

Putting the Trial Penalty on Trial

David S. Abrams[*]

I. INTRODUCTION

The "trial penalty" is a concept widely accepted by all the major actors in the criminal justice system: defendants, prosecutors, defense attorneys, court employees, and judges. The notion is that defendants receive longer sentences at trial than they would have through plea bargain, often substantially longer. The concept is intuitive: longer sentences are necessary in order to induce settlements and without a high settlement rate it would be impossible for courts as currently structured to sustain their immense caseload.

While intuitively appealing, this view of the trial penalty is completely at odds with economic prediction. Since both prosecutors and defendants have the ability to reject unappealing settlements, sentences at trial should be nearly the same as those arrived at through pleas. This is a straightforward application of the "shadow of the law" concept articulated by Mnookin and Kornhauser, as well as others.[1]

This article attempts to answer two questions relating to the trial penalty. Why is belief in its existence so widespread, given

* Assistant Professor of Law, Business Economics, and Public Policy at University of Pennsylvania Law School. Professor Abrams thanks Professor Wesley Oliver and the Duquesne University School of Law for hosting an outstanding conference that led to this piece. I am also grateful to participants at the conference whose questions and comments improve this article. Pearl Li provided excellent research assistance.

1. William M. Landes, *An Economic Analysis of the Courts*, 14 J.L. & ECON. 61 (1971); Robert N. Mnookin & Lewis Kornhauser, *Bargaining in the Shadow of the Law: The Case of Divorce*, 88 YALE L.J. 950 (1979).

that it is at odds with basic economic theory? Which theory does empirical analysis of actual sentencing data support?

I argue that there is a fundamental misunderstanding that is largely responsible for the belief in the trial penalty: the failure to distinguish between conditional and unconditional expected values. I also provide a brief overview of an empirical study that attempts to distinguish between the two theories.[2] The results of that study are surprising and support neither theory—not only is there no evidence for a trial penalty, there appears to be a trial discount!

II. THE SOURCE OF THE CONFUSION: CONDITIONAL VERSUS UNCONDITIONAL EXPECTATIONS

In order to clarify the confusion about the trial penalty, it is first necessary to discuss the mathematical concept of expected value. An expected value is the value of an outcome weighted by its probability. For example, the expected value of a 50-50 coin flip where a head nets the recipient $4 and tails yields $2 will be EV = (0.5)($4) + (0.5)($2) = $3. This is the unconditional expected value.

What does it mean to say that the expected value of this coin flip is $3? Clearly no individual flip will pay off $3; the payment will be either $4 for heads or $2 for tails. The expected value is the amount that will be received *on average* over a large number of trials. The more flips, the closer the average will come to the expected value.[3] For example, 10 flips might yield 6 heads and 4 tails, for an average payoff of $3.20. 100 flips might yield 48 heads and 52 tails, for an average payoff of $2.96. The average approaches the expected value as the number of flips increases.

Of course, sometimes more information can be acquired about events with some randomness in them. For example, consider a game played by choosing a card at random from a deck, where the player is paid according to the suit of the card as follows: club = $1, spade = $2, diamond = $3 and heart = $4. Since the probability of drawing each suit is equally likely at 25%, the expected value of the draw is EV = (0.25)($1) + (0.25)($2) + (0.25)($3) + (0.25)($4) = $2.50. Now consider how this changes with a bit of additional information—in this case, the knowledge that the card drawn is red. There are now only two possible suits that were

2. This paper outlines my findings published in David S. Abrams, *Is Pleading Really a Bargain?*, 8 J. EMPIRICAL LEGAL STUDIES 200 (Dec. 2011).

3. The theorem that states this is called the Law of Large Numbers.

drawn, diamonds and hearts, each with a 50% probability. Thus the *conditional* expected value may be calculated as EV = (0.5)($3) + (0.5)($4) = $3.50. The knowledge that the card drawn is red increased the expected value of the draw substantially.

While the courts are a long way from the card table, they retain a crucial similarity—an element of randomness. No one can tell for certain in advance what a randomly selected jury or a judge will decide for a particular case. Thus, economists treat the question of likely sentencing outcomes in a mathematically identical way to the card draw—as a random variable. This means that we cannot predict the outcome with certainty, but we can say something about average outcomes after a large number of events. Thus for a defendant charged with a crime, we can predict the *unconditional expected sentence* just as we can calculate the unconditional expected value of a card draw. This expected sentence will include a substantial likelihood that there is no prison sentence—due to being found not guilty, dropped prosecution, or other reasons.

Now we may also calculate the *conditional expected sentence*, conditioning on a finding of guilt at trial. This means that the case has proceeded to the guilt/innocence phase and that the defendant has in fact been found guilty. It will not include the large number of cases where this is not the outcome. The conditional expected sentence will necessarily be substantially higher than the unconditional expected sentence because it will exclude many cases that result in zero sentence. Thus, it is crucial to be clear about which type of expected sentence one is concerned with, and this depends on the question at issue.

A defendant who has been found guilty and wishes to have the best guess of the likely sentence he will receive will be interested in the conditional expected sentence. But the defendant who has been offered a plea deal, but has not yet been found guilty should be interested in the unconditional expected sentence when trying to determine whether or not to accept the sentence proffered by the prosecutor. If the defendant is risk-neutral,[4] then he should accept the plea deal if the sentence is lower than the unconditional expected sentence; otherwise he should insist upon a trial, because on average he will receive a shorter sentence by making that choice. Thus the difference between the trial penalty perspective

4. A risk-neutral individual is one for whom the value of a risky event is equal to the expected value. That is, someone who neither likes nor dislikes gambles.

and the shadow of the law perspective comes down to a comparison between the conditional and unconditional expected sentence. The mistake that is frequently made is considering the conditional expected sentence in the context of a defendant with a plea offer, a setting where the unconditional expected sentence is the appropriate one.

The "shadow of the law" concept simply applies the use of expected values to negotiation decisions. [5] Originally applied to divorce settlements, it has become a widely-known theory in the law and economics literature. The essence of the idea is that in a negotiation, both attorneys may use the threat of going to trial if they believe the current bargain is unfair. In the setting of plea bargaining, we assume that the defense attorney is trying to minimize the sentence, the prosecutor is trying to maximize it, and administrative costs for both pleas and trials are small. In this setting, the defense attorney would reject any plea deal with a sentence longer than the expected sentence. Similarly, the prosecutor would reject a deal where the sentence is shorter than the expected sentence. The only solution to these two conditions is that the sentence arrived at via plea is the same as the expected sentence from trial. The name of the theory is due to the fact that while many sentences are achieved through bargaining rather than through direct application of the law, the law still casts a shadow over the negotiation proceedings. I now describe the first empirical test of this theory in the criminal justice setting.

III. Empirical Analysis

To test the shadow of the law and trial penalty theories, it is necessary to evaluate how sentences handed down at trial compare to those arrived at through plea bargain.[6] I obtained data on cases from the early 2000s from Cook County, Illinois, the largest unified criminal court in the United States, which receives the bulk of its cases from the City of Chicago. A simple comparison shows that the average sentence resulting from a plea is 2.4 years, which is 1.1 years longer than that from trial (see Table 1 below).

5. *See supra* note 1.
6. These results were first described in Abrams, *supra* note 2.

Summary Statistics

Variable	Trial	Plea	Difference	t-stat
Case & Offender Characteristics				
Number of charges	2.34	2.11	-0.23	-4.97
Defendant Race (Black = 1)	0.89	0.89	0.00	-0.51
Defendant Sex (Female = 1)	0.14	0.19	0.05	8.96
Defendant Age	29.62	28.91	-0.71	-4.47
Incarceration	0.26	0.59	0.33	45.16
Sentence	1.25	2.36	1.12	32.37
Sentence (non-zero)	2.91	2.44	-0.47	-9.20
Finding of Guilt	0.46	1.00	0.53	196.47
Offense Type				
Drug Possession	0.39	0.46	0.07	8.79
Drug Distribution	0.22	0.15	-0.07	-11.98
Weapons	0.10	0.06	-0.03	-9.22
Theft	0.02	0.07	0.05	12.55
Burglary	0.06	0.06	0.00	0.47
Robbery	0.04	0.04	0.00	1.04
Car Theft	0.03	0.04	0.01	3.48
Assault and Battery	0.05	0.03	-0.02	-7.52

Summary Statistics for 42,552 case-level observations obtained from the Cook County Clerk's Office for cases initiated between 1997 - 2001 and resolved by 2004. Offense characteristics reported for first offense. Data includes 28 judges, each with a minimum of 100 cases. Homicides are excluded due to potential non-random assignment. Additional case selection details are available in Abrams, Bertrand, Mullainathan (2012).

Decomposing the data by type of offense still shows that pleas result in substantially longer sentences on average.

This may seem surprising—that sentence lengths are shorter on average when they result from pleas rather than trials. But it is important to remember that those sentences that do not result in a sentence of incarceration count as zero. If one only compares sentences of incarceration, the average is longer for those from trial. This is exactly the source of the trial penalty confusion. The unconditional expected sentence is longer for plea bargains, but the expected sentence, conditional on having some prison sentence, is longer at trial.

This straightforward comparison, however, does not take into account the self-selection of defendants and prosecutors into plea bargains or trials. Perhaps cases involving less serious crimes are more likely to choose to go to trial, resulting in a low average sentence length that has little to do with the choice of plea versus tri-

al. An ideal study would examine a large sample of otherwise identical cases where defendants are randomly assigned to either take a plea or go to trial. Such a field experiment would not only be impractical, but certainly unethical.

One approach to the challenge of self-selection is a regression analysis that attempts to control for the variables that influence a person's likelihood of pleading versus trial. These may include demographic characteristics of the individuals, the type and severity of the alleged crime, personal characteristics of the attorneys, and others. While preferable to a simple comparison of averages (as above) the method of using control variables is incomplete because one will never have access to all control variables that may influence the decision to plea. The second method attempts to approximate the ideal random assignment and is called the instrumental variables approach. This is a relatively new technique that has become extremely popular among economists since the 1990s. In this method, random assignment to plea or trial is simulated using a proxy, the instrumental variable, with the goal of reducing self-selection bias. Below, I report results from both of these methods. They do not change the central conclusion that there is no evidence for a trial penalty.

Summary statistics of case outcomes separated by case characteristics (Table 1) showed significant differences in outcomes that support the need for controls in the regression. For example, cases that went to trial tended to have 10% more total charges, indicating that these were likely more serious crimes. Cases involving female and younger defendants were more likely to be resolved via plea bargain, as were theft and drug possession cases. Ultimately, the compositions of the cases that went to trial varied substantially from those that were resolved in pleas.

Eight regression analyses were performed, with each controlling for a different combination of variables that could potentially influence plea bargain self-selection. These include the case type, the number of charges, and the year, as well as defendant race, sex, and age. The differences in sentencing length were consistent across the eight regressions, with a standard error of about 0.09 years. The mean difference was 1.15 years, consistent with the finding of 1.12 years in the model without controls. This suggests that even after controlling for self-selection, defendants who enter into plea bargains receive sentences that are on average 14 months longer. Using the same combinations of controls, I also

found that defendants who plead guilty are about twice as likely to be incarcerated as those whose cases went to trial.

I chose judge tenure as the instrument, which should predict likelihood to plea, but in a way not susceptible to selection. Since more experienced judges have a longer history of case adjudications, both attorneys will have a better idea of likely outcome of the case and are thus more likely to agree. This will lead to a greater rate of plea bargains for cases that are assigned to judges with longer tenure. Since cases are randomly assigned to judges this is effectively a randomly determined variable that predicts the likelihood of pleading. Even though the effect may be small, using an instrumental variables regression, one may determine the impact of plea bargaining on sentence length.

To validate the instrument, I analyzed the relationship between the rate of plea bargains and judicial tenure. I found a statistically significant positive correlation between plea rate and judicial tenure. Each additional year on the bench corresponds to a plea rate increase of 4.7 percentage points. The second step was to compare sentence length and judicial tenure. I found a similar positive correlation, but the relatively high standard error indicates that the instrument is relatively weak. Ultimately, the instrumental variables approach yielded a similar conclusion to the ordinary regression: plea bargains result in higher expected sentences than trials.

IV. Discussion

The results from both the regression analysis and the instrumental variables analysis indicate that plea bargains actually result in longer sentences than trials. This conclusion is at odds with the trial penalty concept, which posits that trials generally result in longer sentences. But it is also at odds with the shadow of the law theory, which predicts very similar sentences from trial and plea bargain. This begs the following questions: Why does the shadow of the law prediction appear to be incorrect? And why has the trial penalty theory persisted for so long? I suggest two answers to each of these questions, beginning with the shadow of the law.

First, it is possible that defendants are quite risk-averse and that they prefer a known sentence that is substantially longer than the expected sentence they would receive at trial. This is at odds with most work in the area, which finds that not only are defendants not risk averse, but they may in fact be risk-seeking

(which may help explain why they are in court in the first place). A second potential explanation is that rather than counting fines and sentences of probation as a zero sentence length, they are actually equivalent to non-zero prison sentences. While almost certainly true, this is unlikely to explain much of the large disparity between plea and trial sentences found in this study. Anecdotal evidence suggests that prison sentence is by far the most important driver of defendant decision making.

For the failure of the trial penalty prediction, one explanation may be the availability heuristic, a behavioral phenomenon first studied by Tversky and Kahneman (1973). This phenomenon is essentially the notion that people tend to focus on notable events and ignore more run-of-the-mill ones. While the judges and attorneys who champion the trial penalty have the advantage of personal experience with the incentives of deciding to go trial that academics are further removed from, their memories are also imperfect. The availability heuristic asserts that people are more likely to remember more exceptional events, and are thus apt to overweigh the probability of their occurring again. In our context, law practitioners may deal with dozens of trials that end in short sentences or that ultimately result in no conviction at all, but these more mundane occurrences may take a backseat to memories of especially spectacular trials that resulted in very long sentences. Trials are also more memorable than plea bargains, which are comparatively short and almost never make the news. Laypeople (most importantly, defendants) often never hear about the plea bargains or ordinary trials altogether, so their expectations of going to trial may be even more skewed.

Another explanation for why the trial penalty does not hold is the principal-agent problem inherent whenever a defendant hires (or is assigned) an attorney to represent himself. The incentives of the defendant and the attorney are never perfectly aligned. Defendants want to minimize their sentence length or avoid conviction altogether. Defense attorneys want these things too, of course, but they often have little personal contact with their defendants, which may lead to other incentives coming to the forefront. Their peers are not the defendants, but the prosecutors and judges, with whom they often maintain long-term relationships—providing a potential incentive to avoid creating animosity by going to trial. This distance from their clients is especially important for public defenders, who are involved in a significant por-

tion of the data set. Overworked and underpaid defense attorneys may prefer the brevity of plea bargains to the odyssey of trial.

V. CONCLUSION

The empirical evidence shows little support for the trial penalty, and in fact there appears to be a plea penalty. Expected sentences are at least one year longer in plea bargains than in trials, and incarceration is also about twice as likely to result. These findings remain unchanged after both controlling for observable case characteristics and using judge tenure as an instrumental variable.

One practical implication of this work is that defense attorneys should be very aware of the potential divergence of their incentives from their clients. There are certainly cases in which it is in the defendant's best interest to take a plea deal, but on the average, going to trial should be considered more heavily by both attorneys and defendants. Defense attorneys may wish to consider presenting clients with statistics on the relative outcomes of pleading and going to trial, in order to allow them to make more informed decisions. Greater rotation of court staff may also mitigate the natural desire of attorneys to maintain good collegial relations. By increasing awareness that the trial penalty is a myth, defense attorneys will be able to better inform and serve their clients.