The Economic Perspective on Sentencing

*Joshua B. Fischman**

Although economists have been actively engaged in research on criminal sentencing, the synergies between the two fields are hardly obvious. This Essay considers what economists have to contribute to the study of sentencing. One common explanation—that economists' use of rational choice modeling has applicability to the study of deterrence—does not adequately account for much of the sentencing research that economists are producing.

This Essay considers two alternative explanations. First, empirical research in both fields is predominately observational. Due to practical limits on controlled experimentation, economists have developed a variety of tools for making causal inferences from observational data, many of which have also proved useful in the study of criminal sentencing. Second, both fields are policy-oriented social sciences. Methods developed by economists for relating data to theoretical normative constructs, such as surplus and social welfare, have also proven useful in sentencing research, particularly in the study of inter-judge disparity.

INTRODUCTION 345
I. ECONOMICS AS AN OBSERVATIONAL SCIENCE 349
II. ECONOMIC ANALYSIS OF CRIME AND SANCTIONS.......................... 353
III. ECONOMIC STUDY OF SENTENCING DISPARITY 358
IV. THOUGHTS FOR FUTURE RESEARCH.. 365
CONCLUSION.. 367

INTRODUCTION

Economic analysis has become so pervasive in legal scholarship that

* Associate Professor, Northwestern University School of Law, joshua.fischman@law.northwestern.edu. I thank David Abrams, Tom Miles, Max Schanzenbach, and participants at the "Sentence Structure: The Elements of Punishment" Symposium at Loyola University Chicago School of Law for helpful comments.

it does not seem unusual to hold a panel discussion on "The Economics of Sentencing."[1] But what exactly is "The Economics of Sentencing"? It is quite apparent that economists have made important contributions to sentencing research in recent years, but it is less clear why economists study sentencing. According to the New Oxford American Dictionary, economics is "the branch of knowledge concerned with the production, distribution, consumption, and transfer of wealth."[2] The Dictionary defines criminal sentences as the judicial determinations of "punishment assigned to a defendant found guilty" of crimes.[3] On the basis of these traditional definitions, it is not at all obvious that economists would have much to contribute to the study of sentencing.

In this Essay, I highlight some of the economic literature on crime and sentencing,[4] and consider what is distinctive about economists' perspectives. There are, of course, narrow areas of overlap between economics and sentencing, such as how criminal convictions affect subsequent earnings[5] and how neighborhood crime rates affect the location of businesses.[6] The study of criminal deterrence is arguably a natural fit for economists, insofar as economic models of rational choice

---

1. This Essay is based on discussions from a panel entitled "The Economics of Sentencing" at the *Loyola University Chicago Law Journal* "Sentence Structure: The Elements of Punishment" Symposium, from Friday, April 4, 2014, at Loyola University Chicago School of Law.

2. NEW OXFORD AMERICAN DICTIONARY 550 (Angus Stevenson & Christine A. Lindberg eds., 3d ed. 2010).

3. *Id.* at 1591.

4. I do not attempt to provide a full-length overview of economic research on crime and sentencing. For more comprehensive treatments, see, e.g., Steven N. Durlauf & Daniel S. Nagin, *The Deterrent Effect of Imprisonment*, *in* CONTROLLING CRIME: STRATEGIES AND TRADEOFFS 43, 43–94 (Philip J. Cook et al. eds., 2011) (describing the state of knowledge regarding the deterrent effects of imprisonment, as well as implications for policy); Steven D. Levitt & Thomas J. Miles, *Economic Contributions to the Understanding of Crime*, 2 ANN. REV. L. & SOC. SCI. 147 (2006) (studying incentives, causation, public policy, and costs and benefits of crime from an economic lens); Steven D. Levitt & Thomas J. Miles, *Empirical Study of Criminal Punishment*, *in* 1 HANDBOOK OF LAW AND ECONOMICS 455, 455–95 (A. Mitchell Polinsky & Steven Shavell eds., 2007) (reviewing literature in which economists have empirically evaluated or tested the economic model of criminal behavior); Daniel S. Nagin, *Deterrence: A Review of the Evidence by a Criminologist for Economists*, 5 ANN. REV. ECON. 83 (2013) (highlighting findings regarding the effects of deterrence and proposing areas of future research); Aaron Chalfin & Justin McCrary, Criminal Deterrence: A Review of the Literature (May 9, 2014) (unpublished manuscript) (on file with Journal of Economic Literature) (reviewing recent economics research on the effect of deterrence).

5. *See, e.g.*, Jeffrey R. Kling, *Incarceration Length, Employment, and Earnings*, 96 AM. ECON. REV. 863, 869–72 (2006) (concluding that longer sentences have a positive effect on the employment and earnings of formerly incarcerated individuals one to two years after release, but negligible effects seven to nine years after incarceration began).

6. *See, e.g.*, Stuart S. Rosenthal & Amanda Ross, *Violent Crime, Entrepreneurship, and Cities*, 67 J. URB. ECON. 135, 144–48 (2010) (examining how business owners consider local crime rates in selecting locations for commercial establishments).

can explain how potential offenders respond to the threat of sanctions. Nevertheless, contemporary economic research on crime and sentencing extends far beyond the study of deterrence and these narrow areas of overlap. In my view, economists' interest in sentencing stems more from synergies in empirical methodology. In particular, I focus on the development of methods in empirical economics for making inferences from observational data, which requires finding naturally occurring "experiments" and interpreting empirical findings in ways that can be extrapolated to policy-relevant contexts. Many of these methods have proved useful not only for answering economic questions, but also for addressing questions of relevance to sentencing.

Economists bring multiple perspectives to the study of sentencing. The most obvious is the rational choice perspective—central to much of economic theory—which can model the decisions of potential offenders, law enforcement, and actors within the criminal justice system. This approach contrasts with sociological accounts of crime, which focus on the influences of peers, social groups, and culture; it is also in tension with psychological perspectives on crime, which attribute a larger role to emotion, personality disorders, and mental illness. Building on a seminal article by Gary Becker,[7] many economists have applied the rational choice model to criminal behavior, treating potential offenders as rational actors who weigh the costs and benefits of illegal conduct. Economists have similarly used rational choice theory to model aspects of the criminal justice process, such as interactions between potential offenders and law enforcement,[8] and between prosecutors and defense attorneys.[9]

In my view, however, much of the economic scholarship on sentencing has only a tenuous connection to rational choice. Indeed, such models have serious limits as applied to criminal behavior, especially given the influences of drug use,[10] impulsivity,[11] and mental

7. Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968).

8. *See, e.g.*, John Knowles et al., *Racial Bias in Motor Vehicle Searches: Theory and Evidence*, 109 J. POL. ECON. 203, 209–15 (2001) (using an empirical model to analyze whether the higher rate at which police search African-American drivers is due to racial prejudice or an effort to increase arrests).

9. *See, e.g.*, Jennifer F. Reinganum, *Plea Bargaining and Prosecutorial Discretion*, 78 AM. ECON. REV. 713, 715–23 (1988) (developing a model of plea bargaining in which the prosecution and defense possess asymmetric information).

10. *See* Justin McCrary, *Dynamic Perspectives on Crime*, *in* HANDBOOK ON THE ECONOMICS OF CRIME 83, 83 (Bruce L. Benson & Paul R. Zimmerman eds., 2010) (citing data from the Arrestee Drug Abuse Monitoring ("ADAM") program of the National Institute of Justice that shows that roughly two-thirds of arrestees in the United States test positive for one of five major

illness[12] on criminal activity. Other areas of economic research on sentencing, such as studies of peer effects and inter-judge disparity, have even weaker connections to rational choice. Instead, many contributions by economists appear to be driven by similarities between methodological problems posed by sentencing research and those familiar to economists.

This Essay discusses two key reasons why econometric methodology has found wide applicability in sentencing research. First, economics is predominantly an observational science. While statisticians and some other social scientists operate predominantly within an experimental paradigm, examining the impact of treatments that scientists themselves can manipulate, economists typically study phenomena that cannot be directly manipulated. Production, consumption, and the functioning of markets cannot be easily replicated in a laboratory setting. Analogies to controlled experiments are of limited help in understanding markets, where prices and output levels are determined by equilibrium interactions between producers and consumers. Instead, economists have developed a variety of methods for making causal inferences in observational contexts, particularly in settings involving two-way or multidirectional causation. These methods have also proven useful in studying criminal behavior and the operation of criminal justice.

Second, economics has a strong policy orientation, which has led economists to develop techniques for bridging the gap between empirical findings and policy conclusions. Economists often justify normative claims on the basis of theoretical constructs such as surplus or social welfare, which do not have simple relationships with measureable variables. Rather than letting the data speak for themselves, economists are more inclined to organize empirical findings to support specific policy conclusions. Economists may not be expansive normative theorists—indeed, they sometimes seem to revel in their disregard for non-utilitarian values[13]—but they are often quite

---

drugs).

11. *See generally* Durlauf & Nagin, *supra* note 4, at 73 (discussing psychological literature on crime and impulsivity).

12. *See* Jillian K. Peterson et al., *How Often and How Consistently Do Symptoms Directly Precede Criminal Behavior Among Offenders with Mental Illness?*, 38 L. & HUM. BEHAV. 439, 439 (2014) (estimating that 14–16% of prison inmates suffer from some kind of serious mental illness).

13. *See, e.g.*, RICHARD O. ZERBE, JR., ECONOMIC EFFICIENCY IN LAW AND ECONOMICS 189–99 (2001) (considering whether rape is an inefficient transaction); Elisabeth M. Landes & Richard A. Posner, *The Economics of the Baby Shortage*, 7 J. LEGAL STUD. 323, 328 (1978) (discussing the potential efficiency gains from legalizing the sale of babies); Joel Waldfogel, *The Deadweight Loss of Christmas*, 83 AM. ECON. REV. 1328, 1330–35 (1993) (measuring the deadweight loss of

careful about connecting empirical findings to policy conclusions. As I discuss below, this may explain why economists made key contributions to the measurement of inter-judge disparity, another theoretical construct with complex normative and empirical foundations.

Part I of this Essay discusses the distinction between experimental and observational approaches to causal inference, and explains why empirical economics is primarily observational. Part II examines how observational methods developed by economists have proven useful for studying the impact of sanctions on crime. Part III discusses the economic literature on sentencing disparity, highlighting how economists have taken a rigorous approach to measuring disparity and relating empirical findings to policy conclusions. Part IV offers a brief discussion of current challenges in sentencing research, and considers how we can build upon recent advances.

I. ECONOMICS AS AN OBSERVATIONAL SCIENCE

Empirical social science has long recognized a distinction between experimental and observational approaches to causal inference. In experimental approaches—which are dominant in disciplines such as psychology and biostatistics—researchers assign subjects to treatment and control groups, typically using a randomized process. In observational approaches—which are dominant in economics—researchers measure the effects of interventions by examining naturally occurring changes in variables of interest.

The experimental approach has several clear advantages. The first is that randomization ensures that differences in outcomes can be attributed to the treatment applied and not to differences in the composition of the treatment and control groups.[14] Thus, experimental approaches provide the most credible estimates of causal effects. Second, the researcher may be able to select the type and magnitude of the treatment that is applied in order to best answer a particular research question.[15] Because observational approaches must rely on naturally

---

holiday gift exchange).

14. *See* Daniel E. Ho & Donald B. Rubin, *Credible Causal Inference for Empirical Legal Studies*, 7 ANN. REV. L. & SOC. SCI. 17, 22 (2011) ("Randomization over a large number of units ensures that treatment and control units are comparable in all respects other than the treatment.").

15. *See* Gary Burtless, *The Case for Randomized Field Trials in Economic and Policy Research*, J. ECON. PERSP., Spring 1995, at 63, 69 ("[I]n comparison with most sources of nonexperimental information, experiments permit economists to learn about the effects of a much wider range of prices and policies."); *see also* Jens Ludwig et al., *Mechanism Experiments and Policy Evaluations*, J. ECON. PERSP., Summer 2011, at 17, 30–35 (discussing how to design

occurring sources of variation, the answers generated may not be directly responsive to any policy question of interest.

The primary weakness of the experimental approach is its limited applicability; many important research questions simply cannot be answered using experimental methods.[16] In some cases, there are ethical barriers.[17] For example, one could not test the deterrent effect of the death penalty by randomizing executions. In other instances, the barrier is practical. The behavior studied in laboratory experiments may be too dissimilar from real-world conduct to provide policy-relevant conclusions,[18] or the intended subjects of a study may refuse to cooperate.[19] Experiments conducted in the field may provide greater realism, but because they are harder to control, they also face greater risk of contamination due to "subject attrition, crossover between the treatment and control groups, spillover effects, or even conscious efforts by nonparticipants to undermine the research."[20]

Economists have developed and honed methods for making observational inferences, in large part because many of their research questions cannot be studied experimentally. Often, economists are not seeking to examine a single causal relationship between a treatment and outcome, but rather, equilibrium interactions involving simultaneous causation.[21] In product markets, for example, prices and quantities are determined by interactions between supply curves and demand curves. High prices lead consumers to reduce purchases, but high prices also induce suppliers to increase output. There is no single causal relationship between market prices and output levels, but rather

---

experiments in order to maximize policy-relevant findings).

16. *See* Joshua B. Fischman, *Reuniting 'Is' and 'Ought' in Empirical Legal Scholarship*, 162 U. PA. L. REV. 117, 166–67 (2013) (discussing limitations of experimental methods in empirical legal research).

17. *See id.* at 166 nn.237–38 (referring to sources examining ethical limitations on intentional randomization in the legal process).

18. *See infra* notes 84–85 and accompanying text (discussing concerns about the external validity of laboratory experiments).

19. *See* Fischman, *supra* note 16, at 167 n.243 (discussing efforts by judges and lawyers to undermine randomized studies of the legal system); James J. Heckman & Jeffrey A. Smith, *Assessing the Case for Social Experiments*, J. ECON. PERSP., Spring 1995, at 85, 104 (1995) (describing how program administrators can "subvert any randomization imposed upon them").

20. Fischman, *supra* note 16, at 167; *see also* Michael Abramowicz et al., *Randomizing Law*, 159 U. PA. L. REV. 929, 957–60 (2011) (discussing attrition, crossover, and spillovers in the context of randomized trials).

21. *See* James J. Heckman, *Econometric Causality*, 76 INT'L STAT. REV. 1, 19–20 (2008) (contrasting the statistical approach to causality in which "there is no simultaneity in causal effects," with the econometric approach, which accounts for simultaneous causation (emphasis omitted)).

simultaneous interactions between consumers and producers.

Many statisticians view randomized trials as the "gold standard" for causal inference,[22] viewing observational methods as inferior. When they must work with observational data, they often apply adjustments to "mak[e] them more closely resemble randomized experiments."[23] Many economists, however, do not view the randomized trial as an ideal approach to inference.[24] Rather than "start[ing] from the perspective of a randomized clinical trial, economists start with the notion that individuals receive the treatments they received because they choose to."[25] Often, their goal is to estimate parameters that determine agents' preferences, which can then be used to model their behavior in counterfactual contexts.[26]

There are two primary challenges in making causal inferences in observational contexts. The first is to find naturally occurring sources of variation in the underlying variables. For example, researchers might examine the effects of policies that were implemented at different times in different regions, or sources of arbitrariness—such as just meeting or just missing a threshold that determines eligibility for a treatment—that induce quasi-randomness in variables of interest.[27] The second

---

22. *See* Donald B. Rubin, *For Objective Causal Inference, Design Trumps Analysis*, 2 ANNALS APPLIED STAT. 808, 808 (2008) ("[C]arefully designed and executed randomized experiments are generally considered to be the gold standard.").

23. Guido W. Imbens, *An Economist's Perspective on Shadish (2010) and West and Thoemmes (2010)*, 15 PSYCHOL. METHODS 47, 48 (2010); *see* Rubin, *supra* note 22, at 810–11 (arguing that observational studies should be conceptualized as "approximations of randomized experiments").

24. *See, e.g.*, Angus Deaton, *Instruments, Randomization, and Learning about Development*, 48 J. ECON. LITERATURE 424, 426 (2010) ("[T]he value of econometric methods cannot and should not be assessed by how closely they approximate randomized controlled trials."); Durlauf & Nagin, *supra* note 4, at 57 ("[W]e are sympathetic to concerns that the virtues of randomized experiments have been exaggerated."); Heckman, *supra* note 21, at 20 ("Even under ideal conditions, randomization cannot answer some very basic questions, such as what proportion of a population benefits from a programme . . . ."); Christopher A. Sims, *But Economics Is Not an Experimental Science*, J. ECON. PERSP., Spring 2010, at 59, 59 (2010) (criticizing experiments as "rhetorical devices that are often invoked to avoid having to confront real econometric difficulties"). Many economists, however, advocate greater use of randomized experiments. *See, e.g.*, Abhijit V. Banerjee & Esther Duflo, *The Experimental Approach to Development Economics*, 1 ANN. REV. ECON. 151, 156 (2009) (discussing how experiments can "help[] us answer conceptual questions . . . that could never be reliably answered in any other way").

25. Imbens, *supra* note 23, at 48.

26. *See id.* ("The goal of such analyses is often to infer the preferences of agents in order to predict what would happen if the constraints the agents face were changed. Examples of such changes include imposing taxes on transactions or expanding the set of choices. Underlying this approach is the notion that the preferences are relatively stable and specifically that they do not change in response to changes in the constraints.").

27. *See infra* notes 45–46, 52–56 and accompanying text (examining two studies, one exploiting "features of California's three-strikes law to measure the effect of deterrence," and the

challenge arises out of the fact that these natural sources of variation do not necessarily coincide with the interventions that would be most relevant for policy purposes. This challenge is particularly salient when economists seek to predict the effects of interventions that cannot be tried in advance. This requires using empirical methods that make inferences about causal mechanisms, and building theoretical models that facilitate the extrapolation of these findings to new contexts.[28]

Many experimentalists limit their research to questions that can be analyzed by randomized trials, or to observational contexts that can be plausibly analogized to such trials. Economists do not have this option; such a constraint would put much of the field off-limits to empirical inquiry.[29] Experimental methods have limited value for studying economic phenomena such as supply and demand systems, financial markets, monetary policy, and international trade. As the eminent econometrician Trygve Haavelmo once remarked, "physicists are very clever. They confine their predictions to the outcomes of their experiments. They do not try to predict the course of a rock in the mountains and trace the development of the avalanche. It is only the crazy econometrician who tries to do that . . . ."[30]

Research on the criminal justice system is in many ways like "tracing the development of the avalanche."[31] Due to ethical and practical constraints, much of the study of crime and justice is beyond the scope of experimentation.[32] There would be severe ethical concerns, for example, with studying violent crime in a laboratory setting. Certain forms of experiments can be conducted in the field,[33] but it can be

---

other "exploiting the effect of sentencing enhancements for offenders who possess a gun during the commission of a crime").

28. *See* Heckman, *supra* note 21, at 5 (discussing the challenge of "[f]orecasting the impacts of interventions . . . never historically experienced").

29. *See* Clive Granger, Comment, *Statistics and Causal Inference*, 81 J. AM. STAT. ASS'N 967, 967 (1986) ("[M]any causal questions cannot be tackled within [an experimental] framework, such as most of those arising in history, economics, sociology, meteorology, oceanology, political science, anthropology, or law.").

30. NANCY CARTWRIGHT, THE DAPPLED WORLD: A STUDY OF THE BOUNDARIES OF SCIENCE 46 (1999) (relating personal conversation with Haavelmo).

31. *Id.*

32. *See* Durlauf & Nagin, *supra* note 4, at 56–57 (discussing practical and ethical limitations of randomized experiments in the study of criminal justice).

33. *See, e.g.*, JOHN S. GOLDKAMP & MICHAEL R. GOTTFREDSON, JUDICIAL GUIDELINES FOR BAIL: THE PHILADELPHIA EXPERIMENT 21 (1984) (randomizing assignment of judges to treatment groups, to test effects of proposed bail guidelines); Lawrence W. Sherman & Richard A. Berk, *The Specific Deterrent Effects of Arrest for Domestic Assault*, 49 AM. SOC. REV. 261, 261–62 (1984) (randomizing arrest among eligible domestic violence suspects to measure effect of arrest on subsequent violence).

difficult to secure the cooperation of law enforcement, judges, prosecutors, and defense attorneys.[34] Randomizing the administration of criminal justice also raises serious ethical questions.[35] Because of these limitations, observational methods will inevitably play a primary role in the study of criminal justice. The following Part discusses how econometric methods have proved to be especially useful for studying the deterrent effect of sanctions, the effects of incarceration, and peer effects in criminal activity.

II. ECONOMIC ANALYSIS OF CRIME AND SANCTIONS

Like economics, empirical research on criminal justice presents many challenges involving simultaneous causation. Arguably the most prominent example involves the relationship between crime and sanctions.[36] This two-way causation occurs because sanctions prevent crime, but high levels of crime spur more intensive enforcement and punishment. Of course, the causation is not simultaneous in a literal sense; there are lags between changes in enforcement and changes in criminal activity, and vice versa. Indeed, some early studies sought to use this lag to separately identify the effects of crime on punishment and punishment on crime,[37] although this strategy depends on strong assumptions that are unlikely to be satisfied in practice.[38]

A more credible approach to measuring the effect of sanctions on crime is to find *instrumental variables*—variables that influence the frequency or intensity of punishment, but do not otherwise have any impact on criminal behavior.[39] Several studies of deterrence in the

---

34. *See* Heckman and Smith, *supra* note 19, at 101–04 (discussing institutional limitations on social experiments).

35. *See* Fischman, *supra* note 16, at 166 (discussing the ethical problems presented by intentional randomization with regard to the legal profession and adjudication, because randomization is naturally at odds with “the need for reasoned decisionmaking”); *see also* Adam M. Samaha, *Randomization and Adjudication*, 51 WM. & MARY L. REV. 1, 5 (2009) (discussing judicial opposition to randomization).

36. *See* Nagin, *supra* note 4, at 84 (noting that since the 1960s, “hundreds of studies have tested for deterrent effects”).

37. *See, e.g.*, Thomas B. Marvell & Carlisle E. Moody, *Specification Problems, Police Levels, and Crime Rates*, 34 CRIMINOLOGY 609, 619–22 (1996) (discussing time-series methods used to analyze the relationship between policing and crime).

38. *See* Durlauf & Nagin, *supra* note 4, at 50 (arguing that the Marvell and Moody study, *supra* note 37, “does not, under any interpretation of causality of which we are aware, provide a policy relevant measure of the effects of imprisonment”); *see also* Chalfin & McCrary, *supra* note 4, at 10 (arguing that the methodology used by Marvell and Moody “is subject to the same omitted variables bias issues that plague any least squares regression model and is therefore of dubious value in establishing causality”).

39. *See* Franklin M. Fisher & Daniel Nagin, *On the Feasibility of Identifying the Crime Function in a Simultaneous Model of Crime Rates and Sanction Levels*, *in* DETERRENCE AND

1970s relied on this strategy, but the instrumental variables they used have been criticized for failing to satisfy the necessary assumptions.[40] In recent years, studies have employed more credible instrumental variables. For example, Steven Levitt used the timing of lawsuits challenging prison overcrowding to generate instrumental variables for measuring the impact of prison population on crime levels.[41] He found that an increase in the prison population significantly reduced crime levels. In another study, William Evans and Emily Owens exploited the timing of federal grants to local police agencies, under the Violent Crime Control and Law Enforcement Act, to generate instrumental variables that influence the level of local policing.[42] They found that increases in police grants generated statistically significant reductions in a variety of common crimes.[43]

Economists have also exploited discontinuities in punishment severity to measure the impact of criminal sanctions. For example, criminal penalties increase significantly when juveniles reach the age of majority.[44] Thus, offenders just above the age of majority will face substantially more severe penalties than those just below, although the two groups will otherwise be similar in terms of relevant characteristics. Studies that have employed this strategy have reached conflicting conclusions, with some finding significant deterrent effects[45] and some

---

INCAPACITATION: ESTIMATING THE EFFECTS OF CRIMINAL SANCTIONS ON CRIME RATES 361, 363 (Alfred Blumstein et al. eds., 1978) (noting the need to exclude exogenous variables from one equation in a system of simultaneous equations in order to disentangle the mutual causal effects between crime and sanctions).

40. *See id.* at 372–74 (criticizing several early studies of deterrence for using socioeconomic and demographic variables as instruments, stating that "it is simply not plausible to assume that such . . . variables do not have a direct effect on crime").

41. Steven D. Levitt, *The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation*, 111 Q.J. ECON. 319, 319 (1996). Although Levitt's research design was more credible than prior studies, it has not been immune to criticism. *See* John J. Donohue III & Peter Siegelman, *Allocating Resources Among Prisons and Social Programs in the Battle Against Crime*, 27 J. LEGAL STUD. 1, 13–14 (1998) (raising questions about Levitt's empirical approach and conclusions).

42. *See* William N. Evans & Emily G. Owens, *COPS and Crime*, 91 J. PUB. ECON. 181, 182 (2007) ("[T]he variation in timing and size of grants [were used] to test whether the hiring grants increased the size of police forces.").

43. *See id.* at 183 ("[W]e find that additional officers granted through the COPS program produce statistically significant drops in burglaries, auto thefts, robberies, and aggravated assaults.").

44. *See* Randi Hjalmarsson, *Crime and Expected Punishment: Changes in Perceptions at the Age of Criminal Majority*, 11 AM. L. & ECON. REV. 209, 211 (2009) ("It is generally taken as common knowledge that, conditional on the crime committed, individuals receive a harsher punishment if sentenced in the criminal courts rather than the juvenile courts.").

45. *See* Steven D. Levitt, *Juvenile Crime and Punishment*, 106 J. POL. ECON. 1156, 1181 (1998) ("The evidence suggests that juvenile crime is responsive to harsher sanctions.").

finding negligible effects.[46]

Another study exploited features of California's three-strikes law to measure the effect of deterrence. The study compared two groups of former inmates who had been tried for two strikeable offenses.[47] The first group had been convicted of two strikeable offenses while the second had been convicted of one strikeable offense and one lesser offense.[48] These two groups were roughly comparable in terms of propensity to engage in criminal conduct, but the first group faced much more severe sanctions for an additional felony conviction. This study found significant deterrent effects: those with two strikes were 17% less likely to be rearrested in the three years following their release.[49]

Economists have also developed innovative strategies for disentangling the effects of deterrence and incapacitation. These effects are difficult to distinguish using aggregate crime data because more intensive policing and more severe sanctions will typically deter criminal behavior and also incapacitate a greater number of offenders.[50] Distinguishing between them is important as a policy matter because deterrence is a far less costly form of crime reduction.[51]

An influential study by Daniel Kessler and Steven Levitt examined the effect of a California ballot proposition that mandated sentencing enhancements for certain categories of repeat offenders.[52] Their key

---

46. *See* Hjalmarsson, *supra* note 44, at 245 (finding that at the age of majority, perception of punishment severity is underestimated and evidences little change in delinquent behavior); David S. Lee & Justin McCrary, The Deterrence Effect of Prison: Dynamic Theory and Evidence 32 (July 2009) (unpublished manuscript) (on file with the Journal of Economic Literature) (suggesting that decreased involvement in crime is a function of age, and not a deterrent effect of increased adult criminal sanctions).

47. *See* Eric Helland & Alexander Tabarrok, *Does Three Strikes Deter? A Nonparametric Estimation*, 42 J. HUM. RES. 309, 310 (2007) ("We estimate the effect of the law by comparing the subsequent arrest profiles of criminals who were released with two strikeable offenses with those released with two trials for strikeable offenses but only one conviction for a strikeable offense.").

48. *See id.*

49. *See id.* at 316 ("We estimate that the threat of a third strike reduces arrest rates by . . . 17.2 percent . . . .").

50. *E.g.*, Chalfin & McCrary, *supra* note 4, at 8 (observing that "research on the effect of sanctions typically results in a treatment effect that is a function of both deterrence and incapacitation," although "clever research designs have been used to identify the effect of an increase in the severity of a sanction that is unlikely to result in an immediate increase in incapacitation").

51. *See id.* at 2 ("Deterrence is important not only because it results in lower crime but also because, relative to incapacitation, it is cheap.").

52. *See* Daniel Kessler & Steven D. Levitt, *Using Sentence Enhancements to Distinguish Between Deterrence and Incapacitation*, 42 J.L. & ECON. 343, 343–44 (1999) (separating deterrence from incapacitation by analyzing the immediate effects of increasing incarceration length for certain crimes).

insight was that enhancements are added to sentences that would have been served in any event, so the immediate effect of the ballot proposition must be due to a deterrent effect.[53] By examining short-run trends in crimes eligible for sentence enhancements and comparing them with crimes that are not subject to enhancements, they found that the enhancements resulted in an 8% short-run decrease in crime, which could be attributed to a deterrent effect.[54] More recently, David Abrams employed a similar strategy, exploiting the effect of sentencing enhancements for offenders who possess a gun during the commission of a crime.[55] By examining the adoption of these enhancements in different states at different times, Abrams estimated that these laws generated a short-run 5% decrease in gun robberies due to deterrence.[56]

Economists have developed several other approaches for distinguishing deterrence and incapacitation. The study of the California three-strikes law, discussed above,[57] clearly measures a deterrent effect, because it examines individual-level arrest data on former inmates who are not incarcerated. Similarly, the studies that examine the discontinuous increase in punishment upon the age of majority measure the effect of deterrence.[58] Another study by Emily Owens examined a change in Maryland sentencing guidelines that reduced sentences for many young adult offenders.[59] Because this change in the guidelines was not widely publicized, Owens attributed the subsequent increase in crime among these offenders to a decrease in incapacitation.[60]

---

53. *See id.* at 345 ("[B]y looking at changes in crime immediately following the introduction of a sentence enhancement, it is possible to isolate a pure deterrent effect that is not contaminated by incapacitation.").

54. *See id.* at 357 (finding that crime rates fell by 8.9% immediately after sentence enhancements were instituted).

55. *See* David S. Abrams, *Estimating the Deterrent Effect of Incarceration Using Sentencing Enhancements*, AM. ECON. J.: APPLIED ECON., Oct. 2012, at 32, 32 (exploiting penalty differences produced by "add-on gun laws" to isolate deterrent effects).

56. *Id.* at 45 (finding that gun robbery rates decreased "5 percent within 3 years").

57. *See* Helland & Tabarrok, *supra* note 47, at 310 (the study estimated "the effect of the law by comparing the subsequent arrest profiles of criminals who were released with two strikeable offenses with those released with two trials for strikeable offenses but only one conviction for a strikeable offense").

58. *See* Hjalmarsson, *supra* note 44, at 236–44 (characterizing the impact of reaching the age of criminal majority on incarceration as a deterrent effect).

59. *See* Emily G. Owens, *More Time, Less Crime? Estimating the Incapacitative Effect of Sentence Enhancements*, 52 J.L. & ECON. 551, 552–53 (2009) (evaluating effects of sentence enhancements using differences in punishment severity among a group of twenty-three- to twenty-five-year-olds who received reduced sentences).

60. *See id.* at 556, 558 (taking advantage of unpublicized policy change to estimate amount of crime reduction within each year of added incarceration).

In addition to studying the relationship between crime, deterrence, and incapacitation, economists have also been active in research on peer effects in criminal activity. This is perhaps surprising, given that peer effects are rooted in sociological—rather than economic—models of crime. Here, once again, I believe the explanation lies with the challenges in dealing with simultaneous causation.

The study of peer effects in empirical economics arguably began with an influential article by the econometrician Charles Manski, in which he investigated the challenges in identifying the causal influences of peers and social groups.[61] These challenges are numerous, largely because social influence among peers typically runs in multiple directions. Peer effects are even more difficult to measure when group boundaries are poorly defined,[62] when individuals may self-select into groups,[63] or when group members may be subject to common unobserved influences.[64] Thus, empirical studies of peer effects in criminal behavior must rely on clever strategies to overcome these numerous obstacles in inference. One study, for example, exploited the "Moving to Opportunity" field experiment in which a randomized group of low-income participants were given the opportunity to relocate to low-crime neighborhoods.[65] Because of the random assignment, this research design was able to disentangle the selection effect from the effect of peers. It found no significant impact of peers on crime rates by the study participants.[66]

Economists have also used detailed information about social networks to study the influence of peers. One study exploited detailed surveys in which adolescents identified their closest friends in their peer network,[67] finding that peers have a large influence on an individual's

---

61. *See* Charles F. Manski, *Identification of Endogenous Social Effects: The Reflection Problem*, 60 REV. ECON. STUD. 531, 532 (1993) (discussing difficulties in "distinguish[ing] among competing hypotheses about the nature of social effects").

62. *See id.* at 532 (noting the impossibility of inference when a researcher lacks "prior information specifying the composition of reference groups").

63. *See id.* at 536 (describing the difficulty in measuring peer effects when researchers do not know how individuals form reference groups).

64. *See id.* at 532–34 (noting the difficulty in distinguishing peer effects from "correlated effects" arising from common unobserved influences).

65. *See* Jens Ludwig & Jeffrey R. Kling, *Is Crime Contagious?*, 50 J.L. & ECON. 491, 493 (2007) (examining peer effects in criminal behavior by analyzing arrest rates among similar groups of families randomly assigned to different types of neighborhoods).

66. *See id.* at 500 ("[T]he pattern of results suggests . . . that there are aspects of residential neighborhoods that affect crime, particularly racial segregation, but that the role of neighborhood crime is more limited.").

67. *See* Eleonora Patacchini & Yves Zenou, *Juvenile Delinquency and Conformism*, 28 J.L. ECON. & ORG. 1, 11 (2009) (analyzing the role of conformism on crime using data from the

propensity to commit petty crimes, but a smaller influence on an individual's propensity to commit more serious crimes.[68] Another study examined the assignment of juvenile offenders to correctional facilities, taking advantage of detailed data on prior offenses committed by fellow inmates.[69] This study also found significant evidence of peer effects: individuals were more likely to commit particular offenses following release if they had served with other inmates who had previously committed the same offenses.[70]

III. ECONOMIC STUDY OF SENTENCING DISPARITY

The legal process of sentencing criminal offenders may seem even more remote from the core subject matter of economics. Some research on the sentencing process draws upon positive political theory, using rational choice to model the behavior of government actors. For example, some studies have shown that elected state judges sentence more harshly when they are up for reelection,[71] while others have examined how federal district judges sentence strategically to avoid reversal by circuit courts.[72] Economists' research on the sentencing process is also informed by economic research on litigation and

---

National Longitudinal Survey of Adolescent Health).

68. *See id.* at 20 (estimating the effects of social interactions on specific types of crimes).

69. *See* Patrick Bayer et al., *Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections*, 124 Q.J. ECON. 105, 106 (2009) (studying whether fellow inmates influence juvenile offenders' future criminal behavior).

70. *See id.* at 126–27 (noting the "reinforcing peer effects" on recidivism).

71. *See, e.g.*, Carlos Berdejó & Noam Yuchtman, *Crime, Punishment, and Politics: An Analysis of Political Cycles in Criminal Sentencing*, 95 REV. ECON. & STAT. 741, 742 (2013) ("[S]entencing of serious offenses becomes more severe as elections approach: sentence lengths increase by around 10% between the beginning and the end of a judge's political cycle."); Claire S. H. Lim, *Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges*, 103 AM. ECON. REV. 1360, 1392 (2013) (finding that sentencing patterns of elected judges are more widely varied than those of appointed judges). Political scientists working within the rational choice paradigm authored the first studies on the interplay between sentencing and judicial elections. *See* Sanford C. Gordon & Gregory A. Huber, *The Effect of Electoral Competitiveness on Incumbent Behavior*, 2 Q.J. POL. SCI. 107, 108 (2007) ("[J]udges in partisan competitive systems sentence significantly more punitively than those in retention systems."); Gregory A. Huber & Sanford C. Gordon, *Accountability and Coercion: Is Justice Blind When It Runs for Office?*, 48 AM. J. POL. SCI. 247, 248 (2004) ("[S]entences for . . . crimes are significantly longer the closer the sentencing judge is to standing for reelection.").

72. *See, e.g.*, Joshua B. Fischman & Max M. Schanzenbach, *Do Standards of Review Matter? The Case of Federal Criminal Sentencing*, 40 J. LEGAL STUD. 405, 406 (2011) ("[W]e interpret district court sensitivity to standards of review as evidence that district judges are averse to reversal and respond prospectively to changes in standards of review."); Max M. Schanzenbach & Emerson H. Tiller, *Strategic Judging Under the U.S. Sentencing Guidelines: Positive Political Theory and Evidence*, 23 J.L. ECON. & ORG. 24, 52–53 (2007) (finding that Democratic district judges depart more frequently from U.S. Sentencing Guidelines in circuits with a majority of Democratic appointees, as predicted by strategic models of judicial behavior).

bargaining and by their experience in dealing with simultaneous causation.[73] Because sentences are determined through an interactive process involving judges, prosecutors, and defense attorneys, understanding how any legal change affects sentencing requires careful analysis of the strategic interactions among these actors. Thus, economists have studied prosecutorial motivations,[74] how changes in sentencing law affect plea bargaining,[75] and the impact of defense attorneys.[76]

Yet some of economists' most important contributions relate to the measurement of inter-judge sentencing disparity, which does not involve rational choice, strategic behavior, or welfare maximization.[77] Indeed, it is surprising that economists have been so involved in the study of disparity, given that they do not otherwise take much interest in non-utilitarian concerns. I believe that the connection between economics and the measurement of disparity is primarily methodological, stemming from economists' inclination to formalize

---

73. *See supra* Part I (discussing econometric methodologies used to make causal inferences from observations).

74. *See, e.g.*, David Bjerk, *Making the Crime Fit the Penalty: The Role of Prosecutorial Discretion Under Mandatory Minimum Sentencing*, 48 J.L. & ECON. 591, 593 (2005) (suggesting that prosecutors are motivated to circumvent "three-strikes laws" due to "their own constraints and preferences"); Richard T. Boylan & Cheryl X. Long, *Salaries, Plea Rates, and the Career Objectives of Federal Prosecutors*, 48 J.L. & ECON. 627, 649 (2005) (finding "that in districts with higher private-lawyer salaries, assistant U.S. attorneys are more likely to take cases to trial," supporting the "hypothesis that some lawyers work for the government to accumulate human capital"); Edward L. Glaeser et al., *What Do Prosecutors Maximize? An Analysis of the Federalization of Drug Crimes*, 2 AM. L. & ECON. REV. 259, 288 (2000) (finding that federal prosecutors are more likely to prosecute drug cases involving "high-human-capital individuals," suggesting that they are either maximizing social welfare or selecting cases that "offer the best career returns"); Daniel P. Kessler & Anne Morrison Piehl, *The Role of Discretion in the Criminal Justice System*, 14 J.L. ECON. & ORG. 256, 274 (1998) (suggesting that statutory increases in sentencing length may have "spillover effects" to similar crimes by virtue of "prosecutorial maximization").

75. *See, e.g.*, Ilyana Kuziemko, *Does the Threat of the Death Penalty Affect Plea Bargaining in Murder Cases? Evidence from New York's 1995 Reinstatement of Capital Punishment*, 8 AM. L. & ECON. REV. 116, 140 (2006) ("The findings here suggest that the threat of the death penalty leads more defendants to plead guilty to their original arraignment charges.").

76. *See, e.g.*, David S. Abrams & Albert H. Yoon, *The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability*, 74 U. CHI. L. REV. 1145, 1150 (2007) (finding that Hispanic public defenders and those with more experience secure shorter sentences for their clients, but that defenders who attended higher-ranked law schools do not outperform those who attended lower-ranked schools); James M. Anderson & Paul Heaton, *How Much Difference Does the Lawyer Make? The Effect of Defense Counsel on Murder Case Outcomes*, 122 YALE L.J. 154, 199 (2012) (finding that "public defenders in Philadelphia reduce their clients' murder conviction rate by 19%" and their "overall expected time served in prison by 24%").

77. *See, e.g.*, Fischman, *supra* note 16, at 148 n.163 (discussing how research on disparity is motivated by concerns about "consistency, correctness, determinacy, fairness, predictability, non-arbitrariness, and the rule of law").

normative constructs such as surplus or welfare and to develop methods for making inferences on these constructs from observable data.

Studies of sentencing disparity have traditionally followed one of two research designs. Some studies follow an experimental approach, using surveys to ask judges or lay respondents how they would sentence hypothetical offenders. Other studies rely on observational data, examining judges' decisions in actual cases. The application of these different approaches has divided, to some extent, along disciplinary lines: much of the psychological literature on sentencing disparity follows the experimental approach,[78] while most economic research is observational.[79] Criminologists have employed both observational and experimental methods.[80]

The experimental approach has two primary advantages. First, researchers can directly compare different judges' responses to the same cases. In observational studies, by contrast, researchers can only measure differences in average sentences. This distinction is important; comparisons of average sentences may obscure unequal treatment for individual offenders. Second, researchers can randomly assign different case facts to different respondents in order to measure the impact of these facts on sentences. In studying racial disparity, for example, scholars can manipulate the suspect's race in a hypothetical scenario, while keeping other facts unchanged, to assess how judges would treat white and black offenders differently.[81]

Many studies of sentencing disparity that predated the United States Sentencing Guidelines followed the experimental approach. The

78. *See, e.g.*, William Austin & Thomas A. Williams III, *A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity*, 68 J. CRIM. L. & CRIMINOLOGY 306, 307 (1977) (describing sentencing variation among similarly positioned judges based on five hypothetical legal cases); Shari Seidman Diamond & Loretta J. Stalans, *The Myth of Judicial Leniency in Sentencing*, 7 BEHAV. SCI. & L. 73, 75 (1989) (analyzing sentencing variation among judges, jurors, and students based on four hypothetical criminal cases); Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195, 1204 (2009) (using surveys to measure judges' implicit bias and to test whether it affects judges' decisions in simulated cases); Peter J. van Koppen & Jan Ten Kate, *Individual Differences In Judicial Behavior: Personal Characteristics and Private Law Decision-Making*, 18 LAW & SOC'Y REV. 225, 226 (1984) (testing a written set of protocols that mimicked the decision-making tasks that civil-law trial judges face in order to identify personal factors that result in sentencing discrepancy).

79. *See infra* notes 97–107 and accompanying text.

80. *See, e.g.*, Paul J. Hofer et al., *The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity*, 90 J. CRIM. L. & CRIMINOLOGY 239, 264 (1999) (using observational methodology); Andreas Kapardis & David P. Farrington, *An Experimental Study of Sentencing by Magistrates*, 5 LAW & HUM. BEHAV. 107, 111 (1981) (using experimental methodology).

81. *See, e.g.*, Rachlinski et al., *supra* note 78, at 1211 (describing hypothetical scenarios).

influential Second Circuit Sentencing Study,[82] for example, surveyed fifty district judges about twenty hypothetical cases, directly comparing their responses. Because all of the judges were responding to the same stimuli, the authors could precisely measure the inter-judge variation in each case. The study also divided the judges randomly into two groups, modifying the case facts for one of the two groups. By manipulating the case facts, the authors could measure judges' responses to such factors as prior criminal history, guilty pleas, and drug addiction.[83] Because all other case factors are kept constant, the study could credibly estimate the average causal effect of each manipulation.

There are many limitations of the experimental approach, however, which may explain why such surveys have become less common since the enactment of the United States Sentencing Guidelines. Most notably, there are serious questions about the external validity of experiments involving simulated cases.[84] As I have argued previously, "[s]implified scenarios in written questionnaires may not present the same stimuli as actual cases: judges are not exposed to advocacy from both sides, they are not required to write opinions justifying their decisions, and they do not need to consider the impact of their judgments on actual parties."[85] There is also no guarantee that judges will willingly participate in such studies or that they will approach them with the requisite seriousness.[86] One prominent critique of the Second

82. ANTHONY PARTRIDGE & WILLIAM B. ELDRIDGE, FED. JUDICIAL CTR., THE SECOND CIRCUIT SENTENCING STUDY: A REPORT TO THE JUDGES 1 (1974).

83. *Id.* at 45–53.

84. *See, e.g.*, James M. Anderson et al., *Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines*, 42 J.L. & ECON. 271, 279 (1999) ("It is quite difficult . . . for a simulation to reconstruct the full depth of information available to a judge in a real case."); John M. Conley & William M. O'Barr, *Fundamentals of Jurisprudence: An Ethnography of Judicial Decision Making in Informal Courts*, 66 N.C. L. REV. 467, 474–75 (1988) ("One can never claim with certainty [that experiments] have captured all the elements of a real case, nor can one be sure that subjects will respond to stimuli in the same way as they would in the courtroom."); Shari Seidman Diamond & Hans Zeisel, *Sentencing Councils: A Study of Sentence Disparity and its Reduction*, 43 U. CHI. L. REV. 109, 116 (1975) (describing the differences between sentencing simulations and real decisions); Hofer et al., *supra* note 80, at 264 ("[H]ypothetical situations may be so different from actual sentencing that the results cannot be generalized to the real world."); Vladimir J. Konečni & Ebbe B. Ebbesen, *External Validity of Research in Legal Psychology*, 3 LAW & HUM. BEHAV. 39, 65 (1979) ("We believe . . . that the results of research efforts that deal with the real-world, consequential legal decisions are far more informative than those that deal with simulated decisions."); Gregory C. Sisk et al., *Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning*, 73 N.Y.U. L. REV. 1377, 1394 (1998) (writing that "simulation experiments with judges" suffer from "an inauthenticity that fails to mirror the real world of adjudication and judicial decisionmaking").

85. Joshua B. Fischman, *Measuring Inconsistency, Indeterminacy, and Error in Adjudication*, 16 AM. L. & ECON. REV. 40, 59 (2014).

86. *See id.* (noting that "judges may be loath to cooperate, especially if the research could

Circuit Sentencing Study reports that some of the judges involved in the study treated it as a "joke."[87] Whatever difficulties there may have been in securing judges' participation before the Guidelines, it is much more difficult now that judges are acutely aware of the policy implications of such research. Finally, while surveys can examine how judges would respond to different scenarios, they cannot account for the potential reactions of other actors, such as prosecutors, defense attorneys, or appellate courts.

Observational studies that examine actual judicial decisions can provide greater authenticity, but these studies also have serious limitations. Measuring inter-judge disparity becomes much more difficult because judges' decisions are not simultaneously observable in the same cases. One approach has been to compare judges' sentences for offenders who appear similar in terms of observable characteristics.[88] This approach, however, is less convincing, because judges typically have more information about particular offenders than sentencing researchers. Thus, offenders who appear to be similar in the data may differ in relevant ways that are observable to the judges.

In recent years, many studies have exploited the random assignment of cases to judges in order to compare average sentences among judges. This approach combines the authenticity of real-world cases with the credibility of experimental design, but it also gives rise to several challenging questions. First, how should one measure inter-judge disparity when one cannot observe different judges deciding the same cases? This question has both normative and statistical elements: disparity itself is a normative construct,[89] but researchers must determine how to make inferences about disparity on the basis of observable variables.

Second, how can one determine from observational data whether disparity increased or decreased as the result of a reform? In my view, this is one of the most important challenges in research on sentencing disparity. As long as criminal sentencing involves human judgment, there will inevitably be some degree of inter-judge disparity. We should not ask *whether* such disparity exists; it surely does. Rather, we

---

support reforms that the judges oppose").

87. KATE STITH & JOSÉ A. CABRANES, FEAR OF JUDGING: SENTENCING GUIDELINES IN THE FEDERAL COURTS 109 (1998) (discussing the lack of seriousness with which judges approached experimental study and the lack of detail in the scenarios).

88. *See* Hofer et al., *supra* note 80, at 268–70 (describing studies that measure disparity by matching offenders according to observable characteristics).

89. *See* Fischman, *supra* note 16, at 148–54 (discussing the values motivating research on inter-judge disparity).

should measure the size of such disparities and determine whether particular reforms have succeeded in mitigating them. Reducing inter-judge disparity was one of the primary goals of the United States Sentencing Guidelines,[90] and yet we still do not understand very well if the Guidelines succeeded in this regard, or how disparity changed after *United States v. Booker*, which rendered the Guidelines advisory.[91]

The earliest empirical studies of sentencing recognized the importance of random assignment. A series of annual reports published by New York City magistrates in 1914,[92] for example, compared the magistrates' conviction rates in different types of cases, along with their tendencies to apply various forms of punishment.[93] Although case assignment was not explicitly randomized, the magistrates rotated among the various courts throughout the year, so that "it [could] reasonably be assumed that each magistrate handle[d] practically the same class of cases as those handled by his colleagues."[94] A 1933 study by a team of criminologists[95] applied a similar approach, exploiting randomization to study inter-judge disparity in criminal sentencing in New Jersey.[96]

It was not until the 1990s, however, that sentencing researchers exploited random assignment to apply modern statistical inference in disparity studies. In a study of sentencing in three federal districts, the

---

90. *See* U.S. SENTENCING COMM'N, SUPPLEMENTARY REPORT ON THE INITIAL SENTENCING GUIDELINES AND POLICY STATEMENTS 8 (1987) [hereinafter SUPPLEMENTARY REPORT ON THE INITIAL SENTENCING GUIDELINES] ("The overriding . . . concern with the existing system . . . was directed at the apparent unwarranted disparity and inequality of treatment in sentencing of similar defendants who had committed similar crimes."); Stephen Breyer, *The Federal Sentencing Guidelines and the Key Compromises Upon Which They Rest*, 17 HOFSTRA L. REV. 1, 4 (1988) (describing one of the primary purposes of the Guidelines as "reduc[ing] 'unjustifiably wide' sentencing disparity" (quoting S. REP. NO. 98-225, at 38 (1983), *reprinted in* 1984 U.S.C.C.A.N. 3182, 3221)).

91. United States v. Booker, 543 U.S. 220, 246 (2005) (making the Guidelines advisory by severing two provisions of the Sentencing Reform Act of 1984).

92. N.Y. BD. OF CITY MAGISTRATES, ANNUAL REPORT OF THE CITY MAGISTRATES' COURTS OF THE CITY OF NEW YORK (FIRST DIVISION) FOR THE YEAR ENDING DECEMBER 31, 1914 (1914).

93. *See id.* at 48–67 (detailing and comparing magistrates' discharged cases regarding cruelty to animals, disorderly conduct, intoxication, peddling, motor-vehicle offenses, vagrancy, violation of corporation ordinances, and violation of sanitary law).

94. George Everson, *The Human Element in Justice*, 10 J. CRIM. L. & CRIMINOLOGY 90, 91 (1919).

95. Frederick J. Gaudet et al., *Individual Differences in the Sentencing Tendencies of Judges*, 23 J. CRIM. L. & CRIMINOLOGY 811 (1933).

96. *See id.* at 813 ("Since the rule is that there is no selection of the cases which the judge is to sentence but that the sentencing of a particular prisoner by a particular judge is a matter of chance (the judges rotate), it is obvious that, by chance, each judge should get an equal number of cases whose sentences would normally be long or short.").

economist Joel Waldfogel constructed tests to verify random assignment and to assess the significance of inter-judge disparity.[97] In a second article, Waldfogel proposed an approach for quantifying inter-judge disparity.[98] He started with the assumption that the *average* sentence given by judges in a district was appropriate[99]—an assumption used by the Sentencing Commission in drafting the Guidelines[100]—and then measured the squared deviation between actual sentences and average sentences.[101] This method, however, could not be used for statistical inference or to measure the significance of changes in disparity over time.

A 1999 article by James Anderson, Jeffrey Kling, and Kate Stith ("AKS") developed a sophisticated econometric technique for testing whether changes in inter-judge disparity are statistically significant.[102] The importance of this contribution warrants emphasis: assessing the impact of the Guidelines themselves—or measuring the impact of *Booker*—entails measuring whether there was a significant change in disparity. AKS developed tests for assessing whether changes in the distribution of average sentences were statistically significant,[103] and estimated the significance of these changes for a series of consecutive years from 1982 until 1993.[104] Yet no other study has ever followed the AKS approach. In part, this may be due to the fact that sentencing data with individual judge identifiers has not been widely available.[105] Yet

---

97. Joel Waldfogel, *Aggregate Inter-Judge Disparity in Federal Sentencing: Evidence from Three Districts (D.Ct., S.D.N.Y., N.D.Cal.)*, 4 FED. SENT'G REP. 151, 151 (1991).

98. Joel Waldfogel, *Does Inter-Judge Disparity Justify Empirically Based Sentencing Guidelines?*, 18 INT'L REV. L. & ECON. 293, 294 (1998).

99. *See id.* ("[W]e treat the difference between the overall average sentence for an offender with given circumstances and each judge's average sentence for an offender with the same circumstances as the only source of unwarranted disparity.").

100. *See* SUPPLEMENTARY REPORT ON THE INITIAL SENTENCING GUIDELINES, *supra* note 90, at 13–19 (describing how the Sentencing Commission used empirical estimates of average sentences to determine appropriate sentences under the Guidelines).

101. *See* Waldfogel, *supra* note 98, at 294 ("We evaluate actual sentences by their squared deviation from appropriate sentences, and we measure appropriate sentences by assuming that, except for measurable inter-judge disparity, discretionary sentences are appropriate.").

102. Anderson et al., *supra* note 84, at 279–87 (developing a zero-inflated negative binomial model in which judges are represented by random effects that are correlated across two periods).

103. *See id.* at 282–83 (describing how their approach enables inference on changes in disparity by "incorporat[ing] the estimation of the judge effects directly in a statistical model of the underlying distribution of sentence lengths").

104. *See id.* at 295–96 & fig.2 (reporting estimates of changes in inter-judge sentencing disparity between 1982 and 1993).

105. *See* Max M. Schanzenbach & Emerson H. Tiller, *Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform*, 75 U. CHI. L. REV. 715, 740–43 (2008) (discussing how the U.S. Sentencing Commission has refused to release federal sentencing data with judge identifiers).

it may also be due to the technical complexity of the AKS method, and the fact that it has not been implemented in a user-friendly form.[106] Two more recent studies found evidence that inter-judge disparity increased in the wake of *Booker*, but these studies did not address whether these increases in disparity were significant.[107]

Even the AKS method, which I believe to be the best method developed thus far for measuring disparity, has one important limitation: it only measures changes in *average* disparity. As with all observational approaches to measuring disparity, it is impossible to know how different judges would decide the same case. This is important because comparisons of average sentences could potentially mask a sizeable degree of arbitrariness in sentencing. Two judges could be lenient toward different types of offenders, yet have the same average sentence. In fact, some experimental studies have found substantial disparities in judges' proposed sentences for particular offenders, even among judges with similar average sentences.[108] One promising approach would be to combine observational and experimental methods.[109] Surveys could be used to measure the degree to which judges would sentence similarly in the same cases, while data on actual sentences in randomly assigned cases would provide the most credible estimates of changes in judges' average sentences over time.

IV. THOUGHTS FOR FUTURE RESEARCH

The preceding discussion highlights some important advances in sentencing research, but also serves as a reminder that much work remains to be done. Indeed, it is sobering how little we know about

---

106. The AKS model was implemented in MATLAB. Anderson et al., *supra* note 84, at 293 n.51. MATLAB is a technical computing language that can be used to "analyze data, develop algorithms, and create models and applications." *MATLAB*, MATHWORKS, http://www.mathworks.com/products/matlab/ (last visited Nov. 11, 2014).

107. *See* Ryan W. Scott, *Inter-Judge Sentencing Disparity After* Booker*: A First Look*, 63 STAN. L. REV. 1 (2010) (describing an increase in disparity in one district court, but not formally testing or analyzing change in disparity); Crystal S. Yang, *Have Inter-Judge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from* Booker, 89 N.Y.U. L. REV. 101 (2014) (employing a simplified version of the AKS approach to provide confidence intervals for a measure of disparity, but not testing the significance of changes over time).

108. *See, e.g.*, Fischman, *supra* note 85, at 53 (discussing studies that "[find] that only a small component of sentencing disparities can be attributed to differences in harshness and leniency, and that much of the inter-judge variation is due to heterogeneous reactions to different types of cases").

109. *See id.* at 58–59 ("One approach would entail surveying the judges regarding their responses to hypothetical cases . . . using a set of cases that are representative of [certain types of case characteristics]. Because the judges' responses would be simultaneously observable, it would be possible to estimate measures of association among the judges' decisions.").

many fundamental questions relating to sentencing policy. Although I have highlighted some studies that generated credible estimates of the effect of deterrence, such work is still far too rare, and it is difficult to draw broad conclusions from this body of research.[110] Similarly, we understand very little about how the Guidelines affect inter-judge disparity, even though concerns about disparity were a prime motivation for the Guidelines.[111]

In conclusion, I offer two brief suggestions for sentencing research. First, sentencing scholarship would benefit from greater clarity regarding its normative goals. Empirical research is positive, but the questions that motivate this research are inherently normative.[112] We do not study sentencing out of idle curiosity; we study sentencing to address policy questions, such as whether to reenact binding guidelines or reform existing guidelines. For this reason, we must be precise about how empirical findings relate to policy conclusions. This is an especially important concern in the study of sentencing disparity,[113] where the object of interest—the measure of disparity—is itself a normative construct.

Second, there is a need for richer behavioral models of criminal and judicial behavior. When reforms cannot be tested prior to implementation, we must necessarily rely on models to predict the effects of such reforms. While rational choice has an important role to play in modeling criminal behavior, more realistic models of criminal behavior could incorporate psychological and sociological perspectives.

Similarly, richer models of judicial and prosecutorial behavior are important for predicting the impact of proposed sentencing reforms. In recent years, judges and scholars have debated proposals for modified sentencing guidelines to replace the original guidelines invalidated in *Booker*.[114] Because any new guidelines will necessarily differ from the

---

110. *Cf.* Nagin, *supra* note 4, at 84–87 (arguing that only six studies of deterrence generated convincing results, and that the findings are heterogeneous).

111. *See supra* note 90 and accompanying text.

112. *See* Fischman, *supra* note 16, at 156 & n.196 (noting that empirical social science typically has normative motivations).

113. *See supra* note 89 and accompanying text.

114. *See, e.g.*, *Prepared Testimony of Judge Patti B. Saris, Chair, United States Sentencing Commission, Before the Subcomm. on Crime, Terrorism, and Homeland Security of the H. Comm. on the Judiciary*, 112th Cong. 55–60 (2011), *available at* http://www.ussc.gov/Legislative_and_Public_Affairs/Congressional_Testimony_and_Reports/Testimony/20111012_Saris_Testimony.pdf (recommending steps to strengthen the federal sentencing guidelines); Amy Baron-Evans & Kate Stith, Booker *Rules*, 160 U. PA. L. REV. 1631, 1681 (2012) (arguing against any efforts to enact binding guidelines); William K. Sessions III, *At the Crossroads of the Three Branches: The U.S. Sentencing Commission's Attempts to Achieve Sentencing Reform in the Midst of Inter-Branch Power Struggles*, 26 J.L. & POL. 305, 340–53 (2011) (describing a proposal

original guidelines, evaluating such a proposal entails predicting how judges would sentence under a counterfactual policy. It might be possible to use surveys to ask judges how they would behave under a new proposal, but judges may be loath to participate, and such surveys might lack reliability.[115] An economic approach would involve proposing a judicial utility function, estimating its parameters using historical data,[116] and using these estimates to predict judges' behavior under the proposed guidelines scheme. A richer model could incorporate the strategic responses of prosecutors, defense attorneys, and other relevant actors.

CONCLUSION

The study of sentencing is a multidisciplinary endeavor, involving collaboration among lawyers, criminologists, and social scientists from various disciplines. Economic methods are not necessarily superior to those of other disciplines, but they are well suited for addressing many questions of importance to the study of sentencing. Like economics, the study of sentencing presents many challenges, requiring researchers to disentangle multiple causal influences from observational data. Many pressing policy questions may not have clear answers, but given the stakes involved, we have no choice but to do our best.

for simplified guidelines).

115. *See supra* notes 84–87 and accompanying text (discussing concerns about the reliability of surveys in sentencing research).

116. For a preliminary approach to estimating a sentencing judge's utility function, see generally Todd Sorensen et al., *Race and Gender Differences Under Federal Sentencing Guidelines*, 102 AM. ECON. REV. 256 (2012).