A New Approach to Estimating Sentencing Disparity: Bringing Prosecutors Into the Picture

Sonja B. Starr* & M. Marit Rehavi**

August 31, 2012

* Assistant Professor, University of Michigan Law School.

** Assistant Professor, Department of Economics, University of British Columbia. The research underlying this Article has been presented at the Ninth Circuit Judicial Conference, the National Bureau of Economic Research Summer Institute, the American Law and Economics Association Annual Meeting, the American Society of Criminology Annual Meeting, the Duke Legal Theory Colloquium, the Loyola Law School Faculty Workshop, the UBC Empirical Lunch, the CIFAR-IOG Workshop, and the Fawley, Law and Economics, and Criminal Justice workshops at the University of Michigan Law School. Thanks to participants for their feedback, and thanks to Daron Acemoglu, David Abrams, 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, Sam Gross, Don Herzog, Jim Hines, Jill Horwitz, Thomas Lemieux, Justin McCrary, Julian Mortenson, Brendan Nyhan, JJ Prescott, Eve Brensike Primus, Adam Pritchard, Craig Riddell, Jeff Smith, and Sara Sun-Beale for helpful comments and conversations. Sharon Brett, Michael Chi, Ryan Gersowitz, Michael Farrell, Seth Kingery, Matthew Lee, Midas Panikkar, Art Robiso, Sabrina Speianu, and Adam Teitelbaum provided able research assistance

A New Approach to Estimating Sentencing Disparity: Bringing Prosecutors Into the Picture

In the United States, one of every nine black men under the age of 35 is currently behind bars,[1] and of today's young black men, a predicted one in three will be incarcerated at some point in his life.[2] These rates far exceed those of any other demographic group—for instance, black males of all ages are incarcerated at nearly seven times the rate of white males.[3] These figures are stark, and 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 are there disparities in the way the criminal justice system treats offenders? If it's the latter, can the process be made more equitable through 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*, 543 U.S. 220 (2005), rendered the formerly mandatory Guidelines merely advisory, Justice Stevens' dissent predicted that the "result is certain to be a return to the same type of sentencing disparities Congress sought to eliminate in 1984." 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: that *Booker* and its judicial progeny had quadrupled the black-white sentencing gap among otherwise-similar cases, from 5.5% to 23%.

This Article introduces a new empirical approach, and gives a very different answer. In our view, the Commission's methods are hobbled by two serious flaws that, unfortunately, pervade the rest of the empirical literature on sentencing disparity as well.[5] First, these studies consider the judge's final sentencing decision in isolation, ignoring crucial earlier stages of the justice process—and if there are disparities at those stages, the sentencing disparity estimates will be misleading. The second problem is specific to studies of *changes* in disparity after legal changes

---

[1] PEW CTR. ON THE STATES, ONE IN 100: BEHIND BARS IN AMERICA 2008, at 3 (2008), *available at* http://www.pewstates.org/research/reports/one-in-100-85899374411.

[2] THOMAS BONCZAR, U.S. DEP'T OF JUSTICE, BUREAU OF JUSTICE STATISTICS, PREVALENCE OF IMPRISONMENT IN THE U.S. POPULATION, 1974-2001 (2003).

[3] *See* HEATHER C. WEST, U.S. DEP'T OF JUSTICE, BUREAU OF JUSTICE STATISTICS, PRISON INMATES AT MIDYEAR 2009—STATISTICAL TABLES, 21 tbl.18 (2010). These figures are from 2009 and exclude Hispanics, for whom the rates are between those of whites and blacks. *See id.*

[4] *See, e.g.*, MICHELLE ALEXANDER, THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS (2011); 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] While we focus on race, these weaknesses are also common in research on other "unwarranted disparities" (e.g., inter-district) and much of the non-disparity-related research on sentencing policy.

(like *Booker*): these studies do not effectively disentangle the effects of the legal change from surrounding background trends.

This Article develops both these critiques and presents our own research on racial disparities among federal arrestees, using a method that avoids these problems. We find a substantial black-white gap that appears to be introduced during the criminal process—but it appears to stem mainly from prosecutors' charging choices, especially the use of "mandatory minimum" offenses. We then turn to *Booker*, estimating its effects not only on sentencing, but also on charging, plea-bargaining, and sentencing fact-finding, which no prior studies have assessed. We find no evidence that judges' use of expanded discretion worsens disparity, and suggestive evidence cutting the opposite direction.[6]

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.[7]

Yet the empirical research on sentencing disparity has not tested these claims, and in general fails to account for the role of prosecutorial discretion. In federal courts and other jurisdictions with sentencing guidelines, researchers typically estimate sentencing disparities after controlling for (among other things) the recommended Guidelines sentence. The problem is that 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 and misleading view of the scope and procedural sources of sentencing disparity.

We take a different approach, constructing a dataset that links records from four different federal agencies, allowing us to trace criminal cases from arrest through sentencing. We focus on the gap between black men and white men. 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. This

---

[6] The *Booker* results are new to this Article. We begin, however, by discussing the results of a study of racial disparities in charging and sentencing (which was not focused on *Booker*) that we prepared in for an economics journal: M. Marit Rehavi & Sonja B. Starr, *Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences* (Univ. of Mich. Law & Econ. Empirical Legal Studies Ctr., Paper No. 12-002, 2012) (under review), *available at* http://ssrn.com/abstract=1985377. Here, we explore that study's implications for debates in sentencing policy and legal scholarship.

[7] *See* Lauren O'Neill Shermer & Brian Johnson, *Criminal Prosecutions: Examining Prosecutorial Discretion and Charge Reductions in U.S. Federal District Courts*, 27 JUST. Q. 394, 395 (2010); T. D. Miethe, *Charging and Plea-bargaining Practices under Determinate Sentencing: An Investigation of the Hydraulic Displacement of Discretion*, 78 J. CRIM. L. & CRIMINOLOGY 155 (1987).

method allows us to assess aggregate disparities introduced throughout the post-arrest justice process, from charging through sentencing. Using decomposition methods developed by labor economists for the assessment of wage and income inequality, we can then assess the contribution of each procedural stage (and other underlying case differences) to the total black-white gap.

Our results suggest that the traditional method used by the existing literature risks leading policymakers seriously astray. We begin by analyzing post-*Booker* cases only. After controlling for the arrest offense, criminal history, and other prior characteristics, we find a black-white sentence-length gap of about 10%.[8] 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.

We then turn to *Booker*, using a method that takes prosecutors into account and also corrects another serious flaw that pervades the empirical literature on sentencing-law changes. This second problem is exemplified by the Sentencing Commission's findings. It found that disparities after *Booker* (averaged over a period of years) were larger than disparities before it. But even if so, it is a huge logical leap to conclude that *Booker caused* this increase—a classic confusion of correlation and causation. Many things change over time, and the greater disparity in the post-period 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 make it look like there was greater disparity after *Booker*.

A useful tool for disentangling those underlying trends is called regression discontinuity design. We use it to 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 them—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* changed racial disparity patterns substantially, 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. If judges were inclined to use discretion in ways that widen the black-white gap, one would expect to see disparity jump right after *Booker*, when the departure spike happened.

We do not see such a jump. Right after *Booker*, sentencing disparity did not increase, and even may have modestly dropped. If *Booker* did have any adverse effects on black defendants relative to whites, it was probably a second-order result

[8] This is in non-drug cases; our main sample for the race study excludes drug cases due to data limitations discussed below. When drug cases are added, the gap rises to about 14%.

of *charging* changes. In cases *charged* immediately after *Booker*, there were more mandatory minimums and average sentences increased—but only for black defendants. These findings suggest that prosecutors may have reacted to the Guidelines' weakening by seeking to constrain judges with mandatory minimums instead. However, this effect appears to have been temporary. Graphs of longer-term patterns suggest that mandatory minimum patterns may have soon reverted to their prior trend: a gradual increase for both blacks and whites.

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,"[9] 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 could be tricked by month-to-month fluctuations. And *Booker* was not a clean break in settled law. It came on the heels of a period of serious lower-court confusion, and some of its effects may have stemmed from ending this anomalous period, rather than from ushering in the advisory-guideline era. We conduct tests to examine these threats to causal inference, 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. And indeed, even apart from the challenge of filtering out long-term trends, there does not appear to be any evidence that *aggregate* racial disparity in the post-arrest justice process is actually larger in the post-*Booker* period. Unexplained racial disparities in that process persisted throughout the study period (which runs through 2009), but they actually were slightly smaller by the end of that period than they were just before *Booker*. We do not seek to explain the causes of that change, which is minor in any event; it is unclear whether *Booker* had anything to do with it. But it does appear that the Sentencing Commission's more alarming estimates were an artifact of its methods and its choice to focus on the final slice of the justice process in isolation.

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 Bush Administration advocated mandatory "topless guidelines"—effectively, mandatory minimums but no maximums.[10] The Sentencing Commission this year advanced a multi-prong proposal to strengthen legislative and appellate-court constraints on judicial sentencing discretion.[11] Such "solutions" could be counterproductive. Constraints on judges generally empower prosecutors

---

[9] 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 (2007).

[10] *See* Alberto Gonzales, Prepared Remarks of the Attorney General: Sentencing Guidelines (June 21, 2005), http://www.justice.gov/archive/ag/speeches/2005/06212005victimsofcrime.htm.

[11] *See generally* Public Hearing Transcript, U.S. Sentencing Commission (Feb. 16, 2012), *available at* http://www.ussc.gov/Legislative_and_Public_Affairs/Public_Hearings_and_Meetings/20120215-16/Hearing_Transcript_20120216.pdf. The remarks of Patti Saris, *id.* at 8-9, outline the proposal.

by making their choices more conclusive determinants of the sentence. Our research suggests that mandatory minimums—prosecutors' most powerful tools for constraining judges—are particularly important sources of disparity. 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 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 and the results from our alternative method. We discuss our method's limitations, including sources of disparity that even our broader approach does not measure, and discuss the way race and gender disparities combine to affect black men, drawing on Starr's recent findings in a related gender study.[12] In Part III, we turn to *Booker*'s effects on racial disparity, beginning with our critique of the causal inference methods used by existing sentencing-reform research and proceeding to our regression discontinuity analyses. We conclude with thoughts on the policy implications of this research.

I. Prosecutors, Sentencing, and the "Hydraulic Discretion" Theory

Because we study federal cases, we begin with some background on the federal criminal process. Like state prosecutors, federal prosecutors have always possessed very broad discretion. Prosecutors choose what charges to bring (if any at all), and the complex criminal code often provides a wide range of choices. After that, over 95% of cases result in guilty pleas, and prosecutors control the terms of the plea deals they offer defendants. These can include the charges of conviction ("charge bargaining"), sentence recommendations ("sentence bargaining"), stipulations about sentencing-relevant facts ("fact bargaining"), and prosecutorial requests for "departures" from the usual sentencing range.

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* sentencing requirements were not widespread before the 1980s, and still apply in only a minority of cases (although common for some case types, particularly drug cases). 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 and circumstances 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. Under the Guidelines, sentencing is governed by a complex set of rules determining the "offense level," which is based on the conviction offense plus additional

[12] Sonja B. Starr, Estimating Gender Disparities in Federal Criminal Cases (Univ. of Mich. Law & Econ. Empirical Legal Studies Ctr., Paper No. 12-___, 2012) (under review).

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: "8 to 14 months," for instance. 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 discretion."[13] But under the Guidelines, the outcome of plea-bargaining much more tightly constrained the sentence.[14] The one key 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) allows the judge to base a sentence even on uncharged conduct, so long as it falls within the *statutory* range for the conduct that *is* charged. 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.[15] Judges are not bound by the factual stipulations, and the power to diverge from them (relying on sentencing-stage evidence or on a Probation Office report) is an important aspect of judicial discretion. Still, judges typically lack the incentive, and may lack the information, to diverge from what the parties have agreed.[16] 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.[17] As Nancy King put it, "[e]stablishing facts in an adversarial system without the assistance of adversaries is an awkward business."[18]

---

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

[14] *Id.*

[15] Stephen J. Schulhofer & Ilene H. Nagel, *A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices Under the Federal Sentencing Guidelines*, 66 S. CAL. L. REV. 501 (1992); Stith, *supra* note 15, at 1450.; Nancy King, *Judicial Oversight of Negotiated Sentences in a World of Bargained Punishment*, 58 STAN. L. REV. 293, 295-98 (2005); 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 (2006)

[16] Stith, *supra* note 15, at 1449; Stephen J. Schulhofer & Ilene H. Nagel, *Plea Negotiations Under the Federal Sentencing Guidelines*, 91 NW. U. L. REV. 1284, 1300 (1997); 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 (1995); Brown & Bunnell, *supra* note 17, at 1065.

[17] MOLLY TREADWAY JOHNSON & SCOTT A. GILBERT, FED. JUDICIAL CTR., THE U.S. SENTENCING GUIDELINES: RESULTS OF THE FEDERAL JUDICIAL CENTER'S 1996 SURVEY 10 (1997).

[18] King, *supra* note 17, at 303.

To the Guidelines' many critics, this empowerment of prosecutors was a disastrous flaw, not only leading to harsh results for defendants generally, but also undermining the Sentencing Reform Act's disparity-reduction goals. As Albert Alschuler put it, "the price of whatever success the Guidelines have achieved in reducing judge-created sentencing disparities has been the burgeoning of prosecutor-created disparities."[19] Metaphorically, scholars often refer to discretion 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.[20]

This theory has long been pervasive in theoretical scholarship about the Sentencing Guidelines. As Terence Miethe wrote in 1987, "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."[21]

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 way that they chose to exercise their discretion more conclusively determined the sentence.[22] In a 1996 survey, 86% of judges said that the Guidelines gave prosecutors too much power, and 75% said that prosecutors were now the actor with the *most* influence on sentences—more than judges themselves.[23] 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 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.[24] After adoption of the Guidelines, plea rates went way up in the federal system, from an already-high 87% to 97% by the new millennium.[25]

Since then, however, federal sentencing law has undergone another major change. In January 2005, the Supreme Court decided *United States v. Booker*, which

---

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

[20] Stephanos Bibas, *The Feeney Amendment and the Continuing Rise of Prosecutorial Power to Plea Bargain*, 94 J. CRIM. L. & CRIMINOLOGY 295 (2004); *see* Stith, *supra* note 15.

[21] Miethe, *supra* note 7, at 155-56.

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

[23] JOHNSON & GILBERT, *supra* note 19, at 6-7.

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

[25] Alschuler, *supra* note 21, at 112; *accord* Stith, *supra* note 15, at 1425.

rendered the formerly-mandatory Guidelines merely advisory.[26] The Court held that a mandatory sentencing scheme in which a defendant's maximum sentence could be increased based on judicial findings of fact violated the Sixth Amendment right to a jury trial. The Court could have remedied that defect by requiring more jury fact-finding, but that is not the remedy it chose. Instead, the Court chose to maintain "real offense" sentencing, but to sever the provision of the Sentencing Reform Act that rendered the Guidelines mandatory.[27] The Court's remedial choice remains reversible by Congress, which has so far not taken action.[28] District courts today are free to depart from the Guidelines so long as the ultimate sentence is "reasonable." In a pair of decisions in December 2007, *Gall v. United States* and *Kimbrough v. United States*, the Supreme Court further clarified that courts of appeal should not deem sentences unreasonable merely because they fall far outside the Guidelines,[29] and that judges may depart from the Guidelines on the basis of policy disagreements with them.[30]

Booker was widely seen as an earthquake in federal sentencing law. Note, however, that while *Booker* certainly made a crucial change to the sentencing framework, rendering the Guidelines advisory is not the same as eliminating them. Federal judges are still required to calculate the Guidelines sentence, and although they are then free to depart from it, they usually do not.[31] There are many possible reasons for this continued conformity: federal judges might believe in the goal of reducing disparity,[32] or might be uncomfortable with open-ended, subjective sentencing assessments, or might believe the Guidelines provide insulation from criticism or reversal, or might be treating the Guidelines as an "anchor."[33]

To the extent judges remain compliant with 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* constraints on the

---

[26] Booker, 543 U.S. at 245-46.
[27] *Id.* at 246.
[28] *Id.* at 265 (noting that the "ball now lies in Congress' court.").
[29] Gall v. United States, 552 U.S. 38, 51 (2007).
[30] Kimbrough v. United States, 552 U.S. 85, 101-02, 108 (2007). *See also* Rita v. United States, 551 U.S. 338, 354-55 (2007) (barring appeals courts from treating outside-Guidelines sentences as presumptively unreasonable).
[31] Most federal sentences are still in the Guidelines range. *See, e.g.*, U.S. SENTENCING COMM'N, DEMOGRAPHIC DIFFERENCES IN FEDERAL SENTENCING PRACTICES: AN UPDATE OF THE BOOKER REPORT'S MULTIVARIATE REGRESSION ANALYSIS (2010). Indeed, in our sample, a full 37% of post-*Booker* sentences are exactly at the bottom end of that range.
[32] *See* Alschuler, *supra* note 21, at 97 (noting that even before the Guidelines existed, courthouse norms powerfully influenced sentences).
[33] 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* James J. Prescott & Sonja Starr, *Improving Criminal Jury Decision Making After the* Blakely *Revolution*, 2006 U. ILL. L. REV. 301, 326 (2006) (reviewing the literature).

sentence, which remain mandatory. These constraints are particularly important when prosecutors bring mandatory minimum charges. Second, because they negotiate the factual stipulations accompanying pleas, 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 one cannot assess the likely impact of sentencing law reforms without understanding the tradeoffs between judicial and prosecutorial power. One might expect that this broad consensus would shape empirical research on sentencing disparities and sentencing reforms, but, oddly, it has not, as we demonstrate below.

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 the judge's final sentencing decision while ignoring disparities in the rest of the justice process. In Section A, we review those studies and explain why this problem is so serious. In Section B, we describe the dataset that we constructed to enable broader study of the post-arrest justice process. In Section C, we summarize the results of our alternative strategy. In Section D, we discuss some limitations and implications of our results, and connect them to Starr's related research on gender. Note that in this Part, we focus on disentangling the reasons for the gap between black and white sentences. We do this using the most recent available data (all post-*Booker*), so we do not focus on changes over time. In Part III, we turn to the question of how to assess *changes* in disparity patterns after *Booker*.

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 higher sentences on average than white males do. This is one of the main reasons black males are incarcerated at far higher rates; the other is that black men are much more likely to be criminal defendants in the first place. But is the sentence gap because black defendants have committed more serious crimes, or have more extensive criminal histories? Or is it because they are treated differently in the criminal justice process?

Mass incarceration of black males has serious social consequences regardless of the causes. But if different offending patterns are to blame, the problem might be better addressed with policies focused on addressing crime and its underlying causes, like 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.[34] There are variations on this approach,[35] but all of them estimate differences in the actual sentence relative to what the sentence "should have been" under the Guidelines. Studies in systems without guidelines similarly control for some measure of the severity of the offense of conviction.[36] The problem with these approaches is that the key control variable is only a fairly distant proxy for the seriousness of the underlying conduct. It is the end product of the sequence of discretionary decisions and negotiations described above: charging, plea-bargaining, and sentencing fact-finding. And disparities introduced during those stages may well carry through the process to produce sentencing gaps.

Other scholars have noted this problem.[37] This includes, to their credit, some of those who employ the presumptive sentence approach themselves, who note that their accounts of disparities are incomplete.[38] 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. Absent an account of disparity at the earlier stages of the process, it is difficult to interpret disparities found in the final stage.

For instance, consider the most prominent recent sentencing-disparity study, the Sentencing Commission's post-*Booker* report. The report finds that in the most recent period studied (2008 and 2009), blacks received 23% longer sentences than whites, controlling for the recommended Guidelines sentence.[39] This was an explosive finding, and led to calls (including by the Commission itself) to reinstate stronger constraints on judicial discretion.[40] But how should this result be

---

[34] *E.g.*, U.S. SENTENCING COMMISSION, *supra* note 33; 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 (2011) (following this approach and also reviewing prior literature doing the same).

[35] *See, e.g.*, David B. Mustard, *Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts*, 44 J.L. & ECON. 285 (2001) (including separate dummies for each Guidelines grid cell); Max M. Schanzenbach, *Racial and Gender Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics*, 34 J. LEGAL STUD. 57 (2005) (same); 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 (2008) (treating departures as the outcome variable).

[36] *E.g.*, Darrell Steffensmeier et al., *Gender and Imprisonment Decisions*, 31 CRIMINOLOGY 411 (1993).

[37] *See* Alschuler, *supra* note 21, at 86-87; Engen, *supra* note 24, at 324-28; Shermer & Johnson, *supra* note 7, at 395.

[38] *See, e.g.*, Ulmer et al., *supra* note 36, at 1107-08.

[39] U.S. SENTENCING COMMISSION, *supra* note 33, at 3.

[40] For instance, the former Chair of the Sentencing Commission, Judge William K. Sessions, has warned, citing the Commission's disparity report, that post-*Booker* disparities were likely to lead Congress to adopt more mandatory minimums; Sessions himself proposes a simplified mandatory Guidelines system instead. *At the Crossroads of the Three Branches: The U.S. Sentencing Commission's*

interpreted? Consider just three of many possibilities concerning what might have happened *earlier* in the justice process:

> A. Prosecutors charge whites more harshly and/or offer them worse plea deals, such that the resulting Guidelines recommendation averages 23% higher for whites than for blacks with similar offenses and criminal histories.
>
> B. Prosecutors charge whites more harshly and/or offer them worse plea deals, such that the resulting Guidelines recommendation averages 30% higher for whites than for blacks with similar offenses and criminal histories.
>
> C. Prosecutors charge *blacks* more harshly and/or offer them worse plea deals t, such that the resulting Guidelines recommendation averages 30% higher for blacks than for whites with similar offenses and criminal histories.

Under Scenario A, what looked like a 23% 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" whites *enough*—to sentence based on true culpability, they would have to do more to compensate for prosecutors' favoring blacks. In contrast, under Scenario C, judges are compounding the underlying charging and plea-bargaining disparities; the "true" sentencing disparity is actually much more than 23%. If you don't know which of the scenarios (or others) is true, it is risky to use the 23% figure as a guide to policy.

A recent study by Joshua Fischman and Max Schanzenbach recognizes the problem with the presumptive sentence approach.[41] They instead control for the Guidelines "base offense level"—which, unfortunately, does not really solve the problem.[42] The base offense level is determined 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, crucial determinations such as drug quantity in a drug trafficking case, or, in an assault case, the degree of physical contact and injury, the defendant's intent, and the use of weapons.[43] Disparities in any of those factual determinations, or in the prior charging or plea-bargaining processes, could bias the resulting sentencing-disparity estimates. To avoid this problem, one needs a measure of case severity that precedes those discretionary processes.

The problem with the presumptive sentence control is compounded by a distinct source of potential bias that the existing literature has overwhelmingly failed even to recognize: sample selection affecting the composition of the pool of sentenced cases. Nearly every study of sentencing disparity is confined to a sample

---

Attempts to Achieve Sentencing Reform in the Midst of Inter-Branch Power Struggles, 26 J.L. POL. 305 (2011). The Commission's recent hearings on strengthening the Guidelines centered on the disparity concerns raised by the Commission's report. *See supra* note 11 and accompanying text.

[41] Joshua B. Fischman & Max Schanzenbach, *Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, 9 J. EMP. LEGAL STUD. __ (forthcoming Dec. 2012).

[42] *Id.* at tbl. 5 (listing controls).

[43] U.S. SENTENCING GUIDELINES MANUAL §§ 2A2.3, 2A2.4, 2D1.1 (2011).

consisting of sentenced defendants only—in federal court studies, typically only those sentenced for felonies or Class A misdemeanors ("non-petty offenses"), because that is who the Sentencing Commission collects data on. To make it into the sentencing dataset, cases 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, blacks 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 blacks and whites who get sentenced to be unobservably different: blacks' cases are less serious in a way that controlling for observable variables cannot capture. Sentencing disparity estimates within that sample will be biased because they cannot account for this unobserved difference. Again, without assessing the "funnel," one cannot know whether to expect such a bias to exist, or if so, which direction it should 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 relative severity of initial and final charges.[44] More importantly, few studies (and no federal studies) have assessed disparities in *initial* charging, even though without doing so, it is difficult to give a meaningful interpretation to charge-bargaining results.[45] A few state-level studies have found racial disparities in the use of certain particularly harsh mandatory minimums, including a study of "habitual offender" charges in Florida[46] and Pennsylvania[47] and a Maryland study of add-on mandatory minimums for firearms.[48]

At the federal level, many observers, including the U.S. Sentencing Commission, have pointed to racial gaps in the rate of mandatory minimum convictions.[49] Fischman and Schanzenbach's study provides useful new evidence that mandatory minimums may be an important contributor to sentencing disparities.[50] But these studies raise important further questions. Because they do not control for underlying pre-charge case features affecting a defendant's eligibility

---

[44] *See* Shermer & Johnson, *supra* note 7, at 399 (reviewing the charge-bargaining literature, which mostly finds disparity favoring whites, and presenting their own findings showing no such disparity).
[45] One early study by Spohn et al. found disparities favoring whites in the rate of filing felony charges in Los Angeles County, but did not analyze charge severity within felony charges. *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 note 7 and accompanying text (discussing another charging study).
[46] Charles Crawford et al., *Race, Racial Threat, and Sentencing of Habitual Offenders*, 36 CRIMINOLOGY 481 (1998).
[47] Jeffrey T. Ulmer et al., *Prosecutorial Discretion and the Application of Mandatory Minimum Sentences*, 44 J. RES. CRIME & DELINQ. 427 (2007).
[48] Jill Farrell, *Mandatory Minimum Firearm Penalties: A Source of Sentencing Disparity?*, 5 JUST. RES. & POL'Y 95 (2003).
[49] U.S. SENTENCING COMMISSION, *supra* note 33.
[50] *See supra* note 41.

for mandatory minimums (such as the arrest offense), they do not examine the *reasons* for the mandatory minimum gap. That is, we do not know whether it is just that black defendants have more frequently committed crimes to which mandatory minimums apply, or whether there are racial disparities in prosecutors' exercise of charging discretion.[51]

We do not suggest that sentencing studies following the conventional approach provide no useful information. The final sentencing stage is an important part of the process. Findings that point to disparity in this stage are at least suggestive and should trigger further investigation. Ultimately, though, the objective of reducing disparity should extend well beyond the narrow aim of equalizing sentences for those found in the same Guidelines cell.

The specification of an empirical model of disparity may seem like a purely scientific decision. But as Albert Alschuler has observed, it is in fact bound up with normative questions: what kinds of disparity do we think are important?[52] The choice of control variables determines what kind of disparity one is measuring, and so it should be shaped by a sense of what kinds of disparity policymakers care about (or should care about). There are many reasons one might worry about demographic disparities in the justice process. Such disparities might violate the constitutional principle barring discrimination on the basis of forbidden classifications, exacerbate the social consequences of mass incarceration within particular communities, interfere with retributive or utilitarian punishment objectives, or undermine the credibility of the justice system.

We do not intend in this Article to attempt to resolve what policymakers' objectives in shaping sentencing law 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 Guidelines departure rates. Rather, all of them imply that policymakers should care about whether people with the *same underlying criminal conduct* (including the current offense and, at least arguably, prior criminal history) receive the same sentence. Between the underlying criminal conduct and the Guidelines recommendation, there are many places in the process where disparities could be introduced. Policymakers should care about all of them.

---

[51] One study in 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?, J. Legal Studies (forthcoming). This 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 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 judges' or prosecutors' race and gender and those of defendants also provide interesting insights, but do not 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 34.

[52] Alschuler, *supra* note 21, at 85-88.

A final disadvantage to the "presumptive sentence" approach is simpler: it does not control for differences in crime type, only for differences in crime severity according to the Guidelines. 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 *also* include controls for case type in addition to the presumptive sentence, but only for broad categories like "drugs" or "violent crime," which do not capture much nuance.[53]

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 disparate impact versus disparate treatment distinction is crucial as a matter of constitutional law,[54] although the extent to which it is normatively important is open to debate.[55] We think *all* factors contributing to racial disparity in sentencing—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 generally purport to estimate "legally unwarranted" disparities, and if that is one's objective, one should filter out legally relevant factors like case type.

B. Our Dataset

We take a broader, process-wide approach to the estimation of racial disparities. Doing so 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 of the U.S. Attorneys (EOUSA), the Administrative Office of the U.S. Courts (AOUSC), and the U.S. Sentencing Commission (USSC).[56] The breadth of the dataset allows us to estimate disparities at each post-arrest stage of the process as well as the sentencing consequences of those disparities. It covers two stages of the process that the Sentencing Commission data alone (the sole source for most federal studies) do not include.

---

[53] *E.g.*, U.S. SENTENCING COMMISSION, *supra* note 33, at B1-B2 (using seven categories); Ulmer et al., *supra* note 36, at 1100.

[54] Washington v. Davis, 426 U.S. 229 (1976).

[55] Many critics have argued that the Supreme Court's focus on discriminatory purpose is overly formalistic, instead advocating weighing justifications for policies against the harm they impose on subordinated groups. *See* Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. MIAMI L. REV. 8, 8-10 (2004) (reviewing this literature). We are sympathetic to this view, but this longstanding debate need not be resolved for our purposes; empirical differentiation of the reasons for disparities has practical uses regardless.

[56] 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 6, Data Appendix.

First, it includes the arrest offense, coded with 430 codes, and a string field describing the offense based on the arresting officer's notes. This information allows us to substituting the arrest offense, instead of the presumptive sentence, as the key case-severity control. Doing so 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 after a string of discretionary decisions. That is, we 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.

Second, our dataset includes rich information on *initial* charges, in addition to final charges. Specifically, we know the statutory section(s) under which the defendant was charged and convicted—for instance, 18 U.S.C. § 924(c). In order to assess charges quantitatively, we had to translate 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.[57] 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.[58]

Sometimes, the statutory provisions in the data contained multiple sentencing schemes depending on the facts of the case. Where possible, we resolved such ambiguities based on the *other* charges in the case,[59] or otherwise used reasonable assumptions about which subparagraphs were likely to apply in most cases. However, in drug cases, the ambiguities were extreme—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 could not assess initial charging disparities in drug cases. We subjected drug cases to more limited analysis focusing on disparities in mandatory minimum *convictions*, which are separately recorded in the data. We excluded child pornography cases because of a similar ambiguity. We also excluded immigration cases for different reasons: their stakes typically turn on deportation, and they

[57] This measure 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 a "charge only" measure, allowing the effects of subsequent Guidelines sentencing fact-finding to be separated out.

[58] Per the Guidelines, we assumed concurrent sentencing unless a statute specified otherwise.

[59] For instance, if Charge 1 contained a heightened penalty if a gun was involved and Charge 2 was a gun charge, we applied the heightened penalty for Charge 1. Implementing this system required extensive legal research. In addition to coding all the triggering conditions for statutory or Guidelines 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.

involve different "fast-track" procedural environments.[60] What remained were property and fraud offenses, regulatory offenses, violent crimes, and weapons cases.

We focus on the race gap between black and white U.S. citizen males. Starr's separate gender study, discussed below, also assesses the race gap among women. Other racial groups' outcomes were not analyzed because their numbers were very small. Hispanic defendants are included among the black and white defendants.[61]

B. Our Estimates of Racial Disparities in Charging and Sentencing

We assess disparities introduced throughout the post-arrest justice process and break those aggregate disparities down by procedural source. 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. During our study period, dropping charges required a supervisor's special approval.[62] In *initial* charging, however, the line prosecutor has considerable discretion.[63] In addition, before one could even begin to make sense of plea-bargaining disparities, one has to first know whether the baseline charges already reflected disparities.

The statistical analysis and the resulting estimates are described in detail in Rehavi and Starr (2012), an economics paper; accordingly, we do not re-present the tables or figures here. Instead, we summarize some key findings and explore their legal and policy implications.[64] 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?

---

[60] The Sentencing Commission's *Booker* study includes immigration, but we agree with other scholars who have argued that it should be considered separately. *See* Ulmer et al., *supra* note 36, at 1085.

[61] 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 Hispanics fall somewhere between those of blacks and non-Hispanic whites (as the Sentencing Commission, *supra* note 33, suggests), then our disparity estimates will be somewhat smaller than one would find if one looked only at blacks versus non-Hispanic whites.

[62] Memorandum from Att'y Gen. John D. Ashcroft on Department Policy Concerning Charging Offenses, Disposition of Charges, and Sentencings (Sept. 22, 2003).

[63] DOJ also attempted to constrain charging discretion, *id.*, but this is a weaker constraint in practice, as we explain, *infra* note 129 and accompanying text.

[64] *See* Rehavi & Starr, *supra* note 6. 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. In both the charging and sentencing analyses, we find larger disparities at the top of the distribution, but there are gaps throughout.

By "otherwise-similar," we mean similar in terms of the pre-charge case and defendant characteristics that we can observe and control for. In the charging analysis (Question 1), we controlled for arrest offense, district, age, whether there are multiple defendants in the case, and county-level poverty, unemployment, income, and crime statistics. In the sentencing analysis (Question 2 and 3), we added additional controls that are 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 we added them. These included defense counsel type, marital status, and Hispanic ethnicity, as well as flags for whether certain facts were recorded in the written arrest offense description: guns, other weapons, drugs, conspiracy, racketeering, child victims, and official victims.

On question (1), while we did not find any racial disparity in the *rate* of facing felony charges,[65] we *did* find significant racial disparities in charge *severity*, no matter which of our four charging measures we used. Using the statutory maximum, guidelines, and past-sentence measures, the race gaps were moderate, ranging from 6 to 9%. The disparities in mandatory minimums were much more dramatic. After controlling for the variables above, black men were still nearly *twice* as likely to be charged with a mandatory minimum offense: 8% of white males faced such a charge, compared to nearly 16% of otherwise-similar black males.

Question (2) focuses on the *aggregate* sentencing disparity introduced by the entire post-arrest justice process. We found that black and white defendants were equally likely to be convicted of non-petty offenses (and thus to make it into the sentencing sample), but that among those convicted there were significant unexplained sentencing disparities favoring whites. Most of the large raw sentencing gap (which was 50%) could be explained by the controls, and we used decomposition methods to identify *which* controls were the most important. 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 non-drug cases; when drug cases were added, the unexplained gap rose to 14%. In some subsets of the sample the gap was larger, especially among more serious cases.

Initial charging is an important driver of these sentencing disparities—especially the decision to bring mandatory minimum charges. Out of the 10% otherwise-unexplained sentence gap in non-drug cases, half disappeared when we controlled for mandatory minimum charges. However, 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. When we instead controlled for the mandatory minimum recorded for the crime of *conviction* which is unaffected by the

[65] This was the first stage of the criminal justice "funnel" that we analyzed—filing of felony charges in district court. We found no significant disparity, so we did not worry about sample selection bias stemming from this decision.

coding ambiguities), *all* the otherwise unexplained sentencing disparity disappeared.[66] This latter analysis could be performed for drug cases too. When drug cases were added to the sample, no significant disparity remained after controlling for the mandatory minimum of conviction.

We subjected all 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. For example, the results did not vary substantially by region of the country or based on the arresting agency. Mandatory minimum charging disparities were similar across offense types, but it is worth noting that the most common non-drug mandatory minimum in our data, and the one most responsible for driving sentencing disparities, was 18 U.S.C. § 924(C), the enhancement for crimes involving firearms. 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 30 years) if there is more than one § 924(c) count, which could just mean there were two guns found in a defendant's car. Prosecutors have considerable discretion in applying this statute, especially when the facts make the relationship of a gun to an offense ambiguous (for instance, the gun is simply found in the defendant's car), and a lenient prosecutor may "swallow the gun" entirely.[67] 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.[68]

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

C. Interpretations and Limitations

These results suggest the post-arrest justice process—especially mandatory minimum charging—appears to introduce sizable racial disparities. But are these gaps *really* the result of disparate treatment on the basis of race? Or do they stem from unobserved differences in circumstances that might be appropriate bases for different treatment? As Judge Nancy Gertner has warned, the quest to eliminate

---

[66] 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.

[67] *E.g.*, Erik Luna, Testimony Before the U.S. Sentencing Commission: Mandatory Minimum Sentences Under Law, (May 27, 2010), *available at* http://www.cato.org/publications/congressional-testimony/mandatory-minimum-sentencing-provisions-under-federal-law.

[68] ALEXANDER, *supra* note 4, at 118 (alterations in original).

improper disparities should not lead us to seek "false uniformity" among cases that are actually dissimilar despite superficial similarities.[69]

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.[70] 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 causal theories, and we use them in this Section to explore whether the "unexplained" disparity is truly "unwarranted." 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 is unobserved differences in the underlying crime. This concern is less severe than it might have been because 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.[71] The multi-defendant case variable also captures an important offense characteristic, because multi-defendant cases often involve more serious or complex crimes and often trigger conspiracy charges.

In drug cases (in addition to the limitations to the charge data), the arrest offense codes do contain an important ambiguity: they do not specify drug quantity. The EOUSA suspect investigation files record the drug quantity seized at arrest, but, when we examined patterns in the quantity distribution over time, we discovered a serious problem with this field beginning in 2004, when EOUSA adopted a new data entry system.[72] We therefore reran our drug-case analysis on the last three years of

---

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

[70] 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 or among different decisionmakers), and we use one such approach below to assess *Booker*'s effects. 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.

[71] 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.

[72] After extensive examination, we suspect the problem relates to the addition of a decimal point to the field—perhaps some (but not all) prosecutors are not noticing it. Unfortunately, it was impossible to determine which individual cases were affected. Comparisons to the Sentencing Commission's quantity data do, however, make it apparent that the problem is with the new system,

data before this change (2001-03), and found that adding quantity controls did not substantially reduce the observed racial disparities. We also found racial disparities in the drug quantities found at sentencing fact-finding, after controlling for the seizure quantity and drug type recorded at arrest.[73] 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 (for instance, pickpocketing or vehicle theft), but in others (for instance, wire fraud) they do not. It is unlikely that differences in loss quantity could explain the racial disparities we found, however—in fact, they probably cut in the opposite direction. At least as recorded at sentencing fact-finding, whites 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, i.e., that white defendants tend to be minor players in conspiracies while blacks 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 blacks get only very slightly worse role adjustments on average: a difference of 0.04 offense levels on the 43-level Guidelines scale, after controlling for the observed variables.[74] 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.[75] 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 education level, defense counsel type (an excellent proxy for poverty), marital status, and county-level crime and economic statistics. Perhaps more remarkably, these factors taken together do not contribute significantly to the "explained" share of the racial disparity. This appears to be because poverty itself (as reflected by these indicia) is

---

not the old one. It would be a service to future researchers, and to the public's interest in accurate record-keeping, if EOUSA investigated this problem.

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

[74] The range of possible role adjustments is from -4 to +4.

[75] 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.

not an important predictor of higher sentences.[76] Representation by a public defender is associated with slightly *lower* sentences, all else equal. This absence of socioeconomic disparity is good news, and we return to it below.

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. We estimate 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 that come from the control variables. It is important not to overlook those portions when thinking about what should be done about racial disparity. 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 some discretionary choices.

First, the recorded arrest offenses will be affected by law enforcement choices.[77] This is a key limitation to 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. We believe the arrest offense is a much better proxy for actual conduct than the presumptive Guidelines sentence, but it is not a perfect proxy. 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 bias introduced by law enforcement decisions that determine who lands in the federal criminal justice system at all.[78]

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

---

[76] 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.

[77] To the extent the prosecutor's pre-arrest involvement in the case has influenced the arrest offense, this omission may leave out an aspect of prosecutorial discretion, not just police discretion. When we drop cases with pre-arrest indictments from our samples (the cases with the most extensive pre-arrest involvement), disparity estimates increase.

[78] Blacks are 45% of the sample, a rate far exceeding their general population share; the question is the extent to which this overrepresentation reflects actual crime rates versus policing patterns. This gap is included neither in the "explained" part nor the "unexplained" part of our disparity estimates; we can only decompose disparities within the set of cases we have data on.

system. For arrest-stage disparities instead to explain our results, even partially, one would have to believe that federal law enforcement *favors* blacks. We think this is unlikely. Many criminal justice scholars have argued that black males are disproportionately targeted by law enforcement, while virtually nobody claims that they are disproportionately favored.[79] Blacks are arrested for drug crimes much more often than whites are, even though studies show that blacks self-report both drug use and dealing at equivalent or lower rates.[80] 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.[81] To be sure, federal law enforcement could differ from state and local enforcement, but we are likewise unaware of any anecdotal suggestions that federal agents favor black suspects.[82]

In addition, both the arrest offense and criminal history components of the "explained" disparity reflect subjective policy choices: important sources of disparity may simply be *built into the law*.[83] In the Fair Sentencing Act of 2010, Congress responded to such a concern, softening to some degree the sentencing framework's notoriously harsh treatment of crack cocaine cases.[84] But the crack laws are not the only example of particularly heavy punishments being given to crimes that are disproportionately committed by blacks. The harsh gun enhancements under 18 U.S.C. § 924(C) are another example—because black men are more frequently arrested with guns, these would disparately impact black men even if they were neutrally applied. Similarly, black males are also more frequently arrested for crimes involving violence or threats of violence, and sentencing law is

---

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

[80] *See* Alexander, *supra* note 4, at 99-100 (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).

[81] *See, e.g.*, Andrew Gelman et al., *An Analysis of the NYPD's Stop-and-Frisk Policy in the Context of Claims of Racial Bias*, 102 J. Am. Stat. Assoc. 813 (2007) (finding evidence of disparities against blacks in stop-and-frisk procedures, and reviewing the policing-disparity literature). Much of the existing research focuses on traffic stops, and reaches mixed results. Blacks and males are likelier to be stopped and searched. Some researchers have found a lack of disparity in the "hit rate" of stops and searches (e.g., 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 Police Enforcement, With an Application to Vehicle Searches in Wichita*, 116 ECON. J. F351 (2006). Others find lower hit rates for black and Hispanics, suggesting discrimination. *See, e.g.*, Sarath Sanga, *Reconsidering Racial Bias in Motor Vehicle Searches*, 117 J. Polit. Econ. 1155 (2009). Such studies at least do not suggest irrational *favoritism* toward blacks, however.

[82] Similarly, the processes by which a case is deemed appropriate for federal jurisdiction could introduce sample selection bias. However, if the federal government is simply interested in the types of crimes that blacks tend disproportionately to commit (or vice versa), that difference should be filtered out by the arrest offense controls.

[83] *See* Alschuler, *supra* note 21, at 87-88.

[84] Pub. L. No. 111-220, 124 Stat. 2372 (2010).

often harsher on these crimes than on nonviolent crimes that might reasonably be considered more serious.[85] 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. There are many competing considerations surrounding that judgment. But one consideration is its 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. That is, the criminal history score may be influenced by disparate treatment in past cases.[86] 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.[87] 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.[88] She finds unexplained gender disparities that dwarf the racial disparities our joint study found: men receive sentences that are over 60% longer than those of comparable women, conditional on the arrest offense, criminal history, and other pre-charge observable characteristics. These gaps are much larger than most other studies have estimated because—as with race—they mostly appear to arise prior to the final sentencing decision.[89] In the gender context, 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. 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.

Notably, the gender gap was substantially larger (about 75%) among black defendants. On the other hand, among females, the racial disparities we found for men do not recur—there is no significant unexplained black-white gap in female

---

[85] 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—the same offense level that would have applied had he defrauded Y out of nearly $1 million.

[86] *See* Stith, *supra* note 15, at 1432-33.

[87] 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(e) (2006); U.S. SENTENCING GUIDELINES MANUAL §§ 4B1.1, 4B1.4 (2011).

[88] Starr, *supra* note 13.

[89] In the gender context, an especially large share of the disparity appears to arise in sentencing fact-finding. *Id.* at 4. Mandatory minimums are also important, but only in drug cases, presumably because women are rarely arrested for the kinds of non-drug crimes to which mandatory minimums apply. So mandatory minimums make little contribution to the gender gap in non-drug cases after controlling for the arrest offense.

sentences. 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 or other hidden sources of disparity.

As noted above, black males are incarcerated at extremely high rates in the United States, and in assessing the causes of this problem, policymakers should consider both the race and gender dimensions and their interactions. Black male defendants appear not only to face 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.

III. The *Booker* Question: Does Expanding Judicial Discretion Increase Racial Disparity?

The results above suggest that while the role of prosecutors in producing the black-white sentencing gap has been overlooked by the empirical literature, the role of judges may have been overstated. Still, we do not claim judges' contribution is necessarily trivial. Depending on coding choices, we found up to several percentage points of racial disparity that neither the charges nor underlying case features could explain.[90] Given this fact, one might wonder whether the Supreme Court's recent expansion of judges' discretion was a mistake, and whether disparity could be reduced by restoring tighter constraints, as some have proposed. We thus now turn to the issue 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 more-than-fourfold increase in racial disparity in sentencing, from 5% to 23%. We disagree. In Section A, we explain our methodological objections, which extend beyond the Sentencing Commission's study to the rest of the literature examining the disparity consequences of sentencing law reforms. In Section B, we present the results of an alternative analysis reaching tentative conclusions that contradict the Commission's. Unlike the results above, these findings are completely new to this Article and are therefore presented in more detail. In Section C, we discuss the limitations on our analysis, and why researchers will probably never be able to give an entirely definitive answer to the question of *Booker*'s effects.

A. Studies Assessing the Effects of Sentencing Policy Changes

A subset of the sentencing disparity literature focuses on measuring *changes* in disparity resulting from changes to sentencing law, such as *Booker*, or the initial

[90] Our data suggest that judges' role in producing disparities may be particularly important in more serious cases *See* Rehavi & Starr, *supra* note 6, at 25.

adoption of mandatory sentencing. These studies have two pervasive methodological problems. First, they typically control for the presumptive sentence and thus ignore the possible effects of sentencing reform on pre-sentencing decisionmaking. Second, they fail to effectively disentangle the causal effects of policy changes from underlying trends over time. We explain both problems here.

The problem with the presumptive-sentence approach is largely explained above, but in the sentencing-reform context the reason for its importance is slightly different. 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 differences between groups that the observed variables do not capture.[91] 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, biasing estimates of how much of the racial gap in sentences is actually due to race. In a policy-change study, so long as the *same* broken yardstick is used before and after the policy change, one can get a valid estimate of the policy's relative effects on different groups. This "advantage" is a mixed blessing: estimates of changes in disparity are much 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 first-order "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 Sentencing Commission study discussed above. It found that the black-white gap went from 5% before *Booker* to 15% after, and finally to 23% after *Booker's* successor cases *Kimbrough* and *Gall*. Below, we discuss other problems 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. The problem is that *Booker* may have replaced one broken yardstick with a different one, if it affected charging, plea-bargaining, or sentencing fact-finding in racially disparate ways. Cases with the same presumptive sentences may represent different actual conduct pre- and post-*Booker* in ways that vary by race, biasing the pre/post comparison of disparities. Sample selection bias is also a potential problem: if *Booker* changed which cases are

[91] *See* Fischman & Schanzenbach, *supra* note 41, at 11.

winnowed out by the "funnel," the pool of sentenced cases may look different pre- and post-*Booker.*

There is good reason to worry about these potential biases. One clear lesson from the legal scholarship and practitioner surveys 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 decisionmaking at those earlier stages. After all, consider what happened after the Guidelines were adopted: a drastic increase in guilty pleas, which law scholars (very plausibly) attribute to prosecutors' sharp increase in leverage.

There are many theoretically plausible ways decisionmaking prior to sentencing could change after *Booker.* For example:

- Prosecutors might lose some of their leverage and have to offer more favorable plea deals to induce guilty pleas. This could mean more favorable findings of fact or more dropping of charges, and result in reductions of presumptive sentences and perhaps more trials.[92]
- Prosecutors could respond to the reduction in their power over Guidelines sentences by making more use of their remaining (even more powerful) tool for constraining judges: statutory mandatory minimums.
- Judges might become less willing to make findings of facts 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 more often 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.[93] It is, of course, possible that neither these changes nor any others happened in response to *Booker*, or at least that any changes had racially neutral effects. But in our view, this cannot simply be *assumed*—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

[92] If cases thereby became more resource-intensive, one might expect prosecutors to bring fewer cases or fewer charges per case. *See* James J. Prescott, Empirical Evidence of Prosecutorial Charging Manipulation (2006) (unpublished manuscript) (on file with author) (finding that prosecutors brought fewer charges after an earlier sentencing-procedure Supreme Court decision effectively raised the cost of proving charges).

[93] 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 blacks and whites in response to *Booker*, but their underlying use of mandatory minimums was twice as common for blacks, the doubling would look racially neutral, but would have twice the impact on blacks' sentences.

for the presumptive sentence or something closely related (the Guidelines "base offense level") and are subject to the same concern.[94]

These studies, in short, ignore the "hydraulic discretion" theory that has dominated theoretical scholarship about sentencing reform. Conversely, key aspects of the hydraulic discretion theory remain almost completely untested empirically.[95] 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 did survey judges and public defenders about their *perceptions* of whether aspects of plea-bargaining had changed.[96] However, the researchers did not evaluate these perceptions' accuracy, and the perceptions of judges and defense counsel varied quite substantially.[97]

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 race-related disparities in the dropping of charges at the plea-bargaining stage, yet *increased* racial disparities in sentencing—the opposite of what both mandatory sentencing proponents *and* "hydraulic discretion" theorists might have expected. However, the authors did not evaluate changes in initial charging, without which the results are harder to interpret.[98] 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.[99] No studies have evaluated changes in disparities in sentencing fact-finding.

Beyond the "presumptive sentence" problem, there is another serious methodological concern with these studies: 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

---

[94] *E.g.*, Ulmer et al., *supra* note 36 *passim* (finding a post-*Booker* increase in racial disparity in incarceration rates but not length); Fischman & Schanzenbach, *supra* note 38, at 2-3 (finding mixed results in an analysis of multiple doctrinal changes affecting judicial discretion, but concluding that expanded discretion does not increase racial disparity).

[95] Engen, *supra* note 24, at 324-25.

[96] Jeffrey 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).

[97] *Id.*

[98] 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).

[99] Miethe, *supra* note 7, at 167-71 (1987). 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.

variables.[100] 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 the PROTECT Act of 2003, which contained a rider imposing rules intended to discourage downward departures from the Guidelines. The rider 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.[101] Another important change was the Supreme Court's December 2007 decisions in *Kimbrough* and *Gall* (discussed above), which reinforced the *Booker* holding.

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.[102] 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 earlier periods).[103] It similarly found increases in racial disparity in the post-*Booker* and post-*Kimbrough/Gall* periods, although these were concentrated in the decision to incarcerate defendants rather than in sentence length among those incarcerated.[104]

But comparison of averages across such broad periods is at best suggestive. This kind of comparison 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 at a steady rate from 2003 on, for instance, the average disparity after *Booker* would necessarily be higher than the average in the PROTECT-to-*Booker* period, and the average after *Kimbrough/Gall* highest of all, even if those decisions had no effect on racial disparity. In fact, this

---

[100] The use of separate regressions means that the case mix is not controlled for across time periods, 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 36 (reporting both methods).

[101] PROTECT Act, Pub. L. No. 108–21, 117 Stat. 650 (2003).

[102] Ulmer et al., *supra* note 36, at 1091-94 also include the pre-PROTECT period.

[103] Ulmer et al., *supra* note 34.

[104] *Id.* Much of the authors' criticism focused on the Commission's failure to separate the incarceration decision from the length decision. The principal advantage to separating them is that, without including the zeros, one can log-transform the outcome variables, which may be a better functional fit to the determinants of sentences. We think the Commission's approach was fine on this count. A serious problem with separating the incarceration and length decisions is that if there is disparity in the incarceration decision, it will introduce sample selection bias to the length analysis. Starr, *supra* note 12, examines this problem in detail. In the analysis reported in Part II, we did separate the two stages, which was not a problem 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 selection bias concern is more serious.

would be true even if *Booker* and *Kimbrough/Gall slowed* the rate of increase in disparity.

Sentencing disparity might well be affected by numerous other developments over periods of that length. These could include, for instance, the policies and law enforcement priorities of a Presidential administration changing or taking time to trickle down to line prosecutors; changes in the composition of the judiciary and the U.S. attorneys' offices; or something as simple as administrative changes in supervision of prosecutors.[105] 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. It is also preferable to filter the surrounding trends out of the estimates of the policy's effects by including them in the regression.

The most recent *Booker*-related study is Fischman and Schanzenbach's, discussed above in Part II, and it improves on the standard approach in this regard. 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*.[106] 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 simply 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.

B. Our Method

Our alternative method is called regression discontinuity design ("RD").[107] Rather than comparing racial disparities averaged over periods of years, we create

---

[105] One federal prosecutor that we spoke to suggested that charging might in general have gotten harsher during the 2000s due to the expanded office use of computer-based tracking systems for prosecutors' performance, creating pressure toward toughness. General increases in harshness could have racially disparate consequences; *see supra* note 95. We do not assess the wide variety of possible causal theories for the underlying trends here..

[106] The authors further examine *Kimbrough/Gall* and their predecessor case *United States v. Rita*, 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. *Id.* Because they leave out the five months between *Rita* and *Kimbrough/Gall*, the last pre-period and first post-period are actually nearly a year apart. Nonetheless, this is also a substantial improvement over the Sentencing Commission. But as explained below, 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.

[107] 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* David Lee & Thomas Lemieux, *Regression Discontinuity Designs in Economics*, 48 J. ECON. LIT. 281 (2010); Guido W. Imbens & Thomas Lemieux, *Regression Discontinuity Designs: A Guide to Practice*, 142 J. ECONOMETRICS 615 (2008) .

flexible regression models that filter out month-to-month trends (including non-linear ones) in sentences and other relevant outcomes. We then look for sharp breaks in these trends—discontinuities—immediately after *Booker*. Like other studies, we base our causal inferences on changes over time, and any unmeasured changes that coincide 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*.

We use the same dataset described in Part III, but draw a broader sample from it. The sample runs from fiscal years 2001 to 2009 and does not exclude women and non-citizens. It also includes all non-immigration cases except identity theft, which was subject to other major sentencing-law changes near *Booker*.[108]

Our overall research interest is 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 the PROTECT Act's tightening of the Guidelines. The reason can be seen plainly in Figure 1, which plots departure rates by sentencing month.[109] Note that 96% of these departures are downward.

The vertical lines in Figure 1 mark four key events: the PROTECT Act, and the decisions in *Blakely* (*Booker*'s immediate predecessor), *Booker*, and *Kimbrough/Gall* (the *Booker* successors that clarified and strengthened its holding).[110] As this graph makes clear, *Booker* was a major shock to the amount of sentencing discretion afforded to judges. The number of departures went up immediately and substantially, from about 30% to about 40%. Although there are other month-to-month fluctuations in departures, *Booker* marks by far the most dramatic break. The sharpness of the change around *Booker* helps to alleviate one substantial concern about RD—its inability to capture effects that occur slowly. It is certainly

[108] Including women and non-citizens improves the precision of the estimates by increasing the sample size within each month. However, the results are substantively similar if these groups are excluded. We include controls for gender and citizenship.

[109] 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.

[110] 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 at *Booker* would be even steeper.

possible that the full effects of *Booker* took a while to take hold, and the inability to test that possibility is a disadvantage to our method. But if judges were inclined to use broader sentencing discretion in ways that disadvantage blacks, one would expect to see at least *some* of that effect in the immediate vicinity of *Booker*, because *Booker* immediately changed judges' willingness to depart from the Guidelines.

In contrast, PROTECT and *Kimbrough/Gall* were not 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.[111] The rise before *Kimbrough/Gall* could have been a response to the Supreme Court's decision in *United States v. Rita*, five months earlier, although this is also not obvious because it did not begin until two months after *Rita.* Still, even if *Rita*, *Kimbrough*, and *Gall* collectively led to an increase in departures (which seems likely), the fact that the events were separated by five months makes it too diffuse a change to judges' sentencing discretion to assess with this 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 now turn to the effects of increasing 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*,[112] 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.

---

[111] 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. Both might have influenced district courts indirectly by changing the appellate review standard and (in PROTECT) increasing data collection on departures, but it is unclear how much life-tenured district judges would care. *Kimbrough* did directly affect departure authority, but only in crack cases (where mandatory minimums applied regardless); it was uncontested that judges could depart on policy grounds in other cases. And the crack holding could only have helped black defendants—a logical challenge for studies that point to *Kimbrough* as a source of racial disparity.

[112] The average time from charge to disposition in our sample is 5 months, and the average time from disposition to sentencing is a further 4 months.

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 and their sentencing effects by assessing what happens when the *sentencing* date passes *Booker*. Note that the judicial behavior being measured is not just changes to the final sentencing decision relative to the Guidelines sentence, but also changes to sentencing fact-finding. Separately assessing the conviction date and the sentencing date helps to disentangle judges' contributions to any disparities in sentencing fact-finding from any 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), applying the same Sixth Amendment holding to a state sentencing scheme. It was *Blakely* that was an unexpected "earthquake," throwing federal sentencing into disarray by rendering it fairly obvious that the federal Guidelines were in constitutional trouble.[113] 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, or left them mandatory but shifted fact-finding authority to the jury, or left the whole matter to Congress. The lower courts quickly began weighing in, and the Supreme Court quickly agreed to review *Booker* and scheduled an early argument.

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*, courts and/or the parties anticipated that the mandatory Guidelines would fall and adjusted their behavior ahead of time. 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. If so, those 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 expanding 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

[113] The archives of Douglas Berman's Sentencing Law & Policy blog for this period provide an excellent record of this disarray. *See* http://sentencing.typepad.com/sentencing_law_and_policy/ (last visited Aug. 3, 2012).

federal guidelines.[114] In those circuits, *Booker*'s legal effects were simpler: it changed the governing law from the old regime (mandatory 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. Therefore, as detailed below, we also analyze changes happening at *Blakely* to see whether there is evidence of such anticipation effects.

C. 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 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*.[115]

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. *Booker* reduced 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.[116] As explained above, we could not code the initial charges in drug or child pornography cases. We only know the mandatory minimum for the *conviction* offenses. 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 eventual

---

[114] *See* United States v. Mincey, 380 F.3d 102, 106 (2d Cir. 2004); United States v. Hammoud, 378 F.3d 426 (4th Cir. 2004); United States v. Pineiro, 377 F.3d 464, 473 (5th Cir. 2004); United States v. Koch 383 F.3d 436, 438 (6th Cir. 2004); United States v. Curtis, 380 F.3d 1308(11th Cir. 2004).

[115] 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.

[116] We did not find any additional disparity introduced in charge-bargaining over mandatory minimums. *See* Rehavi & Starr, *supra* note 6 , at tbl. 3, tbl. A3 (showing a slightly *lower* black effect when the conviction mandatory minimum is analyzed instead of the charge).

conviction mandatory minimum when the *charging* date passes *Booker*.[117] 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.[118] This approach allows us to assess all case types.

The results from the formal RD analysis are presented in Table 1, 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 ("Blk-Wht. Diff. Disc") estimates *Booker-related* change for black defendants relative to whites. 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. We estimate regressions that include separate non-linear time trends for blacks and whites, 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 effect of month-to-month variation in arrest offenses and other pre-charge features of the case.[119] The estimated discontinuities represent the break in the curve at *Booker*—that is, the difference between the intercepts of pre-*Booker* curve and the post-*Booker* curve.

Within each panel of Table 1, the four columns show the results of multiple specifications that use different methods of fitting curves to the data (i.e. filtering out trends over time)—we vary the length of the time window used to estimate the curves on each side (12 months versus 18 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 (and thus more trustworthy) if it is consistent across specifications, such that it is not just an artifact of a subjective modeling choice.

---

[117] The "charging date" is the date of the indictment when there is one. Some cases have no formal indictment, in which case 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.

[118] If prosecutors suddenly started charging mandatory minimum offenses more 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; doing so requires a supervisor's special permission. *See* Ashcroft, *supra* note 62.

[119] 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, 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.

We find that when the charging date passed *Booker*, there was a significant, discontinuous increase in the mandatory minimum rate—but only against black defendants (Panel 1A). The increase in the black-white disparity in mandatory minimums is significant and quite large in all specifications, ranging from six to thirteen percentage points.[120] Most of the increase in disparity is because of an increase for black defendants, but there also appears to be a smaller *reduction* in mandatory minimums' use against white defendants.

Figure 2a provides an approximate visual representation of this result.[121] Although the RD is estimated based on a narrower window of time surrounding *Booker*, we show longer surrounding trends in the graph so as to put the estimated discontinuities in context. The hollow circles and dots represent the monthly averages in the *residuals* for whites and blacks, 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 regression on the basis of other observed characteristics. 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 blacks and whites, and the vertical distance between the black and white curves represents the unexplained racial disparity at any given time.

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 is a steady rise in mandatory minimums for both blacks and whites, and that trend continued.

Overall, although there is a significant break, the patterns are much less dramatic than what we saw with 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

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

[121] 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 to it 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 fitted curves on either side at *Booker* (the difference in intercepts) is a visual approximation of the discontinuity estimated by the Regression Discontinuity estimator.

blacks). If so, charging may have reverted to normal when prosecutors saw that *Booker* did not cause a major drop in sentences (which 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.[122]

We next assess whether the ultimate *sentence* length was discontinuously affected by the *charging* date passing *Booker*—that is, did changes in charging translate into sentencing consequences? We find some evidence of this: racial disparity in the sentence jumps by between 4 and 13 months for cases charged immediately after *Booker*, depending on the specification (Panel 1C). Almost all of this consists of an increase for blacks. However, as Figure 2b shows, there noise in the sentence-length data, compared to which the break is not so remarkable. Unsurprisingly, then, the estimates are not very precise, and the statistical significance of the increase in black sentences varies across specifications.

2. Changes in Plea-Bargaining

We next assess *Booker*'s effects on plea-bargaining 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 1, Column 2; Figs. 3a-3c). Mandatory minimum rates for whites are noticeably higher in the post-*Booker* period 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 would mean that 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.

3. Changes in Sentencing Fact-Finding and Sentencing

Finally, we assess changes in judicial decision-making by examining what happens when the *sentencing* date passes *Booker*. We consider 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

[122] Panel 1B also shows a discontinuous increase in disparity in the final offense level for cases charged after *Booker* that is significant in two of the four specifications; the disparity comes mostly from an increase in black offense levels. This might simply reflect the apparent mandatory minimum shift, since the ultimate offense level is affected by any mandatory minimums that apply. In any event, this is further suggestive evidence that charging for blacks got somewhat harsher at *Booker*.

expansion had an immediate effect. In Figure 4a and Panel 3D, we break this effect down by race. The estimates all show a jump in white departure rates of about 8 percentage points, and the jump for blacks is similar, so there is no significant change in disparity in departure rates.[123] Notice that both the black and 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 and clearly had a lasting effect.

Booker's legal holding did not directly affect fact-finding, but it could have affected it indirectly (even setting aside any effects on the 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.[124] Expanded authority to do the latter might make it less necessary to do the former.[125]

Panel 3B of the table gives some support for this theory, but only for whites. In white defendants' cases, the increase in departures appears to be offset to some degree by an increase in the Guidelines sentence from which the judge departs. Estimates of the effect size vary between about 0.1 and 0.9 offense levels; each offense level is associated with a 10-15% increase in the Guidelines sentencing range. The change in the racial disparity in offense levels is less precisely estimated and is thus not always significant in every specification, but the sign is consistently negative, and (adding the two rows together) black offense levels do not seem to go up. Figure 4b shows that after *Booker*, the black and white trends, which had been diverging, converged again. While this may suggest that judicial discretion helps to close racial gaps, an alternative interpretation is less charitable to judges. Before *Booker*, there was a significant racial disparity favoring whites in findings of fact, after controlling for the observable variables. One possible interpretation of *Booker's* effects is that there had been a thumb on the fact-finding scale in favor of whites, and after *Booker*, that thumb became no longer necessary.

Note that the upward shift for whites, although subtle, does not appear to be temporary. There is a long-term trend of gradual increase in the offense level. However, one cannot safely infer that the long-term trend is caused by *Booker*, because RD only estimates the local effect at the discontinuity. Nor can one assume that the trend is necessarily caused by judges—again, the offense level is jointly determined by the parties' negotiations and the judge's findings. As one moves farther away from the discontinuity, one quickly gets to a point in which most cases

---

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

[124] Max M. Schanzenbach & Emerson H. Tiller, *Strategic Judging Under the United States Sentencing Guidelines: Positive Political Theory and Evidence*, 23 J.L. ECON. & ORG. 24 (2006).

[125] 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 reasons they choose to do so in particular cases might be the same kinds of exceptional dissatisfaction with the sentence that the plea agreement would produce that might otherwise motivate departures.

were both pled and sentenced after *Booker*, and it becomes impossible to isolate *Booker*'s effects on judicial fact-finding alone.

Whatever its causes, the trend of increasing offense levels 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. As Figure 4c and Panel 3C of the table show, there appears to have been an immediate drop in black sentences at *Booker*, although there is considerable noise in the data. White sentences did not fall, however (even though white departures increased).[126] Perhaps this is because the departures were offset by fact-finding changes. And perhaps this was also true in the long run for both blacks and whites: Figure 4b shows a sustained trend of increased offense levels, as though the break at *Booker* never reversed itself.

The magnitude of that change looks fairly small, though—perhaps half of one offense level. One might wonder how such a subtle change in fact-finding could cancel out such a large increase in departures. The answer is 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 29 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 2 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 to have been not 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. Sentences continue 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.

D. 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 blacks that we see arise from prosecutors' shift to mandatory minimums, although that shift may have been temporary. Like our results 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. Beyond that, there are a few other things to keep in mind.

---

[126] 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.

First, it is important to understand what our RD analysis does *not* assess. In the initial decompositions 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. 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. 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 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, our results in Part II and 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 stage of the criminal process.

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 be a real shift. But if there are frequent breaks, even at points where there is no other known triggering event, 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 our analyses on twelve other arbitrary breaking points across the study period.[127] These tests were reasonably successful. In the mandatory minimum variable, when the placebo tests were run by charging date, none showed a significant discontinuity in racial disparity in more than one specification (a "false positive"); when run by disposition date, one out of twelve did. This makes us more confident that the spike at *Booker*—even if brief—was likely something real. In the prison variable, which was visibly noisier in the graphs, there was one false positive when the placebos were run by charging month, four when run by disposition month, and two when run by sentence month. In the offense level variable, there were two false positives when the placebos were run by disposition month and one when they were run by sentence month. The departure variable had two false positives (run by sentencing month), but visual inspection makes clear that *Booker* was by far the cleanest break in the study period.

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*, nor are there strange patterns in the six months before *Booker* (Figs. 4a-4c). It appears that these courts really did follow the "business as usual rule." But what about prosecutors?

It does not appear that initial charging changed much—both the mandatory minimum variable and sentence length continue on their existing trajectories after *Blakely* (Figs. 2a-2b). The sharpest effect of *Blakely* appears to have been on plea-bargaining. Just after *Blakely*, there are two anomalous months in which plea severity plunges, in particular for whites (Figs. 3a-3c). This effect is probably because one immediate reaction to *Blakely* in the world of federal criminal practice was anticipation of the possibility that sentencing facts might have to be proven to the jury or pled to by the defendant.[128] That expectation would give the defendant much more leverage in plea-bargaining, because if they refused to plead to adverse facts, it was not at all obvious that the prosecutor would nonetheless be able to argue them at sentencing. Nonetheless, by the third month after *Blakely*—after all five of the circuits we focus on had established the "business as usual" rule—this anomalous period appears to have ended.

This anomaly at *Blakely* could affect the interpretation of what happened at *Booker*, either in plea-bargaining (because it affects the trend in the pre-period) or in sentencing (because cases sentenced around *Booker* might have been plea-bargained around *Blakely*). We do not find any significant effects on plea-bargaining

[127] The tests ran every six months for three years in each direction before *Blakely* and after *Booker*.
[128] The Seventh Circuit had so held in its decision in *Booker*, 375 F.3d 508 (2004), less than three weeks after *Blakely*. *Accord United States v. Ameline*, 376 F.3d 967 (9th Cir. 2004).

at *Booker* anyway, and based on the graphs, it does not appear that any discontinuities would have appeared if those anomalous months had not disturbed the trend. The implications for the sentencing results are harder to interpret. In the sentence-length and offense level graphs (Figures 4b and 4c), one can see a V-shaped pattern in the white trend—a dip right before *Booker.* We suspect that one reason for that dip is the lingering effect of those lenient plea deals that got struck in the two months after *Blakely*, most of which would have been sentenced before *Booker*. The pattern does not exist for black cases, but black cases did not have an anomalous post-*Blakely* period. Perhaps had it not been for those cases, the white sentence levels right before *Booker* would have been higher, and one might have seen a drop in white sentences at *Booker*, paralleling the drop in black sentences. Or one might *not* have seen the jump in white offense levels at *Booker*.

While this interpretation is possible, there is also reason to believe that the *Blakely* effect might not have actually changed the *Booker* discontinuity estimates much. There is enough variation in the time between disposition and sentencing that it is not as though all of those anomalous-period cases were sentenced in the month or two right before *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 come from that period. In other words, whatever effect those plea-bargaining changes had on sentencing should be a roughly continuous trend, and to the extent it is continuous, the polynomials should filter out. To substantially affect the discontinuity estimates at *Booker*, the probability of having been a post-*Blakely* "sweetheart deal" case would have to have plunged *suddenly* in the month of *Booker*. This is a substantial advantage of RD over other methods we could have used to assess *Booker*'s effects.[129]

In addition, the absence of the plea-bargaining anomaly after *Blakely* in black cases is itself substantively interesting—if anything it looks like there may be an *increase* in severity there. That could be suggestive evidence that prosecutors respond to uncertainty and sudden change differently with black and white defendants (and more favorably for whites). This interpretation, while speculative, is alconsistent with the fact that black mandatory minimum use spiked immediately (if briefly) at *Booker*, but white charge severity did not.

A final concern about the interregnum period is that cases could have been delayed until after *Booker*. Causal inference with an RD estimator requires that "individuals—even while having some influence—are unable to *precisely* manipulate the assignment variable," and here, the assignment variable is the exact sentencing date relative to *Booker*.[130] Note that this is a relatively modest requirement—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 try to make it more likely that a case would be sentenced

[129] 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.
[130] Lee & Lemieux, *supra* note 107, at 283.

after *Booker*, such as scheduling the sentencing hearing for a faraway date, this would not be seriously problematic; the scheduling would get the case *near* Booker, but there would still be a chance element as to which side it landed on. The chance element is amplified by the fact that nobody knew when the Supreme Court would rule—legal observers did terribly at predicting *Booker's* release, with many predicting a very fast decision after the October argument.[131]

The case counts do not suggest problematic manipulation. The number of cases sentenced in December 2004 was 1703; the number in January 2005 was 1708. The mean elapsed time since the plea was identical: 4.08 months. The breakdowns by race and case type were almost identical. If anything, there may have been some delaying of cases in November 2004 (1576 cases, the lowest that year), when expectations of an early *Booker* decision were high, but that is a small dip, and when the Court did not release its decision quickly it appears the counts went back to normal. After *Booker*, the number of cases also stayed normal; they were slightly higher in March (1841 cases), 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 these circuits really did do "business as usual," or that any manipulation was too imprecise to have major effects right around the discontinuity.[132]

E. Longer-Term Trends

Relative to the dramatic findings of the Sentencing Commission study, our results seem to tell a very different story about *Booker*. But our results are quite difficult to compare directly to the Commission's—in addition to our use of RD instead of comparisons across broad time periods, we also use a different sample and a very different set of covariates (most notably shifting from the presumptive sentence to the arrest offense). So one might wonder what our results suggest about the Sentencing Commission's findings.[133] Has the Commission identified a real increase in racial disparity over the time period, but simply misattributed it to *Booker*? Or is there no real increase to begin with, once one estimates sentencing disparity properly?

---

[131] *See, e.g.,* Ian Weinstein & Nathaniel G. Marmur, *Federal Sentencing During the Interregnum: Defense Practice as the* Blakely *Dust Settles*, 17 Fed. Sent. Rep. 51, 51 (Oct. 2004) ("It seems likely that the Supreme Court will offer some guidance by Thanksgiving."). The archives of the leading sentencing blog include a series of predictions that *Booker* would be decided the next day. *E.g.*, Douglas Berman, Sentencing Law & Policy, December 7, 2004 ("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."), at http://sentencing.typepad.com/ sentencing_law_and_policy/booker_and_fanfan_commentary/page/27/.

[132] In any event, manipulation would only bias our results if it occurred in a racially disparate way that was not mediated by the control variables—for instance, if courts decided to delay crack cases, our arrest offense controls would filter that out. 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 (risking their withdrawal) while waiting for a Supreme Court decision.

[133] Likewise, even the other two post-*Booker* racial disparity studies, which were skeptical of the Commission's methods and conclusions, did find some increase in disparity, at least after *Kimbrough/Gall*. *See* Ulmer et al., *supra* note 34; Fischman & Schanzenbach, *supra* note 41.

Our tentative answer is that there is no real increase, at least within the subset of (non-immigration) cases that we have examined. We have done only very basic analyses of the longer-term trends, given that our analysis focused principally on *Booker*'s impact, so we cannot give a very nuanced answer. But one simple way to assess whether there has been an increase in racial disparity over a time period is to estimate a linear trend over the whole period.[134] We did this for cases sentenced between the PROTECT Act and the end of FY 2009—the period in which the Sentencing Commission found the purported quadrupling of disparity—and controlled for the arrest offense and other prior characteristics.[135] That is, we estimated the linear trend in the aggregate black-white sentence disparity introduced during the post-arrest justice process. This analysis actually shows a small but significant trend of *declining* racial disparity—the model estimates that the unexplained black-white sentence gap declined from about 14 months to about 11 months over the course of the period.[136]

Why, then, does the Commission find an increase? Notably, when we repeat this linear trend analysis but control for the final Guidelines offense level instead of the arrest offense (the equivalent of the presumptive sentence approach), we also see an increase in disparity, although not the dramatic one found by the Commission; perhaps our other control variables absorb part of it. Black and white sentences do appear to have diverged during this period *relative to the presumptive Guidelines sentence*. But the race gap in Guidelines sentences (relative to the arrest offense and pre-arrest characteristics) declined.[137] In other words, the changes in racial disparity patterns in the judge's final sentencing decision appear to have only partially offset larger changes in the prior stages of the justice process. The presumptive sentence was not the "same broken yardstick" during this period—over time, the yardstick changed.

Note that we make no claims as to the causes of these trends, and we do not claim that *Booker* caused them. Many factors might contribute, and further exploration is necessary to see what part of the earlier process changed. Mandatory minimum use increased slightly over the time period, but given the apparent contribution of mandatory minimums to sentencing disparities, one would not expect this trend to explain a *reduction* in the sentencing gap. Instead, we suspect something else, such as a change in sentencing fact-finding, which could be due to

---

[134] Specifically, the regression includes an *overall* linear time trend as well as an interaction between that time trend and the "black" coefficient.

[135] For this assessment, we did not just use the "business as usual" circuits, because we were not worried about isolating *Booker*'s effects—we used the whole country. As in Part II, we focused on black and white male U.S. citizens. We included all non-immigration offenses except for those that had big unrelated sentencing-law changes (identity theft and child pornography). We controlled for the arrest offense, criminal history, district, education, age, and multi-defendant case structure.

[136] This estimate uses sentence length in months (including zeros) as the outcome variable, tracking the Commission's approach. However, we similarly find a decline if we use a log-linear model of sentence length that excludes the zeros. We find no significant trend in racial disparity in the probability of incarceration, which is the disparity in which Ulmer et al. found an increase.

[137] We estimated this using a linear regression (with time trends) of the final offense level on the arrest offense and the other pre-charge covariates.

changes in plea-bargaining or in judges' willingness to "go behind the plea." Exploring these possibilities further would be a fruitful area of future research.

CONCLUSION

Determining the causes of racial disparities in the criminal justice system is not easy. We believe our dataset and methods offer substantial improvements on existing research, but we do not offer definitive answers, and we doubt that anyone will anytime soon. So what, then, are policymakers to do? We do not seek to resolve that dilemma completely. Even if we *did* have crystalline empirical answers, criminal justice policy does not turn on demographic disparity alone—many competing objectives must be considered. That said, we do think our results have implications for these dilemmas, and we also fear that the contrary results of existing research are at risk of being distorted to support counterproductive "solutions" to disparity. We close with some brief thoughts on these points.

First, despite our concerns about the methods of the Sentencing Commission and others, we believe they are right to be concerned about demographic disparities in the criminal process. The high rate of incarceration of black men is a serious social problem, and examining the possible contribution of criminal justice disparities is important. Our research suggests that in the federal system, disparities in the post-arrest justice process are contributors to this problem. After controlling for the arrest offense, criminal history, and other prior observed characteristics, sentences for black male arrestees diverge substantially from those of white males (by around 10% on average). The disparity that we find does not appear to be growing, but it is persistent, and it provides reason for concern.

Second, the procedural source of this disparity matters, and it is myopic to focus on judicial sentencing alone. Our research suggests that racial disparities in recent years have been mostly 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.[138] Our results suggest that this would not reduce disparities in the justice process. Quite the contrary: we find that prosecutors use their discretion to file mandatory minimums twice as often against black men than against comparable white men. Moreover, for those concerned about mass incarceration of black men, expanding mandatory minimums would be counterproductive for another reason as well. Sentencing law changes that increase severity have a particularly adverse impact on black men because they are disproportionately involved in the criminal justice 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.

[138] *See* Gonzales, *supra* note 10.

Third, we do not advocate attempting to reduce disparity by taking discretion away from prosecutors. Eliminating prosecutorial discretion is probably impossible in any event. The Department of Justice has certainly tried. The 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. If 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 evidence.[139] And even if constraining prosecutorial discretion *did* succeed, one might see another "hydraulic" effect. If prosecutors had to pursue to the fullest every case law enforcement brought them, 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 that would be 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 disparity 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 their *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, such as not bringing charges in the first place. The Fair Sentencing Act of 2010, which reduced crack sentences, showed it is politically possible to reform harsh sentencing laws, and that empirical evidence of racial disparities can help to bring such changes about.[140]

One potential next focus could be the very harsh gun enhancements in 18 U.S.C. § 924(c). These laws hit black men particularly hard both because 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 gun violence concerns. But we wonder, for instance, whether a five-year add-on sentence is truly *always* necessary (such that judicial discretion should be precluded) in cases in which a firearm has merely been "carried"—let alone a mandatory extra 25 years for a second gun, and yet another 25 for a third. And no prosecutor would need to "swallow a gun" if a gun did not automatically trigger a massive additional penalty.

Fourth, there was a piece of good news in our findings that also carries potential policy implications: case outcomes were not correlated with measures of socioeconomic status. This absence cuts against conventional wisdom.[141] Can it

---

[139] *See* Stith, *supra* note 15, at 1470; O'Sullivan, 1425-26; Miller, *supra* note 26, at 1257.

[140] *See supra* note 84 and accompanying text.

[141] The hurdles facing indigent defendants and their counsel have long been the subject of extensive advocacy, scholarship, and policy attention. For a prominent recent example, see the website of

really be that poor defendants do *not* fare worse? We do not conclude that the conventional wisdom is wrong, but it is not readily apparent in the federal courts. This is likely a testament to the unusually high quality of representation provided to indigent federal defendants, especially by the federal public defenders.[142] We suspect that we would not have gotten the same result had we studied many state systems, where indigent representation is notoriously under-resourced and in disarray.[143] 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.[144]

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 found in state courts. More generally, we do not claim to have proven purposeful discrimination by prosecutors or anyone else—it would be difficult 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 with "testers" or manipulation of case files, or qualitative studies involving full case files and interviews.[145] DOJ itself is well positioned to carry out such work. One easy step would be for DOJ, when it tracks prosecutors' performance, to keep statistics on mandatory minimum charging decisions by race. Doing so would not only facilitate research, but could 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 think creatively about solutions and about strategies for answering the empirical questions that remain.

DOJ's Access to Justice Initiative, at http://www.justice.gov/atj/ (describing the need to "address the access-to-justice crisis").

[142] Richard Posner & Albert Yoon, *What Judges Think of the Quality of Legal Representation*, 63 STAN. L. REV. 317 (2011).

[143] *See, e.g.*, Eve Brensike Primus, *Structural Reform in Criminal Defense: Relocating Ineffective Assistance of Counsel Claims*, 92 CORNELL L. REV. 679 (2007)).

[144] *See* Steven B. Bright, *Legal Representation for the Poor,* 75 MO. L. REV. 683 (2010) (noting the superior resources of federal public defenders).

[145] 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 itself uses "testers" (fake applicants) to enforce its discrimination statutes. We are not aware of any currently legal way to conduct similar field experiments in the criminal justice context. A crime staged for research purposes is still a crime, as is submitting fake information to criminal justice authorities. But such studies could be legislatively authorized, under regulated conditions and perhaps to be carried out by DOJ itself.

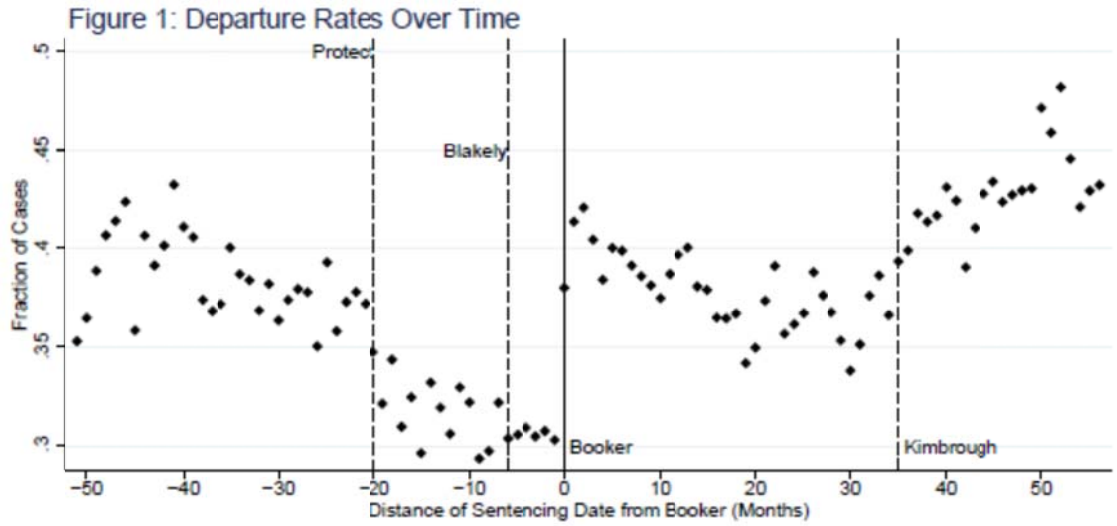
Figure 1: Departure Rates Over Time
Protect
Blakely
Booker
Kimbrough
Fraction of Cases
.3
.35
.4
.45
.5
−50
−40
−30
−20
−10
0
10
20
30
40
50
Distance of Sentencing Date from Booker (Months)

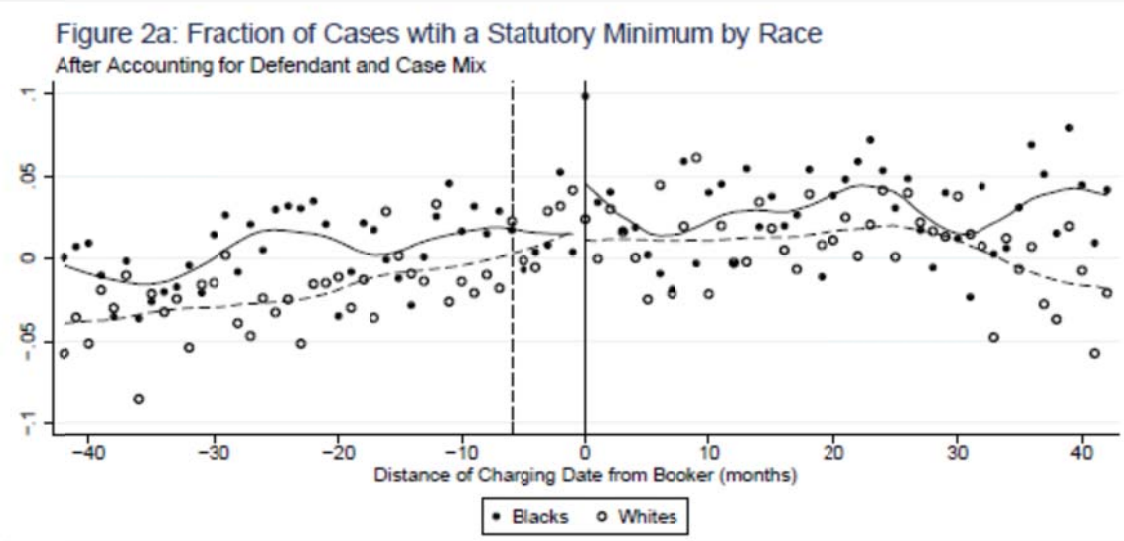
Figure 2a: Fraction of Cases wtih a Statutory Minimum by Race
After Accounting for Defendant and Case Mix
−.1
−.05
0
.05
.1
−40
−30
−20
−10
0
10
20
30
40
Distance of Charging Date from Booker (months)
Blacks
Whites

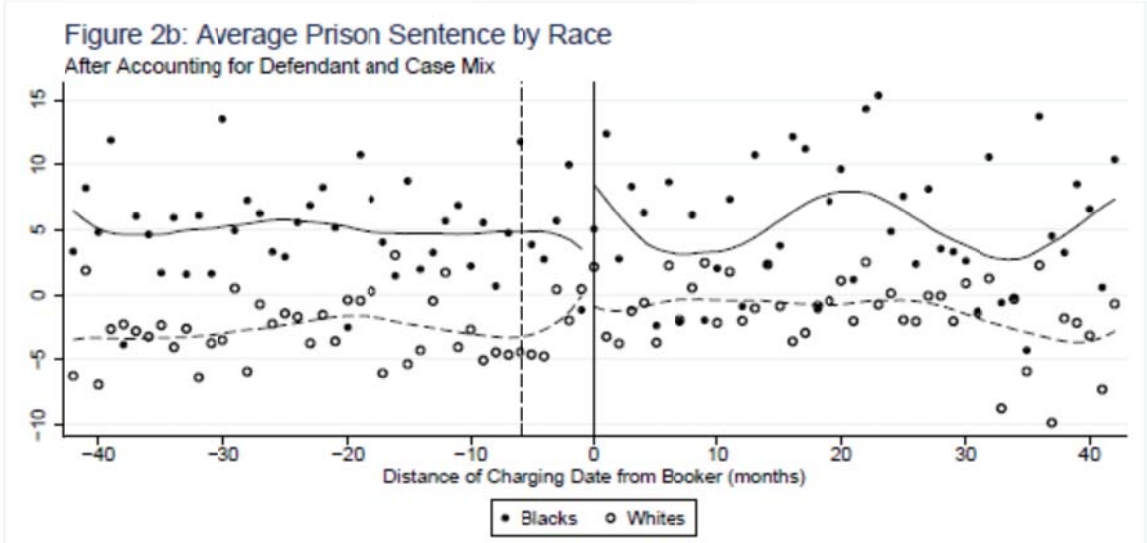
Figure 2b: Average Prison Sentence by Race
After Accounting for Defendant and Case Mix
−10
−5
0
5
10
15
−40
−30
−20
−10
0
10
20
30
40
Distance of Charging Date from Booker (months)
Blacks
Whites

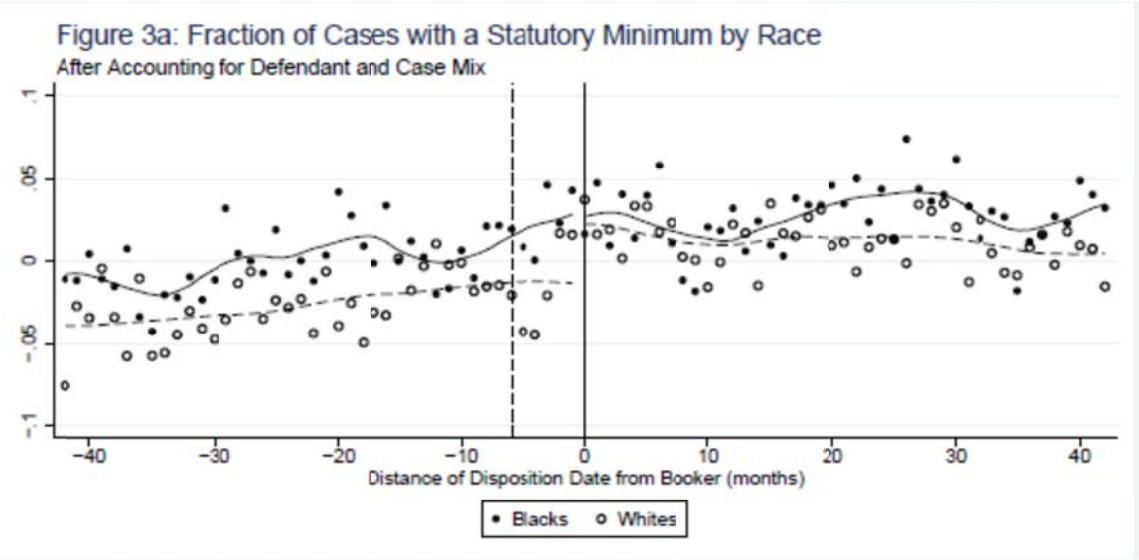
Figure 3a: Fraction of Cases with a Statutory Minimum by Race
After Accounting for Defendant and Case Mix
.1
.05
0
-.05
-.1
-40
-30
-20
-10
0
10
20
30
40
Distance of Disposition Date from Booker (months)
Blacks
Whites

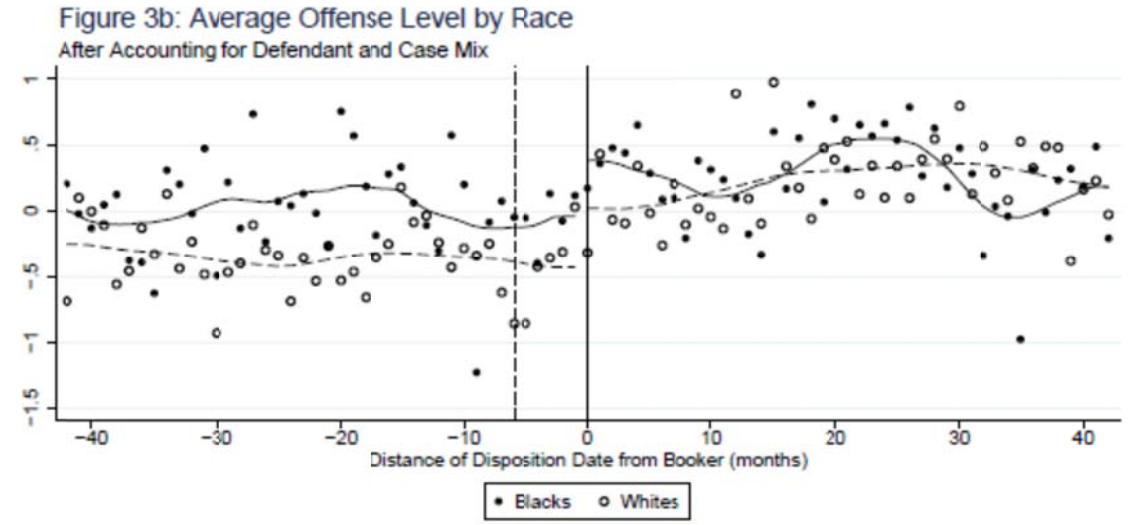
Figure 3b: Average Offense Level by Race
After Accounting for Defendant and Case Mix
1
.5
0
-.5
-1
-1.5
-40
-30
-20
-10
0
10
20
30
40
Distance of Disposition Date from Booker (months)
Blacks
Whites

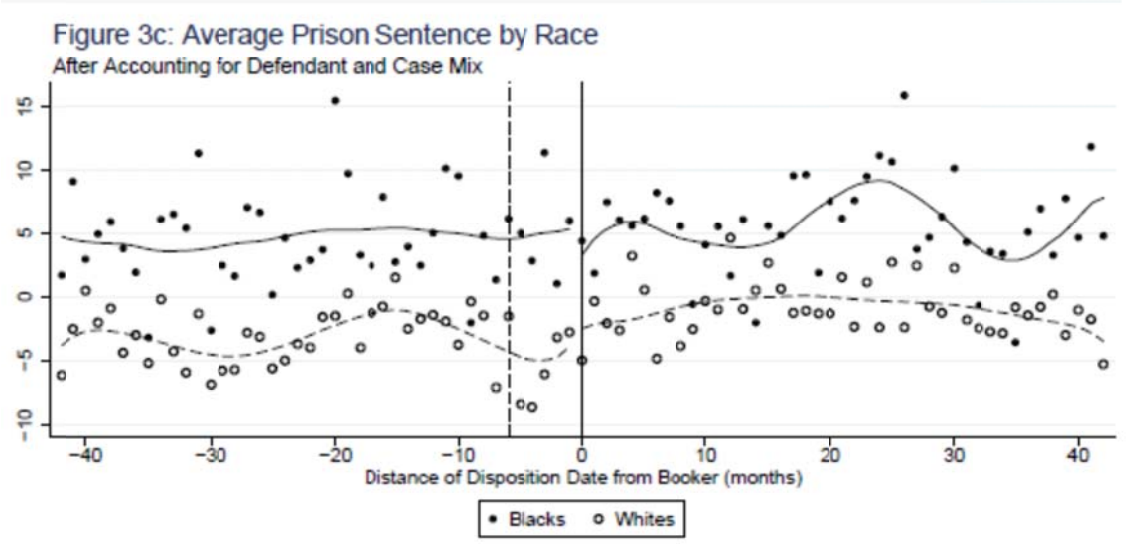
Figure 3c: Average Prison Sentence by Race
After Accounting for Defendant and Case Mix
15
10
5
0
-5
-10
-40
-30
-20
-10
0
10
20
30
40
Distance of Disposition Date from Booker (months)
Blacks
Whites

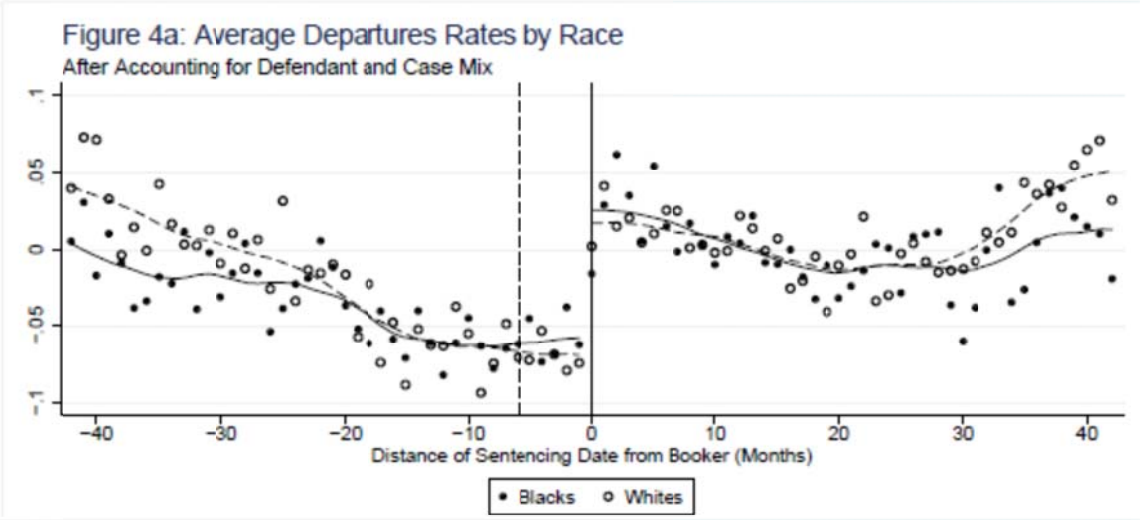
Figure 4a: Average Departures Rates by Race
After Accounting for Defendant and Case Mix
.1
.05
0
-.05
-.1
-40
-30
-20
-10
0
10
20
30
40
Distance of Sentencing Date from Booker (Months)
Blacks
Whites

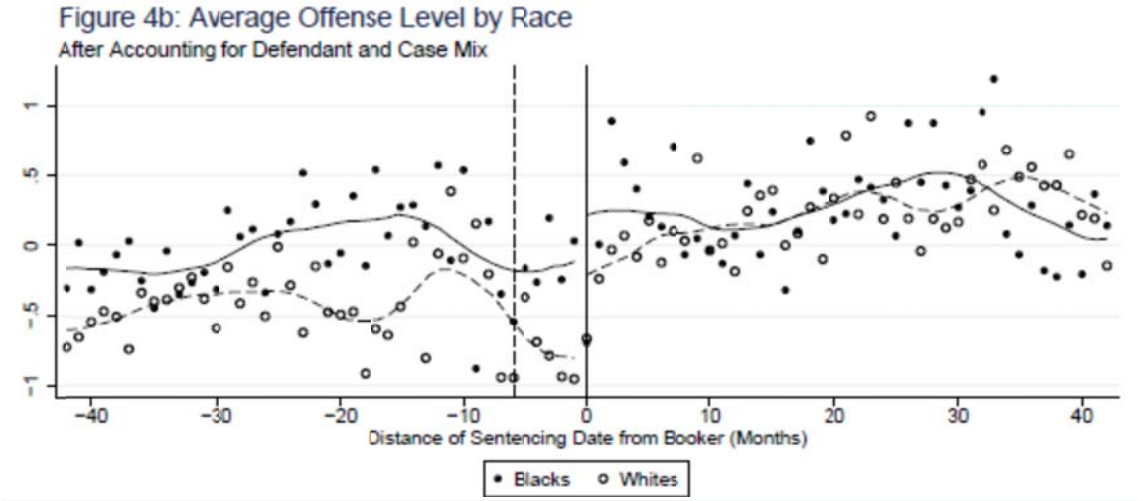
Figure 4b: Average Offense Level by Race
After Accounting for Defendant and Case Mix
1
.5
0
-.5
-1
-40
-30
-20
-10
0
10
20
30
40
Distance of Sentencing Date from Booker (Months)
Blacks
Whites

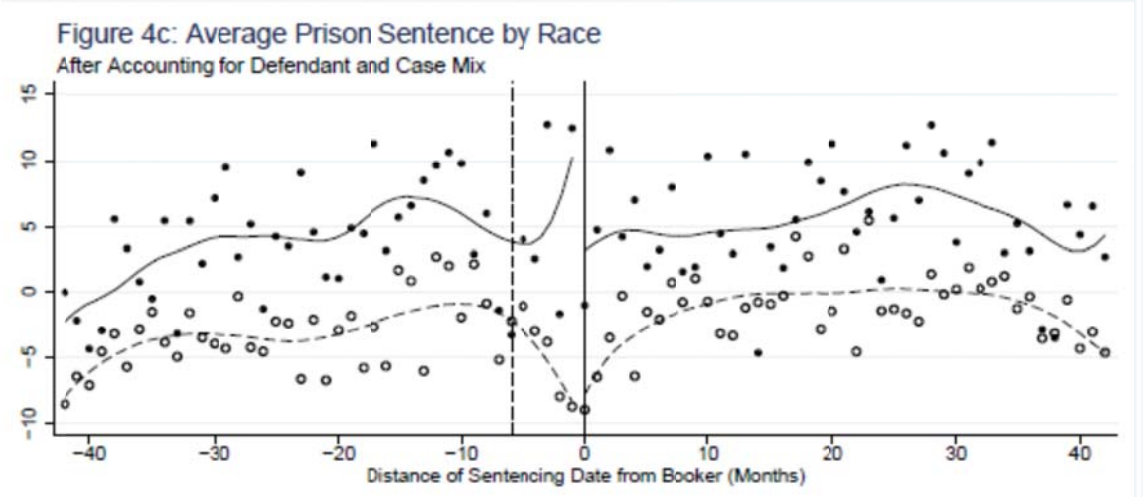
Figure 4c: Average Prison Sentence by Race
After Accounting for Defendant and Case Mix
15
10
5
0
-5
-10
-40
-30
-20
-10
0
10
20
30
40
Distance of Sentencing Date from Booker (Months)

Regression Discontinuity Estimates of *Booker*'s Effects on Charging, Plea-Bargaining, and Sentencing

	1: Charging Date				*2: Disposition Date*				*3: Sentencing Date*			
A: Convicted of Mandatory Minimum Offense												
Overall Discontinuity	-0.023	-0.047**	-0.049**	-0.020	0.037+	-0.008	-0.022	-0.043**				
	(0.014)	(0.016)	(0.016)	(0.019)	(0.022)	(0.013)	(0.014)	(0.015)				
Blk-Wht Diff in Disc.	0.062*	0.104**	0.131**	0.091**	-0.048+	0.006	0.003	0.004				
	(0.027)	(0.029)	(0.029)	(0.034)	(0.026)	(0.021)	(0.021)	(0.021)				
B: Offense Level												
Overall Discontinuity	-0.126	-0.272	-0.318	-0.299	0.179	-0.171	-0.536+	-0.835**	0.924**	0.405*	0.431+	0.079
	(0.248)	(0.231)	(0.245)	(0.233)	(0.296)	(0.301)	(0.298)	(0.249)	(0.221)	(0.194)	(0.237)	(0.282)
Blk-Wht Diff in Disc.	0.377	0.719	1.16*	1.50*	0.253	0.281	0.461	0.913+	-0.614	-0.867+	-1.15**	-0.653
	(0.481)	(0.549)	(0.539)	(0.583)	(0.378)	(0.406)	(0.382)	(0.473)	(0.468)	(0.432)	(0.417)	(0.434)
C: Prison Sentence (months)												
Overall Discontinuity	-1.43	-3.86+	-0.049**	-0.020	2.81	1.15	-1.44	-6.88**	3.92*	-0.161	0.211	3.13
	(2.11)	(2.07)	(0.016)	(0.019)	(2.35)	(2.72)	(2.62)	(2.46)	(1.92)	(1.63)	(1.97)	(1.91)
Blk-Wht Diff in Disc.	4.55	10.08+	0.131**	0.091**	-1.25	-5.04	-2.49	6.10	-4.598	-9.42+	-11.94*	-16.72**
	(4.63)	(5.20)	(0.029)	(0.034)	(4.08)	(4.61)	(4.56)	(4.94)	(5.49)	(5.31)	(5.34)	(5.09)
D: Departures												
Overall Discontinuity									0.075**	0.076**	0.087**	0.086**
									(0.015)	(0.017)	(0.018)	(0.019)
Blk-Wht Diff in Disc.									-0.001	0.001	-0.019	-0.047
									(0.026)	(0.029)	(0.031)	(0.030)
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

Note: Regressions include polynomial time trends estimated separately before and after Booker, interacted with race, and controls 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 with 18 months of Booker (N≈57,000) or within 12 months of Booker (N≈39,000); number of observations varies slightly depending on the outcome variable. The black discontinuity is the sum of the white discontinuity and the discontinuity in the black-white difference. Standard errors are clustered on the running variable (black*month). +p<0.1 *p<0.05 **p<0.01