THE YALE LAW JOURNAL

SONJA B. STARR & M. MARIT REHAVI

Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of *Booker*

ABSTRACT. This Article presents new empirical evidence concerning the effects of *United States v. Booker*, which loosened the formerly mandatory U.S. Sentencing Guidelines, on racial disparities in federal criminal cases. Two serious limitations pervade existing empirical literature on sentencing disparities. First, studies focus on sentencing in isolation, controlling for the "presumptive sentence" or similar measures that themselves result from discretionary charging, plea-bargaining, and fact-finding processes. Any disparities in these earlier processes are excluded from the resulting sentence-disparity estimates. Our research has shown that this exclusion matters: pre-sentencing decision-making can have substantial sentence-disparity consequences. Second, existing studies have used loose causal inference methods that fail to disentangle the effects of sentencing-law changes, such as *Booker*, from surrounding events and trends.

In contrast, we use a dataset that traces cases from arrest to sentencing, allowing us to assess *Booker*'s effects on disparities in charging, plea-bargaining, and fact-finding, as well as sentencing. We disentangle background trends by using a rigorous regression discontinuity-style design. Contrary to other studies (and in particular, the dramatic recent claims of the U.S. Sentencing Commission), we find no evidence that racial disparity has increased since *Booker*, much less because of *Booker*. Unexplained racial disparity remains persistent, but does not appear to have increased following the expansion of judicial discretion.

AUTHORS. Sonja B. Starr is a Professor at the University of Michigan Law School. M. Marit Rehavi is an Assistant Professor of Economics at the University of British Columbia and a Fellow of the Canadian Institute for Advanced Research. For helpful comments and conversations, we thank David Abrams, Daron Acemoglu, Alberto Alesina, Joe Altonji, Alan Auerbach, Nick Bagley, John Bronsteen, Ing-Haw Cheng, Kristina Daugirdas, John DiNardo, Avlana Eisenberg, Leonid Feller, Nicole Fortin, Nancy Gallini, Nancy Gertner, David Green, Sam Gross, Don Herzog, Jim Hines, Jill Horwitz, Thomas Lemieux, Justin McCrary, Julian Mortenson, Brendan Nyhan, J.J. Prescott, Eve Brensike Primus, Adam Pritchard, Jeff Smith, Sara Sun Beale, and participants at the Ninth Circuit Judicial Conference, the National Sentencing Policy Institute, the NBER Summer Institute, the annual meetings of the American Law and Economics Association and the American Society of Criminology, workshops at the University of Michigan, UBC, Duke, and Loyola-Chicago, and the CIFAR-IOG Workshop. Sharon Brett, Michael Chi, Michael Farrell, Ryan Gersovitz, Seth Kingery, Matthew Lee, Midas Panikkar, Art Robiso, Sabrina Speianu, and Adam Teitelbaum provided able research assistance.

ARTICLE CONTENTS

INTRODUCTION

In the United States, one of every nine black men between the ages of twenty and thirty-four is behind bars,[1] and, in 2003, the Bureau of Justice Statistics projected that one in every three young black men could expect to be incarcerated at some point in his life.[2] These rates far exceed those of any other demographic group—for instance, black males are incarcerated at nearly seven times the rate of white males.[3] The impact of demographically concentrated incarceration rates on offenders, families, and communities is a critical social concern.[4] But why do these gaps exist? Can they be explained by differences in criminal behavior, or by differences in how the criminal justice system treats offenders? If it is the latter, can the process be improved by reforms, such as changes to sentencing law?

These questions are not new. For decades, racial and other "legally unwarranted" disparities in sentencing have been the subject of considerable empirical research, which has in turn helped to shape major policy changes. Most importantly, the U.S. Sentencing Guidelines and their state counterparts were adopted with the goal of reducing such disparities. In 2005, when the Supreme Court's decision in *United States v. Booker* rendered the formerly mandatory Guidelines merely advisory, Justice Stevens's dissent predicted that "[t]he result is certain to be a return to the same type of sentencing disparities Congress sought to eliminate in 1984."[5] Whether this prediction was accurate is perhaps the foremost empirical question in sentencing policy today. The most prominent study to date, a 2010 report of the U.S. Sentencing Commission, gave an alarming answer: *Booker* and its judicial progeny had quadrupled the black-white sentencing gap among otherwise-similar cases,

1. Pew Ctr. on the States, *One in 100: Behind Bars in America 2008*, PEW CHARITABLE TRUSTS 3 (Feb. 2008), http://www.pewstates.org/uploadedFiles/PCS_Assets/2008/one%20in%20100.pdf.

2. Thomas P. Bonczar, *Prevalence of Imprisonment in the U.S. Population, 1974-2001*, BUREAU JUST. STAT., U.S. DEP'T JUST. 1 (Aug. 2003), http://www.bjs.gov/content/pub/pdf/piusp01.pdf.

3. Paul Guerino, Paige M. Harrison & William J. Sabol, *Prisoners in 2010*, BUREAU JUST. STAT., U.S. DEP'T JUST. 7, http://www.bjs.gov/content/pub/pdf/p10.pdf (last updated Feb. 9, 2012).

4. *See, e.g.*, MICHELLE ALEXANDER, THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS (2010); TODD R. CLEAR, IMPRISONING COMMUNITIES: HOW MASS INCARCERATION MAKES DISADVANTAGED NEIGHBORHOODS WORSE (2007); IMPRISONING AMERICA: THE SOCIAL EFFECTS OF MASS INCARCERATION (Mary Patillo et al. eds., 2004).

5. 543 U.S. 220, 300 (2005) (Stevens, J., dissenting in part).

from 5.5% to 23.3%.[6] In January 2013, the Commission issued an update with similar figures (revising the latter figure slightly downward, to 19.5%), this time combined with explicit calls for legislation in effect returning the Guidelines to something fairly close to their prior binding status.[7]

This Article introduces a new empirical approach and gives a very different answer. The Commission's methods are hobbled by two serious limitations that also pervade the broader empirical literature on sentencing disparity.[8] First, these studies consider the judge's final sentencing decision in isolation, ignoring crucial earlier stages of the justice process. Those earlier stages have important sentencing consequences, and yet these studies exclude the portions of the ultimate sentence gap that result from earlier-stage decision-making from their estimates. Second, studies of *changes* in disparity after legal changes (like *Booker*) have failed to disentangle the effects of the legal change from surrounding events and background trends.

This Article develops these two critiques and discusses our own research on racial disparities among federal arrestees, which uses a method that avoids these problems. We first highlight some findings from our recent study showing that while a black-white gap appears to be introduced during the criminal justice process, it appears to stem largely from prosecutors' charging choices, especially decisions to charge defendants with "mandatory minimum" offenses. These findings highlight the importance of taking into account the early parts of the justice process. With that in mind, we then present our new findings on *Booker*, estimating its effects not only on sentencing, but also on charging, plea-bargaining, and sentencing fact-finding, an analysis no prior studies have performed. Far from finding evidence that judges' use of expanded discretion worsens disparity, we fail to find an increase in disparity and find suggestive evidence cutting in the opposite direction.[9]

---

6. U.S. SENTENCING COMM'N, DEMOGRAPHIC DIFFERENCES IN FEDERAL SENTENCING PRACTICES: AN UPDATE OF THE *BOOKER* REPORT'S MULTIVARIATE REGRESSION ANALYSIS 3 (2010), http://www.albany.edu/scj/documents/USSC_Multivariate_Regression_Analysis_Report_001.pdf [hereinafter 2010 U.S. SENTENCING COMM'N].

7. U.S. SENTENCING COMM'N, REPORT ON THE CONTINUING IMPACT OF *UNITED STATES V. BOOKER* ON FEDERAL SENTENCING pt. A, at 8-9, 108 (2012), http://www.ussc.gov/Legislative_and_Public_Affairs/Congressional_Testimony_and_Reports/Booker_Reports/2012_Booker/index.cfm [hereinafter 2012 U.S. SENTENCING COMM'N]. The update extended the last period through 2011.

8. While we focus on race, these weaknesses are also common in research on other "unwarranted disparities" (for example, inter-district) and much of the non-disparity-related research on sentencing policy.

9. The *Booker* results are exclusively and fully presented in this Article. We begin, however, by discussing the results of a separate but related study of racial disparities in charging and

Our research seeks to close a surprisingly wide gap that separates two bodies of scholarship: the theoretical and qualitative literature on how the criminal justice system functions (which uniformly recognizes the critical role of prosecutors) and empirical research on sentencing disparities (which effectively ignores that role). The modern criminal justice process is prosecutor-dominated. Prosecutors have broad charging and plea-bargaining discretion, and their choices have a huge impact on sentences. A central claim made by critics of mandatory sentencing is that restricting judicial discretion further empowers prosecutors, who tend to exercise that power in ways that perpetuate or worsen disparity. This "hydraulic discretion" theory has been described as a near-consensus view of sentencing scholars.[10]

Yet the empirical research on sentencing disparity has not tested these claims and fails to account for the role of prosecutorial discretion. Researchers typically estimate sentencing disparities in federal and other courts subject to sentencing guidelines after controlling for (among other things) the recommended guidelines sentence. But the guidelines recommendation is itself the end product of charging, plea-bargaining, and sentencing fact-finding. Controlling for it filters disparities in those processes out of the sentencing-disparity estimates and gives an incomplete view of the scope and sources of sentencing disparity.[11] In effect, the existing literature focuses on disparities in

---

sentencing more generally. In this Article, we discuss that study's motivations, highlight key results, and explain its relevance to current law and policy debates, including the Sentencing Commission's report. The full results, including relevant tables and graphs and a full technical explanation of our methods, can be found in M. Marit Rehavi & Sonja B. Starr, *Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences* (Univ. of Mich. Program in Law & Econ., Working Paper No. 12-002, 2012) (under review), http://ssrn.com/abstract=1985377.

10. 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, 155-56 (1987) (noting that "this 'hydraulic' or 'zero-sum' effect is so firmly entrenched as a criticism of current reform efforts that most researchers begin with the assumption that the displacement of discretion exists"); Lauren O'Neill Shermer & Brian D. Johnson, *Criminal Prosecutions: Examining Prosecutorial Discretion and Charge Reductions in U.S. Federal District Courts*, 27 JUST. Q. 394, 395-96 (2010) (observing that "scholars agree that attempts to curtail judicial discretion are likely to concomitantly increase prosecutorial discretion" and that "prosecutorial discretion . . . may risk the perpetuation of the types of disparities sentencing reforms were intended to reduce").

11. A few studies (most notably the Sentencing Commission's *Booker* study) compound this problem further by also controlling for differences in Guidelines departure rates, filtering out one of the key aspects of the final sentencing decision as well. *See, e.g.*, 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. E, at 7 (including "whether the court determined that a sentence outside the applicable sentencing guideline range was warranted" among a list of control variables); 2010 U.S. SENTENCING COMM'N, *supra* note 6, at 18, app. B at B-1

compliance with the sentencing guidelines. While this is an important piece of the sentence-disparity picture, it is far from the only piece, because decisions made throughout the process ultimately affect the sentence. Moreover, sentencing-stage disparities might either offset or exacerbate disparities arising earlier, making it hard to interpret them in isolation.

We accordingly take a broader, process-wide approach, constructing a dataset that links records from four different federal agencies and allows us to trace criminal cases from arrest through sentencing. We focus on the gap between black men and white men in non-immigration cases. Instead of controlling for the Guidelines sentence, we control for the arrest offense and other characteristics that are fixed at the beginning of the justice process. The arrest offense is an imperfect proxy for underlying criminal behavior, but we believe it is the best proxy available for this purpose. Our method allows us to assess aggregate disparities introduced throughout the post-arrest justice process, from charging through sentencing. Further, it also allows us to analyze the contribution of each procedural stage (as well as underlying case differences) to the total black-white gap.

The problem with the prevailing method is not merely an academic concern. In Part II of this Article, we highlight and discuss key findings of our analyses of charging and sentencing in federal criminal cases from 2007 to 2009.[12] That research shows that after controlling for the arrest offense, criminal history, and other prior characteristics, there remains a black-white sentence-length gap of about 10%. But judges' choices do not appear to be principally responsible. Instead, between half and the entire gap can be explained by the prosecutor's initial charging decision—specifically, the decision to bring a charge carrying a "mandatory minimum." After controlling for pre-charge case characteristics, prosecutors in our sample were nearly *twice* as likely to bring such a charge against black defendants.[13] In other words, studies that focus only on the judicial sentencing decision exclude what appears to be the *most* important procedural source of disparity in sentences.

A proper analysis of *Booker*'s effects on disparity, then, should take the *whole* justice process into account, to the extent possible. In Part III, we present

---

(providing a similar list and explaining that the Commission used separate dummy variables for substantial assistance departures, other downward departures, and upward departures).

12. *See* Rehavi & Starr, *supra* note 9.

13. These results are for non-drug cases. When drug cases are added, the gap after controlling for prior characteristics rises to about 14%, and mandatory minimums similarly appear to explain nearly all of the disparity. The charging study's analysis of drug cases is somewhat more limited. *See infra* text accompanying notes 76-77. In contrast, this Article's study of *Booker* includes drug cases in all its analyses.

the results of such an analysis. We begin that inquiry with a simple linear time-trend analysis, which shows that, when one measures sentence disparity in the broader way that we recommend, unexplained black-white disparity did not grow between 2003 and 2009, the period in which the Sentencing Commission found that it quadrupled. Indeed, our estimate of the disparity trend is *negative*, although imprecise. That is, the gap in sentences for similar black and white arrestees was, if anything, slightly smaller by the end of 2009 than it was just before *Booker*. The Commission's claim that disparity grew over that same period is an artifact of its flawed way of measuring disparity.

Beyond the question of whether disparity has changed during the period surrounding *Booker*, we must further ask whether it has changed *because* of *Booker*. The two questions are not the same, but they are too often confused. In addition to the disparity-measurement question, a second serious flaw pervades the empirical literature on sentencing-law changes: the failure to provide a sound basis for causal inferences. This second problem is exemplified by the Sentencing Commission's analysis. The Commission found that disparities after *Booker* (averaged over a period of years) were larger than disparities before it. Even assuming that were true, it would still be a huge logical leap to conclude that *Booker caused* this increase—a classic confusion of correlation and causation. Many things change over time—for instance, the mix of cases, the composition of the bench and of U.S. Attorneys' and public defenders' offices, substantive criminal legislation and case law, and the Department of Justice's (DOJ's) enforcement priorities and internal policies—and any of these changes could have racially disparate impacts on sentences. The greater disparity in the post-*Booker* period, therefore, could easily have nothing to do with *Booker*. Indeed, even if *Booker* had *slowed* an underlying trend of increasing disparity, the Commission's methods would incorrectly imply that *Booker* led to greater disparity.

Accordingly, we employ a different approach that *can* disentangle the effect of *Booker* from underlying trends: a regression discontinuity-style estimator. Specifically, we assess whether, in the *immediate* aftermath of *Booker*, there is a sharp break in an otherwise continuous trend, which would provide a much stronger basis for inferring causality. Our method focuses on *Booker*'s immediate effects, not its long-term effects, which admittedly is both a strength and a weakness. The long-term effects are presumably what policymakers care most about, but there is no good way to identify *Booker*'s relationship to longer-term trends in disparity—the causal inference problem is too serious. The immediate effects can be more rigorously assessed. Fortunately, there is good reason to believe that if *Booker had* substantially changed racial disparity patterns in judicial decision-making, we would have seen at least part of the effect right away. *Booker*'s effects on Guidelines compliance were not slow or subtle—departure rates immediately and

dramatically spiked. That is, *Booker* was a sudden shock to the scope of judicial discretion, and, if judges were inclined to exercise their discretion in ways that widen the black-white gap, one would expect to see disparity jump in response to that shock, right after *Booker*.

We do not see such a jump. Right after *Booker*, sentencing disparity did not increase, and may have modestly dropped. If *Booker* did have any adverse effects on black defendants relative to white defendants, it was probably a second-order result of *charging* changes: the use of mandatory minimum charges increased for black defendants immediately after *Booker*, but this effect appears to have been quite short-term.

We are very cautious about these findings. Even with our approach, identifying *Booker*'s effects is hard. While *Booker* has been described as a "natural experiment,"[14] as an experiment it leaves much to be desired—it changed the legal regime for every non-petty federal offense at once, leaving no plausible control group. Our method does not require a control group and filters out longer-term trends effectively, but it could be tricked by month-to-month fluctuations. Moreover, *Booker* was not a clean break in settled law; it came on the heels of a period of serious lower-court confusion, further complicating causal inference. We conduct tests to evaluate these problems, but we cannot erase the noise in the data or the complexity of the history. Still, what we can say is that nothing in these data suggests that judges' use of their post-*Booker* discretion exacerbated racial disparity.

Understanding the relative role of prosecutors and judges in producing disparities is important. The specter of increased disparity after *Booker* has been prominently cited to support new constraints on judicial discretion. For instance, the Department of Justice in the George W. Bush Administration advocated mandatory topless guidelines—effectively, mandatory minimums but no maximums.[15] The Sentencing Commission has recently advanced a multi-pronged proposal to strengthen legislative and appellate court constraints on judicial sentencing discretion—a proposal that in effect would restore the Guidelines very nearly to the legal status they enjoyed before *Booker*.[16]

---

14. Paul J. Hofer, United States v. Booker *as a Natural Experiment: Using Empirical Research to Inform the Federal Sentencing Policy Debate*, 6 CRIMINOLOGY & PUB. POL'Y 433, 435 (2007).

15. *See* Alberto Gonzales, U.S. Att'y Gen., Prepared Remarks of Attorney General Alberto Gonzalez: Sentencing Guidelines Speech (June 21, 2005), http://www.justice.gov/archive/ag/speeches/2005/06212005victimsofcrime.htm.

16. The proposal includes requiring judges to give substantial weight to the Guidelines, restricting policy-based departures, and requiring appellate courts to give deference to Guidelines sentences. 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. A, at 111-15; Judge

Such "solutions" could be counterproductive. Constraints on judges generally empower prosecutors by making their choices more conclusive determinants of the sentence. Our research suggests that prosecutorial decisions are important sources of disparity—especially the decision to file mandatory minimum charges, which are prosecutors' most powerful tools for constraining judges. Note that we do not claim our findings prove "discrimination" by prosecutors or anyone else. We are limited to what our data can capture, and unobserved differences between cases could justify different charging decisions or sentencing outcomes. Still, we have rich controls, including detailed arrest offense information; criminal history; and other demographic, geographic, and socioeconomic fields, yet substantial unexplained racial differences remain.

In Part I, we briefly introduce the federal sentencing framework and review the legal scholarship on prosecutorial and judicial discretion. In Part II, we present our critique of the "sentencing only" approach used by the current empirical literature and discuss our preferred process-wide approach, its strengths and limitations, and some insights that can be gleaned from it. In Part III, we present our critique of the causal inference methods used by existing sentencing-reform research. We then pair our process-wide approach to estimating disparity with our regression discontinuity-style approach to causal inference in order to estimate *Booker*'s effects on racial disparity. We conclude with possible policy implications.

I. PROSECUTORS, SENTENCING, AND THE "HYDRAULIC DISCRETION" THEORY

Federal prosecutors, like their counterparts in the states, have always possessed very broad discretion. Prosecutors choose what charges to bring, and the complex criminal code often provides a wide range of choices. Over 95% of convictions result from guilty pleas, and prosecutors control the terms of the deals they offer defendants.[17] These can include the charges of conviction (charge bargaining), sentence recommendations and requests for departures

Patti B. Saris, Chairwoman, U.S. Sentencing Comm'n, Remarks at the Public Hearing Before the United States Sentencing Commission 8 (Feb. 16, 2012), http://www.ussc.gov/Legislative_and_Public_Affairs/Public_Hearings_and_Meetings/20120215-16/Hearing_Transcript_20120216.pdf.

17. *See* Gary Fields & John R. Emshwiller, *Federal Guilty Pleas Soar as Bargains Trump Trials*, WALL ST. J., Sept. 23, 2012, http://online.wsj.com/article/SB10000872396390443589304577637610097206808.html ("Guilty pleas last year resolved 97% of all federal cases that the Justice Department prosecuted to a conclusion.").

from the usual range, and stipulations about sentencing-relevant facts (fact bargaining).

Traditionally, prosecutors' discretion was matched by vast judicial discretion in choosing sentences, which was constrained only by broad statutory ranges—for instance, zero to twenty years. Statutory minimum sentences were not widespread before the 1980s, and still apply in only a minority of cases.[18] Within the statutory ranges, judges were free to tailor sentences to the facts and the offenders' circumstances. The disadvantage was that there was no good way to ensure that similar cases resulted in similar sentences.

In 1984, citing studies finding widespread racial, gender, inter-judge, and inter-district disparities in sentencing, Congress adopted the Sentencing Reform Act, which created a Sentencing Commission to devise binding Sentencing Guidelines.[19] Under the Guidelines, complex rules determine the offense level, which is based on the conviction offense plus additional aggravating or mitigating sentencing facts, such as drug quantity or the defendant's role in a group offense. The offense level is one of two axes of a sentencing grid; the other is the defendant's criminal history category. Within each grid cell is a narrow range: eight to fourteen months, for instance.[20] Prior to *Booker*, departures from this range were permitted only for specified reasons.

By greatly reducing judges' discretion, the Guidelines concentrated tremendous power in prosecutors' hands. As Kate Stith explains, "when judges had discretion to impose any sentence [in the statutory range], prosecutorial power was potentially limited or counterbalanced by the possibility of judicial

---

18. *See* U.S. SENTENCING COMM'N, REPORT TO CONGRESS: MANDATORY MINIMUM PENALTIES IN THE FEDERAL CRIMINAL JUSTICE SYSTEM 23 (2011), http://www.ussc.gov/Legislative_and_Public_Affairs/Congressional_Testimony_and_Reports/Mandatory_Minimum_Penalties/20111031_RtC_Mandatory_Minimum.cfm (reviewing the history of federal mandatory minimums and describing the 1980s "enactment of the mandatory minimums that are most commonly applied today"); *id.* at xxix (finding that only "14.5 percent of all federal offenders were subject to a mandatory minimum penalty at sentencing," although—because mandatory minimums generally lead to longer sentences—39.4% of federal prisoners had been given mandatory minimums).

19. *See* Sentencing Reform Act of 1984, Pub. L. No. 98-473, 98 Stat. 1987 (codified as amended in scattered sections of 18 and 28 U.S.C.); *id.* § 217(a) (codified as amended at 28 U.S.C. § 991) (creating the Sentencing Commission).

20. *See Sentencing Table*, U.S. SENT'G COMM'N (2011), http://www.ussc.gov/Guidelines/2011_Guidelines/Manual_PDF/Sentencing_Table.pdf; *see also* U.S. SENTENCING GUIDELINES MANUAL ch. 5, pt. A (2012) (providing the Sentencing Chart in the Guidelines Manual). *See generally* U.S. SENTENCING GUIDELINES MANUAL chs. 2 & 3 (2012) (laying out guidelines by offense and describing victim-related sentence adjustments).

discretion."[21] But under the Guidelines, plea-bargaining much more tightly constrained the sentence.[22] The one feature of the Guidelines that was intended to limit prosecutorial power was the judge's sentencing fact-finding authority. This system (called "real-offense" sentencing)[23] allows the judge to base a sentence even on uncharged conduct, so long as the sentence falls within the statutory range for the crime of conviction. In principle, this system should reduce prosecutors' ability to offer to understate the defendant's culpability in exchange for a guilty plea.

Still, studies suggest that real-offense sentencing has not constrained prosecutors very much, because in practice prosecutors very strongly influence judges' findings of fact. Plea agreements usually include factual stipulations, and, even though DOJ has long directed prosecutors not to bargain over these facts, many studies have documented the persistence of fact-bargaining.[24] Judges are not bound by the factual stipulations, and the power to diverge from them (relying on sentencing-stage evidence or a probation office report) is an important aspect of judicial discretion. Judges typically lack the incentive, however, and may lack the information, to diverge from what the parties have agreed upon.[25] One 1996 survey found that only 8% of judges said they "go behind" plea agreements "somewhat or very frequently"; 25% said they never do, while the rest said they did so "infrequently."[26] As Nancy King put it,

21. Kate Stith, *The Arc of the Pendulum: Judges, Prosecutors, and the Exercise of Discretion*, 117 YALE L.J. 1420, 1430 (2008).

22. *Id.*

23. U.S. SENTENCING GUIDELINES MANUAL § 1A1.4 (2012); Stith, *supra* note 21, at 1434-36.

24. Mary Patrice Brown & Stevan E. Bunnell, *Negotiating Justice: Prosecutorial Perspectives on Federal Plea Bargaining in the District of Columbia*, 43 AM. CRIM. L. REV. 1063, 1070 (2006); Nancy J. King, *Judicial Oversight of Negotiated Sentences in a World of Bargained Punishment*, 58 STAN. L. REV. 293, 295-98 (2005); Ilene H. Nagel & Stephen J. Schulhofer, *A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices Under the Federal Sentencing Guidelines*, 66 S. CAL. L. REV. 501, 522 (1992); Stith, *supra* note 21, at 1450.

25. *See* Brown & Bunnell, *supra* note 24, at 1068-69; Stephen J. Schulhofer & Ilene H. Nagel, *Plea Negotiations Under the Federal Sentencing Guidelines: Guideline Circumvention and Its Dynamics in the Post-*Mistretta *Period*, 91 NW. U. L. REV. 1284, 1300-01 (1997); Stith, *supra* note 21, at 1449; *cf.* William J. Powell & Michael T. Cimino, *Prosecutorial Discretion Under the Federal Sentencing Guidelines: Is the Fox Guarding the Hen House?*, 97 W. VA. L. REV. 373, 383-84 (1995) (noting "the frustration of federal judges" with the shift in power to prosecutors resulting from the Guidelines).

26. Molly Treadway Johnson & Scott A. Gilbert, *The U.S. Sentencing Guidelines: Results of the Federal Judicial Center's 1996 Survey*, FED. JUD. CENTER 10 (1997), https://bulk.resource.org/courts.gov/fjc/gssurvey.pdf.

"Establishing facts in an adversarial system without the assistance of adversaries is an awkward business."[27]

To the Guidelines' many critics, this empowerment of prosecutors was a serious flaw, leading to harsh results for defendants generally and undermining the Sentencing Reform Act's disparity-reduction goals. As Albert Alschuler argued, "[T]he price of whatever success the Guidelines have achieved in reducing judge-created sentencing disparities has been the burgeoning of prosecutor-created disparities."[28] Scholars often refer to discretion in the criminal justice system as being "hydraulic," such that attempts to constrain it in one place will merely shift it to another. Stephanos Bibas, for example, wrote, "The criminal justice system operates like a toothpaste tube, and departures that are squeezed out of the judge's end of the tube will wind up in the prosecutor's domain. This hydraulic pressure means that departures will still exist, but they will now occur more often on prosecutors' terms."[29] This theory has long pervaded scholarship about the Guidelines. As Terance Miethe wrote in 1987, "[T]his 'hydraulic' or 'zero-sum' effect is so firmly entrenched as a criticism of current reform efforts that most researchers begin with the assumption that the displacement of discretion exists . . . ."[30]

Note that, although scholars' language often refers to shifts in "discretion," this is a slight misnomer; the Guidelines did not really increase prosecutors' *discretion*, which was already almost boundless. Rather, they increased their *power*: the choices prosecutors made more conclusively determined the sentence.[31] In a 1996 survey, approximately 75% of district judges and chief probation officers said that prosecutors were now the actors with the *most* influence on final sentences—more than judges themselves.[32] Prosecutors thereby obtained greater leverage in plea-bargaining—they could nearly promise that defendants would get more lenient sentences if they pled guilty and harsher ones if they refused. In 2004, Marc Miller wrote, "The overwhelming and dominant fact of the federal sentencing system . . . is the

---

27. King, *supra* note 24, at 303.

28. Albert W. Alschuler, *Disparity: The Normative and Empirical Failure of the Federal Guidelines*, 58 STAN. L. REV. 85, 117 (2005).

29. Stephanos Bibas, *The Feeney Amendment and the Continuing Rise of Prosecutorial Power to Plea Bargain*, 94 J. CRIM. L. & CRIMINOLOGY 295, 300 (2004); *accord* Stith, *supra* note 21, at 1427-36 (describing the Guidelines' shift of power from judges to prosecutors).

30. Miethe, *supra* note 10, at 155-56.

31. *See* Rodney L. Engen, *Assessing Determinate and Presumptive Sentencing—Making Research Relevant*, 8 CRIMINOLOGY & PUB. POL'Y 323, 328-29 (2009).

32. Johnson & Gilbert, *supra* note 26, at 6-7.

virtually absolute power the system has given prosecutors . . . . There is a lot of evidence to support this claim, but it can be demonstrated with one simple and awesome fact: Everyone pleads guilty."[33] After the implementation of the Guidelines in the early 1990s, plea rates rose from 87% of all federal convictions to 97% by 2004.[34]

Since then, however, federal sentencing law has undergone another major change. In January 2005, the Supreme Court decided *United States v. Booker*, which rendered the formerly mandatory Guidelines merely advisory.[35] The Court held that a mandatory sentencing scheme in which a defendant's maximum sentence could be increased based on judicial fact-finding violated the Sixth Amendment right to a jury trial.[36] The Court could have remedied that defect by requiring more jury fact-finding, but it chose an alternate remedy: maintaining real-offense sentencing, but severing the provision of the Sentencing Reform Act that rendered the Guidelines mandatory.[37] The Court's remedial choice remains reversible by Congress,[38] which has so far not taken action to reverse *Booker*. District courts today may depart from the Guidelines so long as the ultimate sentence is not "unreasonable."[39] In December 2007, in *Gall v. United States* and *Kimbrough v. United States*, the Supreme Court further clarified that courts of appeals should not deem sentences unreasonable merely because they fall outside the Guidelines,[40] and that sentencing judges may depart from the Guidelines on the basis of policy disagreements.[41]

---

33. Marc L. Miller, *Domination & Dissatisfaction: Prosecutors as Sentencers*, 56 STAN. L. REV. 1211, 1252 (2004).

34. Alschuler, *supra* note 28, at 112; U.S. SENTENCING COMM'N, FIFTEEN YEARS OF GUIDELINE SENTENCING: AN ASSESSMENT OF HOW WELL THE FEDERAL CRIMINAL JUSTICE SYSTEM IS ACHIEVING THE GOALS OF SENTENCING REFORM iv (2004), http://www.ussc.gov/Research_and_Statistics/Research_Projects/Miscellaneous/15_Year_Study/15_year_study_full.pdf (explaining that, although the Guidelines were promulgated in 1987, they were not fully implemented until after *Mistretta v. United States*, 488 U.S. 361 (1989)); *see also* Stith, *supra* note 21, at 1425 (arguing that the Guidelines "provided prosecutors with indecent power relative to both defendants and judges, in large part because of prosecutors' ability to threaten full application of the severe Sentencing Guidelines").

35. 543 U.S. 220, 246 (2005).

36. *Id.* at 232.

37. *Id.* at 247-48.

38. *Id.* at 265 (noting that "[t]he ball now lies in Congress' court").

39. *Id.* at 261.

40. Gall v. United States, 552 U.S. 38, 51 (2007).

41. Kimbrough v. United States, 552 U.S. 85, 108-09 (2007); *see also* Rita v. United States, 551 U.S. 338, 354-55 (2007) (barring appellate courts from treating outside-Guidelines sentences as presumptively unreasonable).

Booker was widely seen as an earthquake in federal sentencing law. Still, rendering the Guidelines advisory is not the same as eliminating them. Federal judges are still required to calculate the Guidelines sentencing range, and, although they are then free to depart from it, they usually do not.[42] There are many possible reasons for this continued conformity: federal judges might believe that the Guidelines meet the goal of reducing disparity,[43] wish to avoid open-ended, subjective sentencing assessments, seek insulation from criticism or reversal, or simply treat the Guidelines as an "anchor."[44]

To the extent that judges continue to follow the Guidelines, the power the Guidelines conferred on prosecutors will presumably remain largely intact. In addition, even if judges felt totally unconstrained by the Guidelines, prosecutors would retain at least two powerful sources of sentencing influence. First, their charging and charge-bargaining choices shape the *statutory* minimum and maximum sentences, which remain mandatory. Second, because they negotiate the factual stipulations accompanying pleas and may introduce evidence at sentencing hearings, prosecutors have enormous influence over the information that gets to judges, and what judges know presumably will influence sentencing regardless of whether they follow the Guidelines. Thus, even in the post-*Booker* era, prosecutors should be expected to play a crucial role in the processes that shape sentencing.

In short, then, legal scholars and justice system participants widely agree both that prosecutorial choices are key drivers of sentences and that sentencing law reforms involve tradeoffs between judicial and prosecutorial power. One might expect that this broad consensus would shape empirical research on

---

42. Most federal sentences are still in the Guidelines range. *See, e.g.*, 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. A, at 3, 5 (finding that in the period between December 11, 2007 and September 30, 2011, "80.7 percent of federal sentences were either within the guideline range (53.9% of sentences) or below the range due to a government motion (26.8% of sentences)").

43. *See* U.S. SENTENCING COMM'N, SURVEY OF ARTICLE III JUDGES ON THE FEDERAL SENTENCING GUIDELINES ch. 2 (2003), http://www.ussc.gov/Research_and_Statistics/Research_Projects/Surveys/200303_Judge_Survey/jschap2.pdf (showing that 52.8% of surveyed judges, when asked to rate the Guidelines' performance in reducing disparities on a 1 to 6 scale, picked either 5 or 6).

44. The "anchoring" literature shows that when people have to translate subjective judgments onto a numeric scale, they are often highly influenced by hearing some number mentioned—even numbers that (unlike the Guidelines) are actually irrelevant to the question posed. *See* J. J. Prescott & Sonja Starr, *Improving Criminal Jury Decision Making After the* Blakely *Revolution*, 2006 U. ILL. L. REV. 301, 326 (discussing "anchoring problems" in mock jury studies of punitive damages awards and criminal jury verdicts).

sentencing disparities and sentencing reforms, but, as we demonstrate below, it has not.

II. ESTIMATING RACIAL DISPARITY IN SENTENCING: A PROCESS-WIDE APPROACH

For decades, unwarranted disparities in sentencing have been a major focus of empirical research. Overwhelmingly, these studies focus exclusively on judges' final sentencing decisions, ignoring the rest of the justice process. In Section II.A, we review those studies and explain why this problem is so serious. In Section II.B, we describe the dataset that we constructed to enable a broader approach, and in Section II.C, we highlight certain key findings of our recent study of racial disparity in charging and sentencing. In Section II.D, we discuss some limitations of this broader approach. Note that this Part does not focus directly on *Booker*'s effects or on changes over time. Rather, we begin by explaining why it is crucial for estimates of sentencing disparity to encompass the pre-sentencing stages of the process: a great deal of the ultimate sentence gap between similar black and white arrestees appears to emerge from decisions made at earlier stages. That insight provides one of the primary motivations for our approach in our analysis of *Booker*, presented in Part III.

A. Studies Estimating the Extent of Unwarranted Sentencing Disparities

Sentencing disparity studies generally begin by pointing to a gap in observed sentence outcomes and asking what generated it. For instance, black male defendants receive much longer sentences on average than white males do—a major contributor to their higher incarceration rates. But does the sentence gap arise because black defendants have committed more serious crimes or have more extensive criminal histories? Or are they treated differently in the criminal justice process?

Mass incarceration of black males has serious social consequences regardless of its causes. But if different offending patterns are to blame, the problem might be better addressed with policies focused on addressing the causes of crime, such as poverty. In contrast, if the criminal justice system is treating like cases differently, then policymakers should focus on fixing that problem. Researchers thus seek to isolate the component of the sentence gap arising in the criminal justice process by controlling for some measure of the underlying severity of the case. But what measure? The answer to that question is the key difference between our approach and those of prior sentencing studies.

When researchers focus on the federal courts or other guidelines-based systems, the typical approach is to control for the "presumptive" or

recommended guidelines sentence—generally, the bottom end of the guidelines range.[45] There are variations on this approach,[46] but all of them estimate differences in the actual sentence relative to what the sentence "should have been" under the guidelines. Most studies also include controls for the statutory mandatory minimum.[47] Studies in systems without guidelines similarly control for conviction severity.[48]

The problem with these approaches is that the key control variables are only distant proxies for the seriousness of the underlying conduct. They are the end product of the discretionary processes described above: charging, plea-bargaining, and sentencing fact-finding. And those processes might also produce disparities. The use of these control variables filters out the share of the ultimate sentencing disparity that comes from those earlier processes. The resulting measure of disparity is thus based on an artificially narrow focus on the final sentencing decision in isolation from all the other processes that produce the sentence. These estimates can be useful in understanding disparities in guidelines compliance, which is one important part of the criminal process. However, we believe that, for most purposes, policymakers likely have a broader interest in the full sentence disparity that an individual faces, regardless of where it originally arose in the justice process. If so, it is important for them to understand that the existing literature is estimating something much narrower.

The specification of an empirical model of disparity may seem like a purely scientific decision. But as Albert Alschuler has observed, it is bound up with

---

45. *See, e.g.*, 2010 U.S. SENTENCING COMM'N, *supra* note 6, app. A, at A-4 to A-5 tbl.A (comparing imposed and presumptive sentences); Jeffrey T. Ulmer et al., *Racial Disparity in the Wake of the* Booker/Fanfan *Decision: An Alternative Analysis to the USSC's 2010 Report*, 10 CRIMINOLOGY & PUB. POL'Y 1077, 1088-90 (2011) (following this approach and also reviewing prior literature doing the same).

46. *See, e.g.*, Brian Iannacchione & Jeremy D. Ball, *The Effect of* Blakely v. Washington *on Upward Departures in a Sentencing Guidelines State*, 24 J. CONTEMP. CRIM. JUST. 419, 420-21 & tbl.1 (2008) (treating upward departures as the outcome variable); David B. Mustard, *Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts*, 44 J.L. & ECON. 285, 297 (2001) (including separate dummy variables for each Guidelines grid cell); Jeffrey S. Nowacki, *Race, Ethnicity, and Judicial Discretion: The Influence of the* United States v. Booker *Decision*, 20 CRIME & DELINQ. 1, 12-13 (2013) (using offense level and criminal history controls); Max Schanzenbach, *Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics*, 34 J. LEGAL STUD. 57, 63 & tbl.1 (2005) (same).

47. *E.g.*, Nowacki, *supra* note 46, at 12-13.

48. *E.g.*, Darrell Steffensmeier et al., *Gender and Imprisonment Decisions*, 31 CRIMINOLOGY 411, 420 (1993) (using a ten-point severity scale).

normative questions: what kinds of disparities do we think are important?[49] The choice of control variables determines what kinds of disparities one is measuring, and so it should be shaped by a sense of the types of disparities policymakers and stakeholders care about. There are many reasons one might worry about demographic disparities in the justice process. For instance, such disparities might violate the Equal Protection Clause, exacerbate the social consequences of mass incarceration within particular communities, interfere with retributive or utilitarian punishment objectives, or undermine the justice system's credibility.

We do not intend in this Article to resolve what policymakers' objectives should be. But *none* of the reasons we can think of for caring about demographic disparities suggest that policymakers should confine their interest to equalizing sentences for cases in the same Guidelines cell. Rather, all imply that the key question is whether people who have committed the *same underlying criminal conduct* (arguably including prior criminal history) receive the same sentence. Between the underlying criminal conduct and the sentence, there are many points in the process where disparities could be introduced. Policymakers should care about all of them.

Other scholars have noted this problem with the prevailing approach.[50] This includes, to their credit, many of those who employ the approach themselves, who note that their accounts of disparities are incomplete.[51] But these caveats generally are not mentioned when the work gets cited, and their importance may well be overlooked by policymakers. This is a serious mistake. The problem is not *just* that these accounts of disparity are insufficiently comprehensive—they are also potentially misleading, at least if one misinterprets them as a measure of whether judges are treating defendants with similar conduct equally. Absent an account of disparity at the earlier stages of the process, it is difficult to interpret disparities found in the final stage.

---

49. Alschuler, *supra* note 28, at 85-88.

50. *See id.* at 86-87; Engen, *supra* note 31, at 324-29; Shermer & Johnson, *supra* note 10, at 395-96.

51. *See, e.g.*, Ulmer et al., *supra* note 45, at 1107-08. The Sentencing Commission itself notes in the 2012 update to its analysis that prosecutorial practices could contribute to sentencing disparities, but it does not connect this insight to its race results, nor does it alter its methods of estimating racial disparity to correct any of the problems identified here with its earlier report. *See* 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. A, at 7-8. In this Article, we focus principally on the 2010 report, which explained its methods of assessing racial disparity in much more detail.

For instance, consider the Sentencing Commission's prominent recent sentencing-disparity report. The report finds that from December 2007 to September 2011, black males received 19.5% longer sentences than white males, controlling among other things for the recommended Guidelines sentence.[52] But how should this result be interpreted? Consider just three of many possibilities concerning what might have happened *earlier* in the justice process:

A. Prosecutors charged white defendants more harshly and/or offered them worse plea deals, such that the resulting Guidelines recommendation averaged 19.5% higher for white defendants than for black defendants with similar offenses and criminal histories.

B. Prosecutors charged white defendants more harshly and/or offered them worse plea deals, such that the resulting Guidelines recommendation averaged 30% higher for white defendants than for black defendants with similar offenses and criminal histories.

C. Prosecutors charged *black defendants* more harshly and/or offered them worse plea deals, such that the resulting Guidelines recommendation averaged 30% higher for black defendants than for white defendants with similar offenses and criminal histories.

Under Scenario A, what looked like a 19.5% sentencing disparity now looks like judges sentencing more or less "correctly," relative to underlying criminal conduct—they are correcting the disparity introduced by prosecutors. Under Scenario B, it actually seems that judges are not favoring white defendants *enough*—to sentence based on true culpability, they would have to do more to compensate for prosecutors favoring black defendants. In contrast, under Scenario C, judges are compounding the underlying charging and plea-bargaining disparities; the "true" sentencing disparity is actually much more than 19.5%. If you don't know which of these scenarios (or others) is true, it is risky to use the 19.5% figure as a guide to policy.

Moreover, even if one were willing to assume that judges were the only relevant source of racial disparity in sentencing, the prevailing method would nonetheless be too limited, because it still filters out part of the judicial sentencing process. Controlling for the presumptive sentence means one is filtering out any disparities in judicial *fact-finding*. And in the Sentencing Commission studies specifically, the problem is even worse. In addition to the presumptive sentence and mandatory minimum, the Commission *also* controls

52. 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. E, at 9.

for whether the judge departed upward or downward from the Guidelines range. In doing so, the Commission is not just considering the final sentencing decision in isolation—it is filtering out a key part of that sentencing decision itself. In effect, the Commission is estimating race gaps in the *size* of departures (and in sentence choices within the narrow Guidelines range), but filtering out *whether* there is a departure and, if so, in what direction. This is, to say the least, a strange choice, and one that could easily produce misleading results. This same problem also appears in the most prominent recent study responding to the Sentencing Commission report, that of Ulmer, Light, and Kramer; the authors critique other aspects of the Commission's methods, but their main analysis of sentencing disparities also controls for departure status as well as the presumptive sentence.[53]

Another recent study by Joshua Fischman and Max Schanzenbach recognizes the problem with the presumptive sentence approach (and also does not control for departure status).[54] Fischman and Schanzenbach instead control for the Guidelines "base offense level." This is an improvement over the presumptive sentence approach; it provides a fuller measure of *judicial* sentencing disparity, and is probably the best approach possible using only the sentencing-stage data from the Sentencing Commission. But it still means that the authors' sentence disparity estimates do not incorporate components introduced by the various *prosecutorial* decisions and negotiations, plus judicial fact-finding, that determine the base offense level.[55] The base offense level is affected not only by charging and charge-bargaining, but also by a large part of the fact-finding required by the Guidelines. It incorporates, for instance, drug quantity in a drug trafficking case,[56] or, in an assault case, the degree of physical contact and injury, the defendant's intent, and the use of weapons.[57]

---

53. *See* Ulmer et al., *supra* note 45. This study does consider downward departure rates separately as an outcome variable, and finds substantial racial disparity in those rates (favoring white defendants). *Id.* That disparity in departure rates is filtered out of the sentence length disparity estimates, however, which presumably biases them downward. Similarly, the Sentencing Commission's recent update to its analysis also assesses non-government-sponsored departure rates as an outcome and likewise finds a disparity favoring white defendants, but continues to filter that disparity out of the sentence length estimates. 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. E, 20-22.

54. Joshua B. Fischman & Max M. Schanzenbach, *Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, 9 J. EMPIRICAL LEGAL STUD. 729 (2012).

55. *Id.* at 754 tbl.5 (listing controls).

56. U.S. SENTENCING GUIDELINES MANUAL § 2D1.1 (2012).

57. *Id.* § 2A2.2, 2A2.4.

Sentence disparities arising from any of those factual determinations, or in the prior charging or plea-bargaining processes, would be filtered out by the use of the base offense level control. To fully avoid the limitations of the presumptive sentence approach, one needs a measure of case severity that precedes all of these discretionary processes.[58]

The problem with the presumptive sentence control is compounded by a distinct source of potential bias that the existing literature has overwhelmingly failed to acknowledge: sample selection shaping the pool of sentenced cases. Nearly every study of sentencing disparity is confined to a sample consisting of sentenced defendants only—in federal court studies, typically only those sentenced for felonies or Class A misdemeanors ("non-petty offenses"), which the Sentencing Commission collects data on. To make it into the sample, defendants must get through the criminal justice "funnel": they must be arrested, charged, and convicted of a non-petty offense.

If these earlier processes are subject to demographic disparities, it could introduce sample selection bias into the estimates of sentencing-stage disparity. Suppose that all else equal, black defendants are more likely to be convicted of a non-petty offense, such that it takes a less serious case to get a black defendant sentenced. If so, we would expect black defendants and white defendants who get sentenced to be unobservably different: black defendants' cases would be less serious in a way that controlling for observable variables cannot capture. Sentencing disparity estimates within that sample would be biased because they cannot account for this unobserved difference. Again,

58. In one portion of their analysis—the assessment of the effects of the Supreme Court's decisions in *Gall* and *Kimbrough* on sentencing disparities—Fischman and Schanzenbach also present alternative results without the base offense level control. Fischman & Schanzenbach, *supra* note 54, at 756 tbl.6. But without the base offense level control, there are *no* controls for the severity of the underlying cases. That approach is plausibly valid for the purposes of their assessment of these Supreme Court decisions' effects if one makes the strong assumption that the underlying case mix did not change in racially disparate ways during their three-and-a-half-year study period. But the approach cannot be used for our purpose here, which is to disentangle the share of the observed racial disparity that appears to be explained by differences in underlying criminal conduct from the share that appears to be introduced in the criminal process. It also does not identify the overall magnitude of the disparity or the proportion of it that is attributable to judges, the key figure for recent policy debates. For that purpose, it is not sufficient simply to drop the problematic controls—one must replace them with a better measure of the severity of the underlying case. As we explain below, we believe the arrest offense is the best available measure. Note that Fischman and Schanzenbach explicitly explain that they do not seek to assess the amount of disparity that is unwarranted by legitimate case differences; they focus only on change over time. *Id.* at 738-39. Some further differences with their approach are discussed in Part III.

without assessing the "funnel," one cannot know whether to expect such a bias to exist and, if it does, which direction it will cut.

Unfortunately, the empirical research on demographic disparities earlier in the justice process is relatively limited. It focuses almost entirely on certain measures of charge-bargaining, such as the rate of dropping charges; studies typically do not assess *severity* reductions.[59] More importantly, few studies (and no federal studies) have assessed disparities in *initial* charging, even though it is difficult to interpret charge-bargaining results without doing so.[60] A few state-level studies have found racial disparities in the use of certain particularly harsh mandatory minimums, including one study of "habitual offender" charges in Florida,[61] another in Pennsylvania,[62] and a Maryland study of add-on mandatory minimums for firearms.[63]

At the federal level, many observers, including the U.S. Sentencing Commission, have pointed to racial gaps in the rate of mandatory minimum convictions.[64] Fischman and Schanzenbach's study provides useful new evidence that mandatory minimums may be an important contributor to sentencing disparities.[65] But these studies raise important further questions. Because they do not control for underlying pre-charge case features affecting a defendant's eligibility for mandatory minimums (such as the arrest offense), they do not examine the *reasons* for the mandatory minimum gap. They do not tell us whether black defendants have simply committed more crimes to which

---

59. *See* Shermer & Johnson, *supra* note 10, at 398-401, 414-21 (reviewing the charge-bargaining literature, which mostly finds disparity favoring white defendants, and presenting their own findings showing no such disparity).

60. One early study by Spohn et al. found disparities favoring white defendants in the rate of filing felony charges in Los Angeles County, but did not analyze charge severity within felony charges. Cassia Spohn et al., *The Impact of the Ethnicity and Gender of Defendants on the Decision to Reject or Dismiss Felony Charges*, 25 CRIMINOLOGY 175 (1987); *see also* sources cited *supra* note 10 and accompanying text (discussing another charging study).

61. Charles Crawford et al., *Race, Racial Threat, and Sentencing of Habitual Offenders*, 36 CRIMINOLOGY 481 (1998).

62. Jeffrey T. Ulmer et al., *Prosecutorial Discretion and the Imposition of Mandatory Minimum Sentences*, 44 J. RES. CRIME & DELINQ. 427 (2007).

63. Jill Farrell, *Mandatory Minimum Firearm Penalties: A Source of Sentencing Disparity?*, 5 JUST. RES. & POL'Y 95 (2003).

64. 2010 U.S. SENTENCING COMM'N, *supra* note 6; *see also Disparate Impact of Federal Mandatory Minimums on Minority Communities in the United States*, FAMS. AGAINST MANDATORY MINIMUMS & NAT'L COUNCIL OF LA RAZA (Mar. 10, 2006), http://www.nclr.org/images/uploads/publications/38367_file_IAHRC_statement_FNLNWQC__2__fnl.pdf.

65. *See* Fischman & Schanzenbach, *supra* note 54.

mandatory minimums apply, or whether there are racial disparities in prosecutors' exercise of charging or charge-bargaining discretion.[66]

A final disadvantage to the "presumptive sentence" approach is simpler: it controls only for differences in crime severity according to the Guidelines, not for differences in crime *type*. Judges might be more likely to depart from the Guidelines for some crimes than others, for reasons that have nothing to do with race. Such tendencies might well have racially disparate impacts, but they are not necessarily "unwarranted"—the nature of the offense is certainly a relevant sentencing consideration. Sentencing studies often do include controls for case type in addition to the presumptive sentence, but only for broad categories such as drugs or violent crime, which do not capture much nuance.[67]

More precise crime-type controls, which we provide, can enable us to better distinguish the disparate impact component of racial disparity (the component that can be explained by non-racial factors like case type) from the component that we cannot explain with the variables we can measure, which could represent disparate treatment on the basis of race. The distinction between disparate impact and disparate treatment is crucial as a matter of constitutional law,[68] although the extent to which it is normatively important is open to debate.[69] We think *all* factors contributing to racial disparity in sentencing—

---

66. One study of Illinois courts evaluates whether judges differ from one another in their racial disparity patterns, finding that they do. *See* David Abrams et al., *Do Judges Vary in Their Treatment of Race?*, 41 J. LEGAL STUD. 347 (2012). That study does not need to control for "presumptive sentence" because it can take advantage of the random assignment of cases to judges. The result interestingly shows that judicial discretion matters to racial disparity patterns. However, it does not answer the more basic question of whether judges are actually treating similar defendants differently based on race, as opposed to varying in their treatment of case features correlated with race. Similarly, studies that evaluate the interaction between the race and gender of judges or prosecutors and those of defendants also provide interesting insights, but cannot squarely address whether or how race or gender affects outcomes. *See* Amy Farrell et al., *Intersections of Gender and Race in Federal Sentencing: Examining Court Contexts and the Effects of Representative Court Authorities*, 14 J. GENDER RACE & JUST. 85 (2010); Schanzenbach, *supra* note 46.

67. *See, e.g.*, 2010 U.S. SENTENCING COMM'N, *supra* note 6, at B1-B2 (using seven categories); Ulmer et al., *supra* note 45, at 1090.

68. Facially neutral government policies and practices will not be deemed unconstitutional unless those challenging them can establish a discriminatory purpose. *See* Washington v. Davis, 426 U.S. 229 (1976).

69. Many critics have argued that the Supreme Court's focus on discriminatory purpose is overly formalistic and have instead advocated a focus on the harms imposed on subordinated groups. *See* Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. MIAMI L. REV. 9, 9-10 (2004) (reviewing this literature). We are sympathetic to this view, but this longstanding debate

whether legally warranted or not—are important for policymakers to understand, a point we return to below. But we believe that disentangling the reasons can help policymakers figure out what to do about them. In any event, studies like the Sentencing Commission's purport to estimate legally unwarranted disparities, and thus they should filter out legally relevant factors like case type.

B. Our Dataset

Our broader approach to the estimation of racial disparities requires something most researchers have not had: a dataset that traces federal cases from arrest through sentencing. We constructed it by linking files from four federal agencies: the U.S. Marshals Service (USMS) (data from arrest and/or booking), the Executive Office for U.S. Attorneys (EOUSA) (prosecutors' investigation and case files), the Administrative Office of the U.S. Courts (AOUSC) (court records), and the U.S. Sentencing Commission (USSC) (sentencing-related data collected from judges).[70] It covers two stages of the process that the Sentencing Commission data alone (the sole source for most federal studies) do not include.

First, our dataset includes the arrest offense, coded with 430 codes, and a text field describing the offense based on the arresting officer's notes. This information allows us to substitute the arrest offense, instead of the presumptive sentence, as the key case-severity control. This substitution means that we are estimating sentencing gaps between black and white defendants who look similar near the *beginning* of the justice process, rather than between those whose cases have come to look similar near the end of it. We can thus estimate the aggregate sentencing disparity introduced by decisions throughout the post-arrest justice process. In addition, the arrest offense codes provide far more detail on crime *type* than sentencing studies typically control for. The arrest offense is not a perfect proxy for underlying criminal activity, to be sure. We discuss its limitations below.[71]

---

need not be resolved for our purposes; empirical differentiation of the reasons for disparities has practical uses regardless.

70. These data are collected by the Bureau of Justice Statistics and made available to researchers under security conditions by the National Archive of Criminal Justice Data. We provide much greater detail on the construction and coding of the dataset in Rehavi & Starr, *supra* note 9, Data App. For more information on the data restrictions and on accessing the data, see *Restricted Data*, NAT'L ARCHIVE OF CRIM. JUST. DATA, http://www.icpsr.umich.edu/icpsrweb/NACJD/private (last visited Sept. 1, 2013).

71. *See infra* Subsections II.D.1, II.D.3.

Second, our dataset includes rich information on *initial* charges, in addition to final charges. Specifically, we know the statutory sections under which the defendant was charged and convicted—for instance, 18 U.S.C. § 924(c).[72] To assess charges quantitatively, we translated each combination of statutory sections into a numeric measure of total charge severity. This is not a simple task, which may be an additional reason prosecutorial decision-making is under-researched. Based on comprehensive research on every federal crime charged during the study period, we developed four different charge severity measures. The first three were grounded in sentencing law: the statutory maximum and minimum and a Guidelines-based measure.[73] The fourth measure was based on sentencing practice: the mean sentence given in a baseline period before the study period. We then calculated the *combined* severity of all charges on all these measures, following the rule laid out in the Guidelines for sentencing in multi-charge cases: we assumed sentences on each charge would run concurrently, unless one of the statutes specified consecutive sentencing.[74]

Sometimes, the statutory provisions in the data contained multiple sentencing schemes depending on the facts of the case; even more often, the Guidelines sentence would vary according to the facts. Where possible, we resolved such ambiguities based on the *other* charges in the case; often, the presence of a second charge would make it evident that the prosecutor was alleging a particular fact that would affect the sentence on the first charge.[75] In

---

72. This section contains the sentencing enhancement for firearms.

73. The statutory maximum and minimum were simply looked up in the listed statutes, although there were sometimes ambiguities to be resolved, as discussed below. The Guidelines measure is more complicated, because a Guidelines sentence normally is not determined based on the statutory charge alone—it depends on additional fact-finding. The measure we used is the Guidelines sentence that would apply if all of the statutory elements of *all* charged offenses were proven, but *no* other aggravating or mitigating facts were proven at sentencing. It is thus intentionally limited to serve as a "charge only" measure, allowing the effects of subsequent Guidelines sentencing fact-finding to be separated out from those of initial charging.

74. Thus, in a two-charge case, unless one of the two statutes specified consecutive sentencing, all of the combined charge severity measures would be determined by the *higher* of the two individual-charge severities that we calculated. If one of the statutes required consecutive sentencing, then the severities would be summed. Note, however, that the presence of the less severe charge still might be important, because it might affect the severity that we calculated for the more severe charge. *See infra* note 75 and accompanying text.

75. For instance, if Charge 1 contained a heightened penalty (under either the statute or the Guidelines) if a gun was involved and Charge 2 was a gun charge, we applied the heightened penalty for Charge 1. Implementing this analysis required extensive legal research. In addition to coding all the triggering conditions for statutory or Guidelines

other cases, we used reasonable, research-driven assumptions about which subparagraphs were likely to apply to most cases brought under that statute.[76] However, in drug cases, the ambiguities were too extreme to resolve with these methods—most cases were charged under omnibus provisions (such as 21 U.S.C. § 841(b)) encompassing all drug types and quantities. We could not meaningfully code the severity of such provisions, and thus cannot assess initial charging disparities in drug cases. It is still possible, however, to analyze drug cases focusing on disparities in the *final* mandatory minimum recorded at sentencing, a separate data field. Child pornography cases must also be excluded from initial-charging analyses because of a similar ambiguity, but they can likewise be included in analyses of the final mandatory minimum.[77] We also excluded immigration cases for different reasons: their stakes typically turn on deportation, making prison sentence length analysis a very incomplete picture of case outcomes, and they involve different "fast-track" procedural environments, which present different policy considerations and also raise concerns about the quality of data.[78]

We focused on the race gap between black and white U.S. citizen males. In a separate study focused on gender disparity, discussed below, Starr also

---

sentencing enhancements for all federal crimes, we also had to code every crime's elements to identify possible factors that could raise the penalty for any other charge in the case. More details on the way we carried out this coding can be found in Rehavi & Starr, *supra* note 9, Data App.

76. For instance, 18 U.S.C. § 1347, the federal health care fraud statute, has a statutory maximum of ten years, but this is increased to twenty years if the fraud results in serious bodily injury and to life if it results in death. A search of case law involving this statute confirmed our intuition that these enhancements are used in only a small fraction of cases, so our default assumption, absent the presence of other charges with elements including serious bodily injury of death, was that the statutory maximum for this crime was ten years. For more discussion of our assumptions, see Rehavi & Starr, *supra* note 9, Data App.

77. As discussed below, in our analysis of overall sentencing disparities as well as the final mandatory minimum, the results were fairly similar when we added drug and child pornography cases to the sample (though the unexplained racial disparities are somewhat larger in drug cases).

78. The Sentencing Commission's *Booker* study includes immigration, but we agree with other scholars who have argued that it should be considered separately. *See, e.g.*, Ulmer et al., *supra* note 45, at 1085-86. Note that the exclusion of immigration cases probably should not be expected to have dramatic effects on black-white disparities because there are few black defendants in immigration cases. For instance, in the Sentencing Commission's 2009 dataset, less than 2% of sentenced immigration defendants were black. In any event, we leave the analysis of immigration cases for another day, and note that other disparities—such as national origin disparities—might be of particular policy interest in the immigration context.

assessed the race gap among women.[79] Outcomes for other racial groups were not analyzed because their numbers were very small. Hispanic defendants are included among the black and white defendants.[80]

C. *Our Research on Racial Disparities in Charging and Sentencing: Some Key Findings*

Our research on the disparities introduced throughout the post-arrest justice process, and their procedural sources, gives us strong reason to believe that the concerns expressed above about sentencing-stage-only estimates are problematic in practice as well as theory. We intend in future research to assess the specific contribution of *every* major stage of the justice process, but we began by focusing on *initial charging* and its role in explaining sentencing disparities. This stage has been almost entirely ignored by existing research, and it is especially important. In most federal cases, the initial charge *is* the final charge; charge-bargaining is the exception, not the rule.[81] In this period, dropping charges once filed required a supervisor's special approval.[82] In *initial* charging, however, the line prosecutor had, and has, considerable discretion.[83] In addition, before one can even begin to make sense of plea-bargaining disparities, one has to first know whether the baseline charges already reflect disparities.

---

79. Sonja B. Starr, *Estimating Gender Disparities in Federal Criminal Cases* 16 (Univ. of Mich. Law Sch. Law & Econ. Research Paper Series, Paper No. 12-018, 2012), http://ssrn.com/abstract=2144002.

80. The USMS data do not identify Hispanic ethnicity. Among sentenced defendants, the Sentencing Commission's data show that almost all persons of Hispanic ethnicity identify as white. If outcomes for Hispanic defendants fall somewhere between those of black and non-Hispanic white defendants (as 2010 U.S. SENTENCING COMM'N, *supra* note 6, at 23 suggests for the time period most closely corresponding to that of our sample), then our disparity estimates will be somewhat smaller than one would find if one looked only at black versus non-Hispanic white defendants.

81. Overall charge severity was reduced in only 10-15% of the cases in our sample (varying depending on the severity measure), and in 85% of cases there was no change to the charge labeled the "lead charge" (the most serious charge) in the AOUSC data.

82. Memorandum from John Ashcroft, U.S. Att'y Gen., to All Fed. Prosecutors, Department Policy Concerning Charging Criminal Offenses, Disposition of Charges, and Sentencing (Sept. 22, 2003), http://www.justice.gov/opa/pr/2003/September/03_ag_516.htm.

83. DOJ also attempted to constrain charging discretion, *id.*, but this is a weaker constraint in practice, as we explain *infra* note 203 and accompanying text.

The statistical analysis and the resulting estimates are described in detail in the study.[84] Here, we highlight some key findings and focus on their implications for legal policy and for assessing the impact of *Booker*. We had three main research questions:

1. Do prosecutors charge otherwise-similar black and white arrestees differently?
2. Do otherwise-similar black and white arrestees ultimately receive different sentences?
3. How much of the sentencing disparity can be explained by the charging disparity?

By "otherwise similar," we mean similar in terms of the pre-charge case and defendant characteristics that we can observe. In the charging analysis (Question 1), we controlled for arrest offense; district; age; whether there were multiple defendants in the case; and county-level poverty, unemployment, income, and crime statistics. In the sentencing analysis (Questions 2 and 3), we added additional controls based on data recorded only for sentenced defendants: criminal history category and education level. Other variables were available only for subsets of the sample, but we checked to make sure that within those subsets, the results did not change when they were taken into account. These included defense counsel type, marital status, and Hispanic ethnicity, as well as dummy variables for whether certain facts were recorded in the written arrest offense description: possession of guns, other weapons, or drugs; conspiracy; racketeering; child victims; and official victims. For all three questions, we used a sample limited to male U.S. citizens.[85]

On Question (1), we did find significant racial disparities in charge severity across all four charging measures. The racial gaps were fairly moderate (less than 10%), but significant.[86] But the disparities in mandatory minimums were much more dramatic. After controlling for the variables above, we found black

---

84. *See* Rehavi & Starr, *supra* note 9, at 13-15, 24. The full analysis includes, for example, an exploration of the marginal effects of race at different points in the charging and sentencing distributions—that is, whether the racial disparities are larger for more or less severe cases—as well as a variety of alternative specifications and estimation strategies. The results discussed here are from the cited working paper version. Note that the figures may differ in the final version due to adjustments to the specification and sample; however, the results should not be substantively different.

85. In a separate study of gender disparity, Starr also explores racial disparities among female defendants. *See* Starr, *supra* note 79.

86. Rehavi & Starr, *supra* note 9, at 12.

men were still nearly *twice* as likely to be charged with an offense carrying a mandatory minimum sentence.[87]

Question (2) focuses on the *aggregate* sentencing disparity introduced by the entire post-arrest justice process. Among those convicted there were significant unexplained sentencing disparities favoring white defendants. Most of the large raw sentencing gap (which was around 50%) could be explained by the observed case and defendant characteristics—that is, the gap declined substantially when we added the controls to the model. We then used decomposition methods to identify *which* controls were the most important in explaining the raw sentencing gap. The factors that could explain by far the largest components of the black-white gap were arrest offense and criminal history. But even after controlling for these and other variables, a gap of about 10% remained unexplained in the main sample, which excluded drug and child pornography cases.[88] The gap was a bit larger in the sample that included drug and child pornography cases (such that the sample consisted of all non-immigration case types). Thus, like other studies, our analysis found significant unexplained racial disparities in sentences.

However, our analysis of Question (3) showed that these gaps do *not* appear to be solely (or even principally) driven by the final sentencing decision. Rather, initial charging—especially the decision to bring mandatory minimum charges—is an important driver of these sentencing disparities. Half of the 10% otherwise-unexplained sentence gap in the main sample disappeared when we controlled for mandatory minimum charges.[89] Furthermore, that estimate almost certainly understates the impact of mandatory minimum charges because of the very conservative coding method we used—when our charge information was ambiguous, we assumed there was no mandatory minimum, which means we missed a substantial number of them.[90] When we instead controlled for the *final* mandatory minimum sentence (which is unaffected by

---

87. *Id.* at 11-12.

88. *Id.* at 3.

89. *Id.*

90. *See supra* notes 75-78 and accompanying text (discussing charging ambiguities). Mandatory minimums are fairly frequently triggered by special factual allegations—such as injury, death, or prior commission of the same crime—that the charge data do not identify. A comparison of our coding of the *final* mandatory minimum to the actual final mandatory minimum recorded by the judge suggests that our initial charge coding is probably missing nearly 40% of the mandatory minimums. Our coding of ambiguous cases was deliberately conservative—that is, we chose to err on the side of understating mandatory minimums' role.

the coding ambiguities, because it is recorded by the sentencing judge), *all* the otherwise-unexplained racial disparity in the average sentence disappeared.[91]

We performed this latter analysis for drug cases and child pornography cases as well; this was possible because it did not require using the ambiguous initial charge data. In a sample consisting of all non-immigration case types, including drug and child pornography cases, no significant disparity remained after controlling for the final mandatory minimum.[92] In short, the results when one includes drug and child pornography cases are consistent with the results when one excludes them: a substantial black-white gap that is unexplained by the control variables, but which appears to be driven largely by differences in the use of mandatory minimums.[93]

We subjected all of these findings to a battery of robustness checks to assess whether varying the control variables, the sample definition, or the estimation method changed the results. Similar disparity patterns appeared in all specifications and subsamples. Mandatory minimum charging disparities were similar across offense types, but the non-drug mandatory minimum that was the most common and the most responsible for driving sentencing disparities was the enhancement for crimes involving firearms, found in 18 U.S.C. § 924(c). This statute has particularly harsh penalties: at least five years, running consecutively to other charges. There are higher minimums if the firearm is brandished or discharged and astonishing minimums (at least thirty years) if there is more than one § 924(c) count, which could simply mean that the defendant was found with two guns.[94] Prosecutors have considerable discretion in applying this statute, especially when the facts make the relationship of a gun to an offense ambiguous (for instance, when the gun is found in the defendant's car trunk), and a lenient prosecutor may "swallow the

---

91. Rehavi & Starr, *supra* note 9, at 19-20. This difference reflects the coding issue, not new disparities introduced by charge-bargaining. Our analyses (using our coding for both the initial and final charges) do not show racial disparities in the rate at which mandatory minimums are dropped during plea-bargaining.

92. *Id.* at 20-21.

93. In its recent report, the Sentencing Commission criticizes our working paper for excluding drug trafficking cases but makes no mention of the additional analyses showing that the results are similar in those cases. *See* 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. E, at 13 n.25. Note, in any event, that the *Booker* analysis, *infra* Part III, which is a more direct counterpoint to the Commission's *Booker* report, uses a broader sample that includes these cases.

94. *See* 18 U.S.C. § 924(c)(1)(A)(ii) (2012) (seven-year minimum if the firearm is brandished); *id.* § 924(c)(1)(C)(i) (twenty-five-year minimum for a second or subsequent conviction). The total minimum sentence exceeds thirty years because the sentences run consecutively.

gun" entirely.[95] Michelle Alexander, in her recent book about race and incarceration, quotes a former U.S. Attorney describing one such incident:

> I had an [assistant U.S. attorney who] wanted to drop the gun charge against the defendant [in a case in which] there were no extenuating circumstances. I asked, "Why do you want to drop the gun offense?" And he said, "He's a rural guy and grew up on a farm. The gun he had with him was a rifle. He's a good ol' boy, and all good ol' boys have rifles, and it's not like he was a gun-toting drug dealer." But he was a gun-toting drug dealer, exactly.[96]

Our results suggest that this incident may not have been an anomaly.

D. Interpretations and Limitations

Our research thus suggests that the post-arrest justice process—especially mandatory minimum charging—introduces sizable racial disparities. But are these gaps *really* the result of racially disparate treatment? Or do they stem from unobserved differences that might be appropriate bases for different treatment? As Judge Nancy Gertner has warned, the quest to eliminate improper disparities should not lead us to seek "false uniformity" among cases that are actually dissimilar despite superficial similarities.[97]

No observational study can fully tease out the causes of demographic disparities because no dataset can ever capture all the subtle ways in which cases can differ.[98] So one must tread cautiously when discussing causation—we speak in terms of "unexplained disparity," rather than claiming to have proven "discrimination." Still, our data are rich enough to shed light on some plausible

---

95. *E.g.*, Erik Luna, *Testimony Before the U.S. Sentencing Commission: Mandatory Minimum Sentencing Provisions Under Law*, U.S. SENT'G COMM'N 3 (May 27, 2010), http://www.ussc.gov/Legislative_and_Public_Affairs/Public_Hearings_and_Meetings/20100527/Testimony_Luna.pdf.

96. ALEXANDER, *supra* note 4, at 116 (alterations in original).

97. *See* United States v. Cabrera, 567 F. Supp. 2d 271, 273 (D. Mass. 2008).

98. In other settings involving potential unobserved variables, economists have developed a variety of useful quasi-experimental approaches, but these are of little help here. Such methods can help to analyze *differences* in disparities (for instance, before and after policies go into effect or among different decision-makers), and we use one such approach below to assess *Booker*'s effects. *See infra* Part III. But they are not of much use in determining whether an apparent racial disparity is "real." Race is inextricable from the rest of the person—there are no clever econometric tricks for isolating the effect of race from the effects of unobserved characteristics that might be correlated with race.

causal theories, as we will briefly discuss in this Section. In addition, we point to some ways in which our disparity estimates may be under-inclusive—they do not encompass every discretionary choice shaping the black-white gap. Finally, we discuss the way these racial disparities appear to interact with gender disparities to produce particularly bad outcomes for black males.

1. *Possible Unobserved Offense Differences*

A first potential concern with the arrest offense control is unobserved differences in the underlying criminal activity. This concern is less severe than it might have been: the detailed USMS offense codes, together with the written offense description field, capture considerable nuance in offense facts. In particular, they seem to effectively capture whether a gun was involved with the offense, which is important because of the substantial contribution of 18 U.S.C. § 924(c) charges to racial disparities.[99] The multi-defendant case variable also captures an important offense characteristic, because multi-defendant cases often involve more serious crimes and often trigger conspiracy charges.

In drug cases, in addition to the limitations to the charge data, the arrest codes also contain an important ambiguity: they do not specify drug quantity, and other sources of initial alleged quantity are only reliable before 2004.[100] But estimates on the most recent years with reliable quantity data (2001-03) were not substantially affected by the addition of quantity controls.[101] There were also racial disparities favoring whites in the drug quantities found at sentencing fact-finding, after controlling for the seizure quantity and drug type recorded

99. Use of guns is usually clear from the arrest codes, and our description flags also included guns, drugs, and the combination thereof. Some cases might have been missed, but we seriously doubt that the number is large enough to explain the large racial disparity in 924(c) charges.

100. The EOUSA suspect investigation files record the drug quantity seized at arrest, but patterns in the quantity distribution over time suggest a serious problem with this data field beginning in 2004, when EOUSA adopted a new data entry system. Our analysis leads us to suspect the problem relates to the addition of a decimal point to the field—perhaps some (but not all) prosecutors did not notice the change. Comparisons to the Sentencing Commission's quantity data do, however, make it apparent that the problem is with the new system, not the old one. It would be a service to future researchers and the public if EOUSA investigated this problem.

101. Rehavi & Starr, *supra* note 9, at 21 n.40.

at arrest.[102] This suggests that white defendants may be negotiating more favorable plea stipulations on quantity.

Similarly, the arrest data do not record the dollar value of losses in economic crimes. In some cases, the arrest codes suggest the scale of the crime (for instance, pickpocketing or vehicle theft), but in others (such as wire fraud) they do not. It is unlikely, however, that differences in loss quantity could explain the racial disparities—in fact, they probably cut in the opposite direction. At least as recorded at sentencing fact-finding, white defendants tend to be involved in significantly *higher*-value property crime cases, after controlling for the other covariates.

Another important factor not captured by the arrest data is the defendant's relative role in group offenses. We do not know of any anecdotal reason to believe that such differences could explain the racial disparities, that is, that white defendants tend to be minor players in conspiracies while black defendants tend to be leaders. If this *were* the basis for the ultimate gaps, one would expect to see a noticeable difference in role adjustments at the sentencing fact-finding stage. But black defendants get only very slightly worse role adjustments on average: a difference of 0.04 offense levels on the forty-three-level Guidelines scale, after controlling for the observed variables.[103] This difference is statistically significant, but it is very small, and suggests that role differences are unlikely to explain much of the black-white sentencing gap.

2. *Possible Differences in Offender Characteristics*

Beyond the offense characteristics, there might be relevant *offender* characteristics that contribute to the race gap. We control for criminal history, the main offender characteristic built into sentencing law.[104] The most obvious other possibility is socioeconomic differences, which are highly correlated with race. While poverty would not be a "warranted" reason for worse case outcomes, it would be a non-racial one and might suggest different policy approaches. However, the unexplained disparities we identify exist even after controlling for a variety of socioeconomic indicators such as education, county-

---

102. Quantities were converted into implied offense levels according to the Guidelines tables to allow comparisons across drug types.

103. The range of possible role adjustments is from -4 to +4. U.S. SENTENCING GUIDELINES MANUAL § 3B1.1 to 3B1.2 (2012).

104. Criminal history was not included in our main charging analysis because it is only recorded for the subset of charged defendants who were eventually sentenced. But within that subset, the charging disparities persisted after controlling for criminal history.

level variables, and defense counsel type (an excellent proxy for poverty because public defenders or other publicly funded counsel are appointed only if the defendant is poor). Perhaps more remarkably, our socioeconomic factors taken together do not contribute significantly to the "explained" share of the racial disparity.[105] This appears to be because poverty itself (as reflected by these indicia) is not an important predictor of higher sentences.[106] Notably, representation by a public defender is associated with slightly *lower* sentences, all else equal.

This absence of socioeconomic disparity is good news, and it cuts against conventional wisdom.[107] Can it really be that poor defendants do *not* fare worse? It is possible that the conventional wisdom might not apply to the federal courts, where indigent defendants generally receive high-quality representation, especially from federal public defenders.[108] We suspect that we would not have gotten the same result had we studied states in which indigent representation is under-resourced and in disarray.[109] We note that this point may have policy implications: the federal example offers a potential model for those states. When a justice system devotes sufficient resources to indigent defense to attract strong lawyers, train them well, and keep caseloads reasonable, poverty need not drive outcomes, and the race gap will likely be smaller than it might otherwise be.[110]

---

105. The decomposition estimators we used allowed us to estimate the combined explanatory value of a group of variables in explaining the black-white gap, and the socioeconomic status indicators did not explain a significant share of that gap.

106. Some of these variables have significant effects on some outcome variables, but these effects are small and inconsistent in sign. There is no overall pattern suggesting that poverty worsens outcomes.

107. The hurdles facing indigent defendants have long been the subject of extensive scholarship and policy attention. *See, e.g.*, *The Access to Justice Initiative*, U.S. DEP'T JUST., http://www.justice.gov/atj/ (last visited Apr. 9, 2013) (noting the "access-to-justice crisis").

108. Richard A. Posner & Albert H. Yoon, *What Judges Think of the Quality of Legal Representation*, 63 STAN. L. REV. 317, 341-42 (2011) ("Federal appellate and district judges in our sample express high regard for prosecutors and public defenders but low regard for court-appointed counsel and retained counsel . . . .").

109. *See, e.g.*, Eve Brensike Primus, *Structural Reform in Criminal Defense: Relocating Ineffective Assistance of Counsel Claims*, 92 CORNELL L. REV. 679, 686-87 (2007) (describing "rampant structural ineffectiveness" resulting from defenders being "incredibly overworked and severely underfunded").

110. *See* Steven B. Bright, *Legal Representation for the Poor: Can Society Afford This Much Injustice?*, 75 MO. L. REV. 683, 685 & n.11 (2010) (noting the superior resources of federal public defenders).

3. Possible Sources of Disparity that Our Estimates Leave Out

Although it is possible that our estimates of "unexplained" racial disparities include components that in fact have legitimate but unobserved explanations, in another sense these estimates are arguably *under*-inclusive. Our process-wide approach estimates disparities across a much broader swath of the criminal justice process than existing studies do, but even our method does not encompass *all* of the key decision points. In addition to prosecutors and judges, other decision-makers shape criminal case outcomes—most notably, law enforcement agents and policymakers.

Any disparities produced by those actors' choices will be found in the "explained" portions of the race gap—that is, the portions attributed to the control variables. It is important not to overlook those portions when thinking about what should be done about racial disparity, however. Rather than simply using regression methods to filter them out, as most studies do, we therefore used decomposition methods that allow us to estimate the relative contribution of each control variable to the total observed black-white gap. These methods showed that the variables with by far the most explanatory value are arrest offense and criminal history. These variables may capture important differences that we *want* sentencing law to reflect, but they also reflect discretionary choices.

First, the recorded arrest offenses will be affected by law enforcement choices.[111] This is a key limitation of our strategy of controlling for the arrest offense. We stated earlier that policymakers should ideally ask whether those who *committed* the same crime end up with the same sentence, but this is a very hard question to answer empirically. Researchers cannot observe what the defendants actually did. The arrest offense is a much better proxy for actual conduct than the presumptive Guidelines sentence, but it is not a perfect one. If it diverges from actual conduct in a racially disparate way, our "unexplained" disparity estimates will not capture that divergence. Nor do our estimates capture sample selection introduced by police decisions that determine who lands in the federal criminal justice system at all.[112]

---

111. If the prosecutor's pre-arrest involvement in the case influenced the arrest offense, this omission may leave out an aspect of *prosecutorial* discretion as well. When we drop cases with pre-arrest indictments (the cases with the most extensive pre-arrest involvement), disparity estimates increase.

112. Black arrestees comprise 45% of our sample, a rate far exceeding the overall share of black people in the general population; the question is the extent to which this overrepresentation reflects actual crime rates or policing patterns. This gap is included neither in the

In theory, these limitations could bias our results in either direction, but we think they probably mean we are understating the total disparities in the justice system. For arrest-stage disparities to explain our results instead, even partially, one would have to believe that federal law enforcement *favors* black suspects. We think this is unlikely. Many criminal justice scholars have argued that black males are disproportionately targeted by law enforcement, while virtually nobody claims the opposite.[113] Black people are arrested for drug crimes at a much higher rate than white people are, even though they self-report both drug use and drug dealing at equivalent or lower rates.[114] Beyond comparing arrest rates to reported crime rates, policing disparities are hard to study empirically because the underlying criminal behavior usually cannot be observed by researchers. But the existing quantitative evidence either supports the conventional wisdom or at least does not cut in the opposite direction.[115] To be sure, federal law enforcement could be different, but we are likewise unaware of any anecdotal suggestions that federal agents favor black suspects.

In addition, both the arrest offense and the criminal history components of the "explained" disparity reflect subjective policy choices: important sources of disparity may simply be *built into the law*.[116] In the Fair Sentencing Act of 2010, Congress responded to such a concern by partially mitigating the sentencing

---

"explained" part nor the "unexplained" part of our disparity estimates; we can only decompose disparities within the set of cases we have data on.

113. *See, e.g.*, ALEXANDER, *supra* note 4. In addition, surveys consistently find white Americans believe the police are fair while black Americans do not. *See* Jon Hurwitz & Mark Peffley, *Explaining the Great Racial Divide: Perceptions of Fairness in the U.S. Criminal Justice System*, 67 J. POL. 762 (2005). These perceptions may reflect real differences in experience.

114. *See* ALEXANDER, *supra* note 4, at 97-98 (reviewing these studies); *see also* William J. Stuntz, *Race, Class, and Drugs*, 98 COLUM. L. REV. 1795 (1998) (observing that drug enforcement targets open-air markets, which are dominated by black men).

115. *See, e.g.*, Andrew Gelman et al., *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*, 102 J. AM. STAT. ASS'N. 813 (2007) (finding evidence that black people in New York City are stopped and frisked at disproportionately high rates, and reviewing the policing-disparity literature). Much of the existing research focuses on traffic stops and reaches mixed results. Black drivers and male drivers are more likely to be stopped and searched, but some researchers have found a lack of disparity in the "hit rate" of stops and searches (for example, the rate of finding drugs), which they argue makes the policing pattern "rational." *See, e.g.*, Nicola Persico & Petra Todd, *Generalising the Hit Rates Test for Racial Bias in Law Enforcement, with an Application to Vehicle Searches in Wichita*, 116 ECON. J. F351, F364 (2006). Others find lower hit rates for black and Hispanic suspects than for white suspects, which suggests discrimination. *See, e.g.*, Sarath Sanga, *Reconsidering Racial Bias in Motor Vehicle Searches: Theory and Evidence*, 117 J. POL. ECON. 1155, 1157 (2009). This literature does not, in any event, suggest irrational *favoritism* towards black people.

116. *See* Alschuler, *supra* note 28, at 87-88.

framework's notoriously harsh treatment of crack cocaine cases.[117] But the crack laws are not the only example of particularly heavy punishments being given to crimes disproportionately involving black defendants. The harsh gun enhancements under 18 U.S.C. § 924(c) are another example—because black men are more frequently arrested with guns, as shown by our data, these enhancements would disparately impact black men even if they were neutrally applied. Similarly, our data show that black males are also more frequently arrested for violent crimes, and sentencing law is often harsher on these crimes than on nonviolent crimes that might reasonably be considered more serious.[118] These sentencing-law features are built into the arrest offense component of the measured disparities.

The criminal history component likewise reflects a subjective policy judgment to assign heavy weight to past crimes, even though those crimes have already been separately punished. While there are many competing considerations surrounding that judgment, it has a racially disparate impact. Moreover, this choice magnifies whatever racially disparate *treatment* exists in the criminal justice system by carrying its impact from one case to the next: the criminal history score may be influenced by disparate treatment in past cases. That past disparity will appear as part of the "explained" disparity, so it is easy to lose sight of it—it will be filtered away by controlling for criminal history.[119] Underlying unwarranted disparity can thus come to appear legally warranted.

4. Race, Gender, and Their Interaction

Finally, another limitation is that we only include men. Starr's related study examines gender disparities and race-gender interactions.[120] She finds unexplained gender disparities that dwarf the racial disparities our joint study found: men receive sentences that are over 60% longer than women's, even

---

117. Pub. L. No. 111-220, 124 Stat. 2372 (codified in scattered sections of 21 U.S.C. and 28 U.S.C.).

118. For instance, suppose *X*, who is unarmed, obtains $20 from *Y* by threatening to hit *Y*, and runs off with it. With no aggravating factors, his Guidelines offense level for robbery would be 20, U.S. SENTENCING GUIDELINES MANUAL § 2B3.1 (2012)—the same offense level that would have applied had he defrauded *Y* out of nearly $1 million, *id.* at § 2B1.1.

119. Through its career offender and armed career criminal provisions, federal sentencing law is particularly harsh on cases that combine violent or (especially) gun cases with extensive criminal history—another structural feature with particularly harsh effects on black men. *See* 18 U.S.C. § 924(c), (e) (2012); U.S. SENTENCING GUIDELINES MANUAL §§ 4B1.1, 4B1.4 (2012).

120. Starr, *supra* note 79, at 1, 16.

after controlling for the arrest offense, criminal history, and other pre-charge observable characteristics.[121] These gaps are much larger than most other studies have estimated because—as with race—they appear to mostly arise prior to the final sentencing decision.[122] The data suggest that differences in offender characteristics not captured by the main control variables may explain substantial shares of this gap, particularly differences in childcare responsibilities and perceived role in group offenses.[123] But Starr finds large unexplained disparities (over 50%) even among non-parents and in one-defendant cases, so these explanations do not appear to come close to explaining the whole gender gap, nor do any of the other theories Starr is able to test.[124]

Notably, the gender gap was substantially larger (about 75%) among black defendants.[125] The racial disparities we found for men do not recur among women; there is no significant unexplained black-white gap in sentences for female defendants. The black female/white female gap appears to be explained entirely by differences in arrest offense and criminal history—although, again, it is possible that these factors build in structural, arrest-stage, or other hidden sources of disparity.

As noted above, black males are incarcerated at extremely high rates in the United States, and, in assessing this problem, policymakers should consider both the race and gender dimensions and their interactions. Black male defendants appear to face not only the harsher side of both the racial and gender disparities, but also an additional interaction effect—an extra apparent penalty for being *both* black and male. Gender disparity need not be seen as being about special treatment of women—rather, one could ask why the criminal justice system appears to treat males so much more harshly. If it did not, Starr's data suggest that many fewer black men would be in prison.

---

121. *Id.* at 17.

122. In the gender context, an especially large share of the disparity appears to arise in sentencing fact-finding. *Id.* at 2. Mandatory minimums are also important, but only in drug cases, *id.* at 11, presumably because women are rarely arrested for the kinds of non-drug crimes to which mandatory minimums apply. Indeed, Starr's analysis excluded gun offenses as well as sex and child pornography offenses entirely because there were so few female defendants in those categories. *Id.* app., at ii.

123. *Id.* at 12-17.

124. *Id.* at 14.

125. *Id.* at 16.

III. THE *BOOKER* QUESTION: DOES EXPANDING JUDICIAL DISCRETION INCREASE RACIAL DISPARITY?

The discussion above illustrates the serious limitations of an empirical approach that focuses on the sentencing decision in isolation. In this Part, we apply that insight to the question that so worried Justice Stevens in his *Booker* dissent: has freeing judges to sentence outside the Guidelines led to an increase in unwarranted disparities? The Sentencing Commission has given the most prominent answer to this question so far, and its answer is a resounding yes. Its race findings have garnered understandable attention, because they are shocking: *Booker* and its progeny appear to have led to a nearly fourfold increase in racial disparity in sentencing, from 5.5% to 19.5%.[126] This was an explosive finding, and it has led to calls (spearheaded by the Commission itself) to reinstate stronger constraints on judicial discretion.[127] However, we show here that the Commission's conclusions are unfounded. Properly analyzed, there is no evidence that unexplained racial disparity in sentences has increased since *Booker*—much less *because* of *Booker*.

There are two core problems with the Commission's analysis of *Booker*—problems that also pervade the rest of the empirical literature examining the disparity consequences of sentencing law reforms. The first is that the studies estimate disparity in a very limited way—the problem discussed in Part II. In Section III.A, we explain why the "presumptive sentence" approach is a particularly poor choice for analyzing *Booker*'s effects, and we present a simple linear trend analysis showing that when disparity is estimated using our broader method, it has *not* increased in the years since *Booker* (and may have declined). In Section III.B, we discuss an additional serious problem with the existing studies: poor causal inference strategies. Even if it were true that disparity had increased after *Booker*, that is, these studies provide no reason to believe *Booker* was the cause. In Section III.C, we introduce a method that *can* be used to assess causation—a regression discontinuity-style approach. In

126. *See* 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. A, at 108.

127. The Commission laid out a detailed plan for restoring force to the Guidelines in its most recent report, 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. A, at 111-15, and has held hearings on the subject. *See supra* note 16 and accompanying text. The former Chair of the Sentencing Commission, Judge William K. Sessions, has cited the Commission's disparity report in warning that post-*Booker* disparities are likely to lead Congress to adopt more mandatory minimums; Sessions himself proposes a simplified mandatory guidelines system instead. William K. Sessions III, *At the Crossroads of the Three Branches: The U.S. Sentencing Commission's Attempts to Achieve Sentencing Reform in the Midst of Inter-Branch Power Struggles*, 26 J.L. & POL. 305, 309-10, 337-56 (2011).

Section III.D, we present the results of this analysis of *Booker*'s effects on sentencing as well as charging and plea-bargaining. Finally, in Section III.E, we discuss the limitations on our analysis and explain why researchers may never be able to give an entirely definitive answer to the question of *Booker*'s effects.[128]

A. The Changing Yardstick Problem

A subset of the sentencing disparity literature focuses on measuring *changes* in disparity resulting from changes to sentencing law, such as *Booker*. Like other sentencing disparity analyses, these studies typically control for the presumptive Guidelines sentence as well as the statutory mandatory minimum. The problem with this approach is largely explained above, but it impacts sentencing-reform studies in a slightly different way. In principle, studies focusing on *changes* in disparities have an advantage over those that estimate the *extent* of "unwarranted" disparity: the ability to ignore the possibility of stable differences between groups that the observed variables do not capture.[129] Suppose the control variables amount to only a "broken yardstick" for measuring the defendant's underlying criminal behavior—for instance, suppose the presumptive sentence variable diverges from true case severity in racially disparate ways. In a policy-change study, so long as the *same* broken yardstick is used before and after the policy change, one can validly estimate the policy's relative effects on different groups. This advantage is a mixed blessing: estimates of changes in disparity are less policy-relevant if we do not know whether the disparity in either the pre- or the post-period is "real." Still, not every study needs to answer every question, and research that brackets the "is this real?" question can be useful.

However, a serious problem arises if one cannot be confident that the yardstick itself has not been affected by the policy change. Consider again the 2012 Sentencing Commission report discussed above. It found that the black-white gap rose from 5.5% before *Booker* to 15.2% after, and finally to 19.5% after *Booker*'s successor cases *Kimbrough* and *Gall*.[130] Other studies have likewise found at least some increase in disparity after *Booker* or after *Kimbrough* and

128. The *Booker* results are presented only in this Article, rather than being further developed elsewhere, so we provide more technical detail in this Part, as well as tables and figures.

129. *See* Fischman & Schanzenbach, *supra* note 54, at 738-39.

130. 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. A, at 108.

Gall (although not as large).[131] Below, we discuss potential confounding factors that make it very problematic to infer that these changes were *caused* by either *Booker* or *Kimbrough/Gall*. But let's start with a more basic question: do these numbers actually tell us that racial disparity in sentences has grown?

In each period, the Sentencing Commission estimates sentencing disparities conditional on the presumptive sentence (likely a "broken yardstick" for the reasons discussed above), and then compares the disparities across time periods. If one were certain that racial disparities in the processes determining the presumptive sentence remained constant pre- and post-*Booker*, then this would be a "same broken yardstick" comparison. Whatever biases were hidden in the presumptive sentence variable would affect the estimates for both time periods similarly, so the comparison would be apples-to-apples.

But the problem is that *Booker* may have replaced one broken yardstick with a different one by affecting charging, plea-bargaining, or sentencing fact-finding in racially disparate ways. In other words, cases with the same presumptive sentences may represent different actual conduct pre- and post-*Booker* in ways that vary by race. Sample selection bias is also a potential problem: *Booker* may have changed which cases are winnowed out by the "funnel" of the criminal process, such that the samples of sentenced cases before and after *Booker* are not fairly comparable.

There is good reason to worry about these potential biases. One clear lesson from the legal scholarship reviewed in Part I is that the stages in the criminal justice process are interrelated. Charging, plea-bargaining, and fact-finding all occur in anticipation of and in an attempt to influence the sentencing consequences. It is not even remotely safe to assume that changes in sentencing law do not affect decision-making at those earlier stages. After all, consider what happened after the Guidelines were adopted: a drastic increase in guilty pleas, which legal scholars have (very plausibly) attributed to prosecutors' sharp increase in leverage.[132]

There are many theoretically plausible ways decision-making prior to sentencing could have changed after *Booker*. For example:

- Prosecutors might have to offer more favorable plea deals to induce guilty pleas, potentially resulting in more favorable findings of fact,

---

131. *See* Fischman & Schanzenbach, *supra* note 54; Nowacki, *supra* note 46; Ulmer et al., *supra* note 45.

132. *See supra* notes 21-34 and accompanying text.

reduced charges and presumptive sentences, and perhaps more trials.[133]

- Prosecutors could respond to the reduction in their power to manipulate the Guidelines to control the sentence by expanding use of their other tool for constraining judges: statutory mandatory minimums.
- Judges might become less willing to make findings of fact that diverge from the plea stipulations, because doing so is no longer necessary to achieve what they perceive as a just sentencing result—they can depart instead.

These changes would only bias estimates of post-*Booker* changes to racial disparity if they had a racially disparate impact on the presumptive sentence or on the composition of the sentenced sample.[134] It is possible that this is not so, of course, but one cannot simply *assume* it is not so—it must be tested. However, all of the existing studies of *Booker* (and prior studies of the initial shift to mandatory sentencing) do assume exactly that, usually implicitly. Other studies have criticized various other aspects of the Sentencing Commission's *Booker* study and have reached different conclusions. But these studies too have taken the sentencing-stage-only approach, controlling either for the presumptive sentence or for something closely related (the Guidelines "base offense level"), and thus are subject to the same concern.[135]

These studies, in short, ignore the "hydraulic discretion" theory that has dominated theoretical scholarship about sentencing reform.[136] Conversely, key aspects of the hydraulic discretion theory remain almost completely untested

---

133. If cases thereby became more resource-intensive, one might expect prosecutors to bring fewer cases or fewer charges per case. *See* J.J. Prescott, Empirical Evidence of Prosecutorial Charging Manipulation (2006) (unpublished manuscript) (on file with authors).

134. This could be the case even if the changes looked superficially equivalent by race. For instance, if prosecutors doubled their use of mandatory minimums for both black and white defendants in response to *Booker*, but their underlying use of mandatory minimums was twice as common for black defendants, the doubling would look racially neutral even as it had twice the impact on black defendants' sentences.

135. *E.g.*, Fischman & Schanzenbach, *supra* note 54, at 730-31 (finding mixed results in an analysis of multiple doctrinal changes affecting judicial discretion, but concluding that expanded discretion does not increase and may mitigate racial disparity); Nowacki, *supra* note 46, at 16-17 (finding a post-*Booker* increase in mean black-white disparity); Ulmer et al., *supra* note 45, at 1108 (finding a post-*Booker* increase in racial disparity in incarceration rates but not in sentence length).

136. *See supra* notes 28-34 and accompanying text.

empirically.[137] No empirical studies have yet used case data to assess changes in disparities in charging, plea-bargaining, or sentencing fact-finding in the wake of *Booker*. One study surveyed federal district court judges and defense attorneys about their *perceptions* of whether aspects of plea-bargaining had changed.[138] However, the researchers did not evaluate these perceptions' accuracy, and the perceptions of judges and defense counsel varied quite substantially.[139]

Just a few studies have looked at changes in charging and plea-bargaining disparities in response to earlier changes to sentencing law and policy. Wooldredge et al. found that Ohio's shift to mandatory sentencing reduced racial disparities in charge-bargaining, yet *increased* racial disparities in sentencing (a surprising result).[140] But the authors did not evaluate changes in initial charging, without which the results are harder to interpret. In a 1987 study of Minnesota's adoption of mandatory sentencing guidelines, Miethe did evaluate initial charging and found a small but significant increase in gender disparity and no significant change in racial disparity; plea-bargaining disparities were unchanged.[141] No studies have evaluated changes in disparities in sentencing fact-finding.

Beyond the failure to account for pre-sentencing stages of the process, recall that the Sentencing Commission's study of *Booker* has an additional problem: it also controls for departure status, thereby also filtering out some of the potential disparities in the sentencing decision as well. This is an especially surprising choice for a study of *Booker*'s effects, because, as we will see below, *Booker* dramatically changed the probability of a departure from the Guidelines by authorizing departures that were previously forbidden. It is odd to compare racial disparities in sentencing before and after *Booker* only after filtering out those mediated by racial differences in departure rates.

In Table 1, we show that the "changing yardstick" problem is neither

---

137. Engen, *supra* note 31, at 324-25.

138. Jeffery T. Ulmer & Michael T. Light, *The Stability of Case Processing and Sentencing Post-Booker*, 14 J. GENDER RACE & JUST. 143 (2010) (finding perceptions of increased detail in factual stipulations and appeal waivers in plea agreements, but also increased entry of "open pleas" with no agreement).

139. *Id.*

140. John Wooldredge et al., *(Un)anticipated Effects of Sentencing Reform on the Disparate Treatment of Defendants*, 39 LAW & SOC'Y REV. 835, 860-64 (2005).

141. Miethe, *supra* note 10, at 167-71. The Miethe and Wooldredge et al. studies are rare examples of studies that focus on "hydraulic" effects of sentencing reform; however, both are subject to the other critique raised below concerning causal inference from changes over time.

merely theoretical nor subtle: the use of these problematic control variables can completely change the apparent trends in racial disparity. We used a simple linear time-trend model to estimate the overall difference in sentences imposed on black and white defendants, as well as the average growth in that gap over time.[142] We included cases sentenced between the PROTECT Act and the end of fiscal year 2009, and focused on black and non-Hispanic white men.[143] Thus, this analysis covers the time period and groups for which the Sentencing Commission found the purported quadrupling of disparity.[144] The sample includes all non-immigration cases except those subject to major substantive sentencing-law changes during the study period: identity theft, obscenity/child sexual exploitation, and sex offender registration. For reasons explained above, we omit immigration cases.[145]

The purpose of Table 1 is to show the contrast in racial disparity estimates and time trends when one uses our preferred method of measuring disparity (described in Part II) as compared to variations on the "presumptive sentence" method. In Column 1, we show the estimated linear trend in average sentence when controlling for the arrest offense and other prior characteristics (our preferred method).[146] That is, Column 1 shows the trend over time in the aggregate black-white sentence disparity introduced during the post-arrest justice process. The estimated trend in racial disparity is insignificant, and its sign is actually *negative*: the model (noisily) estimates that the unexplained

142. Specifically, the regression includes an overall linear (monthly) time trend as well as an interaction between that trend and the "black" coefficient.

143. The PROTECT Act went into effect on April 30, 2003, and included various provisions designed to discourage downward departures from the Guidelines. PROTECT Act of 2003, Pub. L. No. 108-21, § 401, 117 Stat. 650, 667-76 (codified in scattered sections of 18 U.S.C. and 28 U.S.C.); *see infra* note 156 and accompanying text. The Sentencing Commission uses it to demarcate the beginning of the immediate pre-*Booker* period (the period in which the Guidelines were at their most restrictive, and in which they find the lowest disparity), so we also use it as the starting point for this time trend analysis. *See* 2010 U.S. SENTENCING COMM'N, *supra* note 6, at 10-11.

144. We use the time period covered in the main Sentencing Commission report on demographic disparities (from 2010). The Commission's recent update adds 2010 and 2011, but those years of data have not been made available to researchers. In any event, the Commission's original estimates (through 2009) were even more dramatic than those in its later report: it found that during the period from *Kimbrough/Gall* to the end of fiscal year 2009, the disparity was 23.3%, more than four times its post-PROTECT, pre-*Booker* level. *See* 2010 U.S. SENTENCING COMM'N, *supra* note 6, at 23.

145. *See supra* note 78 and accompanying text.

146. We controlled for arrest offense, criminal history, district, education, age, and multi-defendant case structure.

Table 1.

AVERAGE MONTHLY CHANGES IN SENTENCE LENGTH AND FINAL OFFENSE LEVEL

MAY 2003 - SEPTEMBER 2009: COMPARISON OF LINEAR TREND MODELS

	(1)	(2)	(3)	(4)
Outcome	Sentence Length in Months			Final Offense Level
Black (Initial Black-White Disparity)	12.66** (1.15)	3.16** (0.77)	0.91 (0.76)	1.17** (0.11)
Month (Overall Linear Time Trend)	-0.000 (0.013)	-0.089** (0.010)	-0.106** (0.009)	0.014** (0.001)
Black*Month (Trend in Black-White Disparity)	-0.027 (0.023)	0.043** (0.015)	0.044** (0.015)	-0.012** (0.002)
Arrest Offense Controls?	Yes	No	No	Yes
Final Offense Level/ Category and Mandatory Minimum Controls?	No	Yes	Yes	No
Departure Status Control?	No	No	Yes	No
Observations	119,977	119,784	119,784	119,908
Total Change in Black-White Disparity Implied by Estimates in Months	-2.1	+3.3	+3.3	-0.9

Coefficients for sentencing month, black, and black*month from OLS regressions. Regressions also include controls for criminal history, district, age, citizenship, education, and multi-defendant case status, in addition to those controls noted in the table. The implied overall change is calculated by multiplying the monthly disparity trend by seventy-six months (the length of the period). Standard errors clustered on race-month in parentheses. $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$.

black-white sentence gap declined by 2.1 months, from about 12.7 months to about 10.6 months, over the course of the period.[147]

The negative sign of this estimated change is consistent across a variety of estimation strategies and sample definitions. For instance, while Table 1 shows the results when sentence length is estimated in months (including non-incarceration sentences as zeros), we get similar results if we use a log-linear model excluding the zeros. We also get essentially identical results when we estimate yearly rather than monthly trends. Likewise, we see no rise in disparity over time when, instead of estimating linear trends, we estimate the differences in the "black" effect among the three key time periods that the Sentencing Commission study identifies (PROTECT-to-*Booker*, *Booker*-to-*Gall*, and post-*Gall*).[148] And indeed, some reasonable variations on our approach produce significant and much larger estimated downward trends. For instance, the Table 1 results exclude offense categories that were affected by major substantive changes in the law, because we wanted to focus on disparity trends in the *administration* of the law. But had we included these offense categories (as the Sentencing Commission did), the estimated decline in disparity during the study period would have been significant and nearly three times as large—about six months total.[149]

Why, then, does the Sentencing Commission find an increase in disparity during this period? There may be a variety of explanations,[150] but a prime reason appears to be that racial disparity in the processes determining the presumptive sentence declined significantly over the same period. By controlling for the presumptive sentence, the Commission filtered out that

147. This figure is obtained by multiplying the per-month linear trend estimate by the number of months.

148. We obtain this estimate with a single differences-in-differences regression containing time period dummies interacted with the "black" variable. This analysis shows, if anything, slightly *smaller* disparities in the later periods (with the smallest post-*Gall*), although the period-race interaction terms are not significant. Standard errors are clustered on the month to account for possible events affecting many cases at once. All significant results remain so with alternative clustering, such as clustering on the district to account for correlated prosecutorial policies or courthouse practices.

149. This is likely because, as our data show, child pornography and child sexual exploitation arrestees are overwhelmingly white, so increasing sentences for those offenses tends to reduce black-white disparity. Also, had we included Hispanic defendants, the estimated decline in disparity would have been larger and statistically significant.

150. Among other things, these may include the Commission's arbitrary choice of the 0.01-month valuation of non-prison sentences, the lack of detailed offense-type controls, and the inclusion of immigration cases. These issues are discussed *supra* notes 67 and 78 and accompanying text and *infra* note 160 and accompanying text.

reduction in disparity, leaving only a misleading picture. The black-white gap in sentences *relative to the presumptive sentence* may have grown, but that is because the black-white gap in presumptive sentences shrank (after controlling for underlying case characteristics). In other words, when one controls for the presumptive sentence, the disparities look larger in the later period because the presumptive sentence control is filtering out less of the disparity. The presumptive sentence was not the "same broken yardstick" during this period. Over time, the yardstick changed.

Columns 2 through 4 of Table 1 illustrate this point. In the regression shown in Column 2, rather than controlling for the arrest offense, we substituted the final offense level, the mandatory minimum indicator, and broad offense-type categories associated with the offense of conviction. This reflects a fairly typical version of the presumptive sentence approach. Recall that the presumptive sentence is determined by the final offense level (and the criminal history category, which we control for in all regressions). The regression in Column 3 is identical except that we more closely approximate the Commission's approach by also adding departure status controls.[151] After these modifications, both of these regressions show a significant linear increase in racial disparity over time, albeit not as dramatic an increase as the Commission itself found. In the Column 2 version, the unexplained black-white gap increases from about 3.2 months at the beginning of the period to about 6.5 months at the end. When departure status is added as a control in Column 3, the black-white gap is estimated to rise from about 0.9 months at the beginning of the period to about 4.2 months at the end.[152]

Thus, when we use variations on the presumptive sentence approach, we do see significant increases in disparity, just as other studies have found. But that approach is inappropriate and misleading, because of the "changing yardstick" problem. Column 4 focuses directly on that changing yardstick. In Column 4, we show a time-trend regression with the final offense level as the outcome of interest. After controlling for the arrest offense and other pre-charge characteristics, the unexplained black-white disparity in final offense levels declined by nearly one level during this period. For the average case in

151. As explained above, we believe that it is a mistake to control for departure status, even if one intends to focus on judicial behavior alone. *See supra* note 53 and accompanying text. We do so here purely for illustrative purposes.

152. Note that even in the Column 2 and Column 3 regressions, the overall estimated average sentences for black defendants are lower at the end of the period than at the beginning. That is, even with the presumptive sentence approach, the rise in black-white disparity is not large enough to offset the overall monthly trend—other factors equal, sentences for both black and white defendants declined in all of the models over time.

the sample, a change of one offense level is associated with a five-month change in presumptive sentence length—close to the difference between the disparity trend estimate in Column 1 (-2.1 months) and those in Columns 2 and 3 (+3.3 months).[153] That is, the changing nature of the presumptive-sentence yardstick appears to explain nearly all the difference between the disparity decline that we measure using our method (Column 1) and the apparent increase that one sees when one uses a method paralleling those of other studies.

Thus, the overall unexplained racial disparity in the post-arrest justice process certainly does not appear to have increased from 2003 to 2009, and if anything, it seems to have decreased. The linear trend results do suggest that the *procedural sources* of disparity may have shifted over the course of the period, with the earlier stages in the process becoming a bit less important and the judicial sentencing decision becoming a bit more important. However, it bears noting that throughout the time period, the earlier process stages appear to be the dominant procedural sources of disparity. That is, the overall estimated racial disparities are much larger when one controls for the arrest offense—thereby incorporating disparities from those earlier procedural stages—than when one uses either version of the presumptive sentence model (compare the "black" coefficient in Column 1 to the "black" coefficients in Columns 2 and 3). This is consistent with our research on the sources of post-*Booker* disparities,[154] which finds that even in the most recent years, charging decisions appear to be the major driver of sentencing disparity.

Note that we make no claims as to the causes of these longer-term trends, and specifically, we do not claim to have established that *Booker* caused them. As we explain in the next Section, causal inference from changes over lengthy periods of time is a fraught enterprise. Table 1 merely shows that, even setting aside the causal inference concern, racial disparity in the post-arrest justice process is no worse today than it was before *Booker* was decided.

---

153. When we repeat the Column 3 analysis (the one closest to the Commission's) on a sample that includes the excluded obscenity, sex-offender, child sex crimes, and identity theft categories, we obtain a somewhat stronger estimated upward trend in disparity, totaling about four months. Likewise, when we repeat the Column 4 analysis for that sample, we find a larger decrease in offense-level disparity, totaling 1.5 levels. This supports the theory that inclusion of those categories contributed to the Commission's findings of increased disparity in the later periods, largely because the underlying presumptive sentence measure was changing.

154. *See supra* Part II.

B. The Causal Inference Problem

In addition to the use of inappropriate control variables, there is another major methodological problem with previous studies of *Booker*'s effects: they lack a basis for sound causal inference. Causal inferences from changes over time are always risky, because many things change over time. Comparisons of averages between periods before and after a policy change, while appealingly simple, can be misleading.

These studies generally compare the average disparity before and after a policy change. In most, disparities are estimated separately for each period using a regression model that controls for the presumptive sentence and other observed variables.[155] The recent federal studies have focused not just on *Booker*, but also on other recent policy changes affecting judges' sentencing discretion. One such change was Title IV of the PROTECT Act of 2003, which imposed rules intended to discourage downward departures from the Guidelines. It required courts to report to Congress on departure rates, required written justifications for departures, provided for de novo appellate review of departures in some cases, restricted the Sentencing Commission from creating new grounds for downward departures, limited judicially initiated downward adjustments for "acceptance of responsibility," and directed DOJ to adopt an action plan for reducing departures.[156] The Supreme Court's December 2007 decisions in *Kimbrough* and *Gall* (discussed above), which reinforced the *Booker* holding, have also been a focus of the recent research.[157]

The Sentencing Commission focused on three primary time periods, with cases classified by sentencing date: (1) PROTECT-to-*Booker* (nearly two years), (2) *Booker*-to-*Kimbrough/Gall* (nearly three years), and (3) post-*Kimbrough/Gall* (nearly two years). It found the lowest black-white disparities in period (1), when judicial discretion was the most limited, and the greatest in period (3), when discretion was broadest.[158] A competing study by Jeffrey Ulmer, Michael Light, and John Kramer criticized aspects of the Commission's method, but it too compared averages across these time periods (as well as

---

155. The use of separate regressions means that these studies do not control for the case mix across time periods, which is another problem. Some studies report single regressions with race-time period interactions, which is preferable (but does not solve the other problems raised here). *See* Ulmer et al., *supra* note 45, at 1096 (reporting both methods).

156. PROTECT Act of 2003, Pub. L. No. 108-21, § 401, 117 Stat. 650, 667-76 (codified in scattered sections of 18 U.S.C. and 28 U.S.C.).

157. *See supra* notes 40-41 and accompanying text.

158. 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. A, at 2-3.

earlier periods).[159] It similarly found increases in racial disparity in the post-*Booker* and post-*Kimbrough/Gall* periods, although these effects were concentrated in the decision whether to incarcerate defendants rather than in sentence length among those incarcerated.[160] A recent study by Jeffrey Nowacki similarly compares the cases from 2002-2004 (pre-*Booker*) to those from 2005-2008 (post-*Booker*), and finds a fairly small but significant increase in disparity in the latter period, controlling for final Guidelines offense level, criminal history, and other variables.[161]

But comparison of averages across such broad periods is at best suggestive and is too blunt a tool for causal inference. Differences in the averages between periods might merely reflect longer-term trends or other intervening events. If racial disparity were rising steadily throughout the period, for instance, the average disparity after *Booker* would necessarily be larger even if *Booker* had no effect on racial disparity. In fact, this would be true even if *Booker* actually *slowed* the rate of increase in disparity.

Sentencing disparity might well be affected by numerous non-*Booker*-related developments over periods of this length. One possibility is changes in the underlying case mix—including case types, severity, and defendant

---

159. Ulmer et al., *supra* note 45, at 1091-94, also includes the pre-PROTECT period.

160. *Id.* Much of the authors' criticism focused on the Commission's failure to separate the incarceration decision from the length decision. *See id.* at 1093-94. In our view there is no theoretical reason the two decisions necessarily need to be separated; either approach is acceptable. The main problem with the Commission's approach is that even though it kept the zeroes in the main sample, it log-transformed the outcome variable; since the log of zero is undefined, the Commission assigned the value of 0.01 months to the non-incarceration sentences. The problem is that the choice of 0.01 is arbitrary, and the resulting effect estimates are sensitive to this arbitrary choice.

Separating the incarceration and length decisions, as Ulmer et al. did, avoids this problem, but raises another concern: if there is disparity in the incarceration decision, it will introduce sample selection bias to the length analysis. *See* Ulmer et al., *supra* note 45, at 1091-94. Starr examines this problem and various alternative solutions in detail. Starr, *supra* note 79, at 5-7. In the analysis reported in Part II, we did separate the two stages, but it was not problematic because we found no significant disparity in the incarceration decision after controlling for arrest offense. If one *does* find disparity there, as Ulmer et al. did, the sample selection concern is more serious, and it is better to leave the zeroes in the main sample and either not log-transform the outcome or use alternative statistical methods. If the zeroes are included, linear regression (estimated with robust standard errors) can still be used to estimate average disparities; it is an unbiased estimator of the conditional mean even if its distributional assumptions are violated. Note that there is no "censored data" concern here; non-incarceration sentences are known zeroes, not unknown outcomes.

161. Nowacki, *supra* note 46, at 15-16.

characteristics—as underlying crime patterns and federal law enforcement priorities change. Controlling for case characteristics within each time period, as the Commission and similar studies do, only filters out the effects of differences in the distributions of characteristics for black versus white defendants during that period. It does not mean that changes in the case mix between time periods will not affect the disparity estimates. The type of model used by the Commission and in similar studies gives a single estimate for the effect of being black, averaged across all (male) cases, but in practice this average surely hides heterogeneity. That is, the gap between white males and black males might in practice vary depending on the nature of the case or the defendant's characteristics.

For instance, imagine that there are just two kinds of cases in a sample—fraud and robbery—and that the average unexplained black-white gap is 5% for frauds and 20% for robberies. The Commission's approach would produce an average disparity estimate somewhere between 5% and 20%, depending on what fraction of the cases are frauds and what fraction are robberies. If the fraction that are robberies is gradually growing over time, then racial disparity will appear larger in the regressions from the later periods even if nothing else changes (that is, even if the gap remains 5% for fraud cases and 20% for robbery cases).[162]

Changes in the case mix are not the only potentially confounding developments that could occur over time. Other possibilities include the policies and prosecution strategies of the Department of Justice changing or taking time to trickle down to line prosecutors; changes in the composition of the judiciary, U.S. Attorneys' and public defenders' offices; or administrative changes in supervision of prosecutors that shift their incentives. Even if these developments had no racial purpose, they might have had racially disparate impacts. Causal inferences would be more credible if effects were visible in a much shorter time window, such that one could more confidently assume that *Booker* is the only important change that could have driven the outcome. One can also filter the surrounding trends out of the estimates of the policy's effects by including them in the regression.

162. The Commission's latest report does include some results broken down separately by case category, but these categories are broad, such as "drug trafficking." *See* 2012 U.S. SENTENCING COMM'N, *supra* note 7, pt. E, at 23-25. The case-type mix might change in subtler ways (for example, a growing number of methamphetamine cases, more car thefts, etc.), and the distributions of other important variables like criminal history might also change over time.

Among the recent *Booker* studies, Fischman and Schanzenbach's offers an improvement on the standard approach.[163] Their model filters out year-to-year variation in sentencing patterns for different categories of crimes and judicial districts, which captures an important subset of the things that might vary over time. They focus on changes in appellate review of sentencing and find that, in general, looser review has not been associated with increased racial disparity, although (like the Sentencing Commission) they do find a recent increase in disparity after *Kimbrough* and *Gall*.[164] However, their approach only filters out trends in racial disparity if they are mediated by the crime category or district; any trends driven by other factors are left in. Below, we set forth an approach that filters out continuous trends in racial disparity itself (rather than trends in particular factors that contribute to it) and that uses monthly data to capture within-year variation as well.

C. Our Method

In order to disentangle *Booker*'s effects from surrounding trends, rather than comparing racial disparities averaged over periods of years, we create flexible regression models that filter out month-to-month trends (including non-linear trends) in sentences and other relevant outcomes. We then look for sharp breaks in these trends—discontinuities—immediately after *Booker*. This approach is, in effect, a regression discontinuity-style estimator (RD), and, for simplicity, we will use the label RD here.[165] Like other studies, we base our

---

163. Fischman & Schanzenbach, *supra* note 54.

164. *Id.* at 730. The authors further examine *Kimbrough/Gall* and their predecessor, *Rita v. United States*, 551 U.S. 338 (2007), with an event study approach that effectively averages disparities over six-month periods, rather than the Commission's longer periods. Fischman & Schanzenbach, *supra* note 54, at 757. Because they leave out the six months between *Rita* and *Kimbrough/Gall*, the last pre-period and first post-period are actually nearly a year apart. Although this analysis improves on the Sentencing Commission's, we think an even finer-grained approach to time trends yields greater payoffs for causal inference, and we also prefer to focus on *Booker*, the bigger legal change.

165. RD estimators are widely used in the education, public finance, political economy, and labor economics literatures to recover causal estimates when randomized experiments are not possible. *See* Guido W. Imbens & Thomas Lemieux, *Regression Discontinuity Designs: A Guide to Practice*, 142 J. ECONOMETRICS 615 (2008); David S. Lee & Thomas Lemieux, *Regression Discontinuity Designs in Economics*, 48 J. ECON. LITERATURE 281 (2010). Although these estimators most frequently involve discontinuous thresholds in continuous running variables other than time (for instance, distance from a border), the method can be applied to the assessment of policy changes with time as the running variable. *See, e.g.*, Michael L. Anderson, *Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion* 10-11 (Nat'l Bureau of Econ. Research, Working Paper No. 18757, 2013),

causal inferences on changes over time, and any unmeasured changes that coincided with *Booker* could trick us. But because we are looking for *immediate sharp* changes, this concern is less grave. While a lot can change in a couple of years, usually a lot less changes suddenly in a couple of months. In addition, even if continuous background trends did have a noticeable effect on disparities in those couple of months, our method filters the trends out. We are looking only for sharp breaks that coincide with *Booker*. If the surrounding trends are fairly smooth and there is a sudden break at *Booker*, the inference that *Booker* caused the change depends only on the assumption that no other unobserved factor affecting sentencing disparity *suddenly* changed at the time of *Booker*.

Our sample runs from fiscal years 2001 to 2009 and includes women and non-citizens (with controls for gender and citizenship). This broad sample definition is useful in improving the precision of the estimates by increasing the sample size within each month. However, the results are substantively similar if these groups are excluded. The sample includes all non-immigration cases except identity theft, which was subject to other major sentencing-law changes very near *Booker*.[166]

Our overall research interest is in measuring the effect of changes to judicial sentencing discretion on sentencing and case processing disparities. We begin by looking at Guidelines departure rates, not because that is the ultimate outcome of interest, but because departure rates help us determine which legal reforms amounted to important changes to judges' discretion in practice. They directly measure Guidelines compliance and thus are the most logical measure of the extent to which the Guidelines actually constrained judicial behavior at any given time. We focus our attention on *Booker* itself, not on its progeny *Kimbrough* and *Gall* or on the PROTECT Act's tightening of the Guidelines. The reason can be seen plainly in Figure 1, which plots departure rates by sentencing month.[167] Note that 96% of these departures are downward.

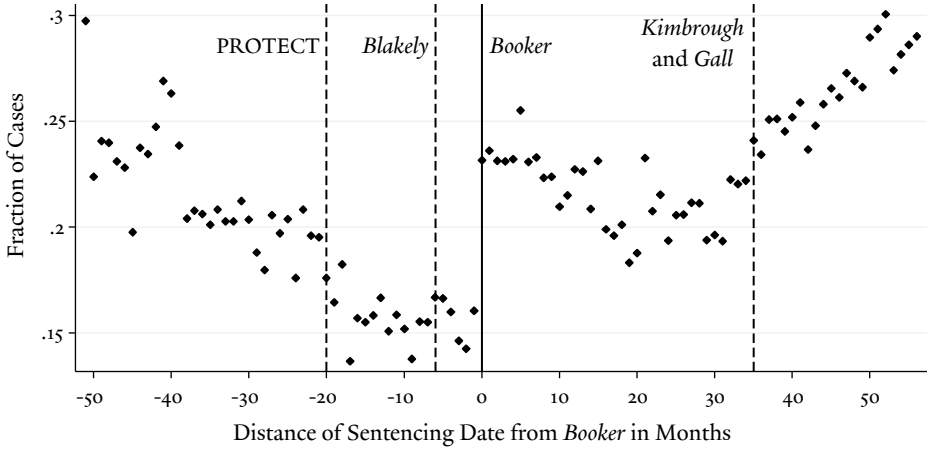
---

http://www.nber.org/papers/w18757.pdf; Wojciech Kopczuk et al., *Do the Laws of Tax Incidence Hold? Point of Collection and the Pass-Through of State Diesel Taxes* 23-24 (Nat'l Bureau of Econ. Research, Working Paper No. 19410, 2013), http://www.nber.org/papers/w19410.pdf.

166. Unlike in the linear trend analysis in Section A, we do not exclude Hispanic defendants from this analysis; they are included among the white and black defendants. However, our results are very similar if we do exclude Hispanic defendants, focusing on the gap between non-Hispanic white and black defendants.

167. For reasons explained below, this graph and all others are limited to district courts in the Second, Fourth, Fifth, Sixth, and Eleventh Circuits. The nationwide departure pattern looks similar.

Figure 1.
DEPARTURES OVER TIME



The vertical lines in Figure 1 mark four key events: the PROTECT Act, *Blakely* (*Booker*'s immediate predecessor), *Booker*, and *Kimbrough/Gall* (which clarified and strengthened *Booker*'s holding).[168] As this graph makes clear, *Booker* was a major shock to the sentencing discretion afforded to judges. Departures increased immediately and substantially, from about 30% to about 40%. Although there are other month-to-month fluctuations, *Booker* marks by far the most dramatic break. After the immediate *Booker* jump, departures continue on a gradually downward trajectory similar to the one that existed before *Booker*—but the whole graph is shifted upward by about 10%, and the departure rate never returns to its pre-*Booker* low. In other words, *Booker*'s effects were sudden, but they were also lasting.

The sharpness of the change at *Booker* helps to alleviate one substantial concern about RD—its inability to capture effects that occur slowly. It is very possible that the full effects of *Booker* took a while to take hold—for example, the size of departures could have grown over time as judges became more comfortable with their newfound discretion. The inability to test that possibility is a disadvantage to our method. Policymakers are of course likely to be interested in *Booker*'s long-term effects. But focusing on the short-term

168. The graph includes *all* departures. In 46% of departure cases, the departure was requested by the government as a reward for "substantial assistance" in another case. If these were excluded, the pattern would look similar, but the rise immediately following *Booker* would be even steeper.

effects can tell us something important about the expected direction of *Booker*'s long-term consequences, even though we cannot directly measure those long-term consequences.

Why, after all, would one ever have worried that *Booker* might increase sentencing disparity, as critics, including the Sentencing Commission, did? The theory is that giving judges more discretion frees them to sentence in ways that turn on their conscious or unconscious sympathies with, or predictions about, particular defendants, and that those sympathies or predictions will differ on the basis of race or other factors correlated with race. That is, the theory assumes that judges have inclinations that effectively favor white defendants over similar black defendants (more strongly than the previous mandatory-Guidelines regime already favored white defendants). *Booker* provides a chance to test whether that theory appears to be true. Legally, making the Guidelines non-mandatory was a sudden and enormous change to judicial discretion. The departure graph shows that this change was not just theoretical and did not take long to have effects in practice. On the contrary, the doctrinal shock to the scope of judicial discretion *immediately* manifested itself in substantially more frequent judicial *exercises* of discretion. If judges were, in fact, inclined to use broader sentencing discretion in ways that disadvantage black defendants, one would expect to see at least *some* of that effect in the immediate vicinity of *Booker*, even if the full effects of the decision took a while to play out. If there is *no* jump in disparity at *Booker*, it suggests that judicial inclinations were not what critics feared they were.

In contrast, the PROTECT Act and *Kimbrough/Gall* did not produce nearly as dramatic a change to the sentencing regime in practice. PROTECT appears to have caused no sudden change at all in departures. *Kimbrough* and *Gall* may have been more important—departure rates did rise afterwards—but the rise continued a trend that began three months before the decisions, and there was no sudden break in the trend (nor was there a sudden break at the time of *Rita v. United States*,[169] five months earlier).[170] Even if *Rita* and *Kimbrough/Gall*

169. 551 U.S. 338 (2007).

170. This is not very surprising. PROTECT and *Kimbrough/Gall* were much subtler changes in the law. PROTECT and *Gall* did not directly speak to judges' legal authority to depart; they affected appellate review standards and data collection procedures. *See* Gall v. United States, 552 U.S. 38, 40-41 (2007). *Kimbrough* did directly affect departure authority, but only in crack cases (in which mandatory minimums applied regardless); it was uncontested that judges could depart on policy grounds in other cases. *See* Kimbrough v. United States, 552 U.S. 85, 101-02 (2007). The crack holding could only have helped black defendants—a logical challenge for studies that point to *Kimbrough* as a source of racial disparity.

collectively led to an increase in departures, the fact that the decisions were separated by five months makes this too diffuse a change to judges' sentencing discretion to assess with our method. And even combined, the change over that whole period is still much smaller than the change at *Booker*. One should not expect small changes to have big effects, and if they appear to, one has to suspect some confounding factor. *Booker*, as the bigger change, is the more logical place to test the effects of changing judicial discretion.

We thus assess the effects of *Booker*'s shock to judges' departure discretion on other stages and outcomes in the justice process. Because criminal cases have several key dates, the RD method can be used to isolate *Booker*'s effect on each key stage in the process. However, it cannot be used to directly estimate the *aggregate* effect of *Booker* on all stages. The Sentencing Commission and other *Booker* researchers have always divided cases by sentencing date, but many cases' processing dates straddle *Booker*, so one cannot simply deem cases "pre-*Booker*" or "post-*Booker*." We assess *Booker*'s effects on charging, as well as the sentencing consequences of those charging changes, by assessing what happens when the charging date passes *Booker*. Cases charged shortly before *Booker* will overwhelmingly have been disposed of and sentenced after *Booker*,[171] so focusing on the immediate effects as the charging date passes *Booker* means that the sentencing effects of changing charging practices can be separated from the sentencing effects of changes to other process stages.

Likewise, we assess plea-bargaining changes and their sentencing effects by assessing what happens when the conviction date passes *Booker*, and we assess changes in judicial behavior by assessing what happens when the sentencing date passes *Booker*. Note that the judicial behavior being measured involves not only changes to the final sentencing decision but also changes to sentencing fact-finding. Assessing the conviction date and the sentencing date separately helps to disentangle judges' contributions to disparities in sentencing fact-finding from disparities in the negotiated plea stipulations.

The most serious complication in drawing causal inferences about *Booker* is that the decision was hardly a bolt from the blue. Rather, *Booker* followed six months after the Supreme Court's decision in *Blakely* (denoted by the second vertical line in Figure 1), which applied the same Sixth Amendment analysis to a state sentencing scheme. It was *Blakely* that was an unexpected earthquake, rendering it fairly obvious that the federal Guidelines were in constitutional

171. The average time from charge to disposition in our sample is five months, and the average time from disposition to sentencing is a further four months.

trouble.[172] What was *not* clear was what the Supreme Court would do to remedy the constitutional defect. Instead of the advisory guidelines approach (which none of the circuits had adopted), the Court could have struck the Guidelines down entirely, left them mandatory but shifted fact-finding to the jury, or left the whole matter to Congress. The lower courts began weighing in, and the Supreme Court quickly agreed to review *Booker*.[173]

The *Blakely* decision raises a dilemma for causal inference for three reasons. First, it could mean that the effects we are looking for happened in a more diffuse manner starting *before Booker*, because courts or parties adjusted their behavior in anticipation of the mandatory Guidelines' fall. In that case, estimating discontinuities at *Booker* alone might understate the effects of moving away from mandatory Guidelines. Second, the anticipation of *Booker* may have affected the mix of cases decided immediately before and after *Booker*, if district courts delayed sentencings while waiting for the Supreme Court's opinion. Such changes in cases could confound estimates of *Booker*'s effects. Third, even assuming *Booker* did cause the measured changes, not all of *Booker*'s effects can necessarily be attributed to the expansion of judicial discretion. In addition to rendering the Guidelines advisory, *Booker* may have affected outcomes by ending the chaotic interregnum period and rejecting the alternative remedies that the Court could have chosen. These problems are not unique to our method—they afflict all studies of *Booker*—but they cannot be ignored.

For this reason, we constrain our analysis to five federal judicial circuits: the Second, Fourth, Fifth, Sixth, and Eleventh. Within two to six weeks of *Blakely*, these five courts of appeals issued decisions holding that *Blakely* did not apply to the federal Guidelines.[174] In those circuits, *Booker*'s legal effects were simpler: it changed the governing law from the old regime (mandatory

---

172. The archives of Douglas Berman's Sentencing Law and Policy blog for the period between the decisions in *Blakely* (June 24, 2004) and *Booker* (Jan. 12, 2005) provide an excellent record of this disarray. *See, e.g.*, *June 26, 2004*, SENT'G L. & POL'Y, http://sentencing.typepad.com/sentencing_law_and_policy/2004/week26/index.html (last visited Sept. 1, 2012).

173. *See* United States v. Booker, 375 F.3d 508 (7th Cir.), *cert. granted*, 542 U.S. 956 (2004); *see infra* note 174 (listing lower court decisions).

174. United States v. Reese, 382 F.3d 1308, 1310 (11th Cir. 2004), *vacated*, 543 U.S. 1114 (2005); United States v. Koch, 383 F.3d 436, 438 (6th Cir. 2004), *vacated*, 544 U.S. 995 (2005); United States v. Hammoud, 381 F.3d 316, 345 (4th Cir. 2004), *vacated*, 543 U.S. 1097 (2005); United States v. Mincey, 380 F.3d 102, 106 (2d Cir. 2004), *vacated sub nom.* Ferrell v. United States, 543 U.S. 1113 (2005); United States v. Pineiro, 377 F.3d 464, 473 (5th Cir. 2004), *vacated*, 543 U.S. 1101 (2005).

Guidelines) to the new one (advisory Guidelines). During the *Blakely*-to-*Booker* period, there was neither legal chaos nor a third legal regime. Figure 1, which is limited to these "business as usual" circuits, shows that nothing happened to departure rates at *Blakely* or during the interregnum—there was no trend break until *Booker*.

Our focus on these circuits is only a partial solution to the *Blakely* problem. While district courts were required to follow the "business as usual" approach, if the *parties* anticipated that the Supreme Court would change the law before sentencing, they were free to let that expectation affect their charging and plea-bargaining decisions.[175] Therefore, as detailed below, we also analyze changes happening at the time of *Blakely* to see whether there is evidence of such anticipation effects.

D. Regression Discontinuity Estimates of Booker*'s Effects*

Here we present our RD estimates for key charge severity, plea-bargaining, and sentencing measures. In addition to the results presented below, we also assessed changes in the criminal justice "funnel," which could have introduced sample selection bias into the RD estimates. However, we found no significant change in the rate of filing charges in district court as the charging date passed *Booker*, nor in the rate of non-petty convictions as the disposition date passed *Booker*.[176]

1. Changes to Charging

The principal charging dynamic that we sought to analyze is whether *Booker* affected prosecutors' use of mandatory minimums, which our (post-*Booker*) findings discussed in Part II show to be a key driver of the black-white gap. There is also a logical causal mechanism for such an effect. *Booker* reduced

---

175. For instance, DOJ directed prosecutors to begin including aggravating factors in indictments, rather than waiting for the plea stipulation or sentencing hearing to allege them. *See* Memorandum from James Comey, Deputy Att'y Gen., to All Fed. Prosecutors, Departmental [sic] of Justice Legal Positions and Policies in Light of *Blakely v. Washington* (July 2, 2004), *reprinted in* 16 FED. SENT'G REP. 357 (2004). This memo only suggested changes in the alleged facts, however, and not in the actual charges.

176. We treat January 2005 as the first month in the post-*Booker* period. There were six business days in January before *Booker* was decided, and the dataset gives dates only in months. Conflating the last week of the pre-period into the post-period is (if anything) likely to mean we slightly understate *Booker*'s effects.

prosecutors' ability to use the Guidelines to control sentencing outcomes, an ability that confers massive leverage in plea-bargaining. Without being able to rely on the Guidelines, it is plausible that prosecutors might turn more often to their other tool for constraining judges: mandatory minimums.

Our findings above also clearly showed that it was the *initial* charging stage in which the mandatory minimum disparity emerged, so that is a key stage to analyze. As explained above, we could not code initial charges in drug or child pornography cases. We only know the *final* mandatory minimum in these cases. Fortunately, unlike in the analysis in Part II, in this part of our analysis there is a solution to this problem. RD allows us to assess changes to the *final* mandatory minimum when the *charging* date passes *Booker*.[177] Even though the outcome variable is measured at the conviction stage, changes in it that are triggered by the timing of the *charge* are probably the result of charging changes.[178] This approach allows us to assess all case types.

The results from the formal RD analysis are presented in Table 2, which shows the estimated discontinuous change in mandatory minimum convictions at *Booker*. Within each panel of the table, the first row ("Overall Discontinuity") estimates the change for the whole population at *Booker*, while the second ("Black-White Difference Discontinuity") estimates the *Booker*-related change for black defendants relative to white defendants. That is, the second row measures the change in racial disparity at *Booker*. To see the estimated change for black defendants at *Booker*, one adds the estimates in the two rows. The estimated change for white defendants at *Booker* is simply the overall discontinuity.

---

177. The "charging date" is the date of the indictment, when there is one. In cases with no formal indictment, we used either the arrest date or the date the prosecutor opened the file on the case, whichever was later (usually they are the same month)—that is, the date that the prosecutor had both the case and the defendant in hand, and declined to add to or change the charges from the complaint.

178. If prosecutors suddenly started charging mandatory minimum offenses more often after *Booker*, that would presumably translate into more *convictions* of mandatory minimum offenses for cases charged after *Booker*, too. As noted above, initial charges usually are not dropped; from 2003 through the end of our study period, doing so required a supervisor's special permission. *See supra* note 82 and accompanying text.

Table 2.

REGRESSION DISCONTINUITY ESTIMATES OF *BOOKER*'S EFFECTS ON CHARGING, PLEA-BARGAINING, AND SENTENCING

	1. Charging Date				2. Disposition Date				3. Sentencing Date			
A. Statutory Mandatory Minimum Indicator												
Overall Discontinuity	-0.0272†	-0.0498**	-0.0518**	-0.0159	0.0328	-0.0148	-0.0276†	-0.0466*	0.0379**	0.0534**	0.0542**	0.0605**
	(0.0139)	(0.0166)	(0.0169)	(0.0180)	(0.0215)	(0.0128)	(0.0148)	(0.0200)	(0.0121)	(0.00997)	(0.0132)	(0.0123)
Black-White Difference Discontinuity	0.0596*	0.0917**	0.114**	0.0641†	-0.0417	0.0115	0.0124	0.0202	-0.0196	-0.0328†	-0.0411†	-0.0148
	(0.0252)	(0.0290)	(0.0285)	(0.0335)	(0.0263)	(0.0208)	(0.0215)	(0.0250)	(0.0207)	(0.0181)	(0.0230)	(0.0204)
B. Offense Level												
Overall Discontinuity	-0.271	-0.412	-0.374	-0.339	0.248	-0.203	-0.522	-0.740**	0.908**	0.410†	0.427†	-0.0244
	(0.263)	(0.249)	(0.278)	(0.239)	(0.312)	(0.306)	(0.335)	(0.272)	(0.232)	(0.214)	(0.254)	(0.268)
Black-White Difference Discontinuity	0.338	0.622	0.926†	1.091†	0.129	0.284	0.422	0.770	-0.551	-0.785†	-1.106*	-0.548
	(0.437)	(0.479)	(0.497)	(0.564)	(0.387)	(0.404)	(0.407)	(0.492)	(0.500)	(0.464)	(0.449)	(0.419)
C. Prison Sentence in Months												
Overall Discontinuity	-1.458	-3.579	-2.088	-1.634	3.373	1.193	-1.141	-6.054*	4.351*	0.467	0.739	2.754
	(2.139)	(2.159)	(2.301)	(1.719)	(2.497)	(2.819)	(2.891)	(2.728)	(2.046)	(1.623)	(1.813)	(1.879)
Black-White Difference Discontinuity	4.377	9.065†	10.29†	10.27†	-0.0829	-3.208	-0.921	8.140†	-3.654	-8.018	-10.95*	-15.25**
	(4.556)	(5.151)	(5.769)	(5.865)	(3.716)	(4.326)	(4.390)	(4.780)	(5.480)	(5.336)	(5.266)	(5.224)
Degree of Polynomial	2nd	2nd	3rd	3rd	2nd	2nd	3rd	3rd	2nd	2nd	3rd	3rd
Window Around *Booker*	±18	±12	±18	±12	±18	±12	±18	±12	±18	±12	±18	±12

Table 2. (Continued)
REGRESSION DISCONTINUITY ESTIMATES OF *BOOKER*'S EFFECTS ON CHARGING, PLEA-BARGAINING, AND SENTENCING

	1. Charging Date				2. Disposition Date				3. Sentencing Date			
D. Departures (Not Including Government-Initiated Substantial Assistance Departures)												
Overall Discontinuity									0.0678**	0.0689**	0.0725**	0.0511**
									(0.00788)	(0.00850)	(0.00917)	(0.00980)
Black-White Difference Discontinuity									0.0270*	0.0144	0.0108	0.0335
									(0.0129)	(0.0130)	(0.0155)	(0.0208)
Degree of Polynomial	2nd	2nd	3rd	3rd	2nd	2nd	3rd	3rd	2nd	2nd	3rd	3rd
Window Around *Booker*	±18	±12	±18	±12	±18	±12	±18	±12	±18	±12	±18	±12

Regressions include polynomial time trends estimated separately before and after *Booker*, interacted with race, and control for arrest offense, criminal history, education, multi-defendant case, gender, age, and citizenship. Coefficients reflect discontinuous changes in average white outcomes and the average black-white disparity as the date of charging, disposition, or sentencing passes *Booker*. Samples are cases within eighteen months of *Booker* (N ≈ 57,000) or within twelve months of *Booker* (N ≈ 38,000); number of observations varies slightly depending on the outcome variable. The black discontinuity is the sum of the overall discontinuity and the discontinuity in the black-white difference. Standard errors are clustered on the running variable (race*month). †p<0.1, *p<0.05, **p<0.01.

We estimate regressions that include separate non-linear time trends for black and white defendants, before and after *Booker*—that is, we filter out both the overall underlying trends and the underlying trends in the black-white disparity. The regressions also filter out the month-to-month variation in arrest offenses and other pre-charge features of the case.[179] The estimated discontinuities represent the break in the curve at *Booker*—that is, the difference between the intercepts of the pre-*Booker* curve and the post-*Booker* curve.

Within each panel of Table 2, the four columns show the results of multiple specifications that use different methods of fitting curves to the data—we vary the length of the time window used to estimate the curves on each side (twelve months versus eighteen months) as well as the degree of the polynomial function of time (quadratic versus cubic). There is no one "right" choice for the window or the polynomial. A result is more robust if it is consistent across specifications, which suggests that it is not just an artifact of a subjective modeling choice.

We find that as the charging date passes *Booker*, there is a significant, discontinuous increase in the mandatory minimum rate—but only for black defendants (Panel 1A). The estimated increase in the black-white disparity in mandatory minimums is quite large in all specifications, ranging from six to eleven percentage points, and is significant in three out of four specifications (and marginally significant in the fourth).[180] Most of the increase in disparity is due to an increase for black defendants, but there also appears to be a smaller *reduction* in the frequency with which white defendants received mandatory minimum sentences.

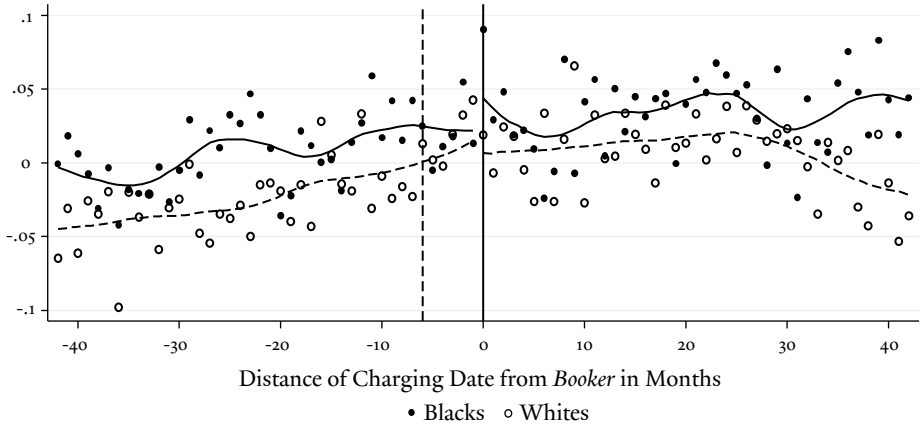
---

179. The controls include arrest offense, criminal history, gender, age, a multi-defendant case flag, U.S. citizenship, criminal history, and education. The results shown exclude district, which was not an important contributor to racial disparity in our initial study; including so many dummy variables was problematic given the sample size per month. District was added in robustness checks, and the results were generally similar but often less precise. Note that controls serve a different function in RD than they do in other regressions—they are mainly there to absorb statistical noise. If there are underlying continuous trends in the effects of the control variables, those will be filtered out by the time-trend variables. Including the controls, however, protects against the possibility of *sudden* changes in underlying case features at *Booker*. In a perfect RD situation—that is, if one could safely assume that other variables changed only in continuous ways—one would not need controls at all, but we do not rely on that assumption.

180. To provide perspective, about 40% of defendants during 2004 faced a mandatory minimum.

Figure 2a provides an approximate visual representation of this result.[181] Although the RD is estimated based on a narrower window of time surrounding *Booker*, the graphs show longer surrounding trends to provide context. The hollow circles and dots represent the monthly averages in the *residuals* for white and black defendants, respectively, from a regression on all the variables from the RD. A residual is the difference between the actual outcome observed for an individual and the outcome predicted by a multivariate regression based on other observed characteristics (for example, arrest offense). Figure 2a thus shows the trends in average black and white charges after controlling for the cases' underlying characteristics other than race. Curves are then fitted to these monthly averages to approximate the month-to-month trends for black and white defendants, and the vertical distance between the black and white curves represents the unexplained racial disparity at any given time.

Figure 2a.
FRACTION OF CASES WITH A STATUTORY MINIMUM
After Accounting for Defendant and Case Mix



181. The curves in the visual representations are fit slightly differently from the formal RD, so the correspondence between the figures and tables is only approximate. The figures contain the monthly average of the variable of interest along with curves fitted using kernel weighted local polynomial smoothing. The curves are fit separately on each side of *Booker*, and capture linear and non-linear trends over time. The vertical distance between the curves on either side at *Booker* (the difference in intercepts) is a visual approximation of the discontinuity estimated with RD.

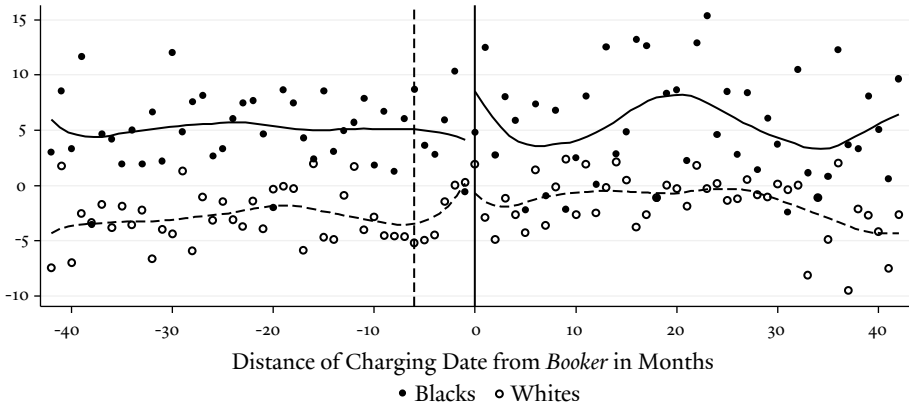
The figure shows that the estimated jump in disparity after *Booker* is heavily influenced by the charging patterns in the first three months after *Booker*, especially the first month. Although there is an unexplained race gap in mandatory minimums through most of the period (the black line is above the white line), the trends had converged in the period leading up to *Booker*. In the month of *Booker*, there was a huge spike in black mandatory minimums. After the first few months, however, things seem to have reverted more or less to the previous trends. The race gap fluctuated somewhat, but the dominant background trend was a steady rise in mandatory minimums for both black and white defendants, and that trend continued.

Overall, although there is a significant break, the patterns are much less dramatic than what we saw with the overall frequency of departures (Figure 1), in which the changes were much larger and stuck. When a trend break is driven largely by a one-month anomaly, one has to wonder if it is due to chance. Here, the divergence from the trend in that one month far exceeds the noise found in the rest of the data, so we suspect that it *is* connected to *Booker*, but, nonetheless, it did not seem to last. Perhaps prosecutors responded to the immediate shock of *Booker* with some degree of panic and hedged their bets against a possible coming wave of Guidelines departures by charging mandatory minimums (in a pattern disparately affecting black defendants). If so, charging may have reverted to normal when prosecutors saw that *Booker* did not cause a major drop in sentences (as we shall see below). This, of course, is only speculation. What we do know is that, despite the significant discontinuity, *Booker*'s longer-term effects on charging look fairly subtle.

We next assess whether the ultimate *sentence* length was discontinuously affected by the *charging* date passing *Booker*—that is, did post-*Booker* changes in charging translate into sentencing consequences? We find only weak evidence on this point. All four specifications estimate that racial disparity in the sentence rose for cases charged immediately after *Booker*, with point estimates varying from four to ten months. However, the estimates are imprecise; three of the four are marginally significant (at the 0.10 level), and the fourth is insignificant. Visually, one can see the reason for the imprecision in Figure 2b: there is considerable noise in the sentence-length data, compared to which the break does not appear particularly clear.[182]

182. Likewise, Panel 1B shows some suggestive, but weak, evidence that the black-white gap in offense levels may have increased in cases charged immediately after *Booker*: the point estimates for the growth in disparity range from 0.3 to 1.1 levels, but these are only even marginally significant in two specifications. If there were an increase in offense-level disparity, it might well be the product of the increase in mandatory minimum charging

Figure 2b.
AVERAGE PRISON SENTENCE IN MONTHS
After Accounting for Defendant and Case Mix



2. Changes in Plea-Bargaining

We now turn to *Booker*'s effects on plea-bargaining, which we assess by examining what happens when the *disposition* date passes *Booker*. Specifically, we assess three outcomes: the conviction mandatory minimum, the final Guidelines offense level, and sentence length. The mandatory minimum and the offense level represent two key subjects of plea negotiations: the charge of conviction and the stipulations of sentencing facts. By assessing the effects of the *conviction* date on the offense level, we can separate out *Booker*'s effects on fact-bargaining from its effects on judicial fact-finding (which will be assessed below). We then turn to the ultimate sentencing consequences of any plea-bargaining changes.

These results can be quickly summarized: nothing dramatic happened, or at least, nothing that can be picked out from the noise of the surrounding data (Table 2, Column 2; Figures 3a-3c). Mandatory minimum rates for white defendants are in general noticeably higher after *Booker* than before it (Figure 3a), but that increase actually occurred several months before *Booker*. Prosecutors, unlike judges, were free to adapt their behavior before the Court ruled, so these changes could have been in anticipation of *Booker*; if so, that

disparity, because the ultimate offense level is affected by any mandatory minimums that apply.

would mean that *Booker* could have increased white mandatory minimums, but too slowly for the RD analysis to detect. *Booker* does not appear to have had any significant discontinuous effects on racial disparity in plea-bargaining or on plea-bargaining outcomes generally.

Figure 3a.
FRACTION OF CASES WITH A STATUTORY MINIMUM
After Accounting for Defendant and Case Mix

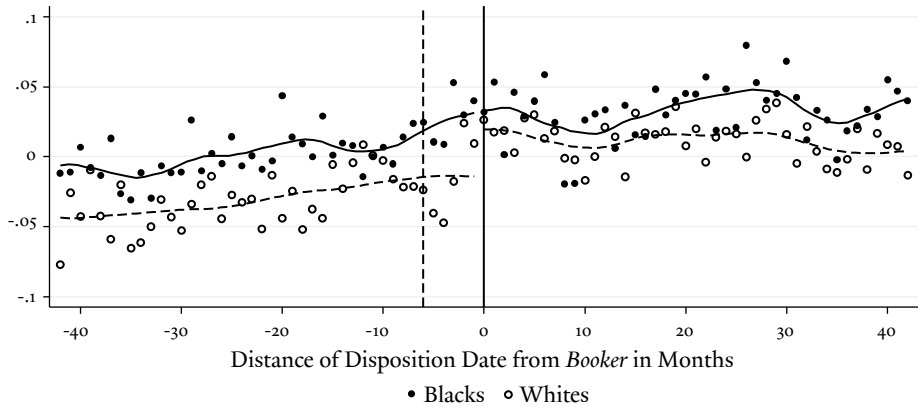


Figure 3b.
AVERAGE OFFENSE LEVEL
After Accounting for Defendant and Case Mix

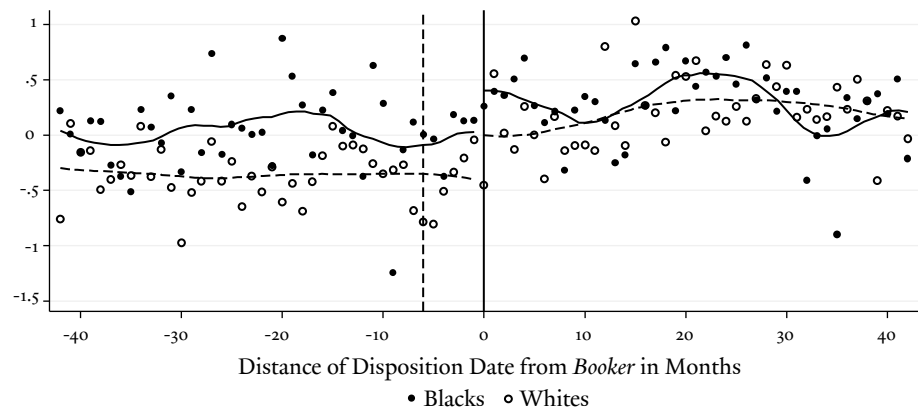
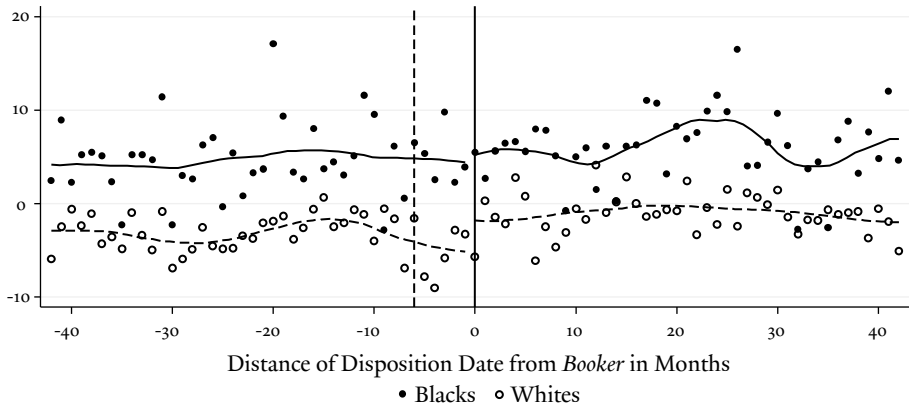


Figure 3c.
AVERAGE PRISON SENTENCE IN MONTHS
After Accounting for Defendant and Case Mix



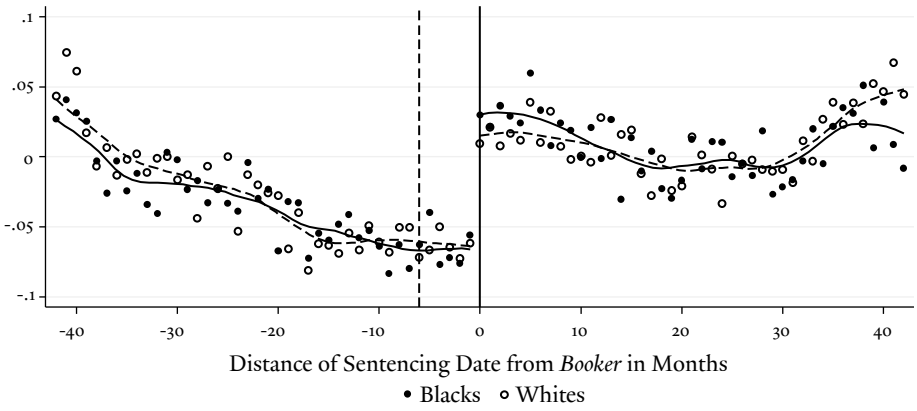
3. Changes in Sentencing Fact-Finding and Sentencing Outcomes

Finally, we assess changes in judicial decision-making by examining what happens when the *sentencing* date passes *Booker*. We focus our analysis on three outcomes: departures, the final Guidelines offense level, and sentence length. *Booker* directly expanded judges' legal authority to depart, and we showed in Figure 1 that this expansion had an immediate effect. In Figure 4a and Panel 3D, we break this effect down by race. We focus here on judicially initiated departures by excluding government-initiated departures for cooperating witnesses in order to examine the use of *judicial* discretion. The patterns are similar if one assesses *all* departures instead.[183] The estimates all show a jump in white departure rates of five to seven percentage points and a slightly larger jump in black departure rates (eight to ten percentage points). If anything, then, black defendants may have benefited more from the increase in departures, but the change in black-white disparity is insignificant in most of the specifications.[184] Notice that in Figure 4a, both the black and the white trends of declining departure rates after *Booker* are identical to the trends before it—but both curves are shifted upward. In other words, *Booker*'s boost to departures occurred immediately, affected black and white defendants quite similarly, and clearly had a lasting effect.

183. If one looks at all departures, there is still no significant rise in disparity.

184. We treat departures as a binary variable here, but one sees similar patterns in departure size.

Figure 4a.
AVERAGE DEPARTURE RATE (NOT GOVERNMENT INITIATED)
After Accounting for Defendant and Case Mix



Booker's legal holding did not directly affect fact-finding, but it could have affected it indirectly (even setting aside any effects on plea negotiations, which our focus on the sentencing date filters out). If a judge believes the sentencing range that follows from the plea agreement is inappropriate, she has two options for altering it: she can make findings of fact that "go behind the plea" or she can depart from the Guidelines.[185] Expanded authority to do the latter might make it less necessary to do the former.[186]

Therefore, in Panel 3B and Figure 4b, we assess whether fact-finding disparities differed in cases sentenced immediately after *Booker*. The results are inconclusive because the estimates are imprecise, but again, if anything, it looks as though changes in judicial decision-making after *Booker* cut in the direction of reducing the black-white gap. The sign of the change in disparity is negative in all four specifications (with point estimates ranging from -0.5 to -1.1 offense levels).[187] The final offense level increases for white defendants in three out of

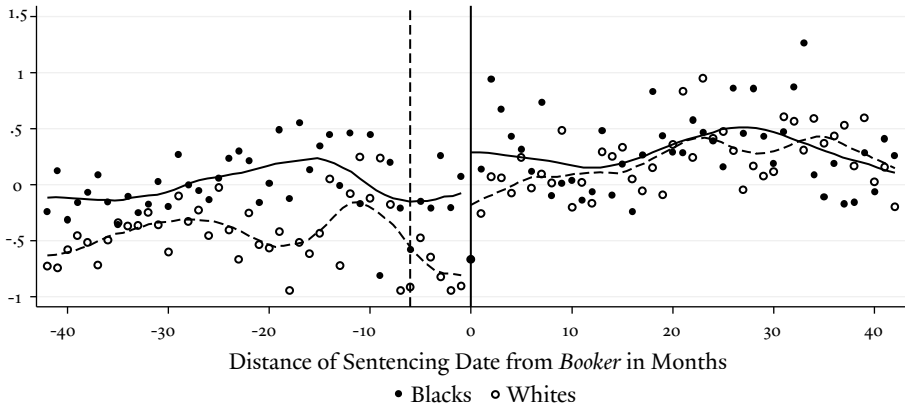
185. *See* Max M. Schanzenbach & Emerson H. Tiller, *Strategic Judging Under the U.S. Sentencing Guidelines: Positive Political Theory and Evidence*, 23 J.L. ECON. & ORG. 24, 28-29 (2006).

186. As discussed above, survey data show that most judges do not diverge from the plea stipulations very often – but that does not mean they never do. The reason they choose to do so in particular cases might be the same reason they might consider departing: dissatisfaction with the sentence that the facts in the plea agreement would produce according to the Guidelines.

187. Although we did not include an additional graph, in Panel 3A, we also show changes in the final mandatory minimum when the sentencing date passes *Booker*. This variable can also be

four specifications, but decreases for black defendants in three out of four specifications. Note that, while Figure 4b shows a fairly clear long-term trend of higher offense levels for white defendants, that increase cannot be safely causally attributed to *Booker* because RD estimates only the local effect right at the discontinuity. We return to the question of assessing long-term trends in Section III.E below.

Figure 4b.
AVERAGE OFFENSE LEVEL
After Accounting for Defendant and Case Mix

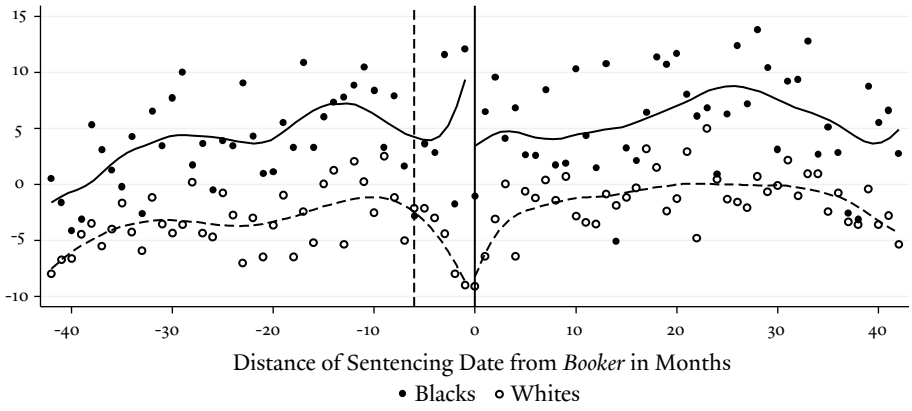


Finally, we look at the effect on sentence length as the sentencing date passes *Booker*—the inquiry that provides the most direct counterpoint to the Sentencing Commission's claims about *Booker*'s effects. As Figure 4c and Panel 3C of Table 2 show, there appears to have been an immediate drop in the length of black defendants' sentences at *Booker*. White sentences did not fall, however, even though white departures increased. Perhaps the increase in

understood to reflect potential changes in fact-finding. The mandatory minimum can be affected by events at the sentencing stage in some contexts, because judicial fact-finding sometimes determines whether mandatory minimums apply. The most important example of this is drug quantity, which is not always stipulated in a guilty plea to a drug trafficking charge. We do not find any significant racially disparate changes in this variable occurring in cases sentenced after *Booker*. If anything, disparity may have declined, but just as with the offense level variable, the change is insignificant in all four specifications. However, there is a statistically significant *overall* increase in the mandatory minimum rate. One theory is that *Booker* led prosecutors to push more aggressively at the sentencing hearing for quantity findings that would trigger mandatory minimums.

departures was offset by the fact-finding changes discussed above.[188] Thus, there is an estimated *reduction* in black-white sentence disparity in cases sentenced just after *Booker* (by between four and fifteen months, depending on the specification). This directly contravenes the conclusion implied by the Sentencing Commission's report. However, the contrary conclusion is only tentative. There is again considerable noise in the sentencing data, and the estimate is only significant in two of the specifications. Still, one can say that these data certainly provide no evidence of an *increase* in sentence disparity at *Booker*.

Figure 4c.
AVERAGE PRISON SENTENCE IN MONTHS
After Accounting for Defendant and Case Mix



Taking Figures 4a through 4c together, one can see that the sustained trend of increasing offense levels seen in Figure 4b may help to explain what otherwise might have been a mystery: why (as Figure 4c shows) sentences did not go down in the long run after *Booker*, even though downward departures went way up and stayed up (Figure 4a). The effect of the departure increase may have been canceled out by the rise in offense levels (for both black and white defendants). The magnitude of the rise in offense levels looks fairly small—perhaps half of one offense level—and one might wonder how such a subtle shift could cancel out such a large increase in departures. The answer is

188. Indeed, if anything, there is a visible upward turn in the white trend at *Booker* (although it does not amount to a discontinuous break), while the black sentence trend stays flat.

that although the increase in departures at *Booker* was a very sharp break in the prior trend, it still only affected a small percentage of cases (about 8%, according to the RD). The average size of a departure from 2005 to 2009 was twenty-nine months, so a back-of-the-envelope calculation suggests that *Booker* brought the average sentence down by only about 2.3 months. An increase of just one-half an offense level, applied to the average case in the sample, would raise the low end of the Guidelines range by two months, enough to cancel out most of that departure effect.

Thus, although *Booker* was the biggest sudden change to federal judges' sentencing discretion since the Guidelines' adoption, it nonetheless was perhaps less of a revolution than various observers either feared or hoped. *Booker* is only what federal judges make of it, and, so far, that appears not to have been much. This post-*Booker* stability should not be taken as especially good news for those concerned about incarceration rates for black men. If *Booker* does not change judicial behavior very much, then it cannot do what critics of the Guidelines hoped: substantially mitigate the Guidelines' harshness. In the long run, sentences have continued to increase, even after controlling for shifts in the pool of offenses and offenders. And with plea levels still over 96%, prosecutors' tremendous leverage appears to remain intact.

E. Limitations and Causal Inference Challenges

Unlike the Sentencing Commission, we find no evidence that *Booker* increased racial disparity in the exercise of judicial discretion; if anything it may have reduced it. The only possibly adverse effects for black defendants that we see arise from prosecutors' shift to mandatory minimums, although that shift appears to have been temporary. Like the results of the charging study discussed in Part II, these findings cut against the case for restoring constraints on judicial discretion. Still, there are some limitations to our method. As we have already discussed, it provides only local estimates of immediate effects, rather than long-term effects. Beyond that, there are a few other things to keep in mind.

1. Limitations of the RD Method

First, it is important to understand what our RD analysis does *not* assess. In the charging study described in Part II, we sought to disentangle the share of the black-white gap that was explained by the disparate impact of factors such as criminal history from unexplained disparities that could represent racially

disparate treatment.[189] Here, in our *Booker* analysis, we only do that in a limited sense. We do control for the arrest offense and the other pre-charge covariates, so in that sense we are measuring changes in (apparently) "unwarranted" disparity. Controlling for those variables means that if the relative composition of the black and white defendant pools (in terms of the observable variables) changed suddenly right around *Booker*—either due to random or seasonal variation in crime or to reaction to *Booker* itself—it should not bias the results.

But the coefficients on those variables—the strength of the relationships between each of them and the outcome variable—are estimated only for the entire time period. While the trends will filter out any smooth (gradual) changes over time, they cannot filter out sharp sudden changes that coincide with *Booker*. We do not separately estimate, for instance, the relationship between criminal history and sentence length before and after *Booker*. If criminal history becomes a stronger predictor of sentence length gradually during the time period, the polynomial trends in our regression would filter that change out. But if the relationship between criminal history and sentence length changes suddenly at *Booker*—if *Booker* changes it—our method will not filter out that change.

In effect, what that means is that we are focused on the question, "Did *Booker* change racial disparity patterns in charging, plea-bargaining, and sentencing?" rather than "*Why* did *Booker* change those patterns?" If, for instance, prosecutors started using mandatory minimums more against black defendants, this need not have been motivated by race—it could have been motivated by wanting to crack down on gun crimes, for instance. In short, we are estimating *Booker*'s racially disparate impacts. We do not filter out the share of those impacts that are mediated by other variables—not just because doing so is impractical with our method but also because it is undesirable. If policymakers care about the effects of sentencing reform on black incarceration rates, filtering out everything that is not racially motivated would not convey those consequences fully. Together, the results of the study described in Part II and our *Booker* results in this Part present a fairly rich picture of the static factors (case features) and dynamic factors (sentencing law reform) that contribute to outcomes at each procedural stage.

Second, while RD effectively filters out long-term trends, it is vulnerable to statistical noise that might generate false positives. If the graph is sufficiently noisy, one might be able to see discontinuities at lots of points. Of course, *Booker* need not have been the *only* shift over the course of the study period to

189. *See supra* notes 84-85 and accompanying text.

be a real shift. But if there are frequent breaks, even at points where there are no known triggering events, then not much can be made of finding a break at *Booker* as well.

We think that with appropriately cautious interpretation, this is not such a serious problem—far less serious than the causal inference problem that pervades other studies. This is why we fit the monthly trends with multiple kinds of functions and do not put stock in an apparent discontinuity that appears only in one version. It is why we do not use even higher-order polynomials, which would likely over-fit the data. It is also why the graphs matter, perhaps more than the numbers. If a discontinuity cannot be picked out with the eye—or if it looks no different from many other unexplained breaks—then it is probably nothing to write home about.

As an additional precaution, we conducted placebo tests on every outcome variable, re-running all of the analyses shown in Table 2, except applied to twelve other arbitrary breaking points across the study period.[190] We deemed the results of these tests "false positives" when, at the breaking point in question, a significant discontinuity appeared in more than one out of the four model specifications. These tests were reasonably reassuring. In the mandatory minimum variable, when the placebo tests were run by charging date or by disposition date, there were false positives at just one out of twelve of these breaking points; when the tests were run by sentencing date, there were no false positives. This makes us more confident that the spike in cases in which mandatory minimum offenses were charged just after *Booker*—although brief—was likely something real, because this variable is not particularly noisy. The sentence length variable was visibly noisier in the graphs and unsurprisingly had more false positives in the placebo tests: just one when the placebos were run by charging month, but four when run by disposition month or by sentence month. In the offense level variable, there were two false positives when the placebos were run by charging month or by disposition month, and one when they were run by sentence month. The departure variable had two

190. These breaking points were in six-month increments every January and July from 2002 through 2008, not including the months immediately following *Blakely* (July 2004), which we analyze separately below, and *Booker* (January 2005). Thus, the tests are also a check against the possibility that what appears to be a discontinuity caused by *Booker* is actually just a regular seasonal variation; if that were true, one would expect similar discontinuities in other Januaries.

false positives (run by sentencing month), but visual inspection makes clear that *Booker* was by far the cleanest break in the study period.[191]

2. Blakely *and Anticipation of* Booker

Finally, we return to the question of *Blakely* and anticipation of *Booker*. *Blakely* is marked with a dotted line in the figures, and we also repeated all the numeric analyses on it. There are no apparent breaks in departures, offense level, or sentence length when the sentencing date passes *Blakely* (Figures 4a-4c). It appears that the courts included in these analyses really did follow the "business as usual rule."

But what about prosecutors? As to plea-bargaining, we find no evidence of discontinuous changes caused by *Blakely* in the "business as usual" circuits. When time trends are estimated based on the disposition month, severity on all measures (mandatory minimum rates, offense levels, and the ultimate sentence) looks relatively low during the first two months after *Blakely*, especially for white defendants (Figures 3a-3c). One might wonder whether this is because *Blakely* increased defendants' plea-bargaining leverage.[192] But a downward trend in these severity measures had already been underway for at least six months before *Blakely*, and the post-*Blakely* months do not represent a significant break from that trend. Moreover, the downward trend turns around again by the third month after *Blakely*. In short, while there are some trend fluctuations when the outcome variables are graphed by disposition date, they do not seem connected to *Blakely* (or *Booker*, as discussed above).

The one thing that does look like it changed discontinuously after *Blakely* is disparity in mandatory minimums, which declined (Figure 2a). The reduction consists mostly of a rise in mandatory minimums for white defendants and is concentrated in drug cases. For cases charged during the whole six-month period between *Blakely* and *Booker*, the black-white gap in mandatory minimums looks quite small, until it jumped in the month of *Booker*. This is

---

191. Note that "false positives" are not necessarily the result of random noise; they could be the result of other influential events that happen to take place around the time of those breaking points.

192. Some observers believed that *Blakely* would be read to mean the prosecutor could not argue aggravating factors at sentencing unless they had been proven to the jury or pled to by the defendant. Two other circuits (outside our sample) had so held. *See* United States v. Ameline, 376 F.3d 967 (9th Cir. 2004); United States v. Booker, 375 F.3d 508 (7th Cir. 2004), *aff'd*, 543 U.S. 220 (2005). Such a rule would be expected to strengthen defendants' plea-bargaining leverage, because a prosecutor would likely often have to offer at least some compromise on aggravating factors in order to avoid a jury trial on them.

another potential reason not to make too much of the spike in charging disparity at *Booker*—in addition to being temporary, it could have been partly the result of the disparity being anomalously small during the *Blakely*-to-*Booker* period. Also, changes in charging disparity around *Blakely* might affect the interpretation of our analysis of changes in plea bargaining or sentencing after *Booker* (since the same cases could be charged near *Blakely* and then either plea-bargained or sentenced near *Booker*).

However, there are two reasons we believe the post-*Blakely* change in mandatory minimum disparity does not pose a serious problem for our interpretation of the *Booker* results. First, the mandatory minimum changes after *Blakely* (unlike those we observed after *Booker*) did not translate into discontinuous changes in *sentence* disparity in cases charged after *Blakely* (Figure 2b). Further analysis suggests that this is because the increase in the presence of mandatory minimums in white defendants' cases was offset by an increase in waivers of those mandatory minimums under the "safety valve" exception that applies in some drug cases.[193] The subset of cases in which mandatory minimum charging patterns changed after *Blakely* were drug cases that were safety-valve eligible, meaning that the mandatory minimum is less consequential than usual.[194]

Second, the time between each key date in a case varies considerably. For instance, it is not as though all of the cases charged in the month after *Blakely* were sentenced in the month right after *Booker*. Rather, from one sentencing month to the next, there is a gradual increase and then a gradual decrease in the probability of the case having been charged right after *Blakely*. In other words, whatever effect changes in charging after *Blakely* had on sentencing should be part of the *continuous* polynomial trends that the RD filters out. To substantially affect the discontinuity estimates at *Booker*, the probability of

193. Under 18 U.S.C. § 3553(f), district courts may waive the mandatory minimum in certain drug trafficking cases involving nonviolent offenders with very little criminal history and no leadership role in a conspiracy, so long as the defendant has given the government a truthful and complete account of the crime.

194. We preferred to use a principal mandatory minimum measure that was determined by the charges of conviction (and findings of fact related to the crime's severity, such as drug quantity), rather than a measure also shaped by the other statutory factors that determine safety-valve eligibility. Our mandatory minimum variable is accordingly based on the sentencing judge's finding that there is an applicable mandatory minimum in the case. In some cases coded as having mandatory minimums, the minimums were ultimately waived under the safety valves or due to cooperation. At *Blakely*, there is no discontinuous change in *non-waived* mandatory minimums for either black or white defendants. In contrast, at *Booker*, there *was* a discontinuous spike in mandatory minimums for black defendants regardless of whether one codes the safety-valve cases as involving mandatory minimums.

having been a post-*Blakely* case would have to have plunged *suddenly* in the month of *Booker*. This is a substantial advantage of RD over other methods.[195]

A final concern about the interregnum period is that some cases could have been delayed until after *Booker*, such that the cases immediately after *Booker* would not have the same characteristics as those immediately before it. Such manipulation could pose a threat to identification using RD. Fortunately, it is not the case that *any* manipulation of timing is fatal to causal inference. As David Lee and Thomas Lemieux explain, "If individuals—even while having some influence—are unable to *precisely* manipulate the assignment variable, a *consequence* of this is that the variation in treatment near the threshold is randomized as though from a randomized experiment."[196] The non-manipulation assumption is thus relatively modest—it only requires that cases sentenced very near *Booker* were not subject to the court's *precise* manipulation of which side of the line they fell on. If a court merely took steps to make it more *likely* that a case would be sentenced after *Booker*, such as scheduling the sentencing hearing for a faraway date, this would not be seriously problematic. The scheduling would have gotten the case near *Booker*, but there would still have been a chance element determining which side it landed on. This chance element is amplified by the fact that nobody knew when the Supreme Court would rule: legal observers performed terribly at predicting *Booker*'s release, with many predicting a very fast decision after the October argument.[197] In addition, sentencing hearings are scheduled months in advance so as to allow the Probation Office time to complete the pre-sentence investigation report and to allow the parties time to prepare (and to arrange for the presence of witnesses in some cases). This delay makes it especially difficult to precisely manipulate the timing of a case relative to a Supreme Court decision.

---

195. This is a principal reason we do not simply use a short-window differences-in-differences approach—for instance, comparing the three months before *Booker* to the three months after. If we had, the *Blakely* effects would have been very different in the pre- and post-periods.

196. Lee & Lemieux, *supra* note 165, at 283.

197. *E.g.*, Ian Weinstein & Nathaniel Z. Marmur, *Federal Sentencing During the Interregnum: Defense Practice as the* Blakely *Dust Settles*, 17 FED. SENT'G REP. 51, 51 (2004) (predicting a decision by Thanksgiving). In the months after *Booker*, the archives of the leading sentencing blog include a series of predictions, citing informed observers in the legal community, that *Booker* would be decided the next day. *E.g.*, Douglas A. Berman, *At Least One More Day to Wait Until* Booker *and* Fanfan, SENT'G. L. & POL'Y (Dec. 7, 2004, 10:15 AM), http://sentencing.typepad.com/sentencing_law_and_policy/2004/12/theyre_here_emb.html ("I have now heard from a large group of insightful folks predicting that tomorrow will (finally) bring the decision. . . . At this point, I will believe it when I see it.").

Still, we analyzed the number and characteristics of cases on either side of *Booker*, looking for any evidence of manipulation. We found none. The number of cases sentenced in December 2004 was 1679; the number in January 2005 was 1682. If sentencings were being delayed, one would expect the mean elapsed time since the plea to be greater for cases after *Booker*, but in fact, the mean elapsed times were nearly identical (indeed, very slightly shorter after *Booker*): 3.99 months before versus 3.96 months after. The breakdowns by race and crime category were likewise essentially identical before and after.[198] If anything, there may have been some delaying of cases in November 2004 when 1566 cases were sentenced, the lowest volume that year. Expectations of an early *Booker* decision were high during November 2004,[199] but the dip was small, and it appears the counts went back to normal once the Court did not release its decision quickly. After *Booker*, the number of cases also stayed normal; it was slightly higher in March (when 1825 cases were sentenced), but this was lower than four other months in 2004 and 2005. In short, there is very good reason to believe that the courts in the circuits included in our analyses really did conduct "business as usual," or at least that any manipulations were too imprecise to threaten RD's assumption of effective randomness in the immediate vicinity of the discontinuity.[200]

CONCLUSION

Determining the causes of racial disparities in criminal justice is not easy. We believe our approach improves substantially on existing research, but we do not offer definitive answers and doubt that anyone will soon. So what are policymakers to do? We do not seek to answer that question completely. Even if we had crystalline empirical answers, criminal justice policy does not turn on

198. Of course, it is theoretically possible that manipulation could have caused the case characteristics to vary only in unobservable ways, but this seems unlikely in practice. If there were substantial manipulation, it seems unlikely that it would not have had *any* effect on the distribution of observable case characteristics like case category and elapsed crime, nor on the case counts.

199. *E.g.*, Douglas A. Berman, *Not Yet for* Booker *and* Fanfan, SENT'G L. & POL'Y (Nov. 9, 2004, 10:20 AM), http://sentencing.typepad.com/sentencing_law_and_policy/2004/11/not_yet_for_emb.html.

200. In any event, manipulation would only bias our results if it occurred in a racially disparate way. And the manipulation concern applies only to our analysis of sentencing responses to *Booker*—there is no reason to believe that any prosecutor would wait to charge or plea-bargain a case until after *Booker*, nor would defendants likely take the large risk of stalling guilty pleas and risking their withdrawal while waiting for a Supreme Court decision.

demographic disparity alone—many competing objectives must be considered. That said, our results have implications for these dilemmas, and we fear that the contrary results of existing research may be distorted to support counterproductive "solutions" to racial disparities. We close with some brief thoughts on these points.

First, despite our concerns about the methods of the Sentencing Commission and others, we agree that the high rate of incarceration of black men is a serious social problem and that examining the possible contribution of disparities in the criminal justice system is important. Our research suggests that, in the federal system, disparities in the post-arrest justice process contribute to this problem. After controlling for the arrest offense, criminal history, and other prior characteristics, sentences for black male arrestees diverge substantially from those of white male arrestees (by around 10% on average). While this disparity does not seem to be growing, it is persistent.

Second, the procedural source of this disparity matters, and it is thus a mistake to focus on judicial sentencing alone. Our research suggests that racial disparities in recent years have been largely driven by the cases in which judges have the *least* sentencing discretion: those with mandatory minimums. Our assessment of *Booker* is more tentative, but we find no evidence that it *increased* racial disparity. The Sentencing Commission's contrary conclusion is based on deeply flawed methods.

For these reasons, we are particularly concerned about proposals to respond to sentencing disparities by restoring tighter constraints on sentencing, especially those that entail expanding mandatory minimums.[201] Our results suggest that this would not reduce disparities in the justice process. Quite the contrary: we find that prosecutors file mandatory minimums twice as often against black men as against comparable white men. Moreover, for those concerned about mass incarceration of black men, expanding mandatory minimums would be counterproductive. Even setting aside racial disparities internal to the criminal justice system, sentencing law changes that increase severity have a particularly adverse impact on black men, who are disproportionately involved in the system in the first place. Making sentencing law more rigid would likely exacerbate this problem even if it led to more equitable administration of the law—and our results suggest that it would likely lead to *less* equitable administration.

Third, we do not advocate attempting to reduce disparity by taking discretion away from prosecutors. Eliminating prosecutorial discretion is probably impossible. The Department of Justice has certainly tried. The

---

201. *See* Gonzales, *supra* note 15.

disparities we found persisted despite the Ashcroft Memo ordering prosecutors to charge and pursue the "most serious, readily provable offense," as well as DOJ bans on fact-bargaining.[202] Taken at their word, these policies would have stripped almost all discretion from line prosecutors. But such policies are very difficult to enforce, because line prosecutors inevitably must subjectively evaluate the available evidence.[203] And even if constraining prosecutorial discretion *did* succeed, one might see another "hydraulic" effect. If prosecutors had to pursue every case law enforcement brought them to the fullest, their current power over case outcomes might shift another step back—to law enforcement, where it might be even harder to monitor. Prosecutors' decision-making is notoriously difficult to observe—unlike judges, they do not publish written reasoning. But law enforcement is even more of a "black box."

Even if *all* discretion could somehow be removed from the justice system, we doubt this would create a justice system anybody would want. Flexibility allows appropriate tailoring of both charges and sentences to the circumstances of individual cases, so as to avoid unduly harsh punishments when they are not justified. Efforts to eliminate unwarranted disparities are important, but they should not come at the cost of unwarranted uniformity. Instead, rather than looking for ways to curtail prosecutorial *discretion*, legislators could consider curtailing prosecutorial *power* by dialing back existing mandatory minimums. If sentencing laws were less rigid, it would be less necessary for decision-makers to find ad hoc means of mitigating their impact. The Fair Sentencing Act of 2010, which reduced crack sentences, showed that it is politically possible to reform excessive sentencing laws, and that empirical evidence of racial disparities can help to bring such changes about.[204]

One potential next focus could be the severe gun enhancements in 18 U.S.C. § 924(c). These laws hit black men particularly hard because, as our data show, they are more frequently arrested for gun crimes and because of large apparent disparities in prosecutors' exercise of charging discretion. Certainly, policymakers must weigh this problem against concerns about gun violence. Notwithstanding these serious concerns, we wonder whether the mandatory minimums in the statute are truly always necessary, such that judicial discretion should be precluded. For instance, is a five-year add-on sentence really necessary in every case in which a firearm has merely been

---

202. Ashcroft, *supra* note 82.

203. *See* Miller, *supra* note 33, at 1257; Julie R. O'Sullivan, *In Defense of the U.S. Sentencing Guidelines' Modified Real-Offense System*, 91 NW. U. L. REV. 1342, 1425-26 (1997); Stith, *supra* note 21, at 1470.

204. *See supra* note 117 and accompanying text.

carried—let alone a mandatory extra twenty-five years for a second gun and yet another twenty-five for a third?[205] Prosecutors would likely feel less need to "swallow a gun" if the gun did not automatically trigger a massive additional penalty.

Finally, while our approach is far more comprehensive than that of prior sentencing studies, there is enormous room for further exploration. For instance, we plan to explore further the possible role of sentencing fact-finding in producing racial disparities. More research is also necessary to see whether patterns like those we found are also present in state courts. More generally, we do not claim to have proven purposeful discrimination by prosecutors or anyone else—it would be impossible to do so with administrative data like ours. Other kinds of studies may be necessary to dig deeper into causal theories for racial disparities: perhaps experimental studies in which race is randomly assigned to otherwise identical prosecutor files, or qualitative studies involving reviews of case files and interviews.[206] DOJ itself is well positioned to carry out such work. One easy step would be for DOJ to keep statistics on mandatory minimum charging decisions by race when it tracks prosecutors' performance. Doing so would not only facilitate research but could also help prosecutors who do not want to contribute to disparities but might not be conscious of them. The government itself should take the elimination of disparities in criminal justice as seriously as other civil rights enforcement matters, and it should think creatively about solutions and strategies for answering the empirical questions that remain.

---

205. 18 U.S.C. § 924(c) (2012).

206. In contexts such as employment and housing, disparity researchers can experimentally manipulate race while leaving other factors identical. *See, e.g.*, Devah Pager, *The Mark of a Criminal Record*, 108 AM. J. SOC. 937 (2003). The federal government uses "testers" (fake applicants) to enforce its discrimination statutes. *See Fair Housing Testing Program*, U.S. DEP'T JUST., http://www.justice.gov/crt/about/hce/housing_testing.php (last visited Sept. 1, 2013). Similar field experiments in criminal justice would generally be illegal: a crime staged for research is still a crime, as is submitting fake information to authorities. But such studies could be legislatively authorized, under regulated conditions, and perhaps carried out by DOJ itself.