2006

Continued Vitality of Structured Sentencing following Blakely: The Effectiveness of Voluntary Guidelines, The

John F. Pfaff
Fordham University School of Law, jpfaff@law.fordham.edu

THE CONTINUED VITALITY OF STRUCTURED SENTENCING FOLLOWING *BLAKELY*: THE EFFECTIVENESS OF VOLUNTARY GUIDELINES

John F. Pfaff*

In two recent opinions, Blakely v. Washington *and* United States v. Booker, *the U.S. Supreme Court effectively invalidated the binding nature of sentencing guidelines used by many states and the federal government over the past thirty years. Not surprisingly, numerous commentators have asserted that* Blakely *and* Booker *profoundly altered the nature of sentencing in the United States. But these claims have been made without any meaningful empirical consideration of whether viable alternatives exist.*

This Article fills that gap. It explores the extent to which voluntary, nonbinding criminal sentencing guidelines influence the sentencing behavior of state trial judges. In particular, it focuses on the ability of such guidelines to encourage judges to sentence consistently and to avoid "impermissibly" taking into account a defendant's race or sex. It also compares such guidelines to the binding guidelines found constitutionally impermissible in Blakely *and* Booker.

In general, the results indicate that voluntary guidelines are able to accomplish much, though not all, that binding guidelines did, especially with respect to sentence variation. For example, voluntary guidelines appear to reduce a measure of variation in sentence length by as much as 35 percent for violent crimes and 21 percent for property crimes. By comparison, the analogous results for binding guidelines are a 57 percent drop for violent crimes and a 54 percent drop for property crimes. For the use of impermissible factors, the results are more ambiguous; binding guidelines appear in general to be slightly more effective than voluntary, but not consistently, and voluntary guidelines still appear to reduce the role of race and sex at sentencing.

*Voluntary guidelines are not the only option available to the states in a post-*Blakely *world: States can rely on sentencing juries or forms of expanded mandatory minimums, for example. And it remains an open empirical question whether, in the end, voluntary guidelines work better than these alternatives. But voluntary guidelines nonetheless appear to be a viable, albeit somewhat less effective, alternative to presumptive guidelines in the wake of* Blakely *and* Booker.

* Associate Professor of Law, Fordham Law School. I would like to thank Ronen Avraham, Gary Becker, Lisa Bernstein, Kenworthey Bilz, Joseph Doherty, Jacob Gersen, Bernard Harcourt, Steven Levitt, Douglas Lichtman, Anup Malani, Tracey Meares, Tom Miles, Eric Posner, Daniel Richman, Max Schanzenbach, Lesley Wexler, and participants at workshops at the law schools of Fordham University, Northwestern University, the University of Chicago, USC, and UCLA for many helpful comments. All errors are my own.

INTRODUCTION

Over the past thirty years, nineteen states and the federal government have attempted to constrain judges' discretion when sentencing criminals by adopting what are called "presumptive" guidelines or "determinate" sentencing rules. Though the particular designs of these systems vary, they all share a common basic structure: A crime is assigned a range of sentences by the legislature, and a judge can deviate from that range only by identifying particular aggravating or mitigating factors. Thus, a crime like kidnapping might warrant a prison term ranging from ten to fifteen years, with a judge having the discretion to lower the minimum only if he finds that the crime was committed in a fit of passion, or to raise the maximum only if he finds that the criminal used a gun during the crime.

The U.S. Supreme Court's recent decision in *Blakely v. Washington*[1] called these presumptive guidelines and determinate sentencing regimes into question. The Court held that a jury (not a judge) must find any fact required to impose a sentence higher than the statute would otherwise allow.[2] Thus, while presumptive guidelines and determinate sentencing regimes formerly empowered the judge to identify on his own any aggravating factor such as the use of a gun, the Court held that the Sixth Amendment denies the judge this sort of authority; juries must find all relevant facts, not only those that define the crime in the first place, but also those that establish the upper boundary of the sentencing range.

Many commentators have heralded (or lamented) *Blakely*, along with its affirmation in *United States v. Booker*,[3] as introducing a sea change in sentencing practices.[4] But these claims overstate the practical importance of *Blakely*'s holding because they pay too little attention to the power of *voluntary* sentencing guidelines. This is my point of departure: Whereas *mandatory* fact-finding by judges is impermissible after *Blakely*, *suggested* fact-finding is perfectly acceptable.[5] In other words, a guideline regime that sets the default penalty for kidnapping at ten to fifteen years and requires a judge to find an aggravating fact, such as the use of a gun, before imposing a higher sentence does not survive *Blakely*, but one that only encourages the judge to impose a sentence of ten to fifteen years for kidnapping unless he finds aggravating factors does survive. This is perhaps an awkward line, but it is one the Court accepts.[6] As discussed in more depth below, the Court's opinion in *Blakely* rests on the idea that presumptive guidelines and determinate sentencing regimes create binding

1. 542 U.S. 296 (2004).

2. The holding in *Blakely* is asymmetric: The jury requirement applies only to upward enhancements, and states can still require judicial fact-finding for downward departures. This is discussed in more depth *infra* Part I.B.

3. 543 U.S. 220 (2005).

4. Kevin Reitz recently noted that "*Blakely* has been likened to a legal earthquake, a forty-car pileup, a bombshell, and a bull in a china shop. It has been called the most significant constitutional decision in criminal justice since *Miranda*—and some have opined that its full force will be greater than any past ruling in the field." Kevin R. Reitz, *The New Sentencing Conundrum: Policy and Constitutional Law at Cross-Purposes*, 105 COLUM. L. REV. 1082, 1086 (2005) (footnote omitted).

5. In *Blakely*, the U.S. Supreme Court stated that "[o]f course indeterminate schemes involve judicial factfinding, in that a judge . . . may implicitly rule on those facts he deems important to the exercise of his sentencing discretion. But the facts do not pertain to whether the defendant has a legal *right* to a lesser sentence—and that makes all the difference insofar as judicial impingement upon the traditional role of the jury is concerned." *Blakely*, 542 U.S. at 309.

6. In *Booker*, the Court "severed and excised" the portion of the federal code that made the federal sentencing guidelines mandatory, making the guidelines "effectively advisory." *Booker*, 543 U.S. at 245.

expectations that only juries can break.[7] Since suggestive guidelines do not create such expectations, they are not vulnerable to *Blakely* concerns.

This line, then, provides a straightforward way to recraft presumptive guidelines so that they comply with *Blakely*: Make them voluntary. In fact, seven states already used these types of guidelines before the Court issued *Blakely*, and the federal guidelines are no longer binding following *Booker*. On the surface, these voluntary guidelines appear to differ strongly from presumptive guidelines and determinate sentencing rules (which I will refer to collectively as "presumptive guidelines").[8] While a judge's failure to adhere to presumptive guidelines is appealable, no such recourse exists under voluntary guidelines. This seems to imply that judges have, and therefore will take advantage of, much more latitude under voluntary guidelines, undermining such guidelines' ability to regulate judicial sentencing behavior, such as the extent to which judges sentence consistently and without "impermissible" reference to the race and sex of defendants. Attorney General Alberto Gonzales expressed this very fear in a speech in Chicago in the summer of 2005.[9]

Thus, it is perhaps not surprising that some commentators such as Michael Tonry have argued that voluntary guidelines do little good.[10] But actually there are several reasons to expect judges to follow voluntary guidelines at least roughly as closely as presumptive guidelines. The first is that voluntary guidelines, like presumptive guidelines, provide information to judges. The guideline ranges serve as benchmarks for judges who wish to sentence consistently and without relying on impermissible factors but who, for various (possibly subconscious) reasons, fail to do so. Furthermore, the guidelines provide judges with a metric by which to evaluate other judges' sentencing behavior, helping the judiciary to develop and enforce informal norms of guideline compliance.

---

7. *See* discussion *infra* Part III.A.

8. This is for ease of discussion. In practice, the differences between presumptive guidelines and determinate sentencing rules are more style than substance. Under both, the law defines ranges of acceptable presumptive sentences, and judges are allowed to deviate from them only when certain conditions are met. That some use grids and other do not, for example, is not substantively important.

9. Specifically, Attorney General Alberto Gonzales said that "I fear it is inevitable over time that, with so many different individual judges involved, exercising their own individual discretion, in so many different jurisdictions, even greater disparities among sentences will occur under a system of advisory guidelines." Alberto Gonzales, U.S. Att'y Gen., Prepared Remarks at the American Bar Association House of Delegates (Aug. 8, 2005), *available at* http://www.usdoj.gov/ag/speeches/2005/080805agamericanbarassoc.htm.

10. Michael Tonry, one of the leading scholars on sentencing law and policy, once dismissed voluntary guidelines as ineffective, claiming that "[e]valuations showed that voluntary guidelines typically had little or no demonstrable effect on sentences imposed"; Tonry did admit, however, that those in Delaware may have had some effect. *See* MICHAEL TONRY, SENTENCING MATTERS 27–28 (1996).

The second possible reason why judges may adhere to voluntary guidelines is that they operate in the shadow of presumptive guidelines. Judges know that legislatures can step in and make voluntary guidelines presumptive. Even in this post-*Blakely* world, such shadows exist, since legislatures can shift the required fact-finding to juries, or they can take advantage of the asymmetry in *Blakely* (namely that the jury requirement applies only to upward, not downward, adjustments) to develop *Blakely*-compliant binding guidelines.[11] To the extent that these approaches are less appealing to legislatures than the more traditional (but now invalid) presumptive guidelines, the threat may not be as powerful as it was before *Blakely*, but it is viable nonetheless.

And the third reason state judges may adhere to voluntary guidelines (and the focus of this Article is on state, not federal, sentencing) is that their careers may depend on it. For example, judges in Virginia, a voluntary guideline state, are appointed and retained by the legislature, and they thus may have an incentive to follow the legislatively created guidelines. It is also possible that promotions to higher courts could be conditioned on whether judges comply with guidelines.

With *Blakely* and *Booker* invalidating presumptive guidelines, it is critical that we examine whether voluntary guidelines in fact work, not with anecdotes and supposition, but with data. To date, no article appears to have provided a rigorous empirical assessment of voluntary guidelines.[12] This Article fills this significant gap in the literature, testing the hypothesis that voluntary guidelines are capable of achieving the goals that presumptive guidelines were designed to accomplish. In particular, I focus on two measures of "success": (1) To what extent do voluntary guidelines reduce the variation in imposed sentence lengths; and (2) to what extent do they reduce the use of "impermissible" factors such as race and sex?

In general, the results I find are consistent with the hypothesis that voluntary guidelines can successfully confront the very problems presumptive guidelines were adopted to address, albeit perhaps not as well. For example, when the states I examine adopt voluntary guidelines, the decline in the variation in sentence lengths is between 9 percent and 35 percent; when they adopt presumptive guidelines, it is between 10 percent and 57 percent. For the

11. Part III.B discusses these alternatives in more depth.

12. This absence has been noted by other commentators as well. Kim Hunt and Michael Connelly, both executive directors of sentencing commissions in voluntary guideline jurisdictions (the District of Columbia and Wisconsin, respectively), recently stated that "[g]iven that advisory systems have rarely undergone rigorous independent review, it goes without saying that advisory guidelines have not been subjected to controlled experiments." Kim S. Hunt & Michael Connelly, *Advisory Guidelines in the Post-*Blakely *Era*, 17 FED. SENT'G REP. 233, 241 n.46 (2005).

sentences imposed on those convicted of property crimes, voluntary guidelines appear to reduce variation by as much as 21 percent, and presumptive guidelines by as little as 13 percent and as much as 54 percent. And similar patterns exist when we look at the sentences for drug offenders. The adoption of voluntary guidelines is correlated with declines in variation between 3 percent and 31 percent, and that of presumptive guidelines with drops between 22 percent and 65 percent.

Similarly, both voluntary and presumptive guidelines appear to reduce the extent to which judges look to a defendant's race and sex when imposing sentence. Again, in general, presumptive guidelines appear to outperform voluntary guidelines, but voluntary guidelines still have salutary effects. In several cases, voluntary guidelines are at least half as effective as presumptive guidelines. In two situations, they appear even three-quarters or equally as useful.

These results better allow us to think through the relative benefits of the proposals being floated following *Blakely* and *Booker* to reform structured sentencing. Making guidelines voluntary is not the only option—as discussed below, legislatures can turn to approaches such as sentencing juries, stricter sentencing laws that allow no upward adjustments by judges, and regimes that look like moving sets of mandatory minimums. And several of these systems have the appeal of retaining some presumptive elements. But these options risk imposing significant costs as well, and without a clear and rigorous understanding of the effectiveness of voluntary guidelines, it is impossible to decide if these options make policy sense.

This Article is organized as follows. Part I provides an overview of modern sentencing reforms and *Blakely*'s likely impact on them. Part II then examines the relative efficacy of voluntary and presumptive guidelines. In particular, it develops two tests: One considers how the variation in sentence length changes following the adoption of voluntary or presumptive guidelines, and the other how judicial reliance on a defendant's race and sex at sentencing changes after such adoption. Part III in turn examines the other alternatives to presumptive guidelines that have been proposed since *Blakely* and discusses some of the tradeoffs between these options and voluntary guidelines.

I. THE SCOPE AND IMPACT OF *BLAKELY*

The dissents in *Blakely* and much of the subsequent commentary suggest that the decision may profoundly alter or undermine the sentencing reforms of the past thirty years. Such claims both understate the diversity of reforms enacted since the 1970s and overstate *Blakely*'s effect on them. This part starts

with a brief history of modern sentencing reform and then examines the particular state structures threatened by *Blakely*.

A. A Brief History of Sentencing Reform

The nature of criminal sentencing in the United States has undergone an important shift since the 1970s. At the start of that decade, sentencing was dominated by the indeterminate approach: Prisoners were sentenced by judges to wide ranges of years (in some states, such as California, a criminal could receive a sentence of one year to life),[13] and parole boards determined when in fact prisoners were released. During the 1970s, however, indeterminacy was attacked on several fronts. First, and perhaps most important, judges and academics alike, led by Andrew von Hirsch and Judge Marvin Frankel, began to view indeterminate sentences as too arbitrary, as "law without order."[14] Similarly situated defendants could receive dramatically different sentences based on which judge each faced or which day each appeared before a given judge. And there was a growing concern that, either consciously or unconsciously, judges were taking into account impermissible factors such as defendants' race and sex when meting out punishments.[15] Second, to the extent that it was justified on reformative or psychotherapeutic grounds, indeterminacy became less appealing as people turned against the rehabilitative ideal.[16] And third, conservative critics feared that indeterminate sentences undermined the deterrent power of the law.[17] Under the weight of these assaults, the old system began to collapse.

---

13. In *Ex parte Jordan*, 212 P. 913 (Cal. 1923), for example, the California Supreme Court acknowledged that the sentence for robbery in California could range anywhere from one year to life in prison.

14. ANDREW VON HIRSCH, DOING JUSTICE (1976); MARVIN E. FRANKEL, CRIMINAL SENTENCES (1973).

15. Susan Klein and Jordan Steiker, for example, argue that a growing literature on the pernicious influence of race, sex, and other factors at sentencing "led to a rallying cry of conservative and liberal judges and policy makers behind Judge Marvin Frankel." Susan R. Klein & Jordan M. Steiker, *The Search for Equality in Criminal Sentencing*, 2002 SUP. CT. REV. 223, 230.

16. Edgardo Rotman notes that open-ended sentences were justified in part by viewing prisoners as patients: Doctors do not release patients until they are healthy. Edgardo Rotman, *The Failure of Reform: United States, 1865–1965*, *in* THE OXFORD HISTORY OF THE PRISON 169, 178–79 (Norval Morris & David Rothman eds., 1995). David Garland discusses the rise of the antirehabilitative "nothing works" school. DAVID GARLAND, THE CULTURE OF CONTROL 61–63 (2001). Interestingly, he demonstrates that the backlash against rehabilitation began before much data on its inefficacy was gathered. *Id.* at 63–65.

17. *See* MICHAEL TONRY, *supra* note 10, at 9 (citing JAMES Q. WILSON, THINKING ABOUT CRIME (1975)).

The result was the fragmenting of correctional policy in the United States, with every state curtailing its indeterminate system to at least some degree, but with each state responding differently.[18] From the 1970s through today, six general types of sentencing reforms gained popularity: (1) the creation of determinate sentencing systems; (2) the development of sentencing guidelines; (3) the imposition of mandatory minimums;[19] (4) the passage of two- and three-strikes laws (which are essentially forms of mandatory minimums for repeat offenders); (5) the abolition of parole; and (6) the adoption of truth-in-sentencing laws. Table 1 shows the wide variation in policies adopted across the United States.

TABLE 1
STATE SENTENCING PRACTICES

State	Abolish Parole	Guidelines	Truth-in-Sentencing	Strike Laws
Alabama				
Alaska		1980		
Arizona	1994	1977	1994	
Arkansas		1994		1995
California		1976	1994	1994
Colorado		1979		1994
Connecticut			1995	1994
Delaware	1990	1987	1990	
District of Columbia			2000	
Florida	1983	1983, abolished 1998	1995	1995
Georgia			1995	1995
Hawaii				
Idaho				
Illinois	1978	1970s	1995	
Indiana	1977	1976		1994
Iowa			1996	
Kansas	1993	1993	1995	1994

18. For works discussing this fracturing, see Michael Tonry, Nat'l Inst. of Justice, *Reconsidering Indeterminate and Structured Sentencing*, SENT'G & CORRECTIONS, Sept. 1999, at 1, *available at* http://www.ncjrs.gov/pdffiles1/nij/175722.pdf; Norval Morris, *The Contemporary Prison: 1965–Present*, *in* THE OXFORD HISTORY OF THE PRISON, *supra* note 16, at 216–18; Kevin R. Reitz, *The Disassembly and Reassembly of U.S. Sentencing Practices*, *in* SENTENCING AND SANCTIONS IN WESTERN COUNTRIES 222 (Michael Tonry & Richard S. Frase eds., 2001); Michael Tonry, *U.S. Sentencing Systems Fragmenting* (Aug. 1999), *reprinted in* PENAL REFORM IN OVERCROWDED TIMES 21 (Michael Tonry ed., 2001).

19. Mandatory minimums are the most widely adopted reforms, employed in some form by the federal government, all fifty states, and the District of Columbia.

State	Abolish Parole	Guidelines	Truth-in-Sentencing	Strike Laws
Kentucky				
Louisiana		1992, abolished 1995	1997	1994
Maine	1975		1995	
Maryland		1983		1994
Massachusetts				
Michigan		1999	1995	
Minnesota	1980	1980	1993	
Mississippi	1995		1995	
Missouri		1997	1994	
Montana				1995
Nebraska				
Nevada				1995
New Hampshire				
New Jersey		1978	1997	1995
New Mexico		1977		1994
New York			1995	
North Carolina	1994	1994	1994	1994
North Dakota			1995	1995
Ohio	1996	1996	1996	
Oklahoma			1998	
Oregon	1989	1989	1995	
Pennsylvania		1982	1911	1995
Rhode Island				
South Carolina			1996	1995
South Dakota				
Tennessee		1989	1995	1995
Texas				
Utah		1979	1985	1995
Vermont				1995
Virginia	1995	1995	1995	1994
Washington	1984	1984	1990	1993
West Virginia				
Wisconsin	1999	1985, abolished 1995, reinstated 2003		1994
Wyoming				

Notes. Dates are the years of implementation. States listed here as adopting truth-in-sentencing laws are those that require prisoners to serve at least 85 percent of their sentences. For a more complete list, see *infra* note 33.

Under determinate sentencing, state legislatures establish either specific sentences or ranges of sentences that judges are required to impose. Though

some discretion might remain with the judge, it is greatly limited. The most rigid of these systems is California's, in which a judge is often given three possible sentences (say, four, six, or eight years) and is expected to impose the middle sentence unless mitigating factors suggest the lower or aggravating factors the higher.[20] In the 1970s, Alaska, Arizona, California, Colorado, Indiana, Illinois, New Jersey, New Mexico, and North Carolina adopted such systems.[21] Though seemingly popular in the 1970s, no state has adopted such an approach since 1980,[22] and at least one state, North Carolina, ultimately replaced its determinate structure with presumptive guidelines.

Instead, like North Carolina, jurisdictions have favored guidelines and sentencing commissions. Minnesota adopted the first guideline system in 1980, and since then nineteen jurisdictions have adopted guidelines, though only seventeen currently employ them; nine use presumptive guidelines, eight voluntary.[23] Five other states currently use commissions though as of 2006 had yet to adopt guidelines.[24] The central difference between the two types is that under presumptive guidelines, judicial failure to adhere to the guidelines is appealable by both the defendant and the government, while no such appeal exists for voluntary guidelines. Besides their ability to rationalize sentencing, guidelines and commissions appeal to legislators for two central reasons: They allow a degree of coordinated policy planning not possible when judges and parole boards make case-by-case decisions, and they provide a measure of political insulation to (at least partially) shield criminal policy from sudden outbursts of public opinion.

---

20. *See* Reitz, *supra* note 18, at 224.

21. *See id.* at 225; Jon Wool & Don Stemen, Vera Inst. of Justice, State Sentencing & Corr., *Aggravated Sentencing: Blakely v. Washington*, POL'Y & PRAC. REV., Aug. 2004, at 3, *available at* http://www.vera.org/publication_pdf/242_456.pdf. These laws are discussed in more depth *infra* Part I.C.

22. Tonry asserts that "no state known to me . . . has adopted a statutory determinate-sentencing system for nearly fifteen years." TONRY, *supra* note 10, at 28.

23. The nine states with presumptive guidelines are Kansas, Michigan, Minnesota, North Carolina, Ohio, Oregon, Pennsylvania, Tennessee, and Washington. The remaining eight jurisdictions with voluntary guidelines are Arkansas, Delaware, the District of Columbia, Maryland, Missouri, Utah, Virginia, and Wisconsin. Florida adopted guidelines in 1983 but abolished them by the end of 1997, and Louisiana adopted guidelines in 1992 but abolished them in 1995.

24. As of June 2006, jurisdictions with sentencing commissions are Alabama, Alaska, Arkansas, Delaware, the District of Columbia, Kansas, Louisiana, Massachusetts, Maryland, Minnesota, Missouri, New Mexico, New Jersey, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Utah, Virginia, Washington, and Wisconsin. *See* National Association of State Sentencing Commissions, Contact List (June 2006), http://www.ussc.gov/states/nascaddr.htm. Rachel Barkow and Kathleen O'Neill provide a good summary of state sentencing commissions, including those that have since been disbanded. Rachel Barkow & Kathleen M. O'Neill, *Delegating Punitive Power: The Political Economy of Sentencing Commissions and Guideline Formation*, 84 TEX. L. REV. tbl.1 (forthcoming 2006), *available at* http://ssrn.com/abstract=905330.

Besides using guidelines, states have sought to restrain judicial discretion at sentencing through mandatory minimums. Despite trenchant theoretical and pragmatic attacks,[25] repeat offenders often face elevated minimum sanctions, and mandatory minimums have also targeted violent offenders and those convicted of violating drug and firearm statutes.[26] The most high profile of these policies are the "three-strikes" laws adopted by twenty-five jurisdictions.[27] These laws require that an offender serve a dramatically increased sentence following a third (or, for two-strikes laws, a second) conviction. In general, strike legislation has been used more for political gain than actual crime control: Of the twenty-five jurisdictions with such laws, only one uses them regularly, California, which is responsible for approximately 90 percent of all sentences under them.[28] The lower-profile mandatory minimums are thus more important.

At the same time that legislatures have limited judicial discretion over sentence length at the front end of the process, they have also sought to constrain discretion, usually that of the parole board, at the back end. Two policies in particular have been implemented to limit discretion over the actual release date. The first is the abolition of parole. Starting with Maine in 1975, fourteen states and the federal government to varying degrees abolished parole boards and their ability to release prisoners early.[29] All but three of these states eliminated discretionary parole release for all offenders. The second is the passage of truth-in-sentencing laws, which require inmates to serve a specific percentage of their sentences.[30] Many truth-in-sentencing laws came into effect following

25. *See, e.g.*, JUSTICE KENNEDY COMM'N, AM. BAR ASS'N, REPORT TO THE HOUSE OF DELEGATES 1 (2004), *available at* http://www.abanet.org/crimjust/kennedy/Punishment.pdf ("[T]he American Bar Association urges that states, territories and the federal government . . . [r]epeal mandatory minimum sentence statutes . . . ."); David A. Dana, *Rethinking the Puzzle of Escalating Penalties for Repeat Offenders*, 110 YALE L.J. 733, 737 (2001).

26. *See* Reitz, *supra* note 18, at 229.

27. The twenty-five are the United States, Arkansas, California, Colorado, Connecticut, Florida, Georgia, Indiana, Kansas, Louisiana, Maryland, Montana, Nevada, New Jersey, New Mexico, North Carolina, North Dakota, Pennsylvania, South Carolina, Tennessee, Utah, Vermont, Virginia, Washington, and Wisconsin. *See* John Clark, James Austin & D. Alan Henry, Nat'l Inst. of Justice, *"Three Strikes and You're Out": A Review of State Legislation*, RESEARCH IN BRIEF, Sept. 1997, at 7–9, *available at* http://www.ncjrs.gov/pdffiles/165369.pdf.

28. *See* Joanna M. Shepherd, *Fear of the First Strike: The Full Deterrent Effect of California's Two- and Three-Strikes Legislation*, 31 J. LEGAL STUD. 159, 159–60 (2002).

29. The fourteen states are Arizona, Delaware, Florida, Illinois, Indiana, Kansas, Maine, Minnesota, Mississippi, North Carolina, Ohio, Oregon, Washington, and Wisconsin. *See* PAULA M. DITTON & DORIS JAMES WILSON, BUREAU OF JUSTICE STATISTICS, SPECIAL REPORT: TRUTH IN SENTENCING IN STATE PRISONS 3 tbl.2 (Jan. 1999), *available at* http://www.ojp.usdoj.gov/bjs/pub/pdf/tssp.pdf.

30. *See* NAT'L INST. OF CORRECTIONS, U.S. DEP'T OF JUSTICE, STATE LEGISLATIVE ACTIONS ON TRUTH IN SENTENCING: A REVIEW OF LAW AND LEGISLATION IN THE CONTEXT OF THE VIOLENT CRIME CONTROL AND LAW ENFORCEMENT ACT OF 1994 (1995). In states that impose a range of years at sentencing, the focus is on the minimum sentence. *Id.* at 3.

the passage of the Violent Crime Control and Law Enforcement Act of 1994, which provides federal funds to states that adopt an 85 percent standard[31] for violent offenders, the focus of most (though not all) state truth-in-sentencing laws.[32] As of 1998, twenty-eight states had adopted the 85 percent level required for federal funds, and thirteen states had imposed other (generally lesser) standards.[33] Though the financial incentives offered by the federal government are often extensive—in 1998, for example, the government awarded over $92 million to just eleven states in the South, including $32 million to Florida alone—many states have declined to adopt truth-in-sentencing laws due to prison capacity constraints.[34]

Finally, it should be noted that the reforms of the past thirty years have focused not only on the issue of how to structure prison sentences, but also on how to develop alternative, nonincarceratory ("intermediate") sanctions. Several options have been considered, including boot camps, intensive supervision, house arrests and electronic monitoring, day fines or day reporting centers, community service, and fines.[35] Faced with limited capacity and limited budgets, states have experimented widely with alternatives to imprisonment. While cogent criticisms have been leveled against many of these policies,[36] they remain important elements of the reform movement of the past three decades.

---

31. 42 U.S.C. § 13704 (2000).

32. See generally, NAT'L INST. OF CORRECTIONS, *supra* note 30, which provides a comprehensive survey of the truth-in-sentencing laws in place in 1995.

33. The twenty-eight jurisdictions that adopted the 85 percent rule are Arizona, California, Connecticut, Delaware, the District of Columbia, Florida, Georgia, Iowa, Kansas, Louisiana, Maine, Michigan, Minnesota, Mississippi, Missouri, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Tennessee, Utah, Virginia, and Washington, with Illinois qualifying in 1996 only. Four states (Indiana, Maryland, Nebraska, and Texas) adopted a 50 percent rule, and three (Idaho, Nevada, and New Hampshire) require that the defendant serve 100 percent of his minimum sentence. Six other states adopted idiosyncratic rules: these are Alaska (100 percent of sentence, at least two-thirds in prison), Arkansas (70 percent for some violent crimes and one drug offense (manufacture of methamphetamine)), Colorado (75 percent for those with two priors, 56.25 percent for those with one), Kentucky (a form of 85 percent), Massachusetts (75 percent of the minimum), and Wisconsin (as of 2000, 100 percent of the term with extended supervision of 25 percent of the sentence). *See* DITTON & WILSON, *supra* note 29, at 2 tbl.1.

34. The data on grants to the South are from Todd Edwards, *Sentencing Reform in Southern States: A Review of Truth-in-Sentencing and Three-Strike Measures*, 72 SPECTRUM 10 (Fall 1999). For a discussion on state unwillingness to qualify for the grants, see NAT'L INST. OF CORRECTIONS, *supra* note 30, at 8, 9, 25.

35. For a detailed discussion of how these alternatives operate and the (often limited) extent to which they have been successful, see TONRY, *supra* note 10, at 108.

36. Tonry, for example, notes the problem of "net-widening," in which an intermediate punishment regime results in more people overall falling under the control of the criminal justice system. *Id.* at 130–31.

B. Sentencing Practices Threatened by *Blakely*

In *Blakely*, the Supreme Court faced only one of the reforms described in Part I.A, namely Washington State's presumptive guideline system. These guidelines specified for each offense a "standard range" that was less than the statutory maximum for that crime; a judge could impose a sentence above that range (and below or equal to the statutory maximum) only if he found certain aggravating factors, and these factors had to be ones "other than those which are used in computing the standard range sentence for the offense."[37] Blakely, for example, was charged with second-degree kidnapping, which carried a statutory maximum sentence of ten years but a "standard range" of forty-nine to fifty-three months. Based on facts he found about the nature of Blakely's crime, however, the judge sentenced him to ninety months, well below the statutory maximum but nearly twice the fifty-three months the judge could have imposed without an additional finding. The Court reversed the enhanced sentence as a violation of the Sixth Amendment, holding that the triggering facts had to found by a jury beyond a reasonable doubt.

Though the decision in *Blakely* struck many observers as surprising, it had in fact been anticipated by at least one federal district court judge, who held the federal guidelines unconstitutional along the same lines as *Blakely* a week before *Blakely* was issued.[38] The outcome in *Blakely* flowed directly from two earlier Court cases, *Apprendi v. New Jersey*[39] and *Ring v. Arizona*.[40] *Apprendi* held that under the Sixth Amendment, "any fact [other than evidence of prior convictions] that increases the penalty for a crime beyond the prescribed statutory maximum must be submitted to a jury, and proved beyond a reasonable doubt."[41] *Apprendi* considered a New Jersey hate-crime statute that allowed the judge to impose a sentence of up to twenty years if a crime was motivated by certain group-based animosities. Although the underlying crime of which Apprendi was convicted carried a statutory maximum of ten years, the judge invoked the hate-crime statute and sentenced Apprendi to twelve years; the Court reversed the sentence.[42] *Ring* took *Apprendi* one step further, requiring that any fact necessary for the judge to increase the sentence

37. Blakely v. Washington, 542 U.S. 296, 299 (2004) (citing State v. Gore, 21 P.3d 262, 277 (Wash. 2001)).

38. *See* United States v. Green, 346 F. Supp. 2d 259, 317 (D. Mass. 2004), *vacated in part by* United States v. Yeje-Cabrera, 430 F.3d 1 (1st Cir. 2005), *vacated by* United States v. Pacheco, 434 F. 3d 106 (1st Cir. 2006).

39. 530 U.S. 466 (2000).

40. 536 U.S. 584 (2002).

41. *Apprendi*, 530 U.S. at 490.

42. *Id.* at 471.

imposed, even *within* the statutory maximum, be submitted to a jury and proved beyond a reasonable doubt.[43] Under the Arizona law challenged in *Ring*, the statutory maximum for murder was death, but a judge could impose that maximum only if he found certain aggravating factors. The Court invalidated the scheme, holding that under the Sixth Amendment such factors had to be found by a jury. The parallels between the sentencing regimes under review in *Ring* and *Blakely* are straightforward: As in Arizona, a judge in Washington could impose a sentence outside the standard range but still within the statutory maximum only if the judge found certain aggravating facts. Like *Ring*, *Blakely* made it clear that only a jury could find such facts.

As a result of *Blakely*, most likely almost all presumptive guidelines and several determinate sentencing regimes will have to change to conform to the Sixth Amendment. In each of these systems judges are not allowed to deviate from recommended sentences without making specific factual findings, and departures made without such findings are appealable. Part I.C examines these laws more closely.

In order to appreciate the full scope of *Blakely*, however, it is also important to look at what the majority said did *not* fall within its reach. The Court distinguished *Blakely* from another sentencing case, *Williams v. New York*.[44] In *Williams*, the Court upheld a death-penalty sentencing scheme that allowed (but did not require) a judge to consider at sentencing additional facts not proven to a jury; at all times, though, the death penalty was available to the judge. As the Court in *Williams* noted, the judge could have "sentenced [the defendant] to death giving no reason at all."[45] In other words, as long as the judge is ultimately free following conviction to set a sentence without any subsequent findings, that sentence is acceptable under the Sixth Amendment, *Williams*, and *Blakely*.

This exactly describes voluntary guideline regimes, which the Court explicitly approved of in *Booker*. Consider, for example, the voluntary guidelines used in Virginia. Virginia clearly states that its guidelines are discretionary, and while it requires judges to provide written explanations for deviations from the guidelines, "failure to follow any or all of the provisions [of the sentencing statute] or the failure to follow any or all of the provisions [of the sentencing statute] in the prescribed manner shall *not* be reviewable on appeal or the basis of any other post-conviction relief."[46] Maryland's voluntary guidelines statute is equally

43. *Ring*, 536 U.S. at 608–09.
44. 337 U.S. 241 (1949); *see also* Blakely v. Washington, 542 U.S. 296, 304–05 (2004).
45. *Williams*, 337 U.S. at 252; *see also Blakely*, 542 U.S. at 304–05.
46. VA. CODE ANN. § 19.2-298.01(F) (LexisNexis 2006) (emphasis added).

direct: The guidelines adopted by the state "are voluntary sentencing guidelines that a court need not follow."[47] And while Maryland requests written explanations for departures, judges comply in only about 25 percent of the cases.[48] Thus, judges under voluntary guidelines are on a footing identical to that discussed in *Williams*: They are ultimately free to impose any sentence up to the statutory maximum without explaining their decisions.

Moreover, it is worth noting that not only are voluntary guidelines immune to *Blakely* challenges, but so too are the other major sentencing reforms of the past thirty years, namely mandatory minimums, strike legislation, the abolition of parole, and truth-in-sentencing laws. For example, neither the abolition of parole nor truth-in-sentencing involves any additional findings of fact; each simply decides when an inmate can apply for parole. Mandatory minimums and strike laws are also safe in general, though problems could arise in particular circumstances. *Blakely* and its precedents held that states have considerable latitude in restricting the floor of sentences,[49] but if the mandatory minimum following a judicial finding is greater than the upper edge of the presumptive range absent that finding, then the imposition of the mandatory sentence may be improper. As long as mandatory minimums fall within the presumptive range, however, they are acceptable following *Blakely*; mandatory minimums in indeterminate jurisdictions are even more secure.

Ultimately, most of the reforms of the past thirty years survive *Blakely*. But it is already becoming apparent in the wake of *Blakely* and *Booker* that the presumptive sentencing regimes used by thirteen states and the federal government likely suffer from important constitutional defects. Since much has already been written about *Blakely*'s (and *Booker*'s) effect on the federal system,[50] the next subpart considers its implications for the twelve affected states (other than Washington).[51]

47. MD. CODE ANN., CRIM. PROC. § 6-211(b) (LexisNexis 2006).

48. *See* Abraham BenMoshe et al., *Issues in Maryland Sentencing—Judicial Compliance with the Maryland Sentencing Guidelines: Review and Recommendations*, MD. STATE COMMISSION ON CRIM. SENT'G POL'Y, May 2001, http://www.msccsp.org/publications/issues_compliance.html.

49. The approach is clearly stated in *Harris v. United States*, 536 U.S. 545 (2002), which upheld mandatory minimums following *Apprendi*. There, the Court argued that "[w]ithin the range authorized by the jury's verdict . . . the political system may channel judicial discretion—and rely upon judicial expertise—by requiring defendants to serve minimum terms after judges make certain factual findings." *Id.* at 567.

50. *See, e.g.*, Stephanos Bibas, Blakely's *Federal Aftermath*, 16 FED. SENT'G REP. 333 (2004); Nancy J. King & Susan R. Klein, *Beyond* Blakely, 16 FED. SENT'G REP. 316 (2004).

51. Washington's guidelines are not discussed here, since *Blakely* itself discusses them at some length.

C. *Blakely*'s Effect on Diverse State Guidelines

While seventeen states (along with the federal government) employ some sort of determinate sentencing law or presumptive guideline system,[52] *Blakely* affects only thirteen of them: Four states employ sentencing regimes that do not violate *Blakely*. The first, Kansas, was held up by the majority in *Blakely* as an example to follow.[53] Following *Apprendi*, the Kansas Supreme Court held the state's sentencing practices unconstitutional,[54] and the legislature responded by introducing bifurcated trials with sentencing juries.[55] Pennsylvania's and Michigan's guidelines also appear to survive *Blakely*, since both use guidelines only to manipulate the minimum, not maximum, sentence. As the Pennsylvania sentencing commission has noted, "[T]he guidelines and most mandatories apply to minimum sentences and do not limit or control maximum sentences" and thus comport with *Blakely*.[56] The Michigan Supreme Court has similarly held that its guidelines do not violate *Blakely*.[57] In a sense, Pennsylvania and Michigan appear to have already adopted a version of Frank Bowman's proposed reforms for the federal system, which are discussed in Part III.B below.[58] Finally, Illinois nominally employs a determinate sentencing regime, but its design shields it from *Blakely*. The prescribed ranges a judge may consider are relatively wide, and mitigating and aggravating factors do not appear to allow the judge to deviate from these ranges, but instead only provide guidance for choosing points *within* them.[59] To limit the discussion, this subpart will not consider in depth the sentencing systems used in these four "immune" jurisdictions, nor the system formerly employed in the state of Washington (since it is discussed in *Blakely* itself).

---

52. As discussed *supra* Part I.A, eight states use determinate sentencing laws and nine presumptive guidelines.

53. Blakely v. Washington, 542 U.S. 296, 309–10 (2004).

54. State v. Gould, 23 P.3d 801, 811 (Kan. 2001).

55. Even Kansas's system might be defective, however. The Kansas Supreme Court held that only durational, not dispositional, upward departures fell within *Apprendi*'s ambit. In other words, a fact necessary to impose a longer sentence must be submitted to a jury, but one necessary to impose incarceration itself (rather than a lesser type of penalty) need not be. Researchers at the Vera Institute of Justice feel that this distinction likely will not stand. *See* Wool & Stemen, *supra* note 21, at 10 n.13.

56. ANNE SKOVE, ESQ., NAT'L CTR. FOR STATE COURTS, *BLAKELY V. WASHINGTON: IMPLICATIONS FOR STATE COURTS* 10 (2004), *available at* http://www.ncsconline.org/WC/Publications/KIS_SentenBlakely.pdf.

57. People v. Claypool, 684 N.W.2d 278, 283–84 (Mich. 2004).

58. Blakely v. Washington *and the Future of the Sentencing Guidelines: Hearing Before the S. Comm. on the Judiciary*, 108th Cong. 35 (2004) (statement of Frank O. Bowman, III, M. Dale Palmer Professor of Law, Indiana University School of Law, Indianapolis, Indiana) [hereinafter *Hearing*].

59. *See* 730 ILL. COMP. STAT. 5/5-8-1 (West Supp. 2006) (defining ranges); *id.* 5/5-5-3.1 to .2 (defining mitigating and aggravating factors).

Thus, twelve states besides Washington currently employ sentencing regimes that likely run afoul of *Blakely*. Of these, five use presumptive guidelines (Minnesota, North Carolina, Ohio, Oregon, and Tennessee), and seven employ determinate sentencing laws (Alaska, Arizona, California, Colorado, Indiana, New Jersey, and New Mexico).[60] Note, however, that this distinction is often one of form, not substance. Kevin Reitz, for example, classifies Ohio's sentencing system as a "presumptive narrative guidelines (no grid)" system, while Jon Wool and Don Stemen file Ohio under the heading "presumptive (non-guidelines) sentencing system[ ]."[61] That Ohio does not use a grid like the federal system or North Carolina does not make its system any less structured.

Sentencing under presumptive guidelines in state courts often differs dramatically from that in federal courts, and these differences often have important policy implications. Perhaps the greatest difference is flexibility, as state judges under presumptive guidelines retain much more discretion than their federal counterparts. Oregon's system, for example, is quite permissive, appearing to limit only upward departures.[62] Minnesota provides even less guidance, allowing judges to depart from the prescribed range for "substantial and compelling" reasons, limited only by the risk of reversal.[63] Even more rigid states, such as North Carolina and Tennessee, adopt less deterministic approaches than the federal guidelines. Like the other states, they allow judges to sentence anywhere in a set upper range if they find that aggravators dominate mitigators, or in a lower range if the opposite occurs, rather than assigning points to every aggravator or mitigator (as is the case with the federal guidelines).[64]

---

60. *See* Wool & Stemen, *supra* note 21, at 3. Jon Wool and Don Stemen also suggest that Michigan, Pennsylvania, and some states using voluntary guidelines may be vulnerable as well. *Id.* at 2–3. As discussed above, the Pennsylvania Sentencing Commission and the Michigan Supreme Court have both stated that their systems comport with *Blakely*'s requirements. Wool and Stemen argue that some voluntary guidelines that require judges to explain their departures could violate *Blakely* if such explanations require any findings; *Blakely* problems would arise only if state appellate and supreme courts interpret reporting requirements in ways that make the guidelines less voluntary and more presumptive. They acknowledge that this outcome is unlikely. *Id.* at 5.

61. Reitz, *supra* note 18, at 226 tbl.6.1; *see also* Wool & Stemen, *supra* note 21, at 3.

62. Departures are governed by OR. ADMIN. R. 213-008-0003 (2006) (durational departures), and OR. ADMIN. R. 213-008-0005 (dispositional departures). OR. ADMIN. R. 213-008-0003(2) restricts upward departures to no more than double the presumptive sentence; the rule imposes no restrictions on downward departures.

63. MINN. SENT'G GUIDELINES II.D (2006) states that, when departing from the presumptive sentence, "it is recommended that the judge pronounce a sentence that is proportional to the severity of the crime for which the sentence is imposed and the offender's criminal history, and take into consideration the purposes and underlying principles of the sentencing guidelines." Though not in violation of *Blakely*, Michigan provides similarly generous departure provisions. *See* MICH. COMP. LAWS ANN. § 769.34(3) (West Supp. 2006) (allowing departures); *id.* § 769.34(11) (regarding appellate review).

64. For North Carolina, see N.C. GEN. STAT. § 15A-1340.17(c) (2005). For Tennessee, see TENN. CODE. ANN. § 40-35-210(c)–(e) (Supp. 2005).

Sentencing in states usually classified as determinate sentencing states is often quite flexible as well. Alaska, for example, sets specific sentences for certain offenses but allows the judge great flexibility if he finds aggravators or mitigators present;[65] most offenses likely do not fall within the guidelines at all.[66] In New Mexico, a judge is free after making the necessary findings to vary a sentence by as much as one-third in the appropriate direction.[67] The Ohio code establishes fairly wide ranges for various felony classes and, while stipulating that judges must sentence at the bottom of the range unless aggravators are found, allows them to choose any sentence but the highest if "[t]he court finds on the record that the shortest prison term will demean the seriousness of the offender's conduct or will not adequately protect the public from future crime by the offender or others."[68] Colorado also establishes generous presumptive ranges and permits judges to move beyond them if aggravators or mitigators are found.[69] Finally, Arizona, Indiana, and New Jersey—like Alaska—fix specific presumptive sentences but allow judges fairly wide discretion if aggravators or mitigators are found.[70]

The one exception to the claim that states have relatively flexible systems is California, which uses a determinate sentencing structure for many felonies.[71] For designated offenses, three possible sentences are provided: a lower, middle, and higher. Judges are required to impose the middle sentence unless they find aggravating or mitigating factors present; if so, they can then move to the

65. ALASKA STAT. § 12.55.125 (2004) establishes presumptive sentences (not ranges, but specific numbers of years) for certain classes of offenses, and *id.* § 12.55.155 provides a list of aggravators and mitigators a judge may consider if established by clear and convincing evidence, as well as the extent to which he can deviate from the presumptive sentence. If the presumptive sentence is fewer than four years, the judge can reduce the sentence to zero or raise it as high as the statutory maximum; if more than four years, the judge can cut the sentence by as much as 50 percent or raise it as high as the statutory maximum.

66. Skove states that it appears only a small fraction of Alaskan cases fall within the structured sentencing regime. SKOVE, *supra* note 56, at 8.

67. N.M. STAT. ANN. § 31-18-15.1C (2000).

68. OHIO REV. CODE ANN. § 2929.14(B)(2)–(C) (West Supp. 2006). Some of the ranges are on the order of six or seven years. *See id.* § 2929.14(A).

69. COLO. REV. STAT. § 18-1.3-401(1)(a)(V)(A) (2005) establishes ranges, and *id.* § 18-1.3-401(6) allows the judge, after the necessary findings, to impose a sentence up to twice the presumptive maximum or half the presumptive minimum.

70. For Arizona, see ARIZ. REV. STAT. ANN. § 13-701 (2001) (fixing exact sentences); *id.* § 13-702 (establishing ranges); *id.* § 13-702.01 (providing for sentences outside of the ranges set in § 13-702). For Indiana, see, for example, IND. CODE § 35-50-2-4 (LexisNexis 2004) (establishing a presumptive sentence of thirty years for a Class A felony, with an upper bound of twenty additional years if aggravators dominate, and a lower bound of ten fewer years if mitigators do). For New Jersey, see N.J. STAT. ANN. § 2C:44-1(f) (West 2005) (establishing specific presumptive sentences); *id.* § 2C:43-6(a) (establishing departure ranges).

71. CAL. PENAL CODE §§ 1170–1170.16 (West 2006).

higher or lower sentence.[72] Judges lack discretion as to the magnitude of the sentence, however, since they can only select one of three fixed points.

Many state guidelines also differ from the federal guidelines in their focus on resources. Unlike the federal government, state governments face much tighter budget constraints and must adjust their policies accordingly.[73] Minnesota's guidelines, for example, list four underlying principles, one of which is the need to conserve correctional resources.[74] Oregon's guidelines similarly give such concerns a prominent role.[75] This attention to capacity and resources can have powerful implications: While the federal prison population has been one of the fastest growing in recent years, the adoption of state sentencing guidelines has often been correlated with slower population growth.[76]

Notwithstanding their flexibility, presumptive state sentencing structures outside of Illinois, Kansas, Michigan, and Pennsylvania are still likely unconstitutional following *Blakely*. Most obviously, the enhancement provisions are no longer constitutional. In each of the twelve states in question, enhancements are found by judges and not beyond a reasonable doubt.[77] This is no longer

---

72. *Id.* § 1170(b).

73. It is possible, for example, that some of the growing leniency observed in state sentencing is due not to changing values but to growing budget constraints. *See, e.g.*, JON WOOL & DON STEMEN, VERA INST. OF JUSTICE, CHANGING FORTUNES OR CHANGING ATTITUDES? SENTENCING AND CORRECTIONS REFORMS IN 2003 1–2 (2004), *available at* http://www.vera.org/publication_pdf/226_431.pdf. Kate Stith and Steven Koh note that the federal guidelines do not pay any meaningful attention to prison resources, and argue that the Guidelines' "so-called 'prison capacity' mandate is no mandate at all." Kate Stith & Steven Y. Koh, *The Politics of Sentencing Reform: The Legislative History of the Federal Sentencing Guidelines*, 28 WAKE FOREST L. REV. 223, 267 (1993).

74. MINN. SENT'G GUIDELINES I.3 (2006) asserts:

> Because the capacities of state and local correctional facilities are finite, use of incarcerative sanctions should be limited to those convicted of more serious offenses or those who have longer criminal histories. To ensure such usage of finite resources, sanctions used in sentencing convicted felons should be the least restrictive necessary to achieve the purposes of the sentence.

75. OR. ADMIN. R. 213-002-0001(3)(a) (2006) states:

> The response of the corrections system to crime, and to violation of post-prison and probation supervision, must reflect the resources available for that response. A corrections system that overruns its resources is a system that cannot deliver its threatened punishment or its rehabilitative impact. This undermines the system's credibility with the public and the offender, and vitiates the objectives of prevention of recidivism and reformation of the offender. A corrections system that overruns its resources can produce costly litigation and the threat of loss of system control to the federal judiciary. A corrections system that overruns its resources can increase the risk to life and property within the system and to the public.

76. Thomas Marvell finds that five of the nine states with presumptive guidelines saw prison growth rates fall after adoption; one exception was Pennsylvania, which rejected earlier guidelines for being too lenient. Thomas B. Marvell, *Sentencing Guidelines and Prison Population Growth*, 85 J. CRIM. L. & CRIMINOLOGY 696, 703–04 (1995).

77. For Alaska, see ALASKA STAT. § 12.55.155(f) (2004) ("Factors in aggravation and factors in mitigation must be established by clear and convincing evidence before the court sitting

possible under *Blakely*. As many commentators predicted, *Blakely* has led to the invalidation of the federal guidelines as well (or at least their binding effect).[78] And several state courts have already found their presumptive guidelines unconstitutional as well.[79] These jurisdictions will thus need new sentencing regimes.

---

without a jury."). For Arizona, see ARIZ. REV. STAT. ANN. § 13-702.01 (Supp. 2005) (allowing the judge to deviate from the presumptive sentence if aggravating or mitigating factors are found). For California, see CAL. CT. R. 4.420(b) ("Circumstances in aggravation and mitigation shall be established by a preponderance of the evidence."). For Colorado, see COLO. REV. STAT. § 18-1.3-401(6) (2005) ("[T]he court shall impose a definite sentence . . . within the presumptive ranges . . . unless it concludes that extraordinary mitigating or aggravating circumstances are present . . . ."). For Indiana, see IND. CODE ANN. § 35-38-1-7.1 (LexisNexis Supp. 2004) (setting forth aggravating and mitigating factors, though not establishing the burden of proof), and *id.* §§ 35-50-2-4 to -6 (establishing sentences for felonies and allowing judge to depart from the fixed sentence if aggravators or mitigators are found). For Minnesota, see MINN. SENT'G GUIDELINES II.D (2006) ("[I]n exercising the discretion to depart from a presumptive sentence, the judge must disclose in writing or on the record the particular substantial and compelling circumstances that make the departure more appropriate than the presumptive sentence."). For New Jersey, see N.J. STAT. ANN. § 2C:44-1(f) (West 2005) (instructing the judge to impose the presumptive sentence unless the "preponderance of aggravating or mitigating factors" suggests otherwise). For New Mexico, see N.M. STAT. ANN. § 31-18-15.1A (Supp. 2000) ("The court shall hold a sentencing hearing to determine if mitigating or aggravating circumstances exist . . . ."). For North Carolina, the burden of proof was initially preponderance of the evidence, but the legislature amended it in response to *Blakely*. *See* N.C. GEN. STAT. ANN. § 15A-1340.16(a) (2005) ("The court shall consider evidence of aggravating or mitigating factors present in the offense that make an aggravated or mitigated sentence appropriate . . . . The State bears the burden of proving beyond a reasonable doubt, that an aggravating factor exists, and the offender bears the burden of proving by a preponderance of the evidence that a mitigating factor exists."). For Ohio, see OHIO REV. CODE ANN. § 2929.14(B)(2) (West Supp. 2006) (The court shall impose the minimum sentence unless it finds such a sentence "demean[ing]" or inadequate to protect the public.). For Oregon, see OR. REV. STAT. § 137.671(1) (2005) ("The court may impose a sentence outside the presumptive sentence . . . if it finds there are substantial and compelling reasons justifying a deviation from the presumptive sentence."). For Tennessee, see TENN. CODE ANN. § 40-35-210 (Supp. 2005) (while providing for the judge to find enhancements, neither this section nor any other appears to establish the burden of proof).

78. *See* United States v. Booker, 543 U.S. 220 (2005).

79. The states that have so far found that their guideline systems violate *Blakely* are Arizona (State v. Brown, 99 P.3d 15 (Ariz. 2004)), Colorado (Lopez v. People, 113 P.3d 713 (Colo. 2005)), Indiana (Smylie v. State, 823 N.E.2d 679 (Ind. 2005)), Minnesota (State v. Shattuck, 689 N.W.2d 785 (Minn. 2004)), New Jersey (State v. Natale, 878 A.2d 724 (N.J. 2005)), North Carolina (State v. Allen, 615 S.E.2d 256 (N.C. 2005)), Ohio (State v. Foster, 845 N.E.2d 470 (Ohio 2006)), and Oregon (State v. Dilts, 103 P.3d 95 (Or. 2004)). And Alaska's legislature has already amended its guidelines to ensure compliance with *Blakely* in S. 56, 24th Leg., 1st Sess. (Alaska 2005), *available at* http://www.legis.state.ak.us/PDF/24/Bills/SB0056A.pdf. The supreme courts in three other states that use presumptive sentencing regimes have sought to distinguish their guideline systems from those discussed in *Blakely* and *Booker*, but it is unclear whether these arguments will ultimately prevail. As of now, however, these states are California (People v. Black, 113 P.3d 534 (Cal. 2005)), New Mexico (State v. Lopez, 123 P.3d 754 (N.M. 2005)), and Tennessee (State v. Gomez, 163 S.W.3d 632 (Tenn. 2005)).

II. THE EFFICACY OF VOLUNTARY GUIDELINES

Following *Blakely*, voluntary guidelines provide a way to retain structured sentencing with judicial fact-finding: Where a presumptive guideline system *requires* a particular finding, a voluntary approach simply *encourages* it. But can such encouragement alone sufficiently influence criminal sentencing? This part examines this question empirically, and the results suggest in general that voluntary guidelines work reasonably well.

To design an empirical test, however, it is first necessary to define a metric by which to measure "success." This Article focuses on two metrics: (1) Do guidelines shrink the variation in sentence length; and (2) do they reduce the role of impermissible factors such as race and sex?

To examine how well guidelines induce consistency, Part II.A.2 explores the effect of guideline adoption on the "mean difference" of sentence lengths, which is the extent to which a defendant's sentence differs (in absolute value) from the average state sentence for his offense.[80] In general, the findings indicate that voluntary guidelines successfully reduce variation, though not necessarily as well as presumptive guidelines.

Part II.A.3 then presents the results for the test looking at how guideline adoption influences the importance of "impermissible" factors, race and sex in particular, to sentence length. The results here are more ambiguous. At least for violent and property crimes, both voluntary and presumptive guidelines are capable of reducing the use of race and sex, but presumptive guidelines generally appear to have stronger effects; for drug offenses, both types of guidelines often appear to have only limited effects. Limitations in the data, however, imply that these results be viewed as tentative.

There are, of course, other measures of success—the Pennsylvania legislature and the U.S. Congress, for example, adopted guidelines in part because they felt judges were too lenient, so for them success depends somewhat on longer average sentences. But reduced variation and less reliance on impermissible factors are among the most commonly advanced justifications for structured sentencing; Part II.A.4 considers in part what the empirical results presented here imply for other possible justifications.

80. For example, if the average defendant convicted of arson in Virginia in 1997 received a sentence of ten years, then a defendant convicted of arson in Virginia in 1997 but sentenced only to three years would have a mean difference score of 7 (10 – 3). Similarly, if the defendant's sentence had been seventeen years, his mean difference score would also be 7—the score is the same whether the defendant's sentence is above or below the mean.

A. Two Tests of the Effectiveness of Voluntary Guidelines

1. The General Model

Both tests discussed in this part draw on the same primary source of data, the National Corrections Reporting Program (NCRP) between the years 1989 and 2000. The NCRP takes the prisoner as the unit of observation: Participating states compile data on each convicted defendant both upon arrival at and departure from prison. Most important, the NCRP provides data on each defendant's offense, the length of the sentence imposed, and demographic information such as race and sex.[81] Sentence length (or a function of it, such as its mean difference) is the dependent variable in the regressions below, and the remaining variables (such as the inmate's race and sex) are factors that, along with the presence of guidelines, should influence sentence variation and length. Some relevant defendant-specific traits, such as the defendant's education, employment history, and socioeconomic status, or the race, age, and sex of any victims, are not included only because the NCRP does not provide usable data on any of these other variables. Appendix 1 provides a more detailed discussion of the NCRP.

Defendant traits are not the only ones that play a role at sentencing, however; those of the judge are also important. A judge's political views, race, sex, and (outside of the federal system) political vulnerability all might affect the sentences defendants receive. An earlier version of this Article attempted to explicitly control for these effects, but the tables presented below do so more indirectly; Appendix 1 discusses the reasons for this shift.

Of the fourteen states providing useful data to the NCRP between 1989 and 2000, two adopted voluntary guidelines during the period (Missouri in 1997 and Virginia in 1995),[82] and two presumptive (Ohio in 1996 and Michigan in 1999).[83] Moreover, four reporting states used guidelines throughout the entire

81. Unfortunately, the National Corrections Reporting Program (NCRP) does not record data about the prisoner's life prior to incarceration—it provides no data on prior criminal history, for example. The NCRP does contain a variable for prior time spent in prison, but this measures only the time the prisoner spent in prison (as opposed to jail) awaiting trial for the offense for which he is currently incarcerated and which can be credited to the sentence he is serving.

82. Virginia initially adopted voluntary guidelines in 1991 and amended them when it revised its sentencing statutes and adopted a truth-in-sentencing law in 1995. For reasons discussed in Appendix 1, this Article uses 1995 as the date of adoption. It should be admitted up front, however, that when 1991 is used, the results for voluntary guidelines are significantly weaker, and are in fact close to zero. But, as argued in the Appendix, this weak effect seems to understate the true effect of voluntary guidelines, at least for Virginia.

83. Michigan's guidelines also raise some empirical concerns, which are discussed in Appendix 2. When the regressions in Part II.A are run distinguishing Michigan from Ohio, the

period: One (Maryland) used voluntary guidelines, two (Tennessee[84] and Washington) presumptive, and one (California) a determinate sentencing law (which is effectively a presumptive system). Finally, six reporting states (Alabama, Illinois,[85] Kentucky, Nevada, Oklahoma, and South Carolina) did not use any sort of guidelines throughout the period; these act as a form of experimental control group.

I ran each regression below separately for each of three classes of crimes, namely violent crimes, property crimes, and drug offenses. This division is justified on two grounds. The first is quantitative: As the summary statistics in Table 2 indicate, the average mean difference and sentence length differ noticeably across the three categories. The second reason is qualitative: The psychological impact of the offenses at sentencing likely varies across categories. Violent crimes, for example, may have the strongest emotional effect on a judge; property crimes, though still clearly malum in se, lack the visceral impact of violent crimes. These differences could influence the extent to which judges adhere to guidelines, look to impermissible and outside factors, or impose consistent sentences even in the absence of guidelines.

results for Ohio alone are close to those for the aggregate measure of presumptive guidelines, if only because Michigan adopted its guidelines so late in the sample period. Part III.B and Appendix 2 examine Michigan's guidelines and what the data in this Article can say about them in more depth.

84. Technically, Tennessee adopted its guidelines toward the end of 1989, but for reasons discussed *infra* note 173, it is not treated as an adopting-in-the-period state.

85. Though nominally a determinate system, the latitude Illinois's sentencing regime affords judges is quite wide. *See supra* note 59 and accompanying text.

TABLE 2
SUMMARY STATISTICS

Variable	No. Obs.	Mean	St. Dev.	Min.	Max.
Dependent Variables					
Mean Difference					
Violent	640,011	35.701	48.464	0	572.971
Property	934,261	22.195	33.059	0	568.733
Drugs	785,789	18.385	30.695	0	571.657
Ln(Mean Difference)					
Violent	640,011	2.978	1.183	0	6.353
Property	934,261	2.600	1.012	0	6.345
Drugs	785,789	2.329	1.122	0	6.35
Sentence Length					
Violent	640,011	83.331	87.630	1	600
Property	934,261	46.002	51.133	1	600
Drugs	784,434	45.395	51.696	1	600
Guidelines					
Voluntary					
Violent	640,011	0.072	0.258	0	1
Property	934,261	0.066	0.249	0	1
Drugs	785,789	0.082	0.275	0	1
Presumptive					
Violent	640,011	0.560	0.496	0	1
Property	934,261	0.504	0.500	0	1
Drugs	784,434	0.609	0.488	0	1
Demographics					
Black					
Violent	640,011	0.464	0.499	0	1
Property	934,261	0.412	0.492	0	1
Drugs	785,789	0.536	0.499	0	1
Women					
Violent	639,985	0.051	0.220	0	1
Property	934,192	0.111	0.314	0	1
Drugs	785,762	0.116	0.321	0	1

Within each class of crimes, I ran two specifications of each regression, differing in the number of states included. The first specification includes just those four states that adopted guidelines during the period covered by the data, and the second includes those four plus the other ten states in the sample. Otherwise, the regressions include the same sets of control variables, including state fixed effects, year fixed effects, and offense fixed effects.[86]

86. State fixed effects control for the possibility that some of the differences in outcomes between guideline and nonguideline states or between presumptive guideline and voluntary

These regressions lag the timing of guideline adoption by one year. In other words, since Ohio adopted its guidelines in 1996, it is not classified as a guideline state until 1997. This reflects the fact that many criminals sentenced the year a guideline system is adopted are sentenced under preguideline rules. For example, any criminal arrested or convicted in Ohio in 1995 but sentenced in 1996 would be sentenced under the preguideline rules. By the time a year has passed, it is likely that most offenders facing sentencing will be subject to the guidelines.

a. Endogeneity

As in any test looking at the effect of legislative changes, I must address the potential problem of endogeneity. No state randomly adopted guidelines or had guidelines forced upon it by an outside power. The timing of guideline adoption is therefore endogenous, implying that the guideline coefficients could be biased. The direction of this bias depends on what forces influence the timing of adoption, and two possible stories can be told. The first is that legislatures adopt guidelines when they feel that problems with sentencing are worsening (when variation is rising, for example). In this case, the bias is upward, pulling the coefficients toward zero: The negative coefficients produced in the regressions discussed below would thus understate the true effect of guidelines.[87] The second story is that legislatures adopt guidelines only when the problems that concerned them are starting to wane; this could simply be the result of legislative inaction or perhaps a public-choice story, in which the targeted group (here, judges) is able to delay implementation of a

guideline states are due not to the guidelines themselves but rather to other, unobserved underlying state factors (for example, even without guidelines, judges in Virginia might sentence offenders differently than those in Michigan). Year fixed effects similarly capture the extent to which different (nationwide) outcomes in different years are due not to the specific policy changes under review but to other, unobserved events that took place in those years. And offense fixed effects, which are discussed in more depth in Appendix 1, take into account the fact that "Violent Crime," "Property Crime," and "Drug Crime" are broad categories incorporating a diverse array of suboffenses. Judges may treat these suboffenses (such as murder and simple assault, both of which are elements of "Violent Crime") differently at sentencing, and offense fixed effects help control for this possibility.

87. The intuition here is the following. On the one hand, the adoption of guidelines should reduce the variation in sentencing. On the other hand, if legislatures adopt guidelines when sentence variation is getting worse (larger), then guideline adoption is positively correlated with increased sentence variation. When variation is regressed on guideline adoption, the resulting coefficient ultimately accounts for both the negative "real" effect and the positive "correlated" effect, resulting in a coefficient less negative than the true effect. In other words, roughly speaking, assume that if guidelines were imposed exogenously, they would on average reduce the mean difference by ten months, but legislatures in fact adopt them only when the mean difference has *risen* by three months. Then the regression will return a result of −7 (−10 + 3) months, not −10 months, understating the true effect.

regulation until the problem has lessened. In this case, the bias is negative, and the negative coefficients returned in the tests below overstate the true effect of guideline adoption.[88]

A simple extension can shed some light on the extent to which endogeneity is a problem.[89] If the passage of guidelines is in part a function of changes in the variation in sentencing, for example, then we should observe a trend in variation prior to guideline adoption; for example, if states adopt guidelines because variation is worsening, then variation should rise during the preadoption years. Including some form of state-specific time trends in the regressions run below addresses this concern. In particular, I consider two adjustments. In the regressions looking at variation, one correction includes a linear state-year trend, while the other includes both linear and quadratic state-year trends.[90] For the impermissible-factors regressions, the corrections are slightly more complex. Besides the linear and quadratic state-year trends used in the variation regressions, the impermissible-factor regressions include four more trend terms, namely linear and quadratic state-year-race and state-year-sex trends.[91] In other words, the regressions control not only for possible trends in overall sentence length, but also for trends in the relative sentences imposed on blacks and women.

Note that from an empirical perspective, *Blakely* provides a useful (future) alternative for addressing the problem of endogeneity: It is a random shock to sentencing practices. Washington State adopted its guideline system endogenously but lost it exogenously; this is true for the other states affected by *Blakely* as well. In several years, when sufficient post-*Blakely* sentencing data are available, the importance of guidelines can be (re)tested by looking at how sentencing practices changed when such structures were unexpectedly removed.

---

88. The intuition here mirrors that in the first case. In this instance, the "real" effect remains negative, but the "correlated" effect is negative as well, since guidelines are adopted when the targeted effect (variation, for example) is declining. To recast the example *supra* note 87, now legislatures adopt guidelines only after the mean difference has fallen by three months, so the regression returns a coefficient of –13 months instead of the true –10 months.

89. This approach is derived from Leona Friedberg's examination of the endogeneity of state unilateral divorce laws. Leona Friedberg, *Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data*, 88 AM. ECON. REV. 608 (1998). Note that because the possibly endogenous variable in this model is a state-level legislative outcome while the data are at the individual level, the instrumental variable approach for addressing endogeneity is inapplicable.

90. In other words, the linear correction includes solely the adjustment $state_k*year_t$, while the linear and quadratic correction (which I will call simply the quadratic correction for ease of discussion) includes both $state_k*year_t$ and $state_k*year_t^2$. Appendix 1 discusses these adjustments in more depth.

91. Technically, these four new variables are $black_i*state_k*year_t$, $black_i*state_k*year_t^2$, $sex_i*state_k*year_t$, and $sex_i*state_k*year_t^2$.

b. A Note About Drug Offenses

Although I run the regressions for all three classes of crimes, the discussion focuses primarily on the results for violent and property offenses. The tables below include the outcomes for drug offenses as well, but these results are not discussed at length. In marked contrast to their treatment of violent and property crimes, states often recodified their drug codes or changed their charging behavior at the same time they adopted their guidelines. For example, after adopting its guidelines, Ohio systematically started charging offenders under possession statutes where earlier it would have sought convictions for trafficking.[92] In some states there is almost no overlap between the drug charges for which inmates are admitted before guideline adoption and those for which they are admitted after. As a result, it is much harder to compare results across time for drug offenses than for violent and property crimes; perhaps not surprisingly, the drug-crime results are consistently less statistically significant and less intuitive than those for the other two types of crimes.

* * *

The next two subparts examine the empirical results. The discussions focus primarily on the results and their interpretations; Appendix 1 provides the technical details. Part II.A.4 considers the limitations of the regressions and their implications.

2. Voluntary Guidelines and Variation

Tables 3 and 4 present the results examining how guideline adoption influences the mean difference in sentences imposed. In particular, I consider two measures of mean difference as the dependent variable: (1) the mean difference itself; and (2) the natural logarithm of the mean difference score. The first approach measures the absolute effect (for example, a coefficient of –5 for the guideline variable implies that the mean difference falls by five months when guidelines are used), and the second the percentage effect (for example, a coefficient of –0.05 for the guideline variable implies that the mean difference falls by 5 percent when guidelines are used). The results in Tables 3 and 4 are consistent with the hypothesis of this paper: Voluntary guidelines can successfully reduce variation in sentencing. It is true that presumptive guidelines consistently do a

92. Thanks to James Lowe, an Assistant Prosecuting Attorney for Franklin County, Ohio, for pointing this out to me. As a result of this change in policy, the fraction of inmates charged with trafficking in Ohio fell from over 98 percent before the guidelines were employed to 37 percent after (according to the NCRP).

better job, but *Blakely* forecloses such options (at least in their current form), and voluntary guidelines appear capable of picking up some of the slack.

Table 3 provides the absolute-effect results. For example, the first two coefficients in Column I imply that, without correcting for endogeneity, the mean difference in sentence variation for violent crimes falls by 9.450 months under presumptive guidelines and 4.704 months under voluntary guidelines when the regressions use only the four changing states; the first two results in Column II imply declines of 13.271 and 10.282 months, respectively, when all fourteen states are used.[93] In Table 4, the dependent variable is the log of the mean difference, so the coefficients reflect the percentage effect of adoption. In other words, the regression results in Column I suggest that, before accounting for endogeneity, voluntary guidelines reduce the mean difference of violent crime by approximately 8.7 percent when using just the four changing states, compared to 39.2 percent for presumptive guidelines. Column II indicates that when considering all fourteen states, the adoption of voluntary guidelines appears to induce a 34.6 percent drop in mean difference, compared to a 57.1 percent drop when presumptive guidelines are adopted.

93. One thing that is immediately apparent here (and in the results that follow) is that the outcomes change noticeably when moving from the four changing states to the complete set of fourteen states. This is likely due to the fact that movement of the mean differences in states that do not change their guideline policies during the period in question (that is, states that either use guidelines throughout or never adopt guidelines) differs systematically from such movement in the adopting states. Results discussed in Appendix 3 indicate that simply omitting California from the regressions causes the full-state results (now full-state-minus-one) to move closer to the four-state results; in other words, dropping California causes the results in Column II in Table 3 to move toward those in Column I (and likewise those in Column IV toward those in Column III), though the convergence is by no means total. California provides nearly half of the observations in each dataset—of the 640,011 offenders measured in the violent-crime dataset, for example, 287,704, or just under 45 percent, are from California—and thus may exert a strong influence over the results. Moreover, not only does California have the largest prison population in the nation during the same period (even larger than the federal prison population), but its "multiple-choice" determinate sentencing law is one of the most rigid in the country.

TABLE 3
THE EFFECT OF GUIDELINE ADOPTION ON DEVIATIONS FROM MEAN SENTENCE IMPOSED (CORRECTED FOR ENDOGENEITY)

	Violent Crimes					
	Uncorrected Results		Corrected Results			
			Linear		Linear & Quad.	
	I	II	III	IV	V	VI
Guidelines						
Voluntary	−4.704	−10.282*	−2.374	−7.435	−3.798	−6.622
	(3.276)	(2.561)	(5.539)	(3.993)	(6.316)	(4.758)
Presumptive	−9.450*	−13.271*	−3.957*	−5.682*	−3.481*	−3.341*
	(1.648)	(3.640)	(0.399)	(1.988)	(0.525)	(1.420)
Demographics						
Black	1.215	1.325*	1.219	1.397*	1.204	1.378*
	(1.112)	(0.520)	(1.114)	(0.499)	(1.119)	(0.494)
Woman	−3.552*	−2.723*	−3.618*	−2.692*	−3.645*	−2.706*
	(0.832)	(0.425)	(0.849)	(0.434)	(0.860)	(0.439)
Offense Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes
All States	No	Yes	No	Yes	No	Yes
R^2	0.090	0.204	0.091	0.208	0.091	0.209
	Property Crimes					
	Uncorrected Results		Corrected Results			
			Linear		Linear & Quad.	
	I	II	III	IV	V	VI
Guidelines						
Voluntary	−2.524	−3.611	−1.393	−1.810	−0.468	−1.408
	(3.501)	(2.252)	(2.128)	(2.498)	(3.122)	(3.287)
Presumptive	−10.453*	−8.918*	−2.075	−1.487	−3.743	−1.346
	(2.554)	(3.667)	(1.315)	(1.156)	(1.554)	(2.246)
Demographics						
Black	2.216*	0.586*	2.244*	0.622*	2.249*	0.644*
	(1.853)	(0.256)	(0.420)	(0.258)	(0.420)	(0.259)
Woman	−2.246	−2.348*	−2.251	−2.291*	−2.229	−2.262*
	(1.853)	(0.649)	(1.844)	(0.624)	(1.845)	(0.619)
Offense Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes
All States	No	Yes	No	Yes	No	Yes
R^2	0.242	0.241	0.245	0.247	0.246	0.249

	Drug Crimes					
	Uncorrected Results		Corrected Results			
			Linear		Linear & Quad.	
	I	II	III	IV	V	VI
Guidelines						
Voluntary	1.259	-5.632*	0.843	-2.631	-2.806	-3.574*
	(2.030)	(1.788)	(2.471)	(1.538)	(1.136)	(1.122)
Presumptive	-9.294	-14.025*	-9.976*	-9.898*	-6.457	-5.706
	(3.363)	(3.024)	(2.643)	(3.105)	(2.964)	(4.119)
Demographics						
Black	-4.873	-2.029	-4.854	-1.714	-4.842	-1.637
	(3.892)	(1.153)	(3.939)	(1.019)	(3.920)	(0.991)
Woman	-4.599	-2.196*	-4.563	-2.174*	-4.560	-2.186*
	(2.033)	(0.511)	(2.025)	(0.493)	(2.047)	(0.503)
Offense Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes
All States	No	Yes	No	Yes	No	Yes
R^2	0.158	0.244	0.158	0.252	0.159	0.255

Notes. The numbers provided in parentheses are robust standard errors, corrected to reflect state-level clustering; coefficients marked with an asterisk are significant at the 95 percent level. "Offense Dummy" refers to whether the regression includes indicator variables for each specific suboffense within the overarching crime category; the first suboffense within each crime category is dropped. "Year Dummy" reflects whether the regression includes year effects (the first year, 1989, is dropped); and "State Dummy" whether the regression includes fixed effects (the first state, Alabama, is dropped). "All States" captures whether the regressions look at data just from the four reporting states that changed their guideline policies during the period under review or all reporting states. All guideline variables are lagged by one year. "Linear" indicates that the linear state-year interaction term is employed to control for endogeneity; "Linear & Quad" indicates that both the linear and quadratic (squaring the year term) corrections are used.

Several patterns emerge from Tables 3 and 4. First, for both violent and property crimes, voluntary guidelines appear to reduce variation, although presumptive guidelines do so more effectively (the coefficients are generally larger numerically and more significant statistically).[94] As mentioned above, the results for drug crimes are less reliable, since many states changed their charging patterns when they adopted guidelines. Briefly, though, voluntary guideline adoption seems to have had less of an effect on drug offenses than on violent or property crimes, at least in percentage terms.

TABLE 4
THE EFFECT OF GUIDELINE ADOPTION ON THE LOG OF THE DEVIATION FROM MEAN SENTENCES IMPOSED

	Violent Crimes					
	Uncorrected Results		Corrected Results			
			Linear		Linear & Quad.	
	I	II	III	IV	V	VI
Guidelines						
Voluntary	–0.087	–0.346*	0.022	–0.091	–0.042	–0.086
	(0.097)	(0.144)	(0.091)	(0.126)	(0.070)	(0.146)
Presumptive	–0.392*	–0.571*	–0.142*	–0.176*	–0.125*	–0.096
	(0.060)	(0.186)	(0.016)	(0.054)	(0.026)	(0.056)
Demographics						
Black	0.003	0.016	0.003	0.022	0.003	0.021
	(0.024)	(0.011)	(0.024)	(0.010)	(0.024)	(0.010)
Woman	–0.021	–0.017	–0.024	–0.014	–0.023	–0.014
	(0.017)	(0.010)	(0.018)	(0.011)	(0.018)	(0.011)
Offense Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes
All States	No	Yes	No	Yes	No	Yes
R^2	0.108	0.236	0.114	0.244	0.114	0.246

94. For property crimes, for example, in absolute terms, voluntary guidelines reduce the mean difference by approximately 2.5 to 3.6 months, compared to 8.9 to 10.5 months for presumptive guidelines. In the log regression, voluntary guidelines reduce mean difference by about 5.4 percent to 21.3 percent, compared to 43.4 percent to 54.0 percent for presumptive guidelines.

	Property Crimes					
	Uncorrected Results		Corrected Results			
			Linear		Linear & Quad.	
	I	II	III	IV	V	VI
Guidelines						
Voluntary	–0.054	–0.213	0.022	0.021	–0.071	–0.088
	(0.151)	(0.180)	(0.185)	(0.256)	(0.164)	(0.263)
Presumptive	–0.434*	–0.540*	–0.149	–0.137	–0.173*	–0.133*
	(0.045)	(0.098)	(0.088)	(0.150)	(0.038)	(0.030)
Demographics						
Black	0.074	0.020	0.076	0.022	0.076	0.023
	(0.027)	(0.015)	(0.027)	(0.016)	(0.027)	(0.015)
Woman	–0.045	–0.041*	–0.045	–0.038*	–0.044	–0.037*
	(0.048)	(0.018)	(0.048)	(0.017)	(0.048)	(0.084)
Offense Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes
All States	No	Yes	No	Yes	No	Yes
R^2	0.262	0.268	0.267	0.277	0.271	0.279
	Drug Crimes					
	Uncorrected Results		Corrected Results			
			Linear		Linear & Quad.	
	I	II	III	IV	V	VI
Guidelines						
Voluntary	0.341	–0.032	–0.076	–0.284*	–0.362	–0.311*
	(0.187)	(0.066)	(0.140)	(0.113)	(0.120)	(0.115)
Presumptive	–0.250	–0.650	–0.611	–0.627	–0.332*	–0.223
	(0.290)	(0.386)	(0.194)	(0.393)	(0.036)	(0.137)
Demographics						
Black	–0.062	–0.138*	–0.060	–0.133*	–0.580	–0.131*
	(0.084)	(0.027)	(0.082)	(0.028)	(0.081)	(0.030)
Woman	–0.046	–0.068*	–0.042	–0.067*	–0.044	–0.068*
	(0.036)	(0.015)	(0.035)	(0.015)	(0.036)	(0.015)
Offense Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes
All States	No	Yes	No	Yes	No	Yes
R^2	0.304	0.299	0.318	0.304	0.327	0.307

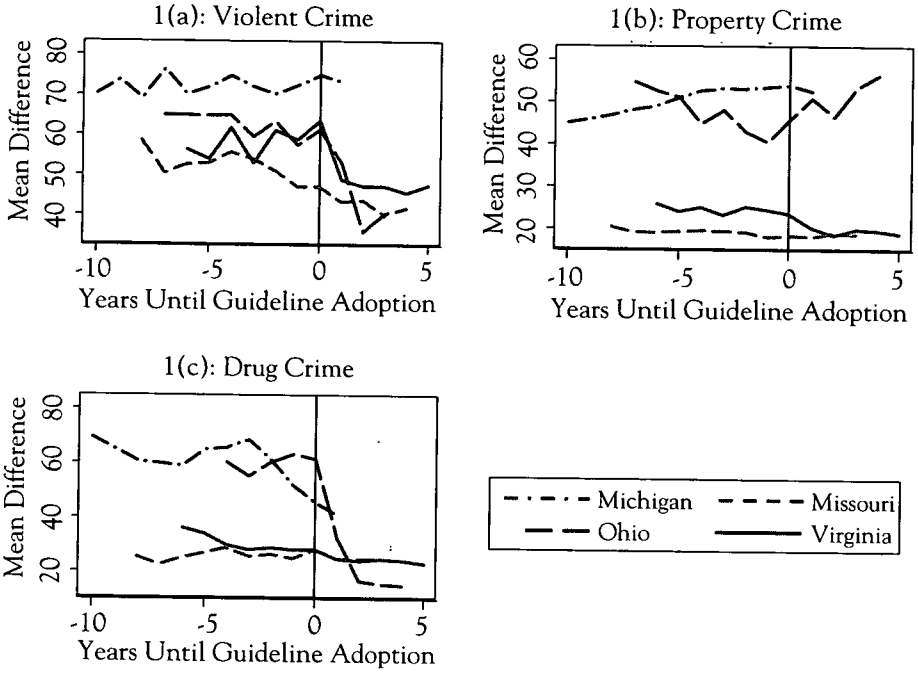
Notes. The numbers provided in parentheses are robust standard errors, corrected to reflect state-level clustering; coefficients marked with an asterisk are significant at the 95 percent level. "Offense Dummy" refers to whether the regression includes indicator variables for each specific suboffense within the overarching crime category; the first

suboffense within each crime category is dropped. "Year Dummy" reflects whether the regression includes year effects (the first year, 1989, is dropped); and "State Dummy" whether the regression includes fixed effects (the first state, Alabama, is dropped). "All States" captures whether the regressions look at data just from the four reporting states that changed their guideline policies during the period under review (Michigan, Missouri, Ohio, and Virginia) or all reporting states. All guideline variables are lagged by one year. "Linear" indicates that the linear state-year interaction term is employed to control for endogeneity; "Linear & Quad" indicates that both the linear and quadratic (squaring the year term) corrections are used.

Second, correcting for endogeneity has noticeable effects. Consider, for example, the results in Table 3: Including the adjustments reduces the impact of both voluntary and presumptive guidelines. It should be noted, however, that while the absolute efficacy of both voluntary and presumptive guidelines falls, in general the *relative* efficacy of voluntary guidelines appears to rise. For example, with respect to violent crimes, correcting for endogeneity reduces the coefficient for presumptive guidelines by nearly three-quarters (from –13.3 to –3.3), while that for voluntary drops by less than half (from –10.3 to –6.6); while the statistical significance for the voluntary guidelines coefficient falls, it remains relatively significant in the full-state regressions (at the 90 percent level when just the linear state-year trend is used, and at the 80 percent level when the quadratic trend is also included). For property crimes, presumptive guidelines fall from being nearly three times as effective as voluntary to equally as effective, albeit at a much lower (and much less statistically significant) level. For drug crimes, the relative efficacy of presumptive guidelines appears to rise (even as its absolute effect falls slightly) when the linear trend is employed, but to fall (from three times as potent to under two) when the quadratic trend is included.

The importance of the time trends can be observed graphically as well. Figure 1(a) plots the average mean difference for violent crimes for the four states that adopted guidelines during the sample period; the *x*-axis is normalized so that for all states "year 0" is the year in which a state adopts its guidelines. Figure 1(b) repeats this for property crimes, and Figure 1(c) for drug offenses. In Figure 1(a), for example, two states have no appreciable preadoption trend while two others experience declining variation prior to adoption. In Figure 1(b) the results are less consistent: Two states exhibit little preadoption movement, one sees the problem of variation worsen, and a fourth again sees a decline prior to adoption. Finally, the patterns in Figure 1(c) are quite similar to those in Figure 1(a): Two states observe little change prior to adoption, and two declines.

FIGURE 1
CHANGE IN MEAN DIFFERENCE OVER TIME



In short, then, it appears that endogeneity may exert a *downward* bias on the coefficients (causing the uncorrected regressions to overstate the true magnitude of adoption's effect). Correcting this bias, however, weakens the apparent absolute effect of both voluntary and presumptive guidelines but *increases* the relative effectiveness of voluntary guidelines.

Voluntary guidelines thus appear capable of reducing the variation in sentences imposed, even if not as effectively as presumptive guidelines. At the very least, the weaker results indicate that voluntary guidelines can continue some of the work done by presumptive guidelines, especially for violent and property crimes (where voluntary guidelines appear to be about 50 percent and 40 percent as effective, respectively). The stronger results in this subpart suggest that voluntary guidelines can accomplish nearly *all* of what presumptive guidelines did.

3. Voluntary Guidelines and Impermissible Factors

Besides reducing variation, guidelines are capable of influencing how judges rely on so-called "impermissible" factors, such as race. This subpart develops an empirical test to examine how effectively they do so. In particular, I consider how the adoption of guidelines changes the importance of race and sex to sentence length.[95] Appendix 1 provides the technical discussion, and Table 5 presents the results. The regressions here use four variables not present in the earlier test; these capture the interactions between race or sex on the one hand and voluntary or presumptive guidelines on the other, and they measure the extent to which the judicial reliance on race and sex changes after guideline adoption.[96]

95. Note that the dependent variable has changed from the first test, from the mean difference in sentence length to the sentence length itself.

96. In Table 5, these are called "Black*Vol," "Black*Pres," "Sex*Vol," and "Sex*Pres," respectively. Thus, for example, in Column II of Table 5, being black adds 11.3 months to a prison sentence for violent crimes, but that that falls by 8.1 months once a state adopts voluntary guidelines.

TABLE 5
THE EFFECT OF GUIDELINE ADOPTION ON THE USE OF IMPERMISSIBLE FACTORS AT SENTENCING

	Violent Crimes					
	Uncorrected Results		Corrected Results			
			Linear Trend		Linear & Quad. Trend	
	I	II	III	IV	V	VI
Race						
Black	7.392	11.327*	9.350	11.282*	14.935	13.884*
	(2.716)	(2.477)	(3.938)	(2.939)	(5.225)	(2.765)
Black*Vol	–3.036	–8.119	6.617	–4.214	0.947	–4.73
	(7.310)	(5.556)	(8.507)	(4.081)	(6.320)	(5.389)
Black*Pres	1.580	–12.017*	–12.822	–15.346*	–10.847*	–15.558*
	(5.439)	(2.244)	(6.002)	(3.636)	(2.296)	(3.127)
Sex						
Sex	–18.039*	–21.979*	–20.963*	–21.395*	–18.970*	–26.227*
	(4.387)	(2.596)	(3.932)	(5.122)	(3.062)	(5.596)
Sex*Vol	2.214	2.323	14.144	7.732	0.539	3.741
	(6.510)	(3.049)	(7.339)	(5.519)	(1.914)	(4.663)
Sex*Pres	6.310*	14.096*	–10.72	12.874	–1.559	17.096*
	(1.669)	(3.176)	(3.609)	(7.231)	(0.967)	(7.186)
Guidelines						
Voluntary	–4.126	–12.196	–9.214	–10.321	–3.265	–7.538
	(11.179)	(6.037)	(3.756)	(5.553)	(4.683)	(7.858)
Presumptive	–64.714*	–64.969*	–24.738	–26.855	–26.197	–26.052
	(6.439)	(19.176)	(8.186)	(18.184)	(8.452)	(17.391)
Offense Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes
All States	No	Yes	No	Yes	No	Yes
R^2	0.356	0.416	0.365	0.423	0.368	0.425

	Property Crimes					
	Uncorrected Results		Corrected Results			
			Linear Trend		Linear & Quad. Trend	
	I	II	III	IV	V	VI
Race						
Black	1.748	0.265	–0.129	–2.478	1.704	0.051
	(0.600)	(0.775)	(3.260)	(1.394)	(3.703)	(1.033)
Black*Vol	2.292	–2.242	–5.28	–5.929*	–1.551	–3.466
	(1.844)	(2.973)	(2.625)	(0.743)	(1.444)	(1.611)
Black*Pres	5.700*	–0.732	9.190	0.693	3.794	–1.994
	(1.971)	(0.890)	(4.129)	(1.901)	(4.301)	(1.672)
Sex						
Sex	–4.98	–9.542*	–7.519	–11.247*	–8.909	–13.689*
	(1.832)	(2.304)	(2.441)	(2.368)	(5.129)	(2.715)
Sex*Vol	6.967	9.212*	–11.623	–6.068	–0.269	2.120
	(7.020)	(3.548)	(6.418)	(4.339)	(2.767)	(3.138)
Sex*Pres	–5.135	6.927*	13.507	14.202*	4.385	14.528*
	(7.259)	(2.605)	(4.931)	(2.401)	(5.178)	(2.963)
Guidelines						
Voluntary	–11.867	–5.546	0.898	–1.619	–1.49	–3.057
	(7.311)	(6.077)	(3.190)	(2.673)	(4.235)	(4.024)
Presumptive	–33.348*	–23.163*	–14.09	–10.958*	–11.563	–8.576
	(7.673)	(9.978)	(6.212)	(5.064)	(7.199)	(6.611)
Offense Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes
All States	No	Yes	No	Yes	No	Yes
R^2	0.350	0.310	0.362	0.319	0.362	0.322

	Drug Crimes					
	Uncorrected Results		Corrected Results			
			Linear Trend		Linear & Quad. Trend	
	I	II	III	IV	V	VI
Race						
Black	3.548	–0.729	–3.645	–2.197	–17.27	–1.07
	(6.846)	(4.270)	(2.891)	(3.999)	(7.529)	(4.742)
Black*Vol	–1.102	0.726	–5.085	–4.680*	–4.131*	–4.269*
	(6.391)	(4.440)	(2.298)	(0.523)	(0.841)	(1.708)
Black*Pres	–0.737	3.006	2.466	1.442	2.082	0.220
	(4.693)	(4.179)	(10.006)	(3.392)	(1.606)	(4.327)
Sex						
Sex	–3.76	–7.327	–2.414	–7.404*	–4.517	–9.073*
	(4.236)	(1.751)	(6.361)	(2.483)	(6.409)	(2.160)
Sex*Vol	–4.158	–0.724	8.257	0.194	5.147	3.277*
	(4.058)	(1.768)	(5.280)	(3.182)	(2.423)	(0.458)
Sex*Pres	–0.63	3.940	–0.426	6.094*	–1.249	7.639*
	(4.406)	(1.900)	(2.489)	(2.600)	(1.115)	(2.599)
Guidelines						
Voluntary	–29.888*	–23.782*	–13.064*	–12.420*	–10.22	–13.667*
	(1.169)	(3.263)	2.884	(1.377)	(4.230)	(1.903)
Presumptive	–17.628	–11.446	–9.933	–5.742	–12.726*	–7.346
	(9.810)	10.305	(10.317)	(5.489)	(2.706)	(4.160)
Offense Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes
All States	No	Yes	No	Yes	No	Yes
R^2	0.563	0.508	0.567	0.516	0.568	0.518

Notes. The numbers provided in parentheses are robust standard errors, corrected to reflect state-level clustering; coefficients marked with an asterisk are significant at the 95 percent level. "Black*Vol" is the interaction between the "Black" and "Voluntary Guideline" indicators; "Woman*Vol" is likewise defined for women, and "Black*Pres" and "Women*Pres" for presumptive guidelines. "Linear" indicates that the linear corrections for endogeneity are used, "Linear & Quad" indicates that both the linear and quadratic corrections are used.

One feature immediately apparent in Table 5 is the ambiguous role of race on sentence length (the overall impact of being black on sentence length is the coefficient for "Black" in that table). While blacks generally face longer sentences, in many cases the effect of race is close to zero: For property crimes, for

example, being black appears to add no more than two months to a prison sentence, and for drug crimes blacks often seem to face shorter sentences. These results are consistent with current studies. Robert Sampson and Janet Lauritsen, in a comprehensive review of the literature on race's role in the criminal justice system, argue that (1) much of the earlier research finding racial bias at sentencing suffers from important methodological flaws, and (2) the current (and improved) work often finds little or no evidence of overt racism.[97] It is true, of course, that sentencing is but one part of the criminal justice system, and the fact that race does not necessarily play a strong role at sentencing does not imply it plays no role in that system overall.[98] This Article, however, concerns itself only with the role of race at sentencing.

Moreover, it is important to note that a positive coefficient for race does not necessarily imply that blacks are being discriminated against by judges. Race is correlated with other factors that legitimately influence sentence length, and unfortunately the spare nature of NCRP data prevents the tests from accounting for them. For example, black defendants tend to have longer criminal histories than their white counterparts,[99] and they thus would likely face longer average sentences even if judges were wholly race blind. Since the NCRP does not provide a measure of past criminal history, the coefficient for race measures the net effect of both issues.[100] Moreover, no ready proxy suggests itself, especially since at present there appears to be no state-level national dataset on prior criminal histories or recidivism rates.[101] In future work, I plan to look at data gathered by

97. Robert J. Sampson & Janet L. Lauritsen, *Racial and Ethnic Disparities in Crime and Criminal Justice in the United States*, *in* ETHNICITY, CRIME, AND IMMIGRATION 311, 346–52 (Michael Tonry ed., 1997). Even when racism seems to play a role, its effect is not always immediately clear. Robert Sampson and Janet Lauritsen cite a study, for example, that found that courts in New York treated low-level black drug offenders better than white ones, but black "kingpins" worse. *Id.* at 347.

98. Sampson and Lauritsen, however, point out that the actual relationship between race and the criminal justice system is highly complex. They argue that research "ha[s] not been kind to a simplistic 'discrimination thesis'" and that "the most compelling evidence concerning racial discrimination in the administration of justice involves community and national constructions of 'moral panics' and political responses to those contexts." *Id.* at 362.

99. *Id.* at 346–47.

100. In fact, the coefficient for race further reflects all other factors that are potentially appropriate at sentencing and are correlated with race (such as age, education, or even the nature of the victim).

101. The only comprehensive study of recidivism seems to be PATRICK A. LANGAN & DAVID J. LEVIN, BUREAU OF JUSTICE STATISTICS, SPECIAL REPORT: RECIDIVISM OF PRISONERS RELEASED IN 1994, at 1 (June 2002), *available at* http://www.ojp.usdoj.gov/bjs/pub/pdf/rpr94.pdf, which considers the recidivism of prisoners released only in the year 1994 and only for a subsample of states (which includes some, but not all, of the states under consideration here). Moreover, information available from the Bureau of Justice Statistics implies that state criminal history data are not available in a single, unified dataset but rather are maintained by each state individually. *See* Bureau of Justice Statistics, Criminal Record Systems Statistics (2006), http://www.ojp.usdoj.gov/bjs/crs.htm ("Criminal records are maintained by each State in a central repository . . . ."); *see also* BUREAU OF JUSTICE STATISTICS,

the individual states to provide sharper insights into both the effect of guidelines on the use of race and sex and the attendant welfare implications.

However, that the NCRP does not allow me to estimate the magnitude (or even the net direction) of the bias from missing correlates is not as problematic as it might initially appear. The focus of this test is not on the importance per se of race and sex at sentencing, but on how that importance *changes* when guidelines are adopted. To the extent that these omitted variables do not change over time (for example, to the extent that the relative differences between black and white criminal histories and their importance at sentencing do not change over time), their omission does not alter the results. Of course, an assumption of no change whatsoever is surely too strong to make. Adopting guidelines that pay explicit attention to past criminal history in score calculations, for example, likely alters such history's importance at sentencing.[102] Given the magnitude of some of the effects discussed below, it is likely that the tests developed here overstate the impact of both voluntary and presumptive guidelines (for example, it is perhaps unlikely that either type of guideline reduces the use of race as strongly as the results below indicate). It is not possible, however, to make any predictions about the effect of the omissions on the *relative* efficacy of the two types of guidelines. These results thus provide a first cut at measuring the importance of presumptiveness, but they are not definitive. Future work with state-gathered data will, I hope, provide more significant results.

Looking at the relationship between guidelines and race in Table 5, several patterns emerge. First, only for violent crimes does race appear to play a meaningful role. For both property and drug offenses, the effect of race on sentence length is numerically small (with one exception never more than 3.6 months) and never significant at the 5 percent level. Not surprisingly, the effects of guideline adoptions are also numerically small and with only a few exceptions statistically insignificant.[103] Second, focusing on the results for violent crimes that consider the full set of reporting states, voluntary guidelines appear to reduce the use of race, albeit not as effectively as presumptive guidelines. Blacks

SURVEY OF STATE CRIMINAL HISTORY INFORMATION SYSTEMS, 2001 (Aug. 2003), *available at* http://www.ojp.usdoj.gov/bjs/pub/pdf/sschis01.pdf.

102. Note, however, that not all states have the same detailed grading of past criminal history that the federal system employs (although some do). Ohio's criminal code, for example, simply permits judges to sentence repeat offenders either to the longest sentence in the guideline range (which is otherwise usually not permitted) or to impose a sentence of up to ten years greater than that otherwise authorized. OHIO REV. CODE ANN. § 2929.14(D)(2) (West Supp. 2006).

103. Interestingly, the guideline coefficients are significant only for voluntary guidelines (in column IV for property crimes and in columns IV, V, and VI for drug offenses), and in each case, blacks face *lower* average sentences and guideline adoption appears to exacerbate this effect. *See* Table 5.

face sentences that are approximately 11.3 months (or 14 percent) longer than those of whites,[104] and voluntary guidelines reduce this effect by 72 percent, compared to 106 percent[105] for presumptive guidelines. The results for presumptive guidelines do not change much with the corrections for endogeneity, while those for voluntary guidelines fall by approximately half. It should be noted that the effects of presumptive guidelines are consistently more statistically significant than those of voluntary guidelines, especially in the presence of the endogeneity adjustments.

For sex, the underlying effect is much less ambiguous: Across all specifications, women face shorter sentences than men, and both types of guidelines consistently reduce this gap, though to varying degrees. For violent crimes, for example, voluntary guidelines reduce the differential by 11 to 36 percent; presumptive guidelines, by 60 to 65 percent. For property crimes, the effect of voluntary guidelines is more scattered, ranging from exacerbating the differential by 54 percent to reducing it by 97 percent. The results for presumptive guidelines are more consistent, reducing the gap by 73 to 126 percent. Finally, for drug crimes, voluntary guidelines generally have little effect—in one specification they reduce the differential by 36 percent, but in the other two the effect is essentially zero both numerically and statistically—while presumptive guidelines reduce the gap by 54 to 84 percent. Note that the endogeneity controls consistently influence the coefficients for voluntary guidelines more noticeably than those for presumptive guidelines.[106] Also, as with race, the effects of presumptive guidelines are consistently more significant statistically.[107]

Thus, at least for violent and property crimes, both voluntary and presumptive guidelines appear to reduce the use of race and sex, often to

---

104. The average sentence length for violent crimes is approximately eighty-three months. Recall that this difference does not necessarily reflect discrimination, since the data do not control for such race-correlated factors as prior criminal history. *See* Table 5.

105. A result greater than 100 percent implies that the use of guidelines leads to a reversal in "bias." In other words, a 106 percent change means that if blacks received sentences one hundred months longer than whites before guideline adoption, they receive sentences six months *shorter* after. *See* Table 5.

106. In general, the clear graphical evidence of endogeneity seen in Figure 1 is not present here.

107. In an earlier version of this Article, I attempted to sharpen the above results by allowing the underlying use of race and sex to vary between voluntary-, presumptive-, and nonguideline states. In particular, I included four new terms that interacted the race or sex indicator with new variables indicating whether a state adopted voluntary or presumptive guidelines at any point in the sample. In other words, the new indicator for presumptive guidelines equaled 1 for all observations from Ohio, not just those that came after the state adopted guidelines.

While this refined model produced seemingly more precise estimates, it is likely that such improvements were spurious. The new indicators were highly collinear with guideline-effect interaction terms, casting doubt on the standard errors. And perhaps more tellingly, the underlying net effects did not change much with the new terms, and the R^2 terms changed by at most 0.001.

(numerically, if not statistically) significant degrees. In most cases, presumptive guidelines work better than voluntary, but voluntary guidelines are nonetheless sometimes effective: The shift from the former to the latter is neither costless nor crippling. These results, however, are not as favorable toward voluntary guidelines as those in the mean difference regressions (although they are also more tentative).

4. Qualifications and Concerns

Though the empirical tests presented above attempt to account for many of the factors that influence sentence variation and sentence length, they do not address all of the possible empirical issues. This subpart considers the implications of six possible problems: (1) The adoption of a particular type of guideline system is an endogenous decision; (2) the data do not indicate whether guidelines affect the decision to incarcerate in the first place; (3) prosecutorial behavior may matter in guideline jurisdictions; (4) guidelines may affect the relative mean sentence length *across* (rather than within) types of crimes in ways correlated with race; (5) jurisdictions have competing definitions of "success"; and (6) jurisdictions that did not adopt guidelines may still change their behavior in response to the examples set in those jurisdictions that did adopt.

a. Endogeneity Revisited

The tests developed above controlled for the fact that states did not adopt guidelines at random times. But there is a second type of endogeneity present here as well. Not only is a state's decision *when* to adopt guidelines not random, but neither is its decision about what *type* of guidelines to adopt, namely, voluntary or presumptive. The legislature in Virginia, for example, may have felt comfortable adopting voluntary guidelines in part because it appoints and retains the judges responsible for implementing them. In this case, endogeneity may be troubling because it suggests that voluntary guidelines work in Virginia because of a factor not easily exportable to other states: the means by which judges acquire and keep their jobs.[108]

108. An earlier version of this Article developed a logit model to examine what factors influenced the adoption of guidelines. Although the model returned some useful results in the binomial case (when voluntary and presumptive guidelines were treated simply as "guidelines"), in the multinomial case (when voluntary and presumptive guidelines were treated as two different choices) the results were weak and ambiguous. Moreover, due to limitations in the data, neither the binomial nor the multinomial model could shed any light on the role of sentencing variation or the use of impermissible factors on the decision to adopt guidelines in general or to choose between voluntary and presumptive guidelines.

And there is some (weak) evidence in the data that this is a potentially valid concern. As suggested by Figure 1, Virginia and Missouri (which elects its trial judges) observe distinctly different results when each adopts guidelines. In regressions not reported here, I disaggregated the effects of Missouri and Virginia on the mean difference of sentence length. For violent and property crimes, the results are consistently stronger for Virginia.[109] Conversely, however, Missouri's guidelines appear more effective at combating variation in drug sentencing, an effect that grows stronger as I control for endogeneity.

These results, at least for violent and property crimes, do not necessarily indicate that Virginia's method of judicial selection is central to its superior output. For example, the data here do not indicate why Missouri, unlike Virginia, faced a downward trend in mean difference prior to adoption, and the method of judicial selection is only one of many differences between Missouri and Virginia. Moreover, as discussed in more depth in Appendix 1, the effectiveness of even Virginia's guidelines appears to have turned in part on their particular design, implying that the effectiveness of guidelines can be sensitive to the way the ranges are set, even keeping judicial selection methods constant. But the differences in coefficients do indicate that it is an issue worthy of further consideration.

b. The Decision to Incarcerate

The second concern is that the data from the NCRP consider the fate only of those defendants admitted to prison. The results here thus speak only to how guidelines influence criminal sentencing *given* that incarceration is the punishment; they do not consider how much guidelines affect the decision to incarcerate in the first place. For example, these results indicate that voluntary guidelines reduce the use of race at sentencing for violent and property crimes, but they do not say anything about how much judges still rely on race when initially choosing between incarceration and probation for such crimes. This is not a minor concern: In 2001, there were nearly four million people on probation—1.96 million of them for felony convictions—compared to just over

109. For example, for violent crimes, the effects on mean difference without controlling for endogeneity are –10.922 for Virginia and –9.462 for Missouri, with both significant at the 5 percent level. With the linear controls, however, the coefficients are –13.613 for Virginia and –2.573 for Missouri; with the quadratic controls, the results are –12.146 and 1.659, respectively. Moreover, only Virginia's results are significant at the 5 percent level. For property crimes, the results are more dramatic. Without controlling for endogeneity, the coefficients are –6.179 for Virginia and –0.995 for Missouri. With the linear controls, the results are –5.745 and 0.904; with the quadratic controls, the results are –5.494 and 3.751, respectively. All of Virginia's results are significant at the 5 percent level, but Missouri's results are significant at 5 percent only with the quadratic controls. *See* Table 5.

1.3 million in prison.[110] However, while the tests here do not address the incarceration-or-probation decision directly, they still might shed light on the issue. No theory immediately suggests itself that can explain why guidelines would successfully regulate judicial behavior when it comes to choosing only the length, and not the type, of sentence imposed. Moreover, the theories that potentially explain why judges adhere to voluntary guidelines, discussed in Part II.B below, point toward guidelines working equally in both cases. That said, the number of defendants on probation indicates that this is an important open question.

c. The Role of Prosecutorial Charging Decisions

The third potential problem with the data used in this Article is that they provide information only on the offense for which the defendant is ultimately convicted, not for which he was arrested or with which he was initially charged. The data thus do not account for the effect of prosecutorial charging decisions and how the adoption of guidelines might change such choices. In other words, even if guidelines induce judges to sentence identically all those charged with and convicted of second-degree kidnapping, for example, the guidelines could also influence how prosecutors choose who gets charged with second-degree kidnapping, as opposed to first-degree, in the first place. The question, then, is whether the adoption of guidelines influences such prosecutorial charging decisions. If it does, then some of the apparently salutary effects of adoption may be ephemeral: While the adoption of guidelines may reduce the effect of race *within* a given offense category, it may change the racial distribution *across* categories, and the data used here measure only the within effect.[111]

It is not clear, though, that this is a particularly serious concern for the results presented here, at least for violent and property crimes. First, Terance Miethe, focusing on sentencing in Minnesota, has compiled evidence that prosecutorial behavior does not necessarily change noticeably following the adoption of guidelines.[112] Miethe notes that numerous formal and informal

110. LAUREN E. GLAZE & SERI PALLA, BUREAU OF JUSTICE STATISTICS, BULLETIN: PROBATION AND PAROLE IN THE UNITED STATES, 2003 (July 2004), *available at* http://www.csdp.org/research/ppus03.pdf.

111. For example, the adoption of guidelines could result in less racial bias within a given crime class (such as second-degree kidnapping), but it could also induce prosecutors to pay more attention to whom they assign to each class (so perhaps relatively more blacks are charged with first-degree as opposed to second-degree kidnapping once guidelines are employed). To the extent that the data look only at the sentences within a class, not the choice between classes, they do not account for this effect.

112. Terance D. Miethe, *Charging and Plea Bargaining Practices Under Determinate Sentencing: An Investigation of the Hydraulic Displacement of Discretion*, 78 J. CRIM. L. & CRIMINOLOGY 155 (1987).

constraints prevent prosecutors from exercising their discretion too freely. For example, differing evidentiary requirements for various offenses can make it hard to substitute across crimes, and "courthouse subculture[s]" can discourage such behavior by prosecutors.[113]

Second, the coarseness of the NCRP data helps: The NCRP uses broad definitions of offenses, such as kidnapping and aggravated assault, not first-degree kidnapping and third-degree aggravated assault. While one can imagine prosecutors changing the degree of kidnapping charged in response to the use of guidelines, it is harder to envision them changing the overall offense (from, say, assault to aggravated assault). And there appears to be little shifting across the subcategories of violent and property crimes; while there is some change in the distribution of offenses before and after guidelines are adopted, it is not particularly dramatic.[114] Of course, as noted above, there are significant changes in the distribution of drug offenses. However, at least at first blush these changes appear to be the result less of selective prosecution tactics and more of systematic changes in the drug codes and in charging practices more generally. Nonetheless, these distributional changes make it harder to draw meaningful implications from the results for this class of crime.

d. Guidelines and Structural Effects on Race

Fourth, even if prosecutors do not change how they charge defendants, guidelines can still in some ways exacerbate the effect of race or sex (even if they reduce its use in others). As pointed out above, the tests in Part II.A.2 consider only how guidelines affect the *variation* in sentence length *within* offense classifications (such as manslaughter, aggravated assault, petty larceny, and so on), but not how they affect the *average* sentence length *across* offenses. For example, assume that arson is committed predominantly by black offenders and assault by white offenders. If guidelines push up the mean sentences for arson and reduce the mean sentences for assault, even as they induce judges to impose consistent sentences *within* each offense category (and even if prosecutors do not change their charging behavior), then the guidelines have reduced the use of race directly (by encouraging more consistent intraoffense sentencing), while increasing it indirectly (by changing the relative average sentence lengths of race-correlated offenses).

---

113. *Id.* at 157.

114. And what changes exist are at least partially due to forces other than the guidelines. Throughout most of the period in question, crime rates dropped as prison populations rose, suggesting that increasingly marginal offenders were being incarcerated. This would imply that the distribution of offenses committed by those incarcerated should change as well.

Note, of course, that such an effect need not imply racism. In the preceding example, if prior to guideline adoption arson is thought to be relatively underpunished compared to assault, then the readjustment in means might be socially beneficial, even as it increases the relative average time a black offender spends in prison. Though differential impact is often read to imply racism,[115] such a claim can be made only following a deeper exploration of *why* such differences exist.[116] The limited data available in the NCRP allow us to examine how relative means shift following guideline adoption and the correlation of these shifts with the racial distribution of offenders, but they do not allow us to draw any further conclusions about whether these results arise from conscious or unintentional racial *bias*.

In general, the data from the NCRP do not point to a single, clear result. In both Michigan and Missouri, for example, sentence lengths for both violent and property crimes appear to rise more for offenses for which blacks were more likely to be convicted. Conversely, in Virginia, the opposite result obtains. And Ohio splits the difference: For violent crimes the relationship between changes in sentence length and relative number of black offenders is negative, but for property crimes it is positive. These results, by no means conclusive, suggest that in some situations (though, as Virginia indicates, not in all) the adoption of guidelines leads to slightly longer sentences for crimes that are relatively more the domain of black offenders. Whether such a change is the result of racism (intentional or otherwise), random coincidence, or the rational response to differences in the social costs of the crimes is an important question, but it is one beyond the scope of this Article.

It should be noted that even if guidelines tend to increase the mean sentence lengths of black-dominated crimes, this is not a result of guidelines per se, but rather of how legislatures opt to design them. The results in Part II indicate that judges respond to the guidelines imposed on them by legislatures.[117] Thus, the credit or blame for the changing distribution of means rests not on the inherent nature of the guidelines, but on how the legislatures and sentencing commissions set the ranges. In other words, nothing in this discussion has

115. Tonry, for example, argues that the war on drugs' disparate effect on black communities reflects, at best, a willful disregard of racial bias. MICHAEL TONRY, MALIGN NEGLECT 82 (1995).

116. Randall Kennedy claims that the problem with the war on drugs is that it does not enforce the drug laws enough, and that the deleterious effects of the crack epidemic on black communities were so strong as to merit profound responses. Randall Kennedy, *The State, Criminal Law, and Racial Discrimination: A Comment*, 107 HARV. L. REV. 1255 (1994).

117. The results from Virginia in particular suggest that this could be an important effect: Sentencing behavior changed markedly after Virginia narrowed its guideline ranges in 1994. This effect is discussed in more depth in Appendix 1.

any bearing on the (abstract) efficacy of guidelines, only perhaps on how legislatures opt in some situations to establish them.

e. Different Measures of Success

A fifth concern with the tests above is their definitions of "success," namely a reduction in either variation or in the use of impermissible factors. Two potential difficulties arise: (1) These goals need not be universally desirable; and (2) there may be other outcomes policymakers wish to accomplish. First, reduced variation and reduced reliance on race or sex are not necessarily inherently positive outcomes. Consider variation first. A sentencing regime that sentences all offenders for all crimes to ten years in prison, for example, would have no variation (and no use of impermissible factors, at least following conviction), but this outcome seems unappealing. Some variation is good: After all, the goal of guidelines is not to sentence all offenders similarly, but to sentence like offenders similarly. Tables 3 and 4 account for this to some degree by calculating mean differences within suboffenses, which assumes that we want variation *between* them: Murderers should face different punishment regimes than arsonists. But there are unaccounted-for factors that argue for some variation even *within* a given suboffense, such as the magnitude of the crime, the nature of the victim, whether the defendant assisted the government, and other idiosyncratic details about the crime and the defendant's behavior. This suggests that guidelines should not necessarily aim to eliminate 100 percent of variation. Moreover, even if guidelines primarily reduce "bad" variation, this reduction introduces possible costs of its own. For those offenders who are risk averse and aware of the general nature of punishment in their state, for example, reducing variation may increase their propensity to commit crime, even if the average sentence remains constant.[118]

Reducing the use of race or sex also need not be a universally accepted outcome. Consider, for example, the impact of sex on a particular sentence. To the extent that mothers may play a relatively more important role in the physical, mental, and emotional development of their children, an argument can be made for sentencing (some) women differently than men. Unfortunately, the NCRP does not contain data on whether a defendant has dependent children, so it is not possible to design a test to see, for example, whether

118. Of course, the opposite result obtains for offenders who are risk loving: Reduced variation undercuts their willingness to engage in crime. For those who are risk neutral or, perhaps more likely, fail to take variation into account when deciding to commit a crime (because, for example, they are unaware of the variation or are incapable of thinking about it due to drug use, mental handicap, or other such impediment), there is no effect either way.

childless women are sentenced more like men than women with children. Nonetheless, there are possible arguments to make for justifying different outcomes for men and women who are otherwise identical.

The second concern with the definition of "success" used here is that other possible outcomes can motivate policymakers. For example, legislatures might be concerned about sentence length; Pennsylvania's legislature rejected an earlier version of its current guideline system, arguing that it was not sufficiently punitive, and the rigid structure of the federal guideline system was driven in part by a fear of judicial leniency.[119] The results in Table 5 indicate that the voluntary and presumptive guidelines considered here (with the exception of Michigan's, as discussed *infra* Part III.B, and Missouri's) have led to shorter sentences, and for some jurisdictions this tradeoff might not be acceptable.[120] Thus, the claim above that voluntary guidelines "succeed" is limited to the specific metrics on which I focus (and the acceptability of these outcomes).

f. The Indirect Effect of Guideline Adoption on Nonguideline States

The sixth concern is that it is possible that the adoption of guidelines in one state (or in an increasing number of states) could influence sentencing behavior in other, nonguideline states. For example, growing guideline adoption could encourage legislatures in nominally nonguideline states to modify their criminal codes to address some of the concerns raised about unguided sentencing (without going as far as formally establishing guidelines). Or even in the absence of legislative action, growing adoption could lead judges in nonguideline states to pay more attention to how they sentence. The first outcome is unlikely—if anything, the statutory conditions in nonguideline states have deteriorated over time. Paul Robinson and Michael Cahill, using Illinois and Kentucky as case studies, document how criminal codes have become increasingly convoluted and, at times, contradictory.[121] For example, the Illinois Supreme Court recently interpreted a statutory provision limiting the imposition of consecutive sentences to trump a separate provision aggravating

119. For a brief discussion of the history of Pennsylvania's guideline system, see Reitz, *supra* note 18, at 228. For a detailed analysis of the legislative history behind the federal guidelines, see Stith & Koh, *supra* note 73.

120. The effects of guidelines on sentence length are given by the coefficients for "Voluntary" and "Presumptive," and they are consistently negative. In regressions (not reported here) run separating the various states, Michigan and Missouri's guidelines seem to lead to longer sentences, while Ohio and Virginia's lead to shorter sentences. States using guidelines throughout the period also appear to have shorter average sentences.

121. Paul H. Robinson & Michael T. Cahill, *The Accelerating Degradation of American Criminal Codes*, 56 HASTINGS L.J. 633 (2005).

sentences based on past criminal history; as a result, a recidivist defendant who pled guilty to five crimes received a sentence less than half that which he would have been eligible for had he pled guilty to just one.[122] Robinson and Cahill also note that the proliferation of statutory crimes often means that prosecutors can prosecute a given act under several different sections, each with different penalties.[123] While it is not clear that this "overcriminalization" has necessarily increased inconsistent sentencing, it is unlikely that it has helped to reduce it.

Even if other factors, such as judges paying more attention to their own behavior, counteracted any negative effects of inconsistent codification, the data available here are unable to observe such an effect. First, the particularly salient guideline adoptions, namely by Minnesota in 1980 (the first state to adopt guidelines) and by the federal government in 1984, with implementation in 1987 (the most high-profile set of guidelines), took place before the data here start, in 1989. In fact, of all the states that had implemented guidelines or determinate sentencing laws by 2000, more than half had acted before 1989. Moreover, trends in sentencing variation in nonguideline states since 1989 do not exhibit any particular pattern between 1989 and 2000.[124] Though the data here cannot say anything definitive, it appears unlikely that nonguideline states changed their behaviors in any significant way.

B. Why Do Voluntary Guidelines Work?

The empirical results in Part II.A suggest that voluntary guidelines in fact work. But that is only half of the issue: *That* they work does not explain *why* they work. And the question why is an essential one. Without an understanding of why voluntary guidelines work, it is difficult to predict how they will function in different situations or what steps states can take to make them more effective. This subpart considers three possible explanations for why voluntary guidelines work: (1) They provide useful information; (2) legislatures can threaten to replace them with more mandatory systems; and (3) judges can be held accountable for noncompliance (outside of appellate reversal).

The first possible explanation is that guidelines, both voluntary and presumptive, provide information—two types in particular. First, they provide useful signposts to judges who wish to sentence uniformly but fail to do so in practice. A judge might sentence two similarly situated defendants differently

122. *Id.* at 639–40 n.23 (citing People v. Pullen, 733 N.E.2d 1235, 1239 (Ill. 2000)).

123. *Id.* at 638–39.

124. The lack of any pattern suggests, perhaps, that the overcodification that concerns Paul Robinson and Michael Cahill has not resulted in significant changes in sentencing behavior.

without realizing it (and despite a desire to act otherwise), perhaps because he was in different moods at each sentencing hearing, because he forgot the exact sentence imposed at the earlier hearing, or because he improperly (but unconsciously) took into account factors like the defendants' race or sex. By providing judges with a sense of the proper sentence for each crime, guidelines (whether voluntary or presumptive) help them combat these unconscious acts.

Furthermore, guidelines provide a metric for evaluating the sentencing practices of judges. While many, perhaps most, judges aspire to sentence uniformly, others consciously account for race or sex, or allow their moods to influence sentencing. For these judges, the first informational effect is not particularly important. But voluntary guidelines might still be effective here, not because they provide information to "misbehaving" judges, but rather to their colleagues. By defining a "proper" sentence, guidelines help judges as a group to identify and informally sanction such misbehaving judges. The institutionalist model of judicial decisionmaking posits that institutional norms and their enforcement can meaningfully restrain judicial behavior.[125] Guidelines of either type can help foster and enforce such norms.

The second possible explanation for why voluntary guidelines work is that that they operate in the shadow of presumptive guidelines. If a state legislature feels that judges are not adhering to voluntary guidelines, it can threaten to replace them with presumptive guidelines; Maryland, for one, made such a threat several years ago (though it uses voluntary guidelines to this day). At one level, *Blakely* appears to deny legislatures this tool. If so, then the results developed here may provide information only on how well voluntary guidelines worked *before Blakely*, which is of little interest. But to argue that *Blakely* eliminates presumptive guidelines in principle overstates its holding. Rather, it only eliminates the ability of such guidelines to rely on *judicial* fact-finding for *upward* adjustments beyond the statutory maximum. This leaves available to legislatures in voluntary guideline states two sticks with which to encourage compliance. The first is to make the guidelines presumptive but to use sentencing juries to make the necessary factual findings. The second is to use a hybrid approach in which the findings needed for downward departures are mandatory, but those for upward departures remain voluntary. Michigan and Pennsylvania already use methods along these lines. Each of these approaches is discussed in more depth in Part III *infra*, but what matters for the discussion

125. *See* Anne Bloom, *The "Post-Attitudinal Moment": Judicial Policymaking Through the Lens of New Institutionalism*, 35 LAW & SOC'Y REV. 219 (2001). The "realist" or "attitudinal" model argues that judges use their decisions as vehicles for implementing their personal policy preferences. The growing "institutional" literature posits a wider range of inputs into the judicial decisionmaking process, focusing in particular on how group membership can alter behavior.

here is that each allows legislatures to continue to threaten to make voluntary guidelines more mandatory. These options may be less appealing than pre-*Blakely* presumptive guidelines, suggesting that the post-*Blakely* shadow may not be as strong as that before *Blakely*, but the shadow nonetheless remains.

The third possible explanation for why voluntary guidelines work is that forces other than the threat of reversal induce judges to adhere to them. For example, judges in Virginia are appointed and retained by the legislature. As long as the legislature believes in the guidelines it has created, it may prefer to retain those judges who follow them. Judges might also adhere to voluntary guidelines if, for example, such practices increased the chances of promotion to a higher court.

Ultimately, these theories will require empirical evaluation themselves, since each suggests different methods for encouraging judges to comply with voluntary guidelines. If the first theory (information provision) is correct, for example, legislatures and the judiciary need only provide judges with training sessions about how the guidelines operate and with internal reports on how well various judges are adhering to the guidelines to induce compliance. But if the second theory (shadow of presumptive) is more accurate, then legislatures need to be able to indicate to judges that a threat to make guidelines more binding is in fact a credible threat. Likewise, the third theory (retention and promotion incentives) posits that states may need to consider more significant changes, such as revamping how judges are selected at various levels (or at least be able to again make a credible threat to institute such changes).

III. THE ALTERNATIVES TO VOLUNTARY GUIDELINES

Though voluntary guidelines have the appeal of retaining structured sentencing with judicial fact-finding, they are not the only form of guidelines available following *Blakely*. Commentators have focused primarily on three other options: (1) the use of sentencing juries; (2) the "uncapping" of guidelines (in which the maximum sentence available in each cell of a guideline chart is the statutory maximum for the offense in question); and (3) the "inversion" of guidelines (in which every aggravator is redefined as the absence of a mitigator). Unlike voluntary guidelines, each of these systems retains presumptive elements. Since presumptive guidelines appear to fight variation more effectively than voluntary guidelines and often work at least as well at addressing the use of impermissible factors, there is some appeal to maintaining binding restraints on judges. It is impossible to state a priori exactly how these systems compare to voluntary guidelines, but this part fleshes out the nature of the tradeoffs policymakers will have to consider.

A. Alternative 1: Sentencing Juries

Sentencing juries[126] are perhaps the clearest way to retain the mandatory nature of presumptive guidelines: Rather than having judges find the necessary sentencing facts, juries find them. Such an arrangement clearly comports with *Blakely*, and it keeps in place the same restrictions on judicial discretion that existed before *Blakely*—in fact, it makes them stronger, since judges no longer get to find the sentencing facts. One state already uses this approach. Anticipating *Blakely* by several years, the Kansas Supreme Court held its guidelines system unconstitutional following *Apprendi*,[127] and the state legislature responded by adopting sentencing juries.[128] The majority in *Blakely* held up Kansas as a model to follow.

Sentencing juries have two arguments in their favor. The first is that they have the potential to preserve the presumptive nature of guidelines. As the results in Subparts II.A.2–.3 indicate, presumptive guidelines appear to reduce variation more effectively than voluntary guidelines, and they roughly do at least as good a job at eliminating the use of impermissible factors. To the extent that jury fact-finding approximates judicial fact-finding (an issue addressed below), sentencing juries preserve the pre-*Blakely* sentencing structure.

A second possible, though perhaps more difficult, argument for sentencing juries is that they better live up to the "spirit" of *Blakely* than voluntary guidelines. At the heart of Justice Scalia's opinion in *Blakely* is a desire to "give intelligible content to the right of jury trial."[129] Voluntary guidelines satisfy *Blakely* not by finding ways to empower juries, but rather by capitalizing on exceptions present in that case and its precedents. Sentencing juries thus seem more consistent with a goal of strengthening juries.

The one challenge with using this argument to support the use of sentencing juries, however, is the apparent reason *why* Justice Scalia wants to strengthen the role of the jury. *Blakely* does not appear to be built around the idea that juries are superior fact-finders—in fact, in a different opinion issued

126. Note that the term "sentencing jury" is ambiguous. It can refer either to a jury that imposes the sentence itself, or a jury that merely finds whether aggravating or mitigating facts exist; in the latter case, sentencing remains with the judge, but he can use only those facts found by the jury. This Article uses "sentencing juries" in the second (fact-finding) sense, although there are six states that allow a jury to impose the sentence itself. *See* Nancy J. King & Rosevelt L. Noble, *Felony Jury Sentencing in Practice: A Three State Study*, 57 VAND. L. REV. 885, 886 (2004). The six states are Arkansas, Kentucky, Missouri, Oklahoma, Texas, and Virginia.

127. *See* State v. Gould, 23 P.3d 801 (Kan. 2001).

128. *See* KAN. STAT. ANN. § 21-4718 (Supp. 2005); *see also* Blakely v. Washington, 542 U.S. 296, 310 (2004).

129. *Blakely*, 542 U.S. at 305.

the same day as *Blakely*, Justice Scalia explicitly notes that juries are not necessarily any better than judges.[130] Instead, the focus of *Blakely* is on the (apparently doctrinal, not pragmatic) need for juries to protect legal rights. Presumptive guidelines create rights for the defendant by making it clear that a sentence above a certain level cannot be imposed without a specific finding; in effect, the defendant has a right to the lesser sentence absent this finding. Justice Scalia makes it clear that only a jury can overturn this right.[131] At the same time, by noting the continuing endurance of *Williams*, Justice Scalia also makes it clear that such rights do not exist absent legislative creation. Voluntary guidelines, by the very volition built into them, thus never create any expectation on the part of a defendant in the first place.

It also remains an open empirical question whether sentencing-jury systems can in fact closely replicate the outcomes seemingly achieved with presumptive guidelines. In other words, it is not immediately clear that jury-based presumptive guidelines will reduce variation (and the use of impermissible factors) as well as judge-based presumptive guidelines. Unlike judges, juries are essentially one-shot players. Not only do the jurors associate with each other only once, but they often are involved in the formal legal system only once (at least in such a capacity). This implies that the informal forces that may induce consensus among judges are less likely to apply to juries. Furthermore, jurors lack any knowledge of how earlier, similar issues were resolved. Even if they wished to act consistently, they lack the necessary experience. And, as Stephanos Bibas warns, jury-imposed sentences would be difficult for appellate courts to review.[132] Thus, sentencing juries might fail to capture all the apparent benefits of presumption.

But this problem with sentencing juries can be overstated, since *Blakely* requires only jury fact-finding, not jury sentencing. This allows a judge to apply

130. In *Schriro v. Summerlin*, 542 U.S. 348 (2004), Justice Scalia states that, "[w]hen so many presumably reasonable minds continue to disagree over whether juries are better factfinders *at all*, we cannot confidently say that judicial factfinding seriously diminishes accuracy." *Id.* at 356.

131. In *Blakely*, Justice Scalia notes that, "[i]n a system that says the judge may punish burglary with 10 to 40 years, every burglar knows he is risking 40 years in jail. In a system that punishes burglary with a 10-year sentence, with another 30 added for use of a gun, the burglar who enters a home unarmed is *entitled* to no more than a 10-year sentence—and by reason of the Sixth Amendment the facts bearing upon that entitlement must be found by a jury." *Blakely*, 542 U.S. at 309. And immediately preceding in that opinion he argues that "[o]f course indeterminate schemes involve judicial factfinding, in that a judge . . . may implicitly rule on those facts he deems important to the exercise of his sentencing discretion. But the facts do not pertain to whether the defendant has a legal *right* to a lesser sentence—and that makes all the difference insofar as judicial impingement upon the traditional role of the jury is concerned." *Id.*

132. Stephanos Bibas notes the importance of finality of a jury acquittal. Bibas, *supra* note 50, at 338.

the facts found by a jury more consistently than the jury itself might do. But the sentencing jury nonetheless changes the way such guidelines would operate. For example, a judge's discretion at sentencing is now asymmetric: He has flexibility in how he uses what the jury finds, but almost none when it comes to what the jury *does not* find. This asymmetry may not hurt a judge's ability to impose enhanced sentences—Nancy King and Rosevelt Noble have found evidence that juries can be quite punitive at sentencing, suggesting that they may be generally willing to find aggravators—but it may restrict his ability to mitigate.[133] And in a system like the pre-*Booker* federal system, in which enhancement and mitigation provisions were effectively mandatory (in marked contrast to state guidelines),[134] a fact-finding jury would act much like a sentencing jury (here, a jury imposing the actual sentence).

Justice Breyer raised a second concern with sentencing juries, namely that they lack the sophisticated legal training of a judge. In *Blakely*, he asked, "How are juries to deal with highly complex or open-ended Sentencing Guidelines obviously written for application by an experienced trial judge?"[135] But Justice Breyer's concern seems far more applicable to the federal system than to state guidelines. The aggravating and mitigating factors used in state guidelines do not seem to require extensive legal training,[136] and state guidelines do not require as much fact-finding as the federal guidelines.[137] In the federal system, the extensive fact-finding could be problematic—Albert Alschuler notes that prosecutors in one post-*Blakely* federal case submitted a special-verdict form that was twenty pages long, leading him to conclude that "the Guidelines were not meant for juries"[138]—but it is unclear whether this would be an issue of much import for the states. Nonetheless, it is possible that the use of lay juries could impose some limitations on the nature of guidelines that states can adopt.

---

133. King and Noble quote numerous prosecutors who appreciate jury sentencing because it encourages defendants to accept pleas out of fear of the punitive nature of such juries. King & Noble, *supra* note 126.

134. In presumptive state systems, judges could not enhance or mitigate without making a specific factual finding, but making that finding did not then compel them to either enhance or mitigate. In other words, the presumptive restraint was a only a necessary, not a sufficient, condition.

135. *Blakely*, 542 U.S. at 346–47 (Breyer, J., dissenting) (citation omitted).

136. *See, e.g.*, N.C. GEN. STAT. § 15A-1340.16 (2005) (listing aggravating factors, all of which a jury should be able to comprehend); TENN. CODE ANN. § 40-35-114 (Supp. 2005) (listing aggravating factors).

137. Much of the fact-finding that takes place during the guilt phase of a state trial is reserved in the federal system for the sentencing phase. As a result, sentencing proceedings are more complex in the latter.

138. Albert W. Alschuler, *To Sever or Not to Sever? Why* Blakely *Requires Action by Congress*, 17 FED. SENT'G REP. 11, 13 (2004) (citing United States v. Medas, 323 F. Supp. 2d 436 (E.D.N.Y. 2004)).

Sentencing juries may also be more expensive than a guideline system. First, the same procedural protections present in jury trials may apply to jury sentencing as well,[139] suggesting that jury sentencing may require discovery and other procedural protections, increasing the cost of sentencing both in terms of money and time. These protections could also limit the resources that are available to make sentencing decisions: Presentence reports, for example, which often contain hearsay evidence, may no longer be admissible.

Moreover, the costs of jury sentencing will rise if bifurcation is required; Justice Breyer notes in *Blakely* that bifurcation has already driven up the cost of capital sentencing.[140] And at least two arguments can be made for required bifurcation. First, the prosecution may not be able to allege certain facts relevant to sentencing in the initial indictment, since such facts would prejudice the jury during the guilt phase. Unified trials could thus prevent the state from raising certain relevant sentencing issues. Second, absent bifurcation, a defendant could find himself in the difficult position of having to argue contradictory positions. As Justice Breyer suggests, if sentencing facts are included in the indictment, some defendants could be forced either to make arguments such as, "I did not assault that person, and if I did, I did not use a gun," or to leave unaddressed certain sentencing facts for the sake of asserting innocence unambiguously. Such an outcome is unappealing—to Justice Breyer, it may even unconstitutionally violate due process.[141]

Note, though, that it is possible to overstate the magnitude of this problem as well, since bifurcation is necessary only in those cases that actually involve a jury trial. When the defendant pleads guilty, the jury is needed only at sentencing (if at all), eliminating the in-the-alternative concerns; when the defendant opts for a bench trial (or waives the need for a jury at sentencing), no jury is needed at any point. Currently, fewer than 4 percent of all criminal cases result in jury trials, suggesting that while costs *per jury trial* may go up, the overall impact may be slight.[142]

A fourth and final concern is an evidentiary problem with sentencing juries. Even with bifurcation, it is unclear why evidence considered prejudicial at trial should be permitted at sentencing: Evidence that is prejudicial at the guilt phase is likely still prejudicial at sentencing. If some evidence is

139. *See* King & Klein, *supra* note 50, at 319; *see also Blakely*, 542 U.S. at 336 (Breyer, J., dissenting) (noting that sentencing juries will likely require formal evidentiary rules to prevent prejudice).

140. *Blakely*, 542 U.S. at 336.

141. *Id.* at 335. Justice Breyer asks, "[H]ow can a Constitution that guarantees due process" force defendants to make these sorts of in-the-alternative arguments? *Id.*

142. Bibas points out that 91 percent of defendants accept guilty pleas, 5 percent appear in bench trials, and 4 percent in jury trials. Stephanos Bibas, *Judicial Fact-Finding and Sentence Enhancements in a World of Guilty Pleas*, 110 YALE L.J. 1097, 1150 (2001).

impermissibly likely to encourage a jury to convict a defendant simply for being a "bad person," why is it not also likely to encourage it to find aggravating facts (or to not find mitigating ones) for the same reason? Moreover, it is unclear that an improper enhancement is inherently any less prejudicial than an improper conviction. Some enhancements, such as the one considered in *Blakely*, have the ability to double the length of time served (if not magnify it even more). An improper enhancement can have significant implications.

The tradeoffs with sentencing juries are thus clear. On the one hand, they preserve the presumptive nature of the guidelines used in many states, and in doing so they may retain the benefits of presumptive guidelines demonstrated in Part II.A. On the other hand, they may result in outcomes less consistent than those made with required judicial fact-finding, they may require less nuanced guidelines, they may be more expensive, and they introduce concerns of prejudice. How these tradeoffs influence the relative efficacy of sentencing juries with respect to voluntary guidelines is ultimately an empirical question.

B. Alternative 2: Uncapped Guidelines

The second alternative to pre-*Blakely* presumptive guidelines is "uncapped" guidelines: Each cell in a guideline grid would contain only a minimum sentence, with the maximum in every case being the statutory maximum for that offense. Such guidelines are essentially sets of almost-mandatory minimums. Based on factual findings made at sentencing, a judge faces a floor below which he cannot go, but otherwise he is free to set any sentence he wishes. Two states, Michigan and Pennsylvania, use this general approach,[143] and Bowman proposed it to Congress as a way to make the federal guidelines comply with *Blakely*. At least one commentator, writing before *Booker*, argued that Congress would likely adopt this approach if it felt compelled to revise the federal guidelines.[144] These uncapped guidelines are

143. In both Michigan and Pennsylvania, the guidelines provide ranges of sentences from which a judge chooses the minimum sentence; the judge remains free to choose any sentence up to the statutory maximum for the maximum sentence. For example, the guidelines in Pennsylvania provide a range of fifty-four to seventy-two months for an offender with an offense score of 11 and a criminal history score of 3. *See* 42 PA. CONS. STAT. ANN. § 9721 (West Supp. 2006). When setting the minimum sentence, then, the judge must choose a sentence within that range, but he is unfettered with his choice for a maximum sentence. The Pennsylvania guidelines include provisions for enhancing or reducing the minimum sentence (in the example here, the judge can set a minimum sentence as low as forty-two months if mitigators dominate aggravators or as high as eighty-four months if the opposite results), but these do not appear to violate *Blakely* since they do not affect the maximum.

144. *See* Bibas, *supra* note 50, at 339 ("Raise the Guidelines Maxima—The Most Likely Response").

permissible under *Blakely* because of the asymmetry in that decision: *Blakely* applies only to upward departures, leaving legislatures free to impose required judicial fact-finding for downward departures.[145]

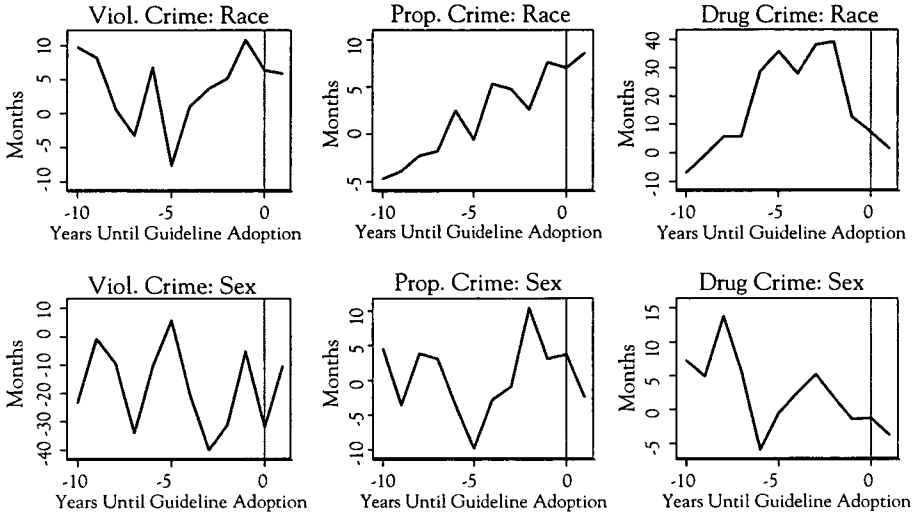
But constitutional permissibility need not imply effective policy. How well do these guidelines work? In regressions not reported here, I disaggregated Michigan from the rest of the presumptive guideline states to examine the effectiveness of its guidelines. These results, however, are tentative, due to a significant limitation in the data. Michigan adopted its present guideline regime only at the beginning of 1999, so the NCRP data used here provide only a single year of (lagged) postadoption data for the state. This temporary lack of data implies that the empirical results for Michigan are still quite provisional.

In general, Michigan's guidelines appear to have had little effect on the mean difference for violent or property crimes. The one exception is that Michigan's guidelines, not Ohio's, seem to drive the drop in mean difference for property crimes when linear corrections for endogeneity are employed; otherwise, Michigan's guidelines have a negligible effect. For drug crimes, however, Michigan's and Ohio's guidelines appear roughly equally effective, except when the quadratic correction for endogeneity is included. In that case, Michigan's guidelines again have no apparent effect.

For impermissible factors, Michigan's lack of extensive postadoption data is even more problematic. As demonstrated in Figure 2, for all three classes of crimes, the differences between either black and white sentences or male and female sentences display significant variation prior to adoption. As a result, it is hard to determine whether any change in the one observation following adoption is due to the guidelines or to the other forces already influencing preadoption trends. Not surprisingly, regressions attempting to isolate the effect of Michigan's guidelines do not produce stable results.

145. *See Blakely*, 542 U.S. at 301 (citing Apprendi v. New Jersey, 530 U.S. 466, 490 (2000)). *Booker* retains this asymmetry. *See* United States v. Booker, 543 U.S. 220 (2005).

FIGURE 2
DIFFERENCE BETWEEN RACE AND SEX
IN SENTENCE LENGTH IN MICHIGAN



Even if future data indicate that Michigan's guidelines are effective at confronting variation and the use of impermissible factors, these guidelines still have the possible drawback of approximating mandatory minimums in an era in which such sanctions are increasingly under attack. While the guidelines used in Michigan (as well as Pennsylvania) are not precisely mandatory minimums, since sufficient mitigating evidence can lower the minimum possible sentence, they nonetheless exhibit much of the (possibly problematic) asymmetry of true mandatory minimums. Judges have more freedom to depart upward than under presumptive guidelines,[146] but not downward; this grants "hanging judges" more latitude without equally compensating their more lenient colleagues. Magnifying this structural feature of mandatory minimums is a possible psychological issue. As Bibas puts it, "Judges who were inclined to split the difference between the minimum and maximum before might be psychologically influenced by the new, higher maxima and move upward. . . . Higher maxima will provide higher mental anchors for defendants, making sentence bargains seem like larger discounts and the prospect of post-trial sentencing even more risky."[147] Thus, even more lenient judges may adjust

146. Technically, an increased sentence would not be a departure under this regime, since it would still fall within the (uncapped) presumptive range.

147. Bibas, *supra* note 50, at 339.

their sentencing practices upward, pushing the average sentence up even higher. And in fact, data from the NCRP support the claim that the Michigan approach leads to higher sentences. Average sentence lengths in Michigan rose after its guideline system was adopted in 1999,[148] and there is some (weak) empirical evidence linking it to that policy change.[149]

Bibas, focusing on the federal response to *Blakely*, suggests that if Congress adopted uncapped guidelines, federal prison sentences would likely not rise noticeably. He notes that few sentences in federal court involve upward departures (fewer than 1 percent) and, more important, that most judges sentence at the bottom of the relevant range, suggesting that they are not itching to impose larger sentences.[150] But this is a distinctive feature of the federal guidelines, which appear to be uniquely more punitive than the judges who apply them. Presumptive ranges in state guidelines are generally less severe and thus more amenable to upward enhancements. In Minnesota, for example, 8.1 percent of sentences in 2002 resulted in upward departures, as did 7 percent in North Carolina in 1996, and 3 percent in Washington in 2003.[151] And the evidence from Michigan suggests that an uncapped system can very well result in longer average sentences.

Of course, a change in policy that results in longer sentences is not per se problematic; what constitutes an "appropriate" sentence length is a difficult normative question. But at least in the current economic climate, such a transition is likely to be unappealing to state legislatures. As states continue to wrestle with large budget deficits, they are increasingly cutting back, or at least trying to restrain, expenditures on corrections.[152] It is also possible that state legislatures are starting to adopt less hard-line views of crime for reasons unrelated to budget difficulties, perhaps due to still-falling crime rates or a growing appreciation of the collateral costs of incarceration.[153] Thus, while

148. For violent crimes, the average sentence imposed in Michigan was 191.5 months during 1989–1998 and 195.1 months for 1999–2000; for property crimes, 120.6 months during 1989–1998 and 140.8 months for 1999–2000; and for drug offenses, 189.0 months during 1989–1998 and 205.4 months for 1999–2000. If the point of change is 1999–2000 rather than 1998–1999 (to account for the one-year lag), then the changes are from 191.6 months to 197.4 months for violent crimes, 122.2 months to 142.9 months for property crimes, and 190.8 months to 200.0 months for drug crimes; in these latter examples, however, there is only one postadoption year (2000). Note that Michigan has not abolished parole, which may explain why the sentences imposed are so long. MICH. COMP. LAWS ANN. § 791.231a (West 1998).

149. In regressions not reported here, the coefficient for the effect of Michigan's guidelines on average sentence length is often positive and both numerically and statistically significant.

150. Bibas, *supra* note 50, at 339.

151. *See* SKOVE, *supra* note 56, at 11–12.

152. *See, e.g.*, Wool & Stemen, *supra* note 73.

153. *See, e.g., id.*

Bibas may be correct that this is the approach Congress may adopt, it is less clear whether state legislatures will be willing to accept the risk of longer average sentences.

The results in Part II.A, however, point to a possible way to address the concerns that uncapped guidelines both (1) do not sufficiently reduce variation and the use of impermissible factors and (2) lengthen sentences; and it is one that Bowman himself has suggested. In his testimony before Congress, Bowman did not suggest a system exactly like that of Michigan and Pennsylvania, but rather a hybrid approach: While the guidelines would set forth binding minimums, they would also propose suggested maximums. In other words, the Bowman proposal weds mandatory minimums with voluntary guidelines for maximum sentences.[154] What, then, is gained (or lost) by adopting this hybrid approach instead of purely voluntary guidelines? The hybrid approach may reduce variation more effectively than voluntary guidelines, since it restricts more firmly the ability for judges to depart in one direction, namely downward. But then it is not immediately clear that such asymmetry is (at least normatively) appealing. Moreover, the hybrid approach may also lead to longer sentences, the benefit of which depends on the financial health of the state, its underlying crime rate, and other highly contextual factors.

C. Alternative 3: Inverting the Guidelines

A third alternative also draws on *Blakely*'s asymmetry. This proposal suggests establishing guidelines with high presumptive sentences and extensive sets of downward departures, possibly turning each aggravator into a mitigator.[155] In other words, if a presumptive guideline system set a base sentence of ten years for robbery with a five-year enhancement for having a gun, the "inverted" presumptive system would set the base sentence for robbery at fifteen years with a five-year reduction for *not* having a gun. Though this approach satisfies the literal holding of *Blakely*, it suffers from at least two possible defects. First, it might not pass constitutional muster. The Court warned in *Apprendi* that it would view such wholesale inversions skeptically.[156] At least one commentator feels that such an approach would be "too gimmicky" to survive

154. *See Hearing*, *supra* note 58, at 35.

155. *See* Bibas, *supra* note 50, at 339.

156. Apprendi v. New Jersey, 530 U.S. 466, 490–91 n.16 (2000). Justice Stevens warned that "if New Jersey simply reversed the burden of the hate crime finding (effectively assuming a crime was performed with a purpose to intimidate and then requiring a defendant to prove that it was not . . .), we would be required to question whether the revision was constitutional under this Court's prior decisions." *Id.*

review,[157] although others wonder how aggressively the Court will monitor such responses to *Blakely*.[158]

Second, even if it were constitutional, an "inverted guideline system" is *not* the mirror image of the current presumptive system: Upward and downward departures are qualitatively different actions, at least in state courts. State court judges, unlike their federal counterparts, are politically accountable. A judge who imposes a twelve-month enhancement on a presumptive twelve-month sentence is likely to be portrayed by the media and viewed by the electorate much differently than one who allows a twelve-month reduction to a presumptive thirty-six month sentence for an identical crime, even though the defendant serves twenty-four months in either case. The former judge comes across tough on crime, the latter weak. States could again find themselves facing the prospect of longer average sentences. In a sense, inverted guidelines could ultimately come to resemble uncapped guidelines like Michigan's. Where Michigan's guidelines prevent judges from setting lower sentences through explicit controls, inverted guidelines might prevent judges from taking advantage of available downward departures through political pressure. As a result, whatever benefits and costs exist with the asymmetry in Michigan's system ultimately may apply to some degree here as well.

D. Other Alternatives

Though sentencing juries, uncapped guidelines, and guideline inversion are the leading candidates to replace presumptive guidelines, they are not the only options. In perhaps the most comprehensive review so far of the alternatives, Bibas lists at least three others besides those discussed above: (1) a return to strict determinacy; (2) a return to indeterminacy; and (3) the expanded use of appellate review of sentencing.[159] This subpart briefly discusses these possibilities.

As Justice Breyer suggests in his dissent in *Blakely*, states may adopt strict determinacy as an alternative to presumptive guidelines.[160] A simple strictly determinate system has the key drawback of treating many cases alike that surely should be treated differently. It reduces variation and limits discretion, but at possibly too great a cost. And there is at least weak evidence that states do not like strictly determinate systems. As discussed *supra* Part I.A, no state

157. *See* Bibas, *supra* note 50, at 339.

158. *See* Ronald J. Allen & Ethan A. Hastert, *From* Winship *to* Apprendi *to* Booker: *Constitutional Command or Constitutional Blunder?*, 58 STAN. L. REV. 195 (2005).

159. Bibas, *supra* note 50, at 338–39.

160. Blakely v. Washington, 542 U.S. 296, 330 (2004) (Breyer, J., dissenting).

has adopted a determinate system since the 1970s,[161] and at least one state that employed a determinate system in the 1970s, North Carolina, replaced it with presumptive sentencing guidelines in the 1990s.

A more viable variant is what Justice Breyer refers to in *Blakely* as a "complex charge offense"[162] system, in which each crime is subdivided into numerous detailed "subcrimes." Justice Breyer provides the example of a statute that defines myriad degrees of robbery, based on what type of institution robbed, on whether a firearm was used, on whether a death threat was made, and so forth.[163] Such a system may require substantial recodification of a state's criminal code, although this is a one-time cost that might not be as great as it might seem. Also, and perhaps more important, a complex-charge-offense scheme shifts sentencing discretion almost wholly to the prosecutor: By choosing which specific subcrime or set of subcrimes to charge, the prosecutor effectively chooses the sentence, or at least the specific sentence range to which a judge is restricted. Thus a complex-charge-offense approach would operate much differently than a guideline system, possibly vesting substantially more power in prosecutors than in judges. Whether this transfer of authority would be a "good" or "bad" outcome is beyond the scope of this Article, but it is a key issue policymakers should consider if thinking about such a system.

At the other end of the spectrum of choices is a return to indeterminacy. As Reitz has noted, indeterminacy is the great unspoken secret of modern American sentencing. Despite all the attacks leveled against it, and despite the extensive sentencing reforms imposed around the nation, it remains the dominant form of sentencing, used in some form in twenty-eight states.[164] The concerns with indeterminacy are well known, namely that a defendant's sentence can depend on the judge's mood that day or, perhaps worse, on the defendant's race or other impermissible factor, and that such idiosyncratic variance in sentences are an affront to the "rule of law." But it is perhaps worth noting that despite such concerns, a majority of jurisdictions have opted to retain indeterminate systems. At some level, legislatures and governors must feel that judges employ their discretion in roughly acceptable ways.

161. *See* TONRY, *supra* note 10, at 28.

162. *Blakely*, 542 U.S. at 334 (Breyer, J., dissenting).

163. *Id.*

164. No state is wholly indeterminate, since every states employs mandatory minimums for some offenses. *See* BUREAU OF JUSTICE ASSISTANCE, 1996 NATIONAL SURVEY OF STATE SENTENCING STRUCTURES 8–9 Exhibit 1-4 (Sept. 1998), *available at* http://www.ncjrs.gov/pdffiles/169270.pdf. And only ten states do not employ any of the major sentencing reforms, namely parole limitations, sentencing guidelines, truth-in-sentencing laws, or strike legislation. But twenty-eight states employ neither guidelines nor determinate sentencing regimes. *See id.* at 10–18.

The efficacy (or lack thereof) of indeterminate sentencing, and the normative arguments for and against it, however, are not directly relevant to the goal of this Article. The question considered here takes as a given the idea that structured sentencing should be retained and asks how to best accomplish this in light of *Blakely*; indeterminacy does not accept this underlying assumption. In their responses to *Blakely*, judges, legislators, and other commentators may find themselves rethinking the merits of structured sentencing regimes, but those questions are not part of the project here.

The third suggestion discussed by Bibas is the use of appellate oversight. Under the system he suggests, guidelines would become "rules of thumb": Judges could deviate from the guidelines, but they would have to provide detailed explanations that would be subject to appellate review.[165] The appeal of this approach is that appellate courts are better able to ensure uniformity than trial courts left to their own devices. This is more true the more review is de novo rather than for abuse of discretion,[166] and the more appellate courts are in fact willing to use their oversight powers.[167] How such a system would work in practice, however, is unclear. Bibas himself seems to indirectly suggest, in a different context, that it might not be particularly feasible. When evaluating the idea of uncapped guidelines, Bibas notes that "searching review" of upward departures "might make the upper bound seem too law-like and trigger *Blakely* protections,"[168] a point forcefully made by Justice Scalia in his dissent in *Booker*.[169] In other words, if appellate review is too strenuous, it creates a situation identical to that of a presumptive guideline system, and the need for jury fact-finding returns. Appellate review then ceases to be an alternative to guideline sentencing but simply changes *who* is imposing the guidelines. Perhaps appellate courts could limit the ability of judges to depart *downward*, but then appellate review begins to mimic uncapped guidelines. And while appellate courts might be able to regulate the procedures used at sentencing, it is unclear whether such review would have much effect on the substantive outcome.

---

165. Bibas, *supra* note 50, at 338.

166. *Id.*

167. Reitz, for example, notes that Pennsylvania's appellate courts appear not to review meaningfully sentences imposed by trial courts. Kevin R. Reitz, *Sentencing Guideline Systems and Sentence Appeals: A Comparison of Federal and State Experiences*, 91 NW. U. L. REV. 1441, 1471 (1997).

168. Bibas, *supra* note 50, at 339.

169. As Justice Scalia states in his dissent in *Booker*, "[A]ny system which held it *per se* unreasonable (and hence reversible) for a sentencing judge to reject the Guidelines is indistinguishable from the mandatory Guidelines system that the Court today holds unconstitutional." United States v. Booker, 543 U.S. 220, 311 (2005) (Scalia, J., dissenting).

CONCLUSION

The Supreme Court's decision in *Blakely v. Washington* will likely compel at least thirteen states to revamp their sentencing regimes. In each of these jurisdictions, the maximum sentence a defendant faced depended, at least in part, on facts found by a judge (rather than a jury); *Blakely* held that such fact-finding violates a defendant's Sixth Amendment jury right, and it thus invalidated a key element of these sentencing regimes. Not surprisingly, many commentators and judges have confronted the issue of how to redesign such sentencing laws to comply with *Blakely*'s requirements.

While much of the discussion has focused on three possible reforms—using juries at sentencing, replacing guidelines with what amount to somewhat-flexible sets of mandatory minimums, and inverting guidelines so that judges need only find mitigators (many of which would be the inverse of earlier aggravators)—one option that has received little attention is the use of voluntary sentencing guidelines. Because *Blakely* imposes the jury right only on required fact-finding, such guidelines pass constitutional muster. Moreover, as this Article demonstrates, these guidelines appear to work: Judges seem to adjust their sentencing behavior in response to voluntary guidelines. In particular, the data examined in this Article suggest that judges operating in voluntary guideline regimes sentence more consistently and with less attention to the race and sex of the defendant than do judges operating without any guidelines at all; this effect appears to take place both between states (comparing voluntary guideline states to nonguideline states) and within states (comparing a state before and after it adopts voluntary guidelines). Voluntary guidelines are not a perfect solution—judges working under now-invalid presumptive guidelines and determinate sentencing laws were even more consistent and paid less attention to race and sex—but they are a viable post-*Blakely* alternative.

Ultimately, legislators and other policymakers will have to compare the benefits (and costs) of voluntary guidelines to those of the other possible alternatives, such as fact-finding sentencing juries and uncapped guidelines. But the results in this Article, if not conclusive, indicate that voluntary guidelines represent a viable replacement for unconstitutional presumptive guidelines.

APPENDIX 1. THE DATA AND EMPIRICAL METHODOLOGY[170]

The primary source of data for this Article is the National Corrections Reporting Program (NCRP).[171] The NCRP is a voluntary program in which states collect detailed information about each admitted inmate both upon entry to and departure from prison. Among other things, the states gather data on a prisoner's offenses, sentences imposed (total maximum, total minimum, and single-longest sentence), and demographic data such as race and sex. Thirteen states have participated in the NCRP since its inception in 1983, and by 1998 thirty-one states were reporting meaningful data; the data have become more reliable as the years have passed.[172]

In this Article, I use data gathered by the NCRP between the years 1989 and 2000. During this period, fourteen states consistently reported data. Of these, two adopted voluntary guidelines during the period (Missouri in 1997 and Virginia in 1995), and two presumptive (Michigan in 1999 and Ohio in 1996). One used voluntary guidelines throughout the whole period (Maryland), two used presumptive throughout (Tennessee[173] and Washington), and one used a determinate sentencing regime throughout (California).[174] The remaining six (Alabama, Illinois,[175] Kentucky, Nevada, Oklahoma, and South Carolina) never employed guidelines during the period.

Note that while I classify Virginia as a voluntary guideline state starting in 1995,[176] Virginia initially adopted voluntary guidelines at the state level in

---

170. These appendices provide a more detailed discussion of the data and empirical techniques used *supra* Parts II.A.2 and II.A.3.

171. The codebooks and data for the NCRP are available at http://www.icpsr.umich.edu/NACJD/ncrp (last visited Oct. 15, 2006).

172. For a brief discussion of the problems with NCRP data, especially in its earlier years, see Ilyana Kuziemko & Steven D. Levitt, *An Empirical Analysis of Imprisoning Drug Offenders*, 88 J. PUB. ECON. 2043 (2004).

173. Because Tennessee implemented its guidelines in 1989, this Article does not count it as a guideline state until 1990. However, it is not included as a "during-the-period" adopter since it has only one preadoption year. The corrections for endogeneity focus on preexisting trends, which cannot be calculated for Tennessee (one year cannot indicate a trend). It therefore makes more sense to include it only in the regressions using the full set of states.

174. Other than Tennessee, these states adopted their sentencing regimes prior to 1983, the first year NCRP data were gathered.

175. For an explanation of why Illinois—nominally a determinate sentencing state—is classified as a nonguideline state for this Article, see *supra* text accompanying note 59.

176. I am not alone in doing this. Reitz similarly classifies Virginia as a voluntary guideline state as of 1995. Reitz, *supra* note 18, at 226 tbl.6.1.

1991.[177] When it developed its truth-in-sentencing law as part of a large sentencing-reform bill in 1994, however, the legislature narrowed the guideline ranges available to judges and expanded the number of categories of crimes covered by the guidelines.[178] For example, the sentence range for selling a schedule I or II drug for an offender with no record and no other offenses fell from over four years to under one, and the range for two counts of grand larceny committed by an offender on parole with a prior for grand larceny fell from four and a half years to two and a half years. For violent crimes, some of the drops were even more dramatic: The sentence range for robbery of a residence with a firearm (but no injury) committed by an offender with no record fell from over ten years to barely over three, and that for a class of rape with a firearm by an offender with a prior fell from twenty-seven years to just over seventeen years. This narrowing appears to have had a strong effect on how consistently judges sentenced, particularly for violent and property crimes, as shown in Figure A.[179] As a result, the regressions in this Article treat Virginia as having adopted its guidelines in 1995.[180] This decision is ultimately an important one. When 1991 is used as the year of adoption rather than 1995, the results for Virginia are much weaker. For violent crimes, for example, the effect of voluntary guidelines on variation when all the states are used falls from –10.282 to –7.204 (with both coefficients significant at at least the 5 percent level); when only the four changing states are used, the drop is from –4.704 to –0.727. It is clear from Figure A why this happens, but it is also clear that this appears to understate improperly the true effect of voluntary guidelines. Moreover, Figure A indicates the importance of how legislatures and commissions *design* guidelines.

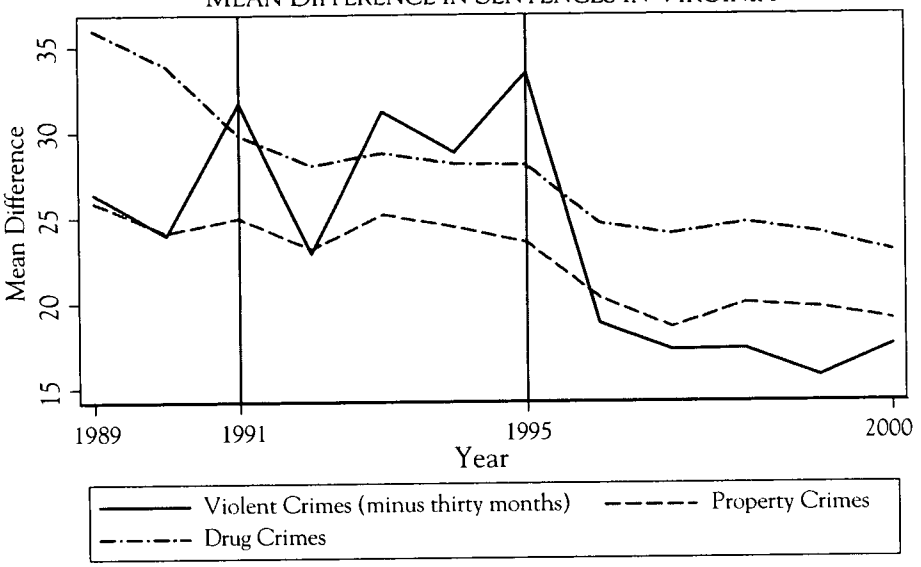
177. For a comprehensive discussion of the development of Virginia's guidelines, see BRIAN J. OSTROM ET AL., TRUTH-IN-SENTENCING IN VIRGINIA 10 (1999), *available at* http://www.ncjrs.org/pdffiles1/nij/grants/187677.pdf.

178. *See id.* at 30–32.

179. To fit all three types of crimes on one graph, Figure A subtracts thirty months from the annual average mean difference for violent crimes.

180. At some level, for the years 1991–1994 this is analogous to treating Illinois, nominally a determinate sentencing state, as indeterminate due to the wide ranges its determinate sentencing law provides.

FIGURE 3
MEAN DIFFERENCE IN SENTENCES IN VIRGINIA



In all the regressions, the dependent variable is "longest sentence imposed" or a function of it (its mean difference).[181] The observations included in this Article, however, are only those for which the longest sentence is less than or equal to six hundred months (fifty years). This affects two types of sentences. First, this drops offenders sentenced to life, life plus years, life without parole, or death. This is partially a concession to coding. The one alternative, using an actuarial approach (if the expected age of death for a twenty-five year old prisoner sentenced to life without parole is sixty-five years, then he has effectively received a forty-year sentence), has the problem of creating the appearance of variation where none exists. A judge who sentences both a twenty-year old defendant and a thirty-year old defendant to life without parole for identical crimes is sentencing consistently (with no variation), but the actuarial approach would result in his sentences appearing inconsistent. If the question is, "Do judges apply the law consistently?" then the actuarial approach can produce false inconsistency. Furthermore, these sentences are also dropped to prevent any small pockets of outliers from skewing the results. Second, even

181. The alternatives are "minimum sentence imposed" and "maximum sentence imposed." Minimum sentence lengths are reported with much less frequency than maximums, and both maximum and minimum sentence lengths are clouded by multiple offenses for which sentences are imposed concurrently. The "longest sentence imposed" variable provides a measure of the sentence imposed for a *single* (and generally the most important) offense.

though sentences greater than six hundred months (but not officially "life" sentences) could be included,[182] it seemed illogical to do so when they often act effectively as life sentences. Moreover, such extreme outliers could again exert undue influence over the regression coefficients. Ultimately, very few offenses result in such draconian punishments: Dropping life and death sentences omits approximately 1 percent of the observations, and only a small number of sentences exceed six hundred months.

In an earlier version of this Article, I attempted to incorporate rough measures of differences across states in judicial behavior as well. Since the NCRP does not provide any information specifically identifying the offender in a given observation, it is impossible to tie the offender to the judge who sentenced him. Instead, a more indirect measure was considered: the means by which a state selects its judges. The approaches states use vary, from partisan and nonpartisan elections, to legislative appointments, to nominations by merit commissions with retention elections.[183] Unfortunately, during the period under review in this Article, no state changed its judicial-selection procedures, which means that indicator variables for judicial-selection methods are redundant in the presence of state fixed effects. As a result, these measures were ultimately dropped.

For the first empirical test in this Article, examining the effect of guideline adoption on the mean difference in sentence length, I use the following regression:

$$dif_{ijkt} = \alpha + \delta_v volguide_{kt} + \delta_p presguide_{kt} + \beta_b black_i + \psi_s sex_i + \varphi_s state_k + \varphi_y yr_t + \varphi_o offense_j + \varepsilon_{ijkt} \quad (1)$$

where dif_{ijkt} is the mean difference score for offender i convicted of crime j in state k in year t. $volguide_{kt} = 1$ if state k employs voluntary guidelines in year t, and $presguide_{kt} = 1$ if it uses presumptive guidelines in that year. $black_i = 1$ if i is black, and $sex_i = 1$ if i is a woman. Finally, $state_k$ is the state fixed effect variable, yr_t the year fixed effect term, and $offense_j$ the offense fixed effect dummy. This last set of fixed effects deserves discussion. The NCRP subdivides each of the three general classes of crimes into subcrimes: There are nineteen types of violent crimes (such as murder, homicide, rape, aggravated assault, and so on), fifteen types of property crimes, and ten types of drug offenses. Without the offense-specific dummies, the regression in equation (1) implicitly assumes that all offenses have the same (unconditional) average mean difference; including

182. The NCRP includes some observations for which the sentence length was apparently over 12,000 months.

183. The American Judicature Society provides a detailed discussion of these various approaches. *See* American Judicature Society, Judicial Selection Methods, http://www.ajs.org/selection/sel_state-select-map.asp (last visited Oct. 15, 2006).

the offense dummies allows each suboffense to have its own average mean difference, which seems more realistic. Since the coefficients of interest in equation (1), δ_v and δ_p, are state-level terms while the unit of observation is the individual, the standard errors are corrected for clustering.

I ran the regression in equation (1) six times, once for each of three classes of offenses—violent crime, property crime, and drug offenses—and then twice within each crime class, once using data taken just from the four states that adopted guidelines during the period 1989–2000, and once using data from the full sample of fourteen states. The regressions for drug offenses differ slightly from those for violent and property crimes. In particular, I included an extra set of dummy variables, interacting the state fixed effect with the offense fixed effect. This is a concession to an aggregation problem that appears to result from the unique recodification of drug offenses following guideline adoption. Without these indicator variables, the aggregate effect of presumptive guideline adoption for all offenses is greater than its effect on any one suboffense. Even with these variables, the aggregate effect of presumptive guidelines appears slightly larger than such guidelines' effect on any one suboffense, but the overstatement is essentially eliminated.

To calculate the percentage effect of guideline adoption on mean difference, I replaced the dependent variable in equation (1), dif_{ijkt}, with its natural logarithm. More precisely, I took the log not of the mean difference itself, but of one plus the mean difference (that is, the dependent variable is now $y_{ijkt} = ln(1+ dif_{ijkt})$).[184] This accounts for the fact that many observations of the mean difference are between zero and one. While $ln(x)$ dampens the distance between two values of x greater than one, it *magnifies* the distance between two xs less than one (and greater than zero). In other words, the distance between $ln(.5)$ and $ln(.3)$ is .5, more than twice the distance between 0.3 and 0.5.[185] This effect is magnified the closer the mean-difference score is to zero. Otherwise, the regression used to calculate the percentage effect is identical to that in equation (1).[186]

The corrected versions of equation (1) presented in Tables 3 and 4 are simple extensions of that regression. To correct for any possible state-specific time trend, I multiplied each state's state fixed effect term, $state_k$, by the year

184. Note that the use of dif_{ijkt} or $1+ dif_{ijkt}$ does not change the results in the unlogged regression given in equation (1) (except for the constant).

185. This is because $ln(.5) = -0.693$, and $ln(.3) = -1.204$.

186. In the end, the transformation has a strong effect only for drug crimes, where under some specifications the unlogged regression returned a negative coefficient for presumptive guidelines while the logged regression (without using the "plus one" correction) returned a positive coefficient, a logical impossibility. In all cases, adding one to the mean-difference score reduces the absolute effect of both types of guidelines, as one would expect.

vector, *year*, where $year_t$ is the value of the *year* vector associated with year t (for example, $year_{1991} = 1991$). Thus, the linear state-year trend regression is:

$$dif_{ijkt} = \alpha + \delta_v volguide_{kt} + \delta_p presguide_{kt} + \beta_b black_i + \psi_s sex_i + \varphi_s state_k + \varphi_y yr_t + \varphi_o offense_j + \xi_k state_k * year_t + \varepsilon_{ijkt} \quad (2)$$

That with the quadratic time trend is then:

$$dif_{ijkt} = \alpha + \delta_v volguide_{kt} + \delta_p presguide_{kt} + \beta_b black_i + \psi_s sex_i + \varphi_s state_k + \varphi_y yr_t + \varphi_o offense_j + \xi_k state_k * year_t + \zeta_k state_k * year_t^2 + \varepsilon_{ijkt} \quad (3)$$

The next test, discussed in Part II.A.3, examines the effect of guidelines on the use of race and sex at sentencing. The first regression presented in Table 5 is:

$$length_{ijkt} = \alpha + \delta_v volguide_{kt} + \delta_p presguide_{kt} + \beta_0 black_i + \beta_v (black_i * volguide_{kt}) + \beta_p (black_i * presguide_{kt}) + \psi_0 sex_i + \psi_v (sex_i * volguide_{kt}) + \psi_p (sex_i * presguide_{kt}) + \varphi_s state_k + \varphi_y yr_t + \varphi_o offense_j + \varepsilon_{ijkt} \quad (4)$$

Here, $black_i * volguide_{kt}$ is the interaction of the race and voluntary guideline indicators, and the other interaction terms are defined accordingly.[187] As with equation (1), I ran the regression in equation (4) six times, once for each class of crimes, and considering either four or fourteen states, with the standard errors adjusted to account for state-level clustering. These regressions also include the additional interaction indicators when considering drug offenses.

The corrected regressions run in Table 5 come from a simple extension to equation (4). The linear correction is:

$$length_{ijkt} = \alpha + \delta_v volguide_{kt} + \delta_p presguide_{kt} + \beta_0 black_i + \beta_v (black_i * volguide_{kt}) + \beta_p (black_i * presguide_{kt}) + \psi_0 sex_i + \psi_v (sex_i * volguide_{kt}) + \psi_p (sex_i * presguide_{kt}) + \varphi_s state_k + \varphi_y yr_t + \varphi_o offense_j + \xi_k state_k * year_t + \theta_{ik} black_i * state_k * year_t + \eta_{ik} sex_i * state_k * year_t + \varepsilon_{ijkt} \quad (5)$$

The quadratic correction then adds to equation (5) the terms $state_k * year_t^2$, $black_i * state_k * year_t^2$, and $sex_i * state_k * year_t^2$.

187. In Table 5, the coefficient for $black_i * volguide_{kt}$ is given by the term for "Black*Vol"; the coefficients for $black_i * presguide_{kt}$, $sex_i * volguide_{kt}$, and $sex_i * presguide_{kt}$ are defined likewise.

APPENDIX 2. AN ASIDE ABOUT MICHIGAN

One of the two states that this Article treats as adopting presumptive guidelines between 1989 and 2000 is Michigan, which imposed a presumptive regime in 1999. The history and structure of Michigan's guidelines, however, raise two important issues. First, it is possible that Michigan employed presumptive guidelines as early as 1984, though these were judicially, not legislatively, created. If so, then Michigan should not be classified as an "adopting" state in the models above. And second, the "uncapped" nature of its (current) guidelines means they survive *Blakely*, as the Michigan Supreme Court has recently held.[188] This could imply that the strong results for presumptive guidelines in the regressions above argue for the adoption of Michigan's system.

The history of Michigan's guideline system is a source of some confusion: Three different commentators tell three different stories. Writing in 1996 (before Michigan began the legislative reforms leading to the 1999 guidelines), Tonry classified Michigan's system as a voluntary guideline approach,[189] while in 2001 Reitz classified Michigan as a presumptive guideline state starting in 1999 with no reference to a voluntary guideline past.[190] During that time, in 1999, a long-serving Michigan judge, Paul Maloney, stated that judges in Michigan had operated under a presumptive regime since 1984.[191] The guidelines discussed by Maloney were created not by the legislature, but rather by a judicial committee, and they were made mandatory not by law, but by an administrative order of the Michigan Supreme Court.[192] As Maloney notes, however, by the mid-1990s the guidelines had become somewhat archaic, covering only a fraction of the crimes for which judges had to impose sentences,[193] and as a result they were becoming less effective.[194]

---

188. People v. Claypool, 684 N.W.2d 278, 286–87 n.14 (Mich. 2004).

189. TONRY, *supra* note 10, at 27.

190. Reitz, *supra* note 18, at 227 tbl.6.1.

191. Paul L. Maloney, *The Michigan Sentencing Guidelines*, 16 T.M. COOLEY L. REV. 13, 16 (1999).

192. *Id*; *see also* Admin. Order No. 1984-1, Sentencing Guidelines, 418 Mich. at lxxx (Jan. 17, 1984).

193. Maloney, *supra* note 191, at 16–17. New crimes could be added to the guidelines only by amendment, and the Michigan Supreme Court added no new offenses after 1988, even as the state legislature continuously added new crimes to the criminal code. *Id.* at 17.

194. Maloney argued:

> Inevitably, when you have significant numbers of offenses not covered by the guidelines, and you have no changes being made to a system of guidelines—and I am not here to suggest that we ought to be tinkering like the Federal Commission does every year or every other year on guidelines—but when the system is not being amended from time to time, I think its efficacy is eroded significantly, and disparity creeps into the system. When home invasion, for example, is not covered, but the old breaking and entering statutes are covered, or when retail fraud is not covered, but the old larceny statute was covered,

How, then, should this Article classify Michigan's guidelines? On one hand, an argument can be made that Michigan has been a presumptive regime since 1984, in which case it should be listed with California, Tennessee, and Washington as a presumptive-throughout-the-period state. On the other hand, the loopholes described by Maloney are sufficiently large that a crafty prosecutor might have been able to avoid the guidelines by choosing what particular offense to charge; in this case, Michigan seems more like a nonguideline state prior to the 1999 reforms. Ultimately, however, little seems to turn on the distinction. Classifying Michigan as a presumptive guideline state throughout the whole period does not change the results in meaningful ways (in part since Michigan adopted its current guidelines so late in the sample period).

The second concern stems from the fact that Michigan's guidelines survive *Blakely*: Do the strong results for presumptive guidelines argue in favor of a Michigan-style guideline system? Or are these results the product of Ohio's more traditional, but now impermissible, approach? When the regressions in Tables 3 and 5 are run again, separating the two presumptive states, the results for Michigan are mixed.[195] With admittedly few postadoption observations, Michigan's guidelines generally have little effect on the mean difference of violent and property crimes, but under some specifications they do reduce the mean difference for drug offenses. As for impermissible factors, the paucity of postadoption data for Michigan makes it impossible to draw even tentative conclusions. With more waves of NCRP data, I hope to generate more reliable estimates in the future.

These results imply, then, that Michigan's guidelines may be capable of addressing some of the concerns raised by this Article, though not all. However, even assuming the results for Michigan are reliable, commentators have raised serious concerns about the continued use of mandatory minimums. Even though Michigan's guidelines (like Pennsylvania's) are not exactly like mandatory minimums, since judges have the ability to depart downward following specific findings, the asymmetry of the system is still problematic for those who are seeking to reduce sentence lengths or who are concerned more generally with the inflexibility of mandatory minimums.

---

inevitably, there is disparity, because the judges are left unconstrained. Again, it is not any malevolence, but the bottom line is that left unconstrained without guidelines, more disparity results.

Id.

195. Part III.B discusses these results in more depth.

APPENDIX 3. THE IMPORTANCE OF CALIFORNIA

One of the fourteen states examined in this Article is California, which throughout the period in question possessed the largest prison population in the nation. Since this Article relies on only thirteen other states (not among which are such prisoner-heavy jurisdictions as Florida, New York, and Texas), California comprises a large fraction of the data: almost 45 percent of the observations for violent crime, almost 42 percent of those for property crimes, and over 52 percent of those for drug offenses. It is possible, then, that California plays a large role in the empirical results in this Article.

To examine the importance of California, I recalculated the results run in Parts II.A.2–.3 above with California omitted. In general, California appears to exert a noticeable influence over the results for violent and property crimes, though not for drug offenses. In other words, for violent and property crimes, the results when California is excluded are closer to those that consider only the four changing states than those that consider all the states. For drug offenses, the reverse holds, suggesting that omitting California does not significantly change those results. For impermissible factors, the one difference from this pattern is that the omission of California does not appear to alter the results for property crimes in any significant way.

***