6 Criminal procedure: empirical analysis

Thomas J. Miles

1. Introduction

The criminal justice system in the United States has two prominent features. The first is that the system is enormous. A common measure of its size is the number of persons currently subject to criminal punishment. For example, in 2007, US prisons and jails held nearly 2.3 million people, and another 5 million were on probation or parole (Bureau of Justice Statistics 2007). But, this stock of punished persons underestimates the number of people who flow through the preliminary stages of the criminal justice system. For example, in 2003 police made more than 13.7 million arrests (Bureau of Justice Statistics 2003), and although an exact figure is not known, the number of people stopped and questioned by police is surely a multiple of arrests. The second important fact about the American criminal justice system is its sharp racial disparities. For example, nearly 20% of black men born between 1965 and 1969 served time in prison by their early thirties, compared to only 3% of whites (Pettit and Western 2004).

The stunning magnitudes of these figures raise difficult questions about what determines how much punishment the criminal justice system delivers, to whom the system metes it out, and whether these choices are the right ones. Regarding the positive questions of the aggregate amount of punishment and its distribution, the content of substantive criminal law and the level of funding for enforcement are surely crucial factors. But, another often-overlooked influence is the law of criminal procedure.

Criminal procedure, the law governing the process of investigating and adjudicating criminal cases, influences the allocation and distribution of punishment in society by affecting the cost of bringing prosecutions. Easterbrook (1983) referred to criminal procedure as a market system that helps set the “price” of committing crimes. By determining how many resources the government must invest in order to obtain a conviction, criminal procedure affects both the level of enforcement and the choice of which crimes to enforce. Legal scholars who usually perceive criminal procedure as mainly pertaining to constitutional protections of individual

---

1 One study of stops by police in New York City showed that only 15% of stops resulted in arrests (Fagan and Davies 2000).
rights may view the economist’s attempt to understand it in terms of costs and benefits as inapposite. But, to the extent that criminal procedure influences the allocation and distribution of punishment, economics is a useful mode of examining criminal procedure.

This chapter surveys the sizable empirical literature on the economic analysis of criminal procedure. It is a diverse literature in which contributions are sometimes not expressly identified as pertaining to criminal procedure. Only rarely have economists attempted to probe the consequences of specific constitutional decisions that safeguard rights in the criminal process. Instead, economists have typically addressed a specific aspect of the criminal justice system or the behavior of a particular actor in that system. As a result, economic work in the area typically focuses on specific subtopics, such as prosecutorial discretion, racial profiling, or sentencing outcomes.

The variety of studies is likely due to absence of a single conceptual framework and limitations on available data. A single conceptual model does not unify empirical studies of criminal justice institutions in the same way that Becker’s (1968) model of deterrence guides the economists’ study of criminal behavior. The Beckerian framework is a model of offending behavior that largely abstracts from the criminal justice system. Legal institutions are relevant to the economic model of crime to the extent that they influence the expected penalty through the probability of punishment or the magnitude of the penalty. In contrast, empirical studies of particular criminal justice institutions typically do not test the deterrence hypothesis. Nor has legal scholarship provided a single conceptual framework. Some legal scholars, such as Stuntz (2001, 2006), have considered the systemic relationships between criminal procedure, substantive criminal law, and crime rates. But, empirical work by economists on criminal justice has not examined these macro-level relationships. Rather, their studies consider the operation of specific legal institutions, actors, or procedures. In lieu of a single unifying conceptual model, these studies typically present a behavioral framework specific to the particular inquiry. The models often lack wider application and are tailored to the particular data set that a researcher possesses.

In addition, a lack of readily available data hampers economic studies of the criminal justice system. No government agency assembles annual data on criminal prosecutions, sentences, and appeals at the state or local level. Although the US Sentencing Commission makes available data on the sentences of federal defendants, many important variables, including the identity of the judge, are not collected. Moreover, federal policies lack the natural cross-sectional variation of state policies, and the absence of cross-state variation hampers identification of causal effects. Economists
interested in empirical analysis of state criminal courts often engage in costly data collection. The use of different data sets limits the comparability of their results even when studies address similar research questions.

In view of these limitations, two important questions are how empirical studies of criminal justice by economists differ from inquiries on similar topics by criminologists and how they differ from economic studies of civil justice. With respect to criminology, the economic approach is distinguished by its use of the rational actor model to develop and test hypotheses about the behavior of particular actors within the justice system. Moreover, some of the economic contributions in this area are effectively studies of specific labor markets that exist within the justice system. Also, economists generally give greater attention to the identification of causal directions than other social scientists. With respect to the economic analysis of civil disputes, these models have much in common with the economic approach to criminal justice because both topics involve adjudication through litigation and importantly the possibility of settlement before trial. But, the criminal context is distinct because the litigants’ objectives are more uncertain than in civil litigation where the parties’ concerns are typically monetary. For example, it is unclear whether a defendant facing the loss of liberty would behave in the same way as a defendant facing even a large damage award.

The remainder of the chapter is organized largely around specific actors in the criminal justice system. Section 2 discusses police behavior, with particular attention to racial profiling and the exclusionary rule. Section 3 reviews research on pretrial release and bail. Section 4 examines prosecutors, their charging decisions, and their career concerns. Section 5 considers judges and their decisions in sentencing criminal defendants. Section 6 concludes.

2. Police Discretion and the Regulation of Investigations

2.1. Racial Profiling

Most economic analyses of police have focused on whether increases in the size of police forces reduce crime. These studies consider the level of policing to be a proxy for the probability of apprehension, and whether the number of police employed affects crime is then a test of Becker’s (1968) deterrence hypothesis. Levitt and Miles (2007) provide a review of these studies. In contrast, criminologists have given less attention to the level of police and instead have given more attention to the efficacy of various policing strategies, such as community policing, “hot spot” policing, and crackdowns. Nagin (1998) and Sherman (2002) review this literature. Economists have infrequently examined which policing strategies
or methods are most effective, but a few exceptions exist. Grogger (2002) studied gang ordinances in Los Angeles and found that violent crime fell substantially in neighborhoods subject to them relative to other neighborhoods. Miles (2005) examined the impact of a specific apprehension device, publicity, and reported that fugitives exposed to it experienced quicker apprehension times.

Other than its effect on offending, the aspect of policing that economists have studied most closely is the racial disparity in enforcement. For example, in city-level data, Donohue and Levitt (2001) found a correlation between the racial composition of police forces and arrests. They reported that increases in the number of minority police officers tracked increases in the arrests of whites, but not the arrests of minorities. Similarly, increases in the number of white police officers correlated with increases in the arrests of minorities, but not the arrests of whites. Both the causes and effects of these patterns were unclear. The estimates suggested that higher own-race policing reduced property crime, but had no appreciable effect on violent crime.

An aspect of the racial disparity in policing that received much attention in recent years is that police disproportionately stop and search black motorists. This fact has led to litigation alleging “racial profiling,” or the use by police of a driver’s race as a criterion in choosing whether to stop or to search, a practice that violates the Equal Protection Clause of the 14th Amendment, such as Chavez v. Illinois State Police, 251 F.3d 612 (7th Cir. 2001).

Racial profiling could arise at both the stop and the search margins. The probability that police stop a minority motorist may be higher than that for a white motorist, and conditional on stopping a motorist, the likelihood police search a minority motorist may be higher than that for a white motorist. A difficulty in assessing whether racial profiling occurs with respect to stops is that researchers do not observe the racial distribution of the population at risk of being searched. That is, the racial distribution of motorists, or in statistical terms the risk set, is not observed. The racial composition of the jurisdiction in which stops occur is a poor measure of the risk set because of racial differences in automobile ownership and travel patterns.

Grogger and Ridgeway (2006) proposed a solution to this problem. They compared the racial incidence of stops occurring at night and during the day. Police are less likely to observe the race of a motorist at night than during the day, and a comparison of the racial distribution of stops occurring during the day and at night is a test for the existence of racial profiling in motorist stops. The test requires the racial distribution of travel to be the same during the day as it is at night, an assumption which may not be
valid. A refinement of their test that addresses this objection exploits the variation in hours of daylight over the seasons. By limiting attention to twilight hours, Grogger and Ridgeway (2006) sought to hold constant the racial distribution of travel, while varying the exposure to daylight. They tested these predictions using data from Oakland, California, and their results did not support the presence of racial profiling in traffic stops. If police engaged in racial profiling, the rate at which police stop minority motorists should be higher during the day than at night, and Grogger and Ridgeway (2006) found the opposite is true. When the sample was limited to twilight hours, the difference was modest, but its direction remained contrary to what the presence of racial profiling would predict.

With respect to the margin of searches, economists sought to distinguish the two economic theories of discrimination. Becker’s (1957) model describes racial prejudice as the decision-maker’s preference for treating members of a particular race less favorably. In contrast, the decision-maker in Arrow’s (1973) model of “statistical discrimination” has no preference for discrimination, but race predicts an outcome of interest. The predictive power of race prompts the decision-maker in Arrow’s model to treat minorities less favorably while maximizing the outcome of interest.2

Knowles et al. (2001) developed a model to test whether police were Beckerian or Arrowian discriminators in conducting motor vehicle searches. The model distinguished whether racial prejudice or the goal of maximizing the number of successful searches explained the racial disparity in motor vehicle searches. The central prediction of the Knowles et al. (2001) model was an “outcomes” test. It compared the rates at which searches of the two racial groups yielded contraband. If police were not racially prejudiced and merely maximized the number of successful searches, the rate at which the marginal white motorist was found to be carrying contraband should have been equal to that for the marginal black motorist. But, if police had a preference for discrimination, they would tolerate a lower rate of successful searches among the disfavored group.

This prediction shifted empirical attention away from the differential treatment of racial groups toward differences in outcomes. Prior research commonly estimated regressions on the likelihood of search as a function of race and other characteristics, and the test for the presence of discrimination was whether the coefficient on race was nonzero. A difficulty of this approach is that researchers cannot be confident that their equations

2 It is worth noting that the economic theory of discrimination is irrelevant to whether the police practice violates the Equal Protection Clause. Either form of discrimination would be illegal.
include all variables used by officers in choosing whether to search, and if variables are missing, the estimated coefficient on race may suffer from omitted variable bias.

An innovation of Knowles et al. (2001) was to direct attention to outcomes rather than treatments. But, their key prediction of equality in hit rates pertains to the marginal motorist who is searched. Absent the full information set of the officers, the researcher cannot identify the marginal motorist and can test only average hit rates. This presents the so-called infra-marginality problem: that average hit rates may differ from marginal rates. A second contribution of Knowles et al. (2001) was to incorporate in their model the motorist’s decision to carry contraband and to show that, in the equilibrium of their model, average and marginal hit rates would be equal. The outcome test for discrimination could then be conducted using average hit rates.

Knowles et al. (2001) tested their hypothesis on a data set obtained from ACLU litigation. The data contained all vehicle searches along an interstate highway in Maryland over a four-year period. The authors could not reject that the rates at which police found contraband during searches were equal across racial groups. They also could not reject equality across racial groups in the rates of successful searches within subsamples of the data, such as age of the vehicle, whether the vehicle is a luxury model, and whether the search occurred during the day or at night. These patterns were consistent with statistical discrimination rather than racial prejudice. Persico and Todd (2006) applied an extended version of the Knowles et al. (2001) model to data from vehicle searches in Wichita, Kansas. In those data, the hypothesis that hit rates were equal across racial groups, age groups, racial groups stratified by age, and gender could not be rejected. These patterns were also consistent with statistical discrimination.

Several economists criticized the Knowles et al. (2001) finding. Dharmapala and Ross (2004) argued that in a more generalized model including varying offenses levels and imperfect observability of offenders, the appropriate test requires stratifying the data by offense severity. Dharmapala and Ross (2004) concluded that, in the absence of these variables, the Knowles et al. (2001) data could not distinguish the presence of discrimination, reverse discrimination, or the absence of discrimination. Harcourt (2007) offered another criticism. He showed that when police are assumed to prefer to minimize crime rather than maximize hit rates, the equilibrium outcome may not be unequal search rates and equal hit rates. Dominitz and Knowles (2006) develop a more formal model of this idea and argue that the extraction of information from hit rates may still be valid under certain assumptions about the joint distribution of criminality by race.
Antonovics and Knight (2009) extended the Knowles et al. (2001) framework to include officers belonging to two racial groups. They predicted that if officers were statistical discriminators who maximized hit rates, the incidence of search should not correlate with an officer’s race. But, they found in data from the Boston Police Department that officers were more likely to search members of the opposite race. The finding is consistent with Donohue and Levitt (2001). Antonovics and Knight (2009) excluded the possibilities that the assignment of officers to different neighborhoods and officers’ ability to search members of their own racial group explained the results. Rather, they attributed the patterns to preference-based discrimination.

Anwar and Fang (2006) further explored whether officers of different races varied in their search behavior. They developed a model centering on whether officers were “monolithic” in their search behavior, meaning that officers of different races searched white motorists at the same rate. Their model predicted that if officers were monolithic, officers of each race would search motorists of a given race at the same rate, and the officers would have the same hit rates in searching motorists of a given race. Anwar and Fang (2006) tested these predictions using two years of data from the Florida Highway Patrol that included information on officer characteristics. They rejected that officers of different races were monolithic in their search behavior. White officers searched motorists of all racial groups more often than black officers did, but white officers had lower hit rates with motorists of all races than black officers did.

As these contributions illustrate, the literature on outcome tests as applied to racial profiling has grown rapidly. Each new article features a model with more sophisticated conceptions of police and motorist behavior and articulates empirical predictions requiring more elaborate data sets. Persico and Todd (2006) present models with officers who are heterogenous in their tastes for discrimination and in their costs of searching and in which motorists can adjust their characteristics in response to monitoring. The methods developed for examining racial profiling are likely to have useful application in the study of racial disparities elsewhere in the criminal justice system.

2.2. The Exclusionary Rule

For lawyers and legal academics, the phrase “criminal procedure” typically refers to the law governing criminal investigations, searches and seizures, and the trial process. The core of the subject is the criminal justice process leading to the adjudication of guilt or innocence. The stages of the criminal process occurring after conviction, such as sentencing, appeals and habeas corpus, also involve complex procedures, are
often not covered in standard “criminal procedure” courses or treatises. Rather, they are cast as specialized, advanced topics. In addition, criminal procedure is effectively a subcategory of constitutional law, because the individual rights secured in the Bill of Rights largely determine how police, prosecutors, and judges must treat criminal defendants or individuals suspected of criminality.

For economists who have studied procedural aspects of the criminal justice system, it is nearly the opposite: the bulk of their work has focused on sentencing rather than the earlier stages of the criminal justice process, and when they have studied other aspects of that process, they have given little attention to the constitutional nuances. Economic research on racial profiling, described above, illustrates the latter pattern. The empirical analysis there sought to distinguish differing economic theories of discrimination. While empirical evidence on this question was illuminating, it may not directly inform the legal question of whether the police’s conduct violated the requirement of equal protection of the law. Thus, inquires of legal academics have not always aligned with interests of lawyers or the legal questions facing courts.

Two exceptions to this pattern are the modest bodies of work on the exclusionary rule and *Miranda* rights. The exclusionary rule refers to the remedy for the government’s obtaining evidence in violation of an individual’s Fourth Amendment right against unreasonable searches and seizures. Such evidence is not admissible at trial, and instead, it must be suppressed or excluded. The purpose of the rule is to safeguard the rights of individuals and to dissuade police from collecting such evidence. A common criticism of the rule is that when incriminating evidence is excluded, the government often cannot carry its burden of proof and must abandon the prosecution. Thus, the guilty defendants escape punishment, the opportunity for incapacitation and deterrence are lost, and crime rates rise.3

A small body of research has tried to measure these hypothesized social costs of the exclusionary rule. Early empirical scholarship on the rule tried to tabulate the frequency with which evidence was excluded under the rule and prosecutions were consequently dropped. Oaks (1970) and Spiotto (1973) examined motions to suppress made in preliminary hearings and at trials in the criminal courts of Chicago in 1969–70 and

---

Both found wide variation across offense categories in the rates at which motions were made, and generally high rates at which motions were granted. For example, Spiotto reported that over 70% of such motions were granted at the preliminary hearing and 20% at the trial stage. Both of these authors noted the high grant rate was likely attributable at least in part to the practical absence of any screening of cases by Chicago’s prosecutors before the preliminary hearing during that period. A decade later, Nardulli (1983) examined a sample of nine midsize counties, and concluded the impact of the rule was “marginal.” He found that motions to exclude physical evidence, identifications, and confessions occurred in fewer than 8% of cases and that these motions infrequently succeeded. Moreover, in cases in which the court suppressed evidence, the government did not necessarily drop its prosecution. Nardulli (1983) concluded that less than 0.6% of cases were lost due to the exclusionary rule. Davies (1983) provided a thorough review of prior studies and in particular criticized the National Institute of Justice’s estimates of the cases lost in California. According to data he separately collected for that state, Davies (1983) found that only 0.8% of felony arrests in California were lost due to the exclusionary rule, and the rate was highest (2.4%) in drug cases.

More recently, a pair of economists attempted to measure the effect of the exclusionary rule on crime rates. Atkins and Rubin (2003) observed that in the early part of the twentieth century, states varied in whether they applied some form of exclusionary rule, and in a 1961 decision, *Mapp v. Ohio*, 367 U.S. 643 (1961), the Supreme Court held that the exclusionary rule applies to state criminal proceedings in addition to federal ones. These variations allowed Atkins and Rubin (2003) to compare crime rates in states before and after they adopted the exclusionary rule relative to states experiencing with no change in the rule. In other words, it afforded the opportunity to estimate the effect of the exclusionary rule using differences-in-differences. Atkins and Rubin’s (2003) main results came from a panel of state crime rates covering the years 1947–1967, and their estimates were sizable. For example, robbery and property crime rates rose by 4% to 8% following the adoption of the exclusionary rule. When these estimates are compared to earlier research showing that lost cases were a relatively rare occurrence, one cannot help but wonder whether such large estimates reflect omitted variable bias. This important legal and public policy question needs further empirical investigation by economists.

A second procedural doctrine that has received significant empirical study is the Supreme Court’s decision in *Miranda v. Arizona*, 384 U.S. 436 (1966). The popularity of innumerable detective-themed entertainments...
Criminal procedure: empirical analysis

has made the four warnings of *Miranda* well known: the right to silence; the possibility that statements will be used against the speaker; the right to counsel; and if the defendant is indigent, the provision of counsel by the state. Under *Miranda*, these warnings are needed to dispel the Court’s presumption that statements made by defendants while in police custody are compelled. Exclusionary rule furnishes the remedy for a violation of the *Miranda*’s framework: the defendant’s statements are inadmissible at trial. Unsurprisingly, the same criticisms that were made of the exclusionary rule were also made of *Miranda*. Namely, that it would permit some guilty defendants to escape punishment and thus increase crime.\(^4\)

*Miranda* immediately inspired a flurry of studies that attempted to measure its effect of police behavior. The first generation of empirical scholarship of *Miranda* sought to measure whether the decision altered how police conducted interrogations and whether defendants invoked the right to silence. Cassell (1996), Schulhofer (1996), and Thomas and Leo (2002) provide convenient guides to this large literature, which consists of direct observations of police interrogations, surveys of interrogators and the interrogated, and reviews of case files. The findings of Wald et al. (1967) and Leiken (1970) are typical of this literature. They observed that police tended to follow the commands of *Miranda* and that suspects, despite receiving the warnings, usually spoke with investigators in the absence of legal counsel. These patterns do not appear to have changed in subsequent decades. Leo (1996, 1998) summarizes more recent studies showing that police faithfully recite the *Miranda* warnings while still persuading over 80% of suspects to waive their rights.

As with the early studies of the exclusionary rule, there have been attempts to calculate the number of convictions lost due to *Miranda*. Cassell (1996) asserted that *Miranda* caused a 3.8% decline in the number of convictions. He arrived at this figure by claiming that *Miranda* lowered the incidence of confession by 16% and nearly 24% of cases depended upon confessions. Schulhofer (1996) challenged these calculations and claimed that a more accurate figure was a loss of only 0.78% of criminal cases. Schulhofer believed that the number of cases in which a confession was instrumental was 19%, but more crucially, he believed *Miranda* lowered confessions by only 4.1%. In addition, Schulhofer hypothesized

---

\(^4\) *Miranda*, 384 U.S. at 542–3 (“In some unknown number of cases the Court’s rule will return a killer, a rapist or other criminal to the streets and to the environment which produced him, to repeat his crime whenever it pleases him. As a consequence, there will not be a gain, but a loss, in human dignity . . . There is, of course, a saving factor: the next victims are uncertain, unnamed, and unrepresented in this case”) (White, J., dissenting).
that these declines were temporary because over time police learn techniques that elicit confessions despite *Miranda’s* warnings.

Cassell and Fowles (1998) examined another metric of law enforcement performance, clearance rates or the ratio of arrests to reported offenses. They regressed a national time series of annual clearance rates on a set of control variables, including an indicator variable for the years 1966 through 1968. The estimated coefficient on that indicator variable was negative and statistically significant for robbery and property offenses, and they claimed that it captured the social cost of *Miranda*. This approach is less convincing than the differences-in-differences estimates of, for example, Atkins and Rubin (2003), because estimates from national time series may be biased by unobserved, contemporary events. Donohue (1998) and Feeney (2000) describe this and many other criticisms in detail. Further study by empirical economists could do much to advance our understanding of the consequences of the Fourth Amendment’s protections.

3. Bail and Pretrial Release

Bail and conditions on a defendant’s pretrial release (such as electronic monitoring) are the primary mechanisms that ensure the defendant’s appearance at trial. A long-standing policy question is what amount of bail and which conditions are appropriate. A difficulty in relating bail to the risk of flight is the endogeneity of pretrial release conditions. Judges are likely to require those with the greatest risk of flight to post the largest amounts of bail or be subject to the most restrictive bail conditions, and ordinary least squares estimates of bail on outcomes such as re-arrest or failure to appear will be biased toward zero.

The first studies by economists of bail and pretrial release often came to conflicting conclusions about the impact of bail on defendant flight because of this endogeneity. For example, Landes (1973) in studying a sample of criminal defendants in New York City found that offense type and bail amounts correlated with the likelihood of jumping bail. But, Clarke et al. (1976) in their study of defendants in Charlotte, NC, reached a different conclusion. According to their estimates, offense type and demographic characteristics were not significantly correlated with the defendant’s failure to appear. Yet, they also found that the strictness of supervision during release and court delay were strongly related to flight risk. Myers (1981) showed that estimates of the effect of bail were sensitive to plea bargaining, and this is perhaps not surprising as a defendant on pretrial release may be reluctant to accept a plea bargain which requires incarceration. In studying New York defendants, Myers (1981) found that after controlling for plea bargains, the risk of flight correlated weakly
with court delay, but it was strongly and negatively related to larger bail amounts. Helland and Tabarrok (2004) demonstrated that bounty hunters are much more effective than public law enforcement in reducing bail jumping and catching those who skip bail.

Abrams and Rolfs (2011) attempted to overcome the endogeneity problem by studying a randomized experiment in bail setting in Philadelphia in 1981. The experiment required a group of judges to follow a set of bail-setting guidelines that had the effect of lowering the amounts of bail required, and when combined with the random assignment of cases to judges, the experiment induced plausibly random variation in bail amounts. The resulting estimates implied that bail reduced the probability of re-arrest by a greater degree than was apparent in cross-sectional comparisons. Using these estimates, Abrams and Rolfs (2011) calculated that the average defendant would be willing to pay roughly $1,000 for 90 days of freedom, and they concluded that judges set bail amounts close to the socially optimal levels even in the absence of bail guidelines.

Just as they have examined the influence of race in other parts of the criminal justice system, economists have made careful study of racial disparities in bail. Ayres and Waldfogel (1994) studied criminal defendants in New Haven, Connecticut, and found that bail amounts for African-American and Hispanic men were substantially higher than for whites and women, even after controlling for observable characteristics. Their estimates also indicated that African-Americans and Hispanics paid significantly lower interest rates on their bail bonds. Ayres and Waldfogel (1994) argue that because bond dealers willingly accepted lower interest rates from minorities for use of their capital, the possibility that minorities pose greater flight risks could not justify their higher bail amounts. The approach here – asking whether racial differences in flight risk justify the racial disparity in bail amounts – is an early example of the sort of “outcome analysis” that was subsequently developed in the racial profiling literature. Ayres and Waldfogel (1994) concluded their estimates were initial evidence of racial prejudice in setting bail.

4. The Behavior of Prosecutors
The seminal model of prosecutorial behavior is Landes (1971), in which prosecutors maximized the sum of expected sentences of cases subject to a resource constraint. Prosecutors and defendants differed in their expectations of victory at trial, and Landes derived conditions for settlement and the amount of resources invested in trial. Using several years of state and federal court data, Landes tested these predictions, and found that trials were more common in jurisdictions where defendants more often secured pre-trial release and where a defendant’s legal fees were subsidized.
A sizable theoretical literature has extended Landes’ model (see, e.g., Grossman and Katz 1983; Reignanum 1988; Bjerk 2007). In addition, it inspired a flurry of early empirical investigations into the determinants of prosecutor behavior and plea bargaining. Rhodes (1976) reported that in federal courts the frequency of plea bargains correlated positively with the leniency of sentencing offers. Forst and Brosi (1977) found in data from the District of Columbia trial court that the length of time prosecutors left a case pending was positively related to the strength of the evidence against the defendant. Weimar’s (1978) estimates from a California court showed that the sentence prosecutors offered in plea negotiations rose with a proxy for the expected sentence at trial. Also, he found that the likelihood that the defendant insisted on a trial fell with the magnitude of the discount offered in a plea.

More recent empirical economic studies have given close attention to the effect of harsher penalties on prosecutorial behavior. These studies are highly relevant to policy because many legislatures recently passed mandatory minima in an effort to impose tougher penalties and to constrain judicial discretion. Under Landes’ conception of prosecutorial behavior, mandatory minimums should raise the length of the average sentence. But prosecutors may not favor higher penalties in all instances. The cost to a prosecutor of obtaining a conviction may be higher when a larger penalty induces a defendant to litigate rather than plea. Also, Tonry (1996) claimed that some prosecutors perceive mandatory sentences as unduly harsh and prosecutors avoid them by charging offenses that are similar but do not trigger application of the mandatory sentence. This response highlights a particular form of plea bargaining, called “charge bargaining,” in which prosecutors agree to seek conviction for an offense carrying a lesser penalty rather than offering a reduced penalty on the originally charged offense. Discretion to charge bargain gives prosecutors the opportunity to avoid large penalties and particularly mandatory minima.

Yet, several recent studies find that when penalty enhancements are enacted, punishments increase. This pattern suggests that either prosecutors’ discretion is not so wide that they can ignore the higher penalties, or that they prefer to impose higher penalties. Kessler and Piehl (1998) studied the consequences of California’s Proposition 8, a referendum passed in 1982 that imposed mandatory minimum sentences for particular violent offenses. Kessler and Piehl (1998) compared sentences before and after the passage of the referendum for three groups of defendants: those convicted of offenses subject to the mandatory minimum; those convicted of offenses factually similar but not subject to the mandatory minimum; and those convicted of offenses factually different and not subject to the mandatory minimum. Among their findings was that the average sentence
for robbery, a crime subject to the mandatory minimum, increased by 50% following the referendum. In contrast, the average sentence for grand larceny, which was factually similar to robbery but not subject to a mandatory minimum, increased slightly. The crime of drug possession provided a still sharper contrast. It is a crime that is factually different than robbery, and it is not subject to the mandatory minimum. Following passage of the referendum, its average sentence fell slightly. Kessler and Piehl saw their results as evidence that prosecutorial discretion is not as unconstrained as some thought and that when prosecutors exercise their discretion, they do so to raise rather than reduce sentences.

In the 1990s, several states passed three-strikes laws, which are statutes imposing enhanced penalties on repeat offenders. The cross-sectional variation in the presence of these laws and the differential timing among states that adopted them permitted Bjerk (2005) to employ a differences-in-differences methodology to estimate their effect on sentencing. He found that when an eligible defendant was arrested on a felony charge requiring application of the third-strike enhancement, prosecutors were roughly twice as likely to charge a misdemeanor instead. But, because relatively few defendants were eligible for the third strike, the fraction of cases in which prosecutors exercised this discretion remained modest. Despite these mitigating decisions by prosecutors, the three-strike law increased the average sentence of defendants eligible for the penalty enhancement. Bjerk (2005) discussed other evidence suggesting that his estimates reflected the moderating decisions of prosecutors rather than resource constraints or the behavior of other actors in the criminal justice system.

Kuziemko (2006) studied the reintroduction of the death penalty in New York in the mid-1990s. Using a differences-in-differences methodology, she found that death-eligible felony defendants accepted more severe plea bargains after the death penalty was re-established. But, the frequency of plea bargains did not change. Kuziemko interpreted these results as indicating that the presence of the death penalty altered the terms of trade in plea bargains. Her results also showed that prosecutors did not completely undo this increase in the potential penalty by declining to charge death-eligible offenses.

Piehl and Bushway (2007) examined how prosecutors exercise discretion in response to sentencing guidelines rather than enhanced penalties. Numerous legal scholars have observed that sentencing guidelines reallocate the discretion to determine sentences from judges to prosecutors. Piehl and Bushway (2007) predicted that prosecutors have greater power to charge bargain when the guidelines are mandatory for judges and when the sentencing grid of the guidelines has narrower ranges. To capture this, they measured the reduction in sentence from a charge bargain rather
Procedural law and economics

than the incidence of charge bargains, and they compared sentences in two states: Maryland, which had voluntary guidelines with wide sentencing ranges and Washington, which had mandatory guidelines with narrow ranges. The results were complicated by the fact that Washington used incarceration less frequently than Maryland. But after controlling for this difference, the guidelines in Washington had a larger impact on sentencing outcomes in plea bargaining than those in Maryland. The estimates are consistent with the view that prosecutors exercised more discretion in the state with the mandatory, narrow guidelines.

These studies might leave the impression that discretion is inevitable in the criminal process, that it may be shifted between actors in the system but not eliminated. Wright and Miller (2002) examined the district court in New Orleans, where a local prosecutor implemented an early screening process in selecting which cases to prosecute. The screening process also sought to reduce charge bargaining by discouraging changes in charges once they had been made. Although the article is not a test of an economic model, the empirical patterns were pronounced. Wright and Miller (2002) found that, under the policy, prosecutions were declined more often and plea bargains occurred less often. Consequently, the incidence of trials rose slightly, and defendants pled guilty to the initial charge more often. The results suggest that prosecutorial offices can adopt policies that limit the discretion of their line prosecutors.

Economists have increasingly given attention to broader aspects of prosecutorial preferences, such as career concerns and their productivity. These studies examine prosecutors as actors in a broader labor market rather than examine the effect of a specific criminal procedure on prosecutors. Glaeser et al. (2000) compared the characteristics of prison entrants in the federal and state systems to learn which cases prosecutors preferred. They found that federal defendants were on average older and more highly educated and that they were more likely to be female, white, Hispanic, and married. Federal defendants were also more likely to have experience in managerial or technical employment and less likely to have a prior conviction. They were more often charged with drug distribution rather than possession and hired private attorneys with greater frequency. These patterns suggest that prosecutors gained valuable human capital from prosecuting cases involving more sophisticated defendants and more complex offenses.

Boylan (2005) evaluated federal prosecutors’ career patterns by assessing how their performance in office related to success in their subsequent careers. In a sample of former US Attorneys who served between 1969 and 2000, a federal judgeship was the most preferred career outcome. The length of prison sentences correlated positively with later career success,
but conviction rates did not. Boylan’s (2005) evidence supports the Landes (1971) view that prosecutors maximize expected sentence lengths. Boylan and Long (2005) examined the career paths of line prosecutors in the federal system. They found that federal prosecutors were more likely to take a case to trial in districts where private attorneys received higher salaries. They hypothesized that lawyers seek positions as federal prosecutors in order to obtain trial experience that is valued in private practice. They documented that turnover among prosecutors was higher in districts with higher private compensation, and that departing prosecutors in higher income districts were more likely to take positions with large law firms.

The variation in experience and career paths that Glaeser et al. and Boylan and Long documented suggests that attorneys differ in their ability. Abrams and Yoon (2007) attempted to measure attorney ability in criminal cases by exploiting the random assignment of defense attorneys to cases in the Las Vegas public defender’s office. They observed wide variation in outcomes for clients that was uncorrelated with the case characteristics. But, it correlated with some attorney characteristics. Perhaps most importantly, there appeared to be returns to experience. They estimated that a public defender with ten years of experience reduced the average client’s sentence by 17% relative to a public defender in her first year on the job. Although they found no effects of gender or quality of legal education, Hispanic attorneys secured shorter sentences on average than other attorneys. The estimates of Abrams and Yoon (2007) provide strong support that the quality of lawyering can affect outcomes in criminal cases.

5. The Imposition of Punishment

5.1. The Regulation of Judicial Discretion
The central issue in sentencing policy over the past two decades has been the efficacy and desirability of sentencing guidelines. Before the advent of guidelines in the 1980s, sentencing judges had wide discretion to sentence defendants to terms falling anywhere within the statutory range, which often encompassed not only the length of sentence, but also the choice between probation and prison. The advantage of discretionary sentencing was that it afforded judges the opportunity to tailor sentences to individual defendants and to extend leniency where appropriate. But, discretionary sentencing had two prominent disadvantages. Broad judicial discretion risked compromising the principle of equal treatment under law by producing dissimilar treatment of similar defendants (Frankel 1973). Under discretionary sentencing, the sentences of defendants convicted of the same offense and having similar criminal histories exhibited wide variation. In addition, these variations were often correlated with race.
Minority defendants received more severe sentences than whites convicted of similar crimes and with similar criminal histories.

In response to these problems, states and the federal government introduced sentencing guidelines, which in many jurisdictions (including at the federal level) were mandatory rather than mere recommendations. Sentencing guidelines attempted to restrain judicial discretion in sentencing while retaining sufficient flexibility to permit individual sentences in each case (Breyer 1988; Stith and Cabranes 1998). At the federal level, the guidelines narrowed the range of permissible sentences and allowed deviations from those ranges only for a certain specified reasons. Under the guidelines, a grid largely determines sentences (US Sentencing Commission 2003). The vertical axis measures the severity of the defendant’s criminal history, and the horizontal axis measures the severity of the current offense. A complex set of rules governs calculation of the criminal history category and offense level. Each box on the grid contains a sentencing range that is much narrower than the range of punishment prescribed in the corresponding criminal statute. A judge may “depart” from the guidelines, or impose a sentence outside the range specified on the grid, only in an atypical case that presents factors or circumstances that the guidelines do not consider. By establishing this formal process for determining a defendant’s sentence, guidelines sought to reduce the influence of the identity of the sentencing judge and the demographic characteristics of the defendant in the determination of the sentence.

Criminologists and sociologists have developed a large empirical literature on sentencing, and the topic drew the interest of economists only since the enactment of sentencing guidelines at the federal level in 1989. Albonetti (1997), Hofer et al. (1999), and Bushway and Piehl (2001) provide extensive reviews of the earlier literature. Economists have generally been interested in empirically evaluating whether the guidelines succeeded in their stated goals. Much of the work by economists on sentencing falls into two categories: examinations of whether the guidelines reduced inter-judge disparities in sentencing and tests of whether they eliminated racial disparities. Recently, a third category has arguably emerged, and it analyzes whether judicial ideology influences sentences.

With respect to inter-judge disparity, a threshold question is how to measure it. Although the principle that judges should impose similar sentences in similar cases is intuitively clear, it is difficult to express in practice. Most studies take as the benchmark the average sentence conditional on offense of conviction and criminal history within a district and year. But, researchers take different approaches to calculating this disparity. Anderson et al. (1999) measure it as the mean difference in prison sentences for each judge relative to the mean prison sentence in the district
in a given year. In contrast, Hofer et al.’s (1999) metric of inter-judge disparity was the proportion of the total variation in sentences attributable to judges, as measured by whether the R-square statistic rises when the regression on sentence lengths includes fixed-effect controls for the identity of each judge.

The studies typically evaluate whether inter-judge disparity has fallen by using the introduction of guidelines as source of “before and after” variation in judicial discretion. But, the studies differ in their samples. Waldfogel (1998) examines a single federal court in California, LaCasse and Payne (1999) study two federal courts in New York, Payne (1997) analyzes two federal courts in New York and one in Pennsylvania, and Hofer et al. (1999) and Anderson et al. (1999) examine national samples of federal courts. Perhaps because of their different metrics and samples, the studies did not reach similar conclusions about the amount of inter-judge disparity and whether it fell following the introduction of guidelines. Waldfogel (1998) estimated that 9% of the variation in discretionary sentences was attributable to inter-judge disparity, but Payne (1997) found that it was only 5%. Anderson et al. (1999) estimated that the expected difference in sentence between any two judges was 17% or 4.9 months before the guidelines and 11% or 3.9 months, after the implementation of the guidelines. In contrast, LaCasse and Payne (1999) found that the amount of variation in sentences attributable to the identity of the judge did not change for plea bargains and even rose for cases that proceeded to trial. Payne (1997) described the reduction in inter-judge disparity after the introduction of the guidelines as “negligible,” but observed that the effect varied by district.

A criticism of the before-after comparisons is that they cannot control for contemporaneous but unobserved changes in the distribution of cases. A specific concern is that the introduction of sentencing guidelines at the federal level occurred at the same time as the establishment of mandatory minima for many drug crimes, which were then the most commonly prosecuted offenses in federal courts. For example, Anderson et al. (1999) found that over their study period, the average sentence rose from 24 to 35 months. Payne (1997) observed an increase in the length of drug sentences subject to mandatory minima and little change in sentence length for other offenses. The increase in the average sentence length makes it difficult to evaluate whether the reduction in inter-judge disparity was socially meaningful. For instance, Hofer et al. (1999) found that 2.32% of the variation in sentences was attributed to judge identity before the guidelines, and that this proportion fell by 1.08% after the guidelines. In terms of sentence length, the reduction was modest. The portion of the sentence attributable to judge identity was 7.87 months in the pre-guidelines period, and it was 7.61 months in the post-guidelines period.
Another important policy change that was contemporaneous (or nearly so) with the introduction of sentencing guidelines was the abandonment of indeterminate sentencing. Previously, parole boards rather than judges determined the actual length of time served by defendants. Many states switched to fixed-sentence regimes by adopting so-called truth-in-sentencing laws, and the federal government eliminated parole when it adopted its sentencing guidelines. Economists have given relatively less attention to this institutional change, and in the few instances when they have, their interest has been in its effect on the incidence of crime rather than the behavior of other actors in the criminal justice system.

Shepherd (2002) examined the effect of truth-in-sentencing laws by exploiting differential timing in the adoption of these laws across states. She concluded that determinate sentencing substantially reduced violent crimes, but caused increases in burglary and motor vehicle theft. Kuziemko (2007) used micro-level data and variation in Georgia’s parole system to estimate the impact of parole across several margins. She found that longer prison terms reduced recidivism, and that restrictions on a parole board’s discretion led inmates to decrease investments in rehabilitative activities and to experience higher rates of recidivism. Her evaluation implied that the social benefits of parole exceed its costs.

The second major line of inquiry in criminal sentencing by empirical economists has been whether “extrajudicial factors,” specifically race of the defendant, influenced judges in assigning sentences. Hagan (1974) and Kleck (1981) surveyed the very large criminological literature on this question, and a small number of economists have recently made contributions. Rather than drawing pre- and post-comparisons as much of the literature on inter-judge disparities does, these economists have examined whether a defendant’s race correlates with his sentence even during the post-guidelines period. The key prediction is that it should not, but Mustard (2001) and Bushway and Piehl (2001) found that it did.

Mustard (2001) analyzed the most comprehensive data set of federal sentences in the literature and tested whether a host of demographic characteristics correlated with sentences, even after controlling for the defendant’s criminal history and offense severity. Mustard found that males, blacks, non-citizens, and persons with less than high school education received longer sentences. The sentences of drug-trafficking defendants exhibited the widest variation. The disparities resulted principally from

---

5 Data limitations prevent these studies from examining whether the race of the defendant and the race of the victim interact in influencing the sentence in the manner found by Baldus et al. (1983) in death penalty cases.
“departures” from the guidelines rather than within-range adjustments. A departure occurs when a judge decides that the sentence recommended by the guidelines fails to take into consideration an aggravating or mitigating factor and chooses to impose a sentence different than the guidelines recommendation. Mustard (2001) found that black defendants were significantly less likely to receive downward departures. Moreover, in an equation for whether a defendant received an upward departure, the estimated coefficient was positive, implying that a black defendant had a greater likelihood of receiving an upward departure, but this difference was not statistically significant.

Bushway and Piehl (2001) studied sentences in the state courts of Maryland, a state with sentencing guidelines analogous to the federal system. These authors described in great detail the numerous differences in functional form among studies in the literature, such as whether sentence length should be estimated separately from the decision to incarcerate and whether the non-linearities in the sentencing grid warrant expressing sentence length in logarithms (and hence percentages) rather than levels. They also provided a more elaborate test of how legal and extra-legal factors influence sentencing.

The primary test in Mustard (2001), for example, was whether the coefficients on extra-judicial variables, such as the defendant’s race, were zero, and he controlled for the defendant’s recommended sentence by including a series of fixed effects for the defendant’s position in the sentencing grid. In contrast, Bushway and Piehl (2001) included the recommended sentence as a control variable, and predicted that if judges followed the guidelines, the coefficient on the recommended sentence should be unity. Their estimate for this coefficient exceeded one, which implied that as the recommended sentence increased, judges in their discretion punished these offenders more severely. In some specifications, these authors constrained its coefficient to equal one in order to help isolate the behavior of the sentencing judge from that of other actors in the criminal justice system. The choice to constrain this coefficient made little difference on the coefficient for race. They found that African-Americans receive sentences 20% longer after controlling for the sentence recommended under the guidelines. This estimate meant that although judges tended to sentence defendants with longer recommended sentences more harshly and although black defendants tended to have longer recommended sentences, race continued to exert a significant direct influence on the sentence received.

Studies of this type have two important limitations. The first is that by focusing on the post-guidelines period, these studies do not provide any insight into whether the guidelines have reduced racial disparities in sentencing. They convincingly demonstrate that race continues to influence
sentencing even under a set of mandatory guidelines, and in so doing, they provide evidence that the guidelines are not a complete success. But, they do not reveal whether racial disparities were worse before the imposition of the guidelines, and thus, they do not speak to whether the guidelines achieved partial success along that dimension.

A second criticism is that these analyses of sentence lengths are analogous to the first-generation studies of racial profiling in policing. Unless the researcher possesses all of the information of the sentencing judge, a condition which is unlikely to be satisfied, estimates of the effect of race may suffer from omitted variable bias. In addition, sentence lengths are a sort of “treatment” the judge imposes, and recidivism may be, to a substantial degree, the outcome that concerns the judge. For these reasons, a likely next step in economic studies of racial disparities in sentencing is to apply outcome tests of the sort used in the racial profiling and bail literatures to sentencing, with recidivism as perhaps the relevant outcome. A possible prediction would be that a judge engaging in statistical discrimination would impose longer sentences on minority defendants than white defendants, and recidivism rates of white and black offenders would be equal. In contrast, a judge with a preference for discrimination would impose longer sentences on minority defendants despite their lower recidivism rates. These predictions await further development and testing.

Recently, a third category of sentencing studies by economists has emerged. These articles are more closely related to the political science literature on judicial politics in that they examine not just whether judges vary in their sentencing patterns, but also whether particular judicial characteristics predict their sentencing patterns. Schanzenbach and Tiller (2007, 2008) predicted that judges differ in their policy preferences with regard to criminal cases, and consistent with that view, they found that Republican-appointed federal judges gave longer sentences on average for many crimes than Democratic appointees. Also, they predicted that the legal basis on which a judge sets a sentence will partly depend on the political composition of the appellate court that reviews the sentences for compliance with existing law. In their model, a judge chooses between rendering a decision on the basis of law or fact. Law-based decisions have precedential effect, but appellate courts review them de novo. Fact-based decisions have little or no precedential effect, but appellate courts review them with some deference. Schanzenbach and Tiller (2007, 2008) predicted that, when the majority of the appellate court shares a judge’s political affiliation, the risk of reversal is lower, and she will render the sentence on the basis of law in order to influence future cases. But when the majority of the appellate court does not share a judge’s political affiliation, the risk of reversal is higher, and she will render the sentence on the basis of fact.
In a large sample of federal sentencing cases, Schanzenbach and Tiller (2007, 2008) found evidence to support this strategic prediction. Under the federal sentencing guidelines, a judge may alter a sentence by making an adjustment, which is based on the facts of the case and has few implications for future cases, or by making a departure, which is based on law and has precedential effect for future cases. Schanzenbach and Tiller (2007, 2008) found that when a judge’s political party aligned with that of the majority of the appellate court, the judge moved the sentence in her preferred policy direction by more often making departures, a law-based criterion. But when a judge’s political party did not align with that of the majority of the appellate court, she was more likely to move the sentence in her preferred policy direction by making adjustments, a fact-based criterion.

The strategic theories of judging show that the strictness of appellate review is a potentially important constraint on a judge’s discretion in sentencing. Fischman and Schazenbach (2011, forthcoming) present further evidence on this point. They hypothesize that differences between the sentences assigned by Republican and Democratic appointees should be greater when appellate review is more deferential. Since the passage of the federal Sentencing Guidelines, the Supreme Court and Congress clarified (or outright changed) the standard of review for criminal sentences several times, and Fischman and Schazenbach (2011, forthcoming) used the fluctuations in the standard to test their hypothesis. They found that the average gap between the sentences assigned by Democratic and Republican appointees was smaller when appellate courts conducted a (stricter) de novo review rather than a (more lenient) review for an abuse of discretion. Their results suggested that standards of review constrained judicial discretion in federal sentencing.

An interesting caveat on their finding was that judges who were appointed to the bench before the passage of the Sentencing Guidelines were more likely to depart from the Guidelines and were not responsive to changes in the standard of review. The latter of these two patterns suggests that judges vary in their sensitivity to the threat of appellate review and that the ability of standards of review to cabin judicial discretion may be limited in some circumstances. The reasons for the variation in judges’ responsiveness to the standard of review remain an area for future research.

Perhaps the most important development in sentencing since the passage of the federal Guidelines has been a series of decisions which culminated in the Supreme Court rendering the Guidelines advisory. These decisions have inspired a burst of empirical research by economists on whether advisory rather than mandatory guidelines constrain the decisions of judges, and these studies represent a fourth potential category of sentencing literature by economists. The first of these decisions, Apprendi v.
New Jersey, 530 U.S. 466 (2000), held that any fact, other than the fact of a prior conviction, that raised a criminal penalty above the statutory maximum must be proven to a jury beyond a reasonable doubt or admitted by the defendant. Most commentators believed that the ruling would benefit criminal defendants by raising the cost of bringing a prosecution, and others thought that juror hostility toward criminal defendants implied Apprendi would not favor defendants. Prescott (2010) hypothesized that defendants with longer criminal histories would be closer, all else equal, to an applicable statutory maximum under the Guidelines and thus the Apprendi decision was more likely to affect their sentences. He compared the sentences received by these defendants to a group of defendants with less extensive criminal histories in the three years before and the three years after Apprendi. His differences-in-differences estimates showed that the expansion of jury trial rights under the Apprendi decision reduced the average sentence of a defendant with a lengthy criminal history by about 5% or put differently, about six months. This result supports the view that stronger jury trial rights favor defendants.

In United States v. Booker, 543 U.S. 220 (2005), the Supreme Court applied the principle of Apprendi to the federal Sentencing Guidelines and concluded that the Guidelines violated the Sixth Amendment right to a jury trial because the Guidelines permitted a judge to find a fact (other than a prior conviction) that would raise a sentence above an otherwise applicable Guideline maximum. The remedy for this constitutional problem was to make the Guidelines effectively advisory rather than mandatory. The Booker decision provided another opportunity for researchers to examine whether Guidelines succeeded in reducing the overall variation in sentences and in the disparity of sentences across judges. If mandatory guidelines had constrained judicial discretion, variation in the sentences should increase following Booker. Pfaff (2006) predicted that Booker would not produce a return of disparities of the same degree that existed before the implementation of the Guidelines. Pfaff based this prediction on the experience of state courts with voluntary guidelines in the years before Booker. Over the period of Pfaff’s study, 1989–2000, several states adopted mandatory or voluntary guidelines, and this legal variation permitted the use of a differences-in-differences identification strategy. According to Pfaff’s estimates, the adoption of mandatory guidelines reduced the variation in state sentence lengths by roughly 55% for both violent and property offenses, relative to sentences in states without any guidelines. In states that adopted voluntary guidelines, these declines were 35% for violent offenses and 21% for property offenses. Pfaff gave three possible explanations for the constraining effect of voluntary guidelines: they furnish a useful metric for judges, they represent an implicit threat...
from the legislature to impose mandatory guidelines, or that they provide a safe-harbor for judges fearing appellate reversal. The validity of these intriguing explanations awaits further investigation.

In the years since the Booker decision, enough data on federal sentences have accumulated to permit statistical analysis, and initial studies are finding increases in inter-judge disparities. For example, Scott (2010) examined the variation in sentences before and after Booker in a single federal district, Massachusetts, where the identities of individual judges could be determined. The analysis did not include many of the control variables that are now standard in the literature, such as offender demographics and criminal history. Still, the results suggested that the identity of the judge influences criminal sentences and that this effect has become larger since Booker. According to Scott’s (2010) estimates, the identity of the judge explained 31% of the variance in sentence lengths before Booker and 6.1% in late 2007 and early 2008. These percentages imply that the average sentence given by the most lenient judge differs from that of the most severe judge by more than two years. More empirical scholarship on the implications of Booker is sure to follow in the years to come.

5.2. Post-conviction Litigation

With a stock of more than seven million people currently subject to criminal supervision and a large flow of new felony convictions in state and federal courts in recent years, appeals are likely to be an important feature of the criminal justice system. Systematic data on criminal appeals in state courts are not available, but data from the federal system indicates that there has been a substantial growth in criminal appeals mirroring the upward trend in prison populations. In 1980, there were 4,405 appeals in criminal cases in federal courts, and by 1990, this figure had more than doubled to reach 9,493. It peaked at 16,060 in 2005, and in 2010, the latest year for which figures are available, 12,797 criminal appeals were filed in federal courts (Bureau of Justice Statistics, 2010). These figures represent 25% to 30% of the federal appellate courts’ annual caseload.

Despite their substantial growth and their importance in the workload of the courts, empirically-inclined economists and legal scholars have given scant attention to criminal appeals. Heise (2009) made an initial examination of criminal appeals in federal courts. He observed that appeals filed in 2006 had a reversal rate of about 12% and that his rate was lower than the corresponding rate for civil appeals in federal courts.

---

6 These figures do not include prisoner petitions, such as habeas petitions or suits over prison conditions.
Procedural law and economics

(18%) and much lower than the rate for civil appeals in state courts (32%). He also noted that the overall reversal rate masked substantial geographic variation across federal circuits. Reversal rates ranged from a low of less than 6% in one circuit to a high of just over 20% in another.

Other researchers have closely studied the reversal rates in a particular class of cases: death penalty cases. Leibman et al. (2000) gathered information on more than five thousand death sentences imposed after 1973 and reviewed by 1995. They traced the rates at which courts reversed death sentences at three different stages of appeal: direct appeal in state courts, post-conviction in state courts, and federal habeas appeal.

Cumulatively, the reversal rate for cases in which the defendant received a capital sentence was 68%. Gelman et al. (2004) examined an updated version of the Leibman et al. (2000) data and estimated the probability a capital sentence is reversed as a function of local demographic characteristics and the court’s workload. Among their findings was that reversal was more likely where death sentences were imposed more often and where the rate of incarceration per FBI Index crime was lower.

Recently, empirical scholars have also investigated the determinants of federal court civil lawsuits by jail and prison inmates. Schlanger (2003) presented a detailed portrait of prison inmate litigation and offered initial evidence that the Prison Litigation Reform Act (PLRA) of 1996 greatly reduced the volume of such litigation. Piehl and Schlanger (2004) predicted that the volume of litigation on a per-inmate basis would rise with the number of inmates as more crowding could worsen conditions. They also hypothesized that the filing rate would rise more quickly with prison populations than with jail populations in part because the flow of inmates is faster through jails, which reduces opportunities for grievances to arise. Examining a panel of states covering 1981–2001, Piehl and Schlanger (2004) found support for these predictions. They also observed that the PLRA weakened the relationship of the filing rate to prison populations but had little effect on its relationship to jail populations.

Future scholarship will likely uncover other regularities about criminal appeals and inmate litigation, and it is probable that the study of these forms of litigation will eventually involve testing behavioral theories about various actors and institutions in the criminal justice system.

6. Conclusion
This chapter surveyed the diverse and growing empirical literature by economists on the criminal justice system and the law of criminal procedure in particular. Empirical economics may on first impression have little relationship to criminal procedure because it is a topic primarily concerned with individual rights that are not well suited to quantification.
But when procedure’s influence in setting the amount and distribution of criminal punishment is seen, empirical economics offers new insights into questions such as racial profiling, plea bargaining, and sentencing.

References


Procedural law and economics


