

Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing

Crystal S. Yang

ABSTRACT

The federal sentencing guidelines were created to reduce unwarranted sentencing disparities among similar defendants. This paper explores the impact of increased judicial discretion on racial disparities in sentencing after the guidelines were struck down in *United States v. Booker* (543 U.S. 220 [2005]). Using data on the universe of federal defendants, I find that black defendants received 2 months more in prison compared with their white counterparts after *Booker*, a 4 percent increase in average sentence length. To identify the sources of racial disparities, I construct a data set linking judges to defendants. Exploiting the random assignment of cases to judges, I find that racial disparities after *Booker* were greater among judges appointed after *Booker*, which suggests acculturation to the guidelines by judges with experience sentencing under a mandatory-guidelines regime. Prosecutors also responded to increased judicial discretion after *Booker* by charging black defendants with binding mandatory minimum sentences.

1. INTRODUCTION

Sentencing disparities by race, gender, education, and socioeconomic status are prevalent in the federal criminal justice system. Black defendants are sentenced to 5 months longer in prison than white defendants who commit similar offenses and have similar observable demographic traits

CRYSTAL S. YANG is Assistant Professor of Law, Harvard Law School. Support was provided by Harvard Law School's John M. Olin Center for Law, Economics, and Business and the Multidisciplinary Program on Inequality and Social Policy at Harvard University. I am very grateful to Lawrence Katz for detailed feedback on this project. I also thank Raj Chetty, Will Dobbie, Josh Fischman, Claudia Goldin, Louis Kaplow, Ilyana Kuziemko, Jeff Miron, J.J. Prescott, Shayak Sarkar, Max Schanzenbach, Steven Shavell, and Kate Stith; participants in the Harvard University Labor Economics/Public Finance Seminar, Harvard Law School Law and Economics Seminar, University of Virginia Law and Economics Colloquium, and American Law and Economics Association

[*Journal of Legal Studies*, vol. 44 (January 2015)]

© 2015 by The University of Chicago. All rights reserved. 0047-2530/2015/4401-0003\$10.00

and criminal histories. Male defendants are sentenced to over 5 months more in prison than similar female defendants, and defendants with lower educational attainment and income receive significantly longer sentences than otherwise similar offenders (Mustard 2001). Even in the same court, judges appear to vary significantly in their treatment of defendant race (Abrams, Bertrand, and Mullainathan 2012), which suggests that racial disparities in the criminal justice system may be a source of the overrepresentation of blacks in the prison population.

In response to concerns that judges were introducing unwarranted disparities in sentencing (Frankel 1973), Congress adopted the US sentencing guidelines under the Sentencing Reform Act (SRA) of 1984. While the guidelines reduced interjudge sentencing disparities in their early years (Anderson, Kling, and Stith 1999), many criticized them for being rigid (for example, Freed 1992; Stith 2008) and for shifting power to prosecutors in their charging and plea-bargaining decisions (Stith and Cabranes 1998; Alschuler 1978; Nagel and Schulhofer 1992).

After almost 2 decades of mandatory-guidelines sentencing, the guidelines were struck down in *United States v. Booker* (543 U.S. 220 [2005]). *Booker* greatly increased the degree of judicial discretion afforded to judges (see, for example, Berman 2005), with subsequent cases further increasing judicial discretion by reducing the degree of appellate scrutiny. Empirical work on the impact of *Booker* suggests increases in interjudge sentencing disparities (Scott 2010; Yang 2014) but has yielded mixed results on racial disparities, with some researchers finding large racial disparities in the aftermath of *Booker* (USSC 2006, 2010a) and others finding no significant impact on racial disparities in sentence length (Ulmer, Light, and Kramer 2011).

Most recently, Starr and Rehavi (2013) find no change in racial disparities in the immediate aftermath of *Booker*. In contrast, Fischman and Schanzenbach (2012) find evidence of increased racial disparities after later Supreme Court decisions that further encouraged judges to depart from the guidelines, although they attribute the disparities to the increased relevance of mandatory minimum sentences. In light of possible evidence of increasing disparities after *Booker*, the US Sentencing Commission (USSC) and policy makers have considered ways to constrain ju-

annual meeting; and anonymous referees and the editors for helpful comments and suggestions. The Transactional Records Access Clearinghouse (TRAC) at Syracuse University, in particular Sue Long, generously provided sentencing data for use in this project in my role as a TRAC Fellow. All errors are my own.

dicial discretion, such as resurrecting the mandatory guidelines (Sessions 2011).

This paper estimates the impact of increased judicial discretion via *Booker* on racial disparities in federal sentencing using data on the universe of defendants sentenced between 1994 and 2010. Comparing the sentence disparities between similar defendants within a district court before and after *Booker*, I find that *Booker* significantly increased racial disparities after controlling for extensive offender and crime characteristics. The black-white sentencing gap increased by 2 months in the post-*Booker* period, a 4 percent increase in the average sentence length and a doubling of the baseline racial gap. Increased racial disparities in sentence length can be attributed to black defendants being more likely to be sentenced above the guidelines-recommended range and less likely to be sentenced below the guidelines-recommended range, compared with similar white offenders. Even conditional on being sentenced within the guidelines range, black defendants received significantly longer sentences than similar white defendants after *Booker*.

After *Booker*, disparities also emerged by defendants' educational attainment and gender, but racial disparities persist even after accounting for differential treatment of offenders based on other observable traits. Results are also robust to controlling for offense determined at the time of arrest, which suggests that differential fact finding at the sentencing stage cannot fully explain the increase in racial disparities. I also consider the longer-term effects of *Booker* after the Supreme Court reduced the degree of appellate scrutiny for sentencing decisions in late 2007. I find evidence that the racial sentencing gap expanded most prominently after periods of more deferential appellate review, which indicates that judges are responsive to changes in the likelihood of appellate reversal.

I also examine some sources of increasing disparities after *Booker* by studying how different types of judges respond to increased judicial discretion. Many scholars suggest that judges have different sentencing philosophies (for example, Hofer, Blackwell, and Ruback 1999), which may be affected by the standard of appellate review (Fischman and Schanzenbach 2011), with correlations between sentencing practices and judicial characteristics such as race, gender, and political affiliation (Welch, Combs, and Gruhl 1988; Schanzenbach 2005; Schanzenbach and Tiller 2007, 2008). However, prior empirical research on interjudge disparity and the impact of judicial demographics on sentencing practices has been hampered by the lack of judge identifiers. Relying on aggregate district-level variation in judicial demographics can lead to biased esti-

mates if districts with different judicial compositions differ in ways that affect all judges in the district court.

I surmount these issues by utilizing a novel data set constructed for this study. Matching three data sources, I construct a data set of almost 400,000 criminal defendants linked to sentencing judges from 2000 to 2010. Given that cases are randomly assigned to judges in a district court, any difference in sentencing practices across judges can be attributable to judge differences rather than case composition. I find that increases in racial disparities after *Booker* were larger among post-*Booker*-appointed judges, even after accounting for the fact that these judges were appointed by George W. Bush. Nor are the sentencing patterns of post-*Booker*-appointed judges explained by the fact that these judges were relatively new to federal sentencing. My findings suggest that judges with experience sentencing under the guidelines may have become relatively acculturated to the guidelines regime compared with newer judges who began their tenure in a post-*Booker* regime.

I conclude by considering the impact of judicial discretion on other actors in the criminal justice system. Arrest, charge, trial, and plea-bargaining decisions made earlier in the process are all ripe avenues for unwarranted bias (Anwar, Bayer, and Hjalmarsson 2012; Rehavi and Starr 2014). After *Booker*, prosecutors commented that they were far less willing to forgo charging mandatory minimum sentences when judges ultimately sentence defendants to terms far below the guidelines-recommended minimum sentence. Consistent with this story, I find evidence that increased judicial discretion via *Booker* changed the prosecutorial treatment of statutory mandatory minimum sentences, which *Booker* left intact. Black offenders are generally more likely to be charged with mandatory minimum sentences than are similar white offenders. After *Booker*, black defendants were significantly more likely to face binding mandatory minimum sentences than were white defendants and subsequently more likely to be sentenced at the mandatory minimum, consistent with the findings of Fischman and Schanzenbach (2012). Accordingly, prosecutorial charging is likely a substantial contributor to recent increases in racial disparities.

The paper is structured as follows. Section 2 provides a brief legal background of the guidelines and the *Booker* decision. Section 3 describes the data and presents summary statistics. Section 4 describes the empirical methodology. Section 5 presents results, and Section 6 concludes.

2. LEGAL BACKGROUND

2.1. Adoption of the US Sentencing Guidelines

For over a century prior to the adoption of the guidelines, judges had virtually unfettered discretion to determine the lengths of sentences. A 1977 study of Virginia state district court judges revealed that judges applied radically different sentences to identical offenders (Austin and Williams 1977). A 1988 study of federal courts similarly found that white-collar offenders who committed similar offenses received very different sentences depending on the court in which they were sentenced (Wheeler, Mann, and Sarat 1988).

By the 1970s, the legal community and public expressed alarm at the widespread disparities in criminal sentencing that resulted from this indeterminate-sentencing regime (Frankel 1973). Some members of the public argued that judges and parole boards endangered public safety with lenient sentencing of offenders (Tonry 2005), while others decried the inequitable and arbitrary treatment of criminals. The American Friends Service Committee (1971) claimed that decreasing discretion among judges was the only way to eliminate racial discrimination in the criminal justice system.

Policy makers also recognized that judges were often “left to apply [their] own notions of the purposes of sentencing,” which led to “an unjustifiably wide range of sentences to offenders convicted for similar crimes” (US Senate Committee on the Judiciary 1983, p. 31). In order to eliminate unwarranted sentencing disparities “among defendants with similar records who have been found guilty of similar criminal conduct” (28 U.S.C. sec. 991[b][1][B]), Congress created the USSC to adopt and administer the guidelines. Part of the SRA of 1984, the guidelines were applied to all federal offenses committed after November 1, 1987, and prohibited courts from using race, sex, national origin, creed, religion, and socioeconomic status in sentencing decisions.

Under the guidelines, federal district court judges assign each defendant’s crime to one of 43 offense levels and each defendant to one of six criminal history categories. The more serious and harmful the offense, the higher the base offense level. For instance, trespass offenses are assigned a base offense level of 4, while kidnapping is assigned a base offense level of 32. From the base offense level, adjustments are made for applicable offense and defendant characteristics in order to obtain the final offense level. Under chapter 2 of the guidelines, adjustments are made on the basis of characteristics such as the amount of loss involved in the offense,

use of a firearm, and the age or condition of the victim. Under chapter 3 of the guidelines, further adjustments are made on the basis of aggravating or mitigating factors, such as obstruction of justice or a defendant's acceptance of responsibility.

The criminal history category reflects the frequency and severity of a defendant's prior criminal convictions, predictive of recidivism risk. To determine a defendant's criminal history category, a judge adds points for prior sentences in the federal system, 50 state systems, systems in all territories, and foreign or military courts. Three points are added for each prior prison term exceeding 1 year and 1 month, and 2 points are added for each prior prison term of at least 60 days but less than 1 year and 1 month. Two points are also added if the defendant committed the instant offense under any criminal justice sentence. These points are then converted to a criminal history category.

The intersection of the final offense level and criminal history category yields a narrow guidelines-recommended sentencing range (see Table A1 in the online appendix for the guidelines sentencing chart). If a judge determines that there are aggravating or mitigating circumstances that warrant a departure from the guidelines, she would have to justify her reasons for departure to the appellate court. However, the guidelines were treated as sufficiently mandatory prior to *Booker*, and judges could only consider factors such as a defendant's age, education, and employment history in deciding the sentence length for within-range sentences. The government is permitted to appeal a sentence resulting in a departure below the guidelines range, and the defendant can appeal an above-range departure.¹

2.2. Challenges to the Mandatory-Guidelines Regime

The constitutionality of mandatory-sentencing guidelines was first questioned in reference to Washington State's sentencing guidelines. In *Blakely v. Washington* (542 U.S. 296 [2004]), the Supreme Court held that the Sixth Amendment right to a jury trial prohibited judges from increasing a defendant's sentence beyond the statutory maximum on the basis of facts other than those decided by the jury beyond a reasonable doubt. As a result, Washington's mandatory-sentencing guidelines were

1. Congress has also attempted to limit unwarranted disparities in sentencing through other means. In the 1980s, Congress enacted a series of mandatory-minimum statutes directed at drugs and firearms offenses. In 2003, Congress also passed the Prosecutorial Remedies and Other Tools to End the Exploitation of Children Today (PROTECT) Act to curtail judicial departures, particularly for child sex offenses.

struck down. Shortly afterward, the reasoning of *Blakely* was applied to the US sentencing guidelines.

In *United States v. Booker*, the mandatory federal guidelines were also found unconstitutional under the Sixth Amendment. The *Booker* ruling, however, did not apply to mandatory minimum sentences enacted by Congress. Instead of invalidating the guidelines, the Supreme Court held that the guidelines would be “effectively advisory,” as opposed to mandatory. The Court explained that “district courts, while not bound to apply the Guidelines, must consult those Guidelines and take them into account when sentencing” (543 U.S. 264).

In the aftermath of *Booker*, circuit courts reached a consensus that sentencing must begin with the calculation of the applicable guidelines range. Today, after a sentencing judge has calculated the guidelines range, she must consider seven factors under 18 U.S.C. sec. 3553(a) before imposition of punishment: the nature and circumstances of the offense and the history and characteristics of the defendant, the need for the sentence imposed, the kinds of sentences available, the kinds of sentences and the sentencing ranges established, any pertinent policy statement issued by the USSC, the need to avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct, and the need to provide restitution to any victims of the offense.

Subsequent Supreme Court decisions further weakened the effect of the guidelines on criminal sentencing by reducing the degree of appellate review. In *Rita v. United States* (551 U.S. 338, 350 [2007]), the Court held that a sentence within the guidelines-recommended range could be presumed “reasonable” because a “judge who imposes a sentence within the range recommended by the Guidelines thus makes a decision that is fully consistent with the Commission’s judgment in general.” In *Gall v. United States* (552 U.S. 38 [2007]), the Court held that federal appeals courts could not presume that a sentence outside the range recommended by the guidelines was unreasonable. Concurrent with the *Gall* decision, the Court in *Kimbrough v. United States* (552 U.S. 85 [2007]) held that federal district court judges have the discretion to impose sentences outside the recommended guidelines range because of policy disagreements with the USSC, such as the disparate treatment of crack and powder cocaine offenses—the so-called 100:1 ratio.

How might *Booker* and subsequent cases that reduced the degree of appellate scrutiny affect sentencing disparities? Judges have preferences for sentencing according to their tastes but are constrained by the costs of exercising discretion (Posner 2005). As the guidelines became advisory

after *Booker*, and the standard of appellate review more deferential after *Rita*, *Gall*, and *Kimbrough*, one might see the emergence of larger disparities.

3. DATA

This paper utilizes data from three sources: the USSC, the Transactional Records Access Clearinghouse (TRAC), and the Federal Judicial Center. I describe each data set in turn.

3.1. US Sentencing Commission

I use data from the USSC on records of all federal offenders sentenced pursuant to the sentencing guidelines and policy statements of the SRA of 1984 in fiscal years 1994–2010 (October 1, 1993–September 30, 2010).² These data include demographic, guidelines application, and sentencing information on federal defendants, but defendant and judge identifiers are redacted. This information is obtained from numerous documents for every offender: indictment, presentence report, report on the sentencing hearing, written plea agreement (if applicable), and judgment of conviction.

Demographic variables include the defendant's race, gender, age, citizenship status, educational attainment, and number of dependents. Data are also provided on the primary offense type, with a total of 35 offense categories. Offense-level variables include the base offense level, the base offense level after chapter 2 adjustments, and the final offense level after chapter 3 adjustments. Criminal history variables include whether the defendant has a prior criminal record, the total number of criminal history points applied, and the final criminal history category.

For each offender, there is a computed guidelines range and a guidelines range adjusted for applicable mandatory minimum sentences. From these variables, I construct indicator variables for above-range and below-range departures from the guidelines. Information is also provided on whether the offense carries a mandatory minimum sentence under various statutes and whether departures from the statutory minimum are granted under a government substantial assistance motion. Sentencing

2. For 2002–10 data, see US Sentencing Commission, Commission Datafiles (<http://www.ussc.gov/research-and-publications/commission-datafiles>); data for 1994–2001 are on file with the author. Over 90 percent of felony defendants in the federal criminal justice system are sentenced pursuant to the Sentencing Reform Act of 1984, and all cases are assessed to be constitutional.

characteristics include the district court in which sentencing occurred (94 total) in addition to the sentencing month and year.³ Data are also available for whether a case is settled by plea agreement or trial. Sentencing outcomes include incarceration or probation, sentence length, receipt of supervised release, and length of supervised release.

I apply two sample restrictions. First, I exclude individuals with missing or invalid criminal records (offense level, criminal history category, and offense type), about 6 percent of the sample. Second, I exclude individuals missing an indicator for race, about .2 percent of the sample. Table 1 presents summary statistics for the main variables from the USSC data. Table 1 indicates that 83 percent of the defendants in the data set are incarcerated versus receiving probation. The average unconditional sentence length is approximately 50 months. Approximately 30 percent of cases carry a statutory minimum sentence, and only 4 percent of cases go to trial.

Before *Booker*, 34 percent of defendants were white, 27 percent were black, and 35 percent were Hispanic, with the share of Hispanic defendants increasing after *Booker*.⁴ The share of defendants who are not US citizens increased from 30 percent to 38 percent between the two time periods. Defendants have an average of 1.6 dependents, and almost a majority have less than a high school degree. Over 85 percent of the defendants are male. Defendants average approximately 34 years of age. Most of the defendants have had some previous interaction with the criminal justice system, as approximately 75 percent have some criminal history.

The most common offense is drug trafficking, followed by immigration offenses. Before *Booker*, drug trafficking represented about 40 percent of the cases, followed by immigration offenses, which were 14 percent of the cases. After *Booker*, the share of immigration offenses rose to over 25 percent, which likely explains the increasing share of Hispanic defendants and non-US citizens. In terms of guidelines range calculations, defendants have an average final criminal history score of 2.3 and a final offense level of 18.7. This criminal history category and offense level combination yields an average guidelines-recommended range of 30–37 months in prison.

3. The US Sentencing Commission (USSC) data prior to 2004 include information on the exact sentencing day, but this variable is not available in later years.

4. The remaining race category is for defendants whose race is classified as “other,” which is composed primarily of Native Americans.

Table 1. Summary Statistics

| Variable | Before <i>Booker</i> | After <i>Booker</i> |
|---------------------------------|----------------------|---------------------|
| Incarceration | .832 (.375) | .835 (.371) |
| Sentence length (months) | 50.608 (81.276) | 50.301 (64.717) |
| Statutory minimum applied | .296 (.457) | .298 (.457) |
| Settled by trial | .043 (.203) | .043 (.203) |
| White | .336 (.472) | .279 (.449) |
| Black | .271 (.445) | .239 (.426) |
| Hispanic | .352 (.478) | .440 (.496) |
| Non-US Citizen | .297 (.457) | .375 (.484) |
| Dependents (<i>N</i>) | 1.574 (2.162) | 1.661 (1.757) |
| Less than high school education | .445 (.497) | .495 (.500) |
| Male | .852 (.355) | .871 (.356) |
| Age (years) | 34.447 (10.782) | 35.186 (10.805) |
| Criminal history | .722 (.448) | .784 (.412) |
| Drug-trafficking offense | .405 (.491) | .351 (.477) |
| Immigration offense | .144 (.351) | .253 (.435) |
| Criminal history category (1–6) | 2.287 (1.674) | 2.503 (1.733) |
| Final offense level (1–43) | 18.698 (9.021) | 18.908 (8.845) |
| <i>N</i> | 528,076 | 372,924 |

Source. For 2002–10 data, see US Sentencing Commission, Commission Datafiles (<http://www.ussc.gov/research-and-publications/commission-datafiles>); data for 1994–2001 are on file with the author.

Note. Data are means, with standard deviations in parentheses.

3.2. Transactional Records Access Clearinghouse

Sentencing data obtained through Freedom of Information Act requests are provided by TRAC. The data do not contain defendant demographics, offense characteristics, and guidelines-application information, but defendants are linked to the sentencing judge. To link the defendant and

crime characteristics to sentencing judge, I match sentencing records from the USSC to data provided by TRAC. Matching is conducted by district court on several key variables: sentencing year, sentencing month, sentence length in months, probation length in months, amount of monetary fine, whether the case ended by trial or plea agreement, and whether the case resulted in a life sentence. For defendants sentenced prior to fiscal year 2004, I also match on the sentencing day. I successfully match approximately 60 percent of all cases from fiscal years 2000–2010.

3.3. Federal Judicial Center

To provide information on judges' characteristics, I match the USSC and TRAC combined data to judges' biographical data from the Federal Judicial Center. Federal district judges are Article III judges who serve life-term tenures. New appointments are generally made when a judge retires or dies.⁵ As of 2014, there are a total of 677 federal district court judgeships. The largest district court is the Southern District of New York, with 28 authorized judgeships. The majority of other district courts have between two and seven judgeships.

I obtain information on judges' race, gender, political affiliation of appointing president, and commission year. After applying the same sample restrictions described in Section 3.1, the final matched data set consists of 381,361 cases resulting in prison sentences from fiscal years 2000–2010.⁶ This unique data set permits an examination of judicial demographic characteristics on sentencing practices in the wake of increased judicial discretion via *Booker*. Of judges active between 2000 and 2010, 20 percent are female, and over 80 percent are white. Black judges represent approximately 9 percent of the share of all judges. Judges appointed by Democratic presidents represent 45 percent of all judges.

4. EMPIRICAL METHODOLOGY

4.1. Estimation Specification

The *Booker* case was decided on January 12, 2005, and applied immediately to all future cases and prior cases that had not reached sentencing. This paper exploits the timing of this decision and subsequent changes in

5. On a few occasions, Congress has also increased the number of judgeships in a district in response to changing population or caseload.

6. The Federal Judicial Center does not collect demographic information on judges in three districts: Guam, the Virgin Islands, and the Northern Mariana Islands.

appellate review to estimate the effect of increased judicial discretion on racial disparities in sentencing outcomes. Intuitively, I compare sentence differences between similar defendants sentenced within a district court before and after *Booker*.

The main specification is of the form

$$Y_{ijkdm} = \beta_0 + \beta_1 \times \text{Booker} \times \text{Race}_i + \beta_2 \times \text{Booker} + \beta_3 \times \text{Race}_i + \beta_4 \times \mathbf{X}_i \\ + \beta_5 \times Z_i + \text{Guide}_{ijk} + \text{Offtype}_i + \gamma_d + \delta_t + \lambda_m + \varepsilon_{ijkdm}, \quad (1)$$

where Y_{ijkdm} is a sentencing outcome for defendant i with criminal history category j and offense level k sentenced in district court d in year t and month m . Main outcomes include sentence length measured in months, a binary indicator for whether the defendant received an above-range departure, and a binary indicator for whether the defendant received a below-range departure. Additional outcomes include a binary indicator for incarceration, probation length, receipt of supervised release, term of supervised release, application of a statutory minimum, and departures from statutory minimum sentences.

The main coefficient of interest is β_1 , which captures the impact of *Booker* on racial gaps in sentencing outcomes. The term *Booker* is an indicator variable for defendants sentenced after the *Booker* decision.⁷ The term Race_i is a set of dummy variables for defendant i 's race: white, black, Hispanic, or other. The term \mathbf{X}_i comprises a vector of demographic characteristics of the defendant including gender, age, age squared, educational attainment (less than high school, high school graduate, some college, college graduate), number of dependents, and citizenship status. The term Z_i is an indicator variable for whether the offender is charged with an offense carrying a mandatory minimum sentence.⁸

The term Guide_{ijk} includes dummy variables for criminal history category j and offense level k and each unique combination of criminal history category and offense level. The interaction captures differential sentencing tendencies at each cell of the guidelines grid (258 total). To proxy for underlying offense seriousness and all aggravating and mitigating factors, I control for final offense level rather than base offense level. I also

7. For defendants sentenced in January 2005, the USSC data contain a variable denoting whether the case was heard prior to or after the *Booker* decision.

8. Controlling for the application of a mandatory minimum sentence is important because of large differences by race. Moreover, the application of mandatory minimum sentences is not endogenous to *Booker* (see Table 8). In unreported results, findings are robust to excluding controls for mandatory minimum sentences.

control for final criminal history category. The term Offtype_i is a dummy variable for offense type.

The specification also includes district court fixed effects (γ_d), sentencing-year fixed effects (δ_t), and sentencing-month fixed effects (λ_m). Because sentencing-year fixed effects are included, the Booker indicator is identified off variation in the dependent variable during the 11 days in January 2005 prior to when the case was decided. However, the coefficient on $\text{Booker} \times \text{Race}$ estimates the average effect over the entire post-*Booker* period since sentencing year is not fully interacted with the defendant's race.⁹ Race-specific trends are included to account for pre-existing trending differences in sentencing outcomes between defendants of different races. All standard errors are clustered at the district court level to account for serial correlation.

To analyze the differential sentencing practices of certain types of judges, I compare how judges differ in their relative treatment of similar black and white defendants in response to increased judicial discretion, compared with other judges in the same district court. Because cases are randomly assigned to judges in a district court, judge identifiers allow me to compare judges in the same court, which captures judges' differences in sentencing rather than different caseloads.¹⁰

I identify the sources of increasing racial disparities after *Booker* using the following specification:

$$\begin{aligned} Y_{ijkdm} = & \beta_0 + \alpha_1 \times \text{Judge}_i \times \text{Race}_i \times \text{Booker} + \alpha_2 \times \text{Judge}_i \times \text{Race}_i \\ & + \alpha_3 \times \text{Judge}_i \times \text{Booker} + \alpha_4 \times \text{Judge}_i + \beta_1 \times \text{Booker} \times \text{Race}_i \\ & + \beta_2 \times \text{Booker} + \beta_3 \times \text{Race}_i + \beta_4 \times \mathbf{X}_i + \beta_5 \times \mathbf{Z}_i + \text{Guide}_{ijk} \\ & + \text{Offtype}_i + \gamma_d + \delta_t + \lambda_m + \varepsilon_{ijkdm}, \end{aligned} \quad (2)$$

where Judge_i includes judicial demographics such as race, gender, political affiliation, and an indicator for pre- versus post-*Booker* appointment. The coefficient α_1 captures the impact of particular judicial characteristics on racial disparities in the wake of *Booker*.

4.2. Potential Threats to Identification

There are two primary threats to identification. The first is if *Booker* was associated with a change in defendant characteristics that affects sentenc-

9. In unreported results, findings are robust to the exclusion of sentencing-year fixed effects.

10. Most courts use a random drawing (see Administrative Office of the US Courts, Frequently Asked Questions [<http://www.uscourts.gov/Common/FAQS.aspx>]). I also test for random assignment in Section 5.5.

Table 2. Defendants' Criminal Characteristics

| | Criminal History (1) | Total Criminal Points (2) | Criminal History Category (3) | Base Offense (4) | Final Offense (5) |
|-------------------|-------------------------|---------------------------------|-------------------------------------|---------------------|----------------------|
| Booker × Black | .0124 (.00687) | .0962 (.0794) | .0442 (.0238) | .326 (.246) | -.165 (.211) |
| Booker × Hispanic | -.00898 (.00786) | .289 (.360) | -.0192 (.0417) | .628 (.414) | -.315 (.295) |
| Booker × Other | .0175 (.0106) | .190 (.181) | .0236 (.0490) | .948* (.365) | .567 (.328) |
| Black | .0738** (.00659) | 1.714** (.108) | .591** (.0268) | 2.616** (.303) | 1.831** (.253) |
| Hispanic | -.0221* (.00857) | -.533** (.183) | -.0605* (.0263) | 2.308** (.428) | 1.702** (.306) |
| Other | -.0588** (.0152) | -.950** (.189) | -.255** (.0523) | .528 (.472) | .218 (.430) |
| Booker | -.00567 (.0135) | -.280 (.221) | -.0420 (.0391) | .360 (.380) | 1.153** (.355) |
| N | 720,189 | 906,331 | 908,189 | 905,494 | 908,189 |
| R ² | .223 | .195 | .304 | .144 | .139 |

Source. For 2002–10 data, see US Sentencing Commission, Commission Datafiles (<http://www.ussc.gov/research-and-publications/commission-datafiles>); data for 1994–2001 are on file with the author. The criminal history indicator is available only for 1999–2010.

Note. Regressions for criminal history, total criminal history points, and criminal history category contain controls for final offense level and offense type. Regressions for base and final offense level control for criminal history category. Regressions also contain district, sentencing-year, and sentencing-month fixed effects. Standard errors, in parentheses, are clustered at the district level.

* Significant at the 5 percent level.

** Significant at the 1 percent level.

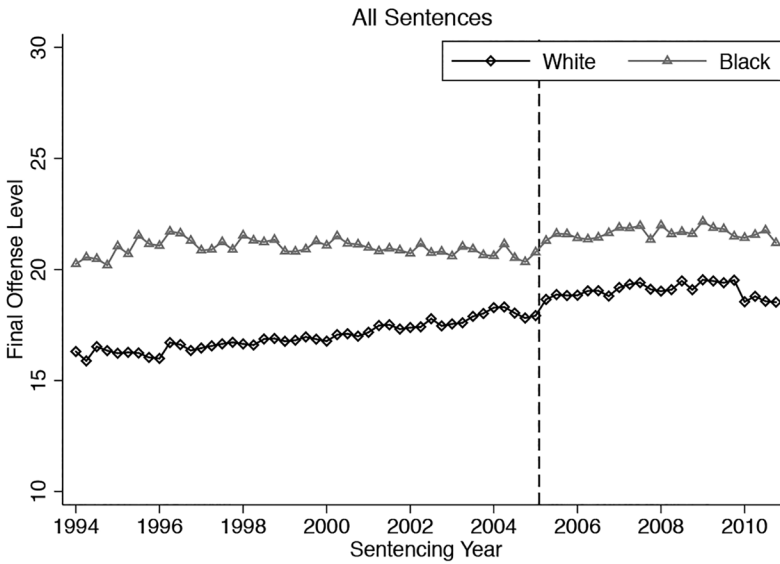


Figure 1. Defendants' offense severity by race

ing decisions. In Table 2, I test whether *Booker* affected defendants' criminal history and find no change in criminal history measures or any differential change by defendant race after *Booker*.

Similarly, a threat to identification arises if offense levels changed after *Booker* with no change in real offense severity. If judges were less concerned with deflating white defendants' offense levels in order to justify lower sentences, relatively lower offense levels for black defendants compared with white defendants after *Booker* may mechanically generate the appearance of racial disparities. To address this potential endogeneity, previous researchers have either excluded any control for offense severity or controlled for base offense level rather than final offense level (Fischman and Schanzenbach 2012). Excluding any measure of offense level as a control is potentially problematic if there are underlying trending differences in offense severity by race over the time period.

Indeed, Figure 1 documents the presence of differential trends in final offense levels, with offense severity for black defendants remaining relatively constant and offense severity for white offenders increasing over time.¹¹ Thus, failing to control for exogenous trending racial differences

11. Data points in Figure 1 are quarterly averages. See the online appendix for trends in base offense levels, which follow a similar pattern.

in offense severity would bias any finding of racial disparity. Controlling for differential trends, I find that offense-level severity did not change differentially by offender race after *Booker* in columns 4 and 5 of Table 2. However, final offense levels increased after *Booker* for all offenders, which is potentially reflective of less judicial concern with deflating offense levels. In preferred specifications, I control for final offense levels to capture important differences in crime severity, such as the use of a weapon during the underlying offense. As discussed in Section 5, results are robust to not controlling for offense severity, controlling for base offense levels, and measures of offense determined at the time of arrest.

A second potential threat to identification arises if *Booker* was associated with changes in the types of defendants that reached the sentencing stage. For instance, if prosecutors disproportionately dropped or dismissed charges against marginal black defendants after *Booker*, the remaining black defendants at the sentencing stage might receive longer sentences compared with similar white offenders. To address potential changes in selection prior to the sentencing stage, I test the likelihood of guilty pleas, dropped charges, and deferred prosecutions against black defendants compared with similar white defendants after *Booker* using data on all federal arrests and bookings from 1994 to 2009.¹² Table A2 in the online appendix indicates no significant changes in the rates at which black offenders plead guilty or the likelihood of dropped charges or deferred prosecution compared with similar white offenders, which suggests no significant changes in selection prior to sentencing.¹³

5. RESULTS

5.1. Sentence Length

Figure 2 presents trends in sentence length by defendants' race in the raw data using quarterly averages. Figure 2 indicates preexisting trending differences in sentence lengths across defendants of different races. Average sentence lengths were relatively constant for black defendants before *Booker* but increased over time for white defendants. The gap in sentence length between black and white defendants changed after *Booker* as

12. Data are from US Department of Justice (1994–2009), which covers all offenders in the custody of the US Marshals Service.

13. A deferred prosecution occurs when a prosecutor agrees to not file charges in exchange for the defendant taking specified actions, such as payment of fines and continued cooperation during the investigation.

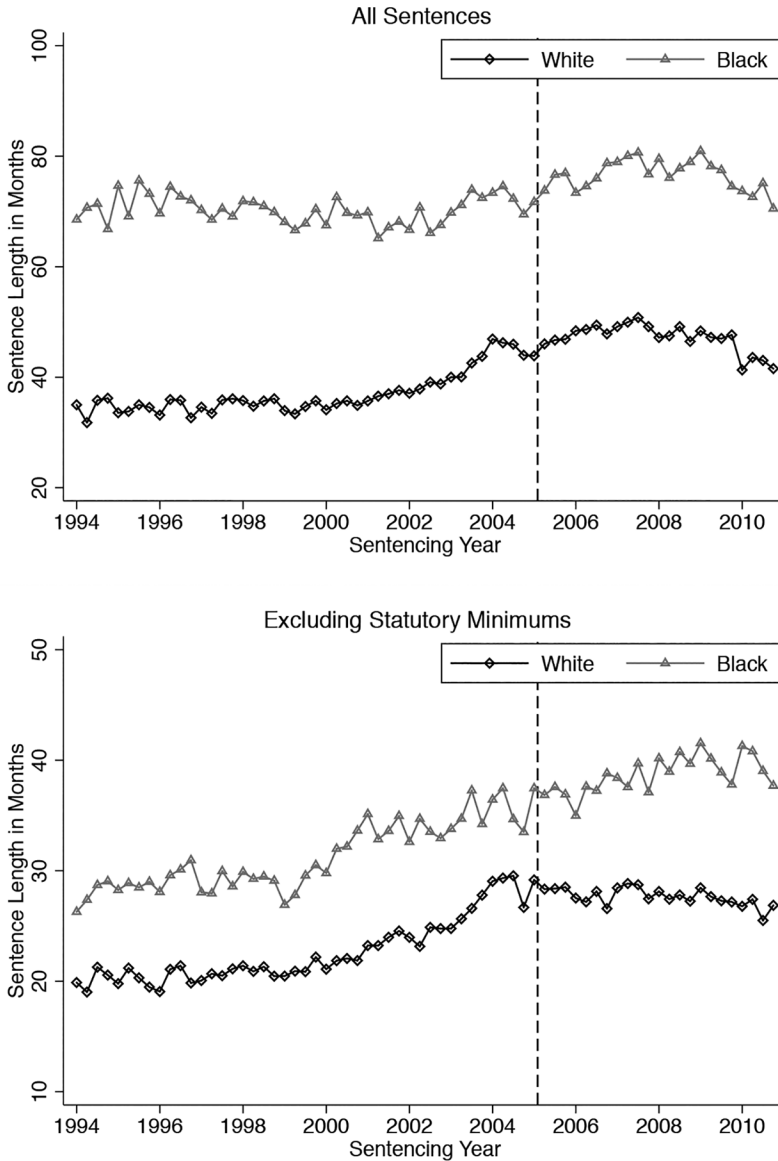


Figure 2. Defendants' sentence lengths in months by race

sentence lengths for black and white defendants diverged. The evidence is even more striking when excluding cases with mandatory minimum sentences, as it is apparent that sentence lengths for white defendants de-

creased after *Booker* while sentence lengths for black defendants continued to rise, increasing racial disparities in sentence length.

Table 3 presents regression results for the impact of increased judicial discretion via *Booker* on disparities in sentence length. Column 1 presents results controlling only for the Booker indicator, defendant race, an interaction between the two, and race-specific trends; column 2 adds demographic controls. In addition to controls for offense type, column 3 includes all interactions between criminal history category and final-offense-level severity as well as an indicator for the application of a mandatory minimum sentence. Column 4 presents results for the preferred specification.

Across all specifications, the coefficients on Booker interacted with defendant race suggest growing racial disparities after *Booker*.¹⁴ The preferred specification indicates that black offenders received sentences that are 1.9 months longer than those of white offenders after *Booker*, an increase of over 75 percent in the racial gap in sentence length and a 4 percent increase in the average sentence length for all offenders. The coefficients on defendant race are consistent with prior findings regarding disparities in sentencing. On average, black offenders receive sentences that are approximately 2.5 months longer than those of comparable white offenders, who are the omitted category. Hispanic offenders receive prison sentences that are over 1.9 months longer than those of similar white offenders.

Table 4 investigates the sensitivity of the results to alternative specifications that include district-by-sentencing-year fixed effects to capture unobserved geographically correlated shocks, control for base offense level instead of final offense level, utilize log sentence as the dependent variable, and exclude immigration cases, which increased in the post-*Booker* period. Racial disparities are robust across all these specifications. A specification using *Blakely* as the treatment date (column 5) tests whether judges anticipated the holding in *Booker* and endogenously responded before the decision, which potentially biases the main findings downward. Results using the timing of *Blakely* are almost identical to the main *Booker* results.¹⁵

14. The coefficient on *Booker* \times Black is almost significant at the 10 percent level in column 2 of Table 3 ($p = .14$) and is not statistically different from the coefficient under the other specifications.

15. In Table A3 in the online appendix, I also replicate the preferred specification for eight placebo periods prior to *Booker*. None of the coefficients are statistically significant at the 5 percent level, which suggests that increases in racial disparities in the aftermath of *Booker* are unlikely spurious.

Table 3. Sentence Length in Months

| | (1) | (2) | (3) | (4) |
|---|--------------------|--------------------|--------------------|-------------------|
| Booker × Black | 3.592* (1.555) | 2.151 (1.440) | 1.696* (.664) | 1.919** (.652) |
| Booker × Hispanic | -.152 (1.263) | -.329 (1.295) | .0373 (.792) | -.699 (.780) |
| Booker × Other | .735 (2.169) | -.175 (2.069) | -1.106 (1.067) | -1.088 (1.002) |
| Black | 27.82** (2.477) | 24.57** (2.258) | 2.652** (.861) | 2.477** (.607) |
| Hispanic | -3.661 (3.213) | 1.149 (2.399) | 1.318 (1.011) | 1.867** (.611) |
| Other | -3.236 (2.387) | .667 (1.985) | 1.604 (1.013) | 2.810** (.858) |
| Booker | 2.032 (1.064) | 2.410* (1.007) | -3.065** (.494) | -1.627 (1.043) |
| Demographic controls | No | Yes | Yes | Yes |
| Offense controls | No | No | Yes | Yes |
| District, year, and month fixed effects | No | No | No | Yes |
| N | 915,048 | 901,780 | 901,000 | 901,000 |
| R ² | .036 | .075 | .556 | .560 |

Source. For 2002–10 data, see US Sentencing Commission, Commission Datafiles (<http://www.ussc.gov/research-and-publications/commission-datafiles>); data for 1994–2001 are on file with the author.

Note. All regressions include race trends. Standard errors, in parentheses, are clustered at the district level.

* Significant at the 5 percent level.

** Significant at the 1 percent level.

Table 4. Sentence Length: Robustness Checks

| | District × Year Fixed Effects (1) | Base Offense (2) | Log Sentence (3) | Excludes Immigration (4) | Blakely (5) | Arrest Offense (6) |
|-------------------|---|---------------------|---------------------|--------------------------------|-------------------|-----------------------|
| Booker × Black | 1.407* (.610) | 2.145** (.769) | .0335** (.0125) | 1.842** (.656) | 1.851* (.728) | 2.583* (1.148) |
| Booker × Hispanic | -.326 (.578) | -1.216 (.779) | .0192 (.0100) | -.476 (.972) | -.840 (.861) | -3.919** (1.101) |
| Booker × Other | -.680 (.864) | -.427 (1.162) | -.0223 (.0220) | -1.012 (1.059) | -1.073 (1.055) | .451 (2.035) |
| Black | 2.812** (.585) | 2.526** (.674) | .0513** (.00972) | 2.613** (.605) | 2.448** (.662) | 4.297** (.941) |
| Hispanic | 1.685** (.543) | .308 (.633) | .0365** (.00925) | 1.809** (.664) | 1.980** (.664) | 4.382** (.823) |
| Other | 2.977** (.721) | 2.984** (.908) | .0661** (.0197) | 2.749** (.951) | 2.839** (.981) | 2.861 (1.525) |
| Booker | -1.659 (.991) | 1.009 (1.135) | -.0718** (.0167) | -2.302 (1.200) | -1.041 (.591) | -4.622** (1.743) |
| N | 901,000 | 899,460 | 749,394 | 735,458 | 901,000 | 389,401 |
| R ² | .563 | .449 | .733 | .571 | .560 | .337 |

Source. For data in columns 1–5 for 2002–2010, see US Sentencing Commission, Commission Datafiles (<http://www.ussc.gov/research-and-publications/commission-datafiles>); data for 1994–2001 are on file with the author. Data in column 6 are from the linked arrest-through-sentencing data set for 1994–2009 (US Department of Justice 2012).

Note. All regressions contain district, sentencing-year, and sentencing-month fixed effects. Standard errors, in parentheses, are clustered at the district level.

* Significant at the 5 percent level.

** Significant at the 1 percent level.

Table 4 also tests the robustness of my results using a more plausibly exogenous measure of offense severity—the arrest offense. Using data from the US Marshals Service, the Executive Office for US Attorneys, the Administrative Office of the US Courts, and the USSC and linking files provided by the Bureau of Justice Statistics, I match federal defendants from the arrest through sentencing stage from 1994 to 2009.¹⁶ Using this linked data set, I determine for each sentenced offender a highly detailed offense type determined at the time of arrest (399 total), exogenous to the sentencing stage.¹⁷ Column 6 replicates the preferred specification controlling for arrest offense rather than final offense level. Results are robust to controlling for arrest offense, which indicates that recent increases in racial disparities are likely not driven by endogenous offense-level determinations.

While racial disparities in sentence length have increased as a whole, a more disaggregated analysis reveals that the growing racial disparities after *Booker* do not appear uniformly across all offenses. Table A4 in the online appendix presents results on sentence lengths disaggregated into the most prevalent seven offenses, which make up 84 percent of all offenses in the data set. Racial disparities increased significantly among defendants convicted of drug-trafficking offenses, controlling for primary type of drug.¹⁸ Black defendants convicted of drug offenses received sentences that are an additional 2.2 months longer than those of white defendants after *Booker*. Given that almost 70 percent of drug offenders receive a mandatory minimum sentence, the increase in racial disparities in drug offenses after *Booker* may reflect differential application of mandatory minimum sentences, explored further in Section 5.6.

5.2. Departures from the Guidelines

Table 5 presents results on how *Booker* impacted departures from the guidelines. Column 1 replicates the preferred specification from Table 3.

16. For descriptions of the data, see Bureau of Justice Statistics, Program Resource Guide (<http://www.icpsr.umich.edu/icpsrweb/content/NACJD/guides/fjsp.html>); for the linking files, see US Department of Justice (2012). Because fewer than 50 percent of all sentenced defendants can be linked from arrest to sentencing, I do not employ the linked sample in the main results because of potential sample selection bias. However, the main results in Table 3 are similar in this subset of linkable cases.

17. I am unable to test whether arrest offense is endogenous to *Booker* because detailed offense types are categorical rather than a continuous measure of offense severity.

18. Because fiscal year 2010 is included in the data, results on drug offenses may capture some of the early effects of the Fair Sentencing Act of 2010, which reduced the crack-to-powder-cocaine sentencing disparity to 18:1.

Table 5. Sentencing Departures from the Guidelines

| | Sentence (1) | Below Range (2) | Above Range (3) | Sentence within Range (4) |
|--------------------------------|-------------------|----------------------|----------------------|---------------------------------|
| <i>Booker only:</i> | | | | |
| Booker × Black | 1.919** (.652) | -.0115 (.00610) | .00712** (.00219) | .829** (.241) |
| Booker × Hispanic | -.699 (.780) | -.0272 (.0159) | .00203 (.00174) | .187 (.103) |
| Booker × Other | -1.088 (1.002) | .00388 (.0118) | -.00318 (.00333) | .0357 (.281) |
| Black | 2.477** (.607) | -.0221** (.00411) | -.00129 (.00121) | -.366** (.121) |
| Hispanic | 1.867** (.611) | -.000985 (.0130) | -.00218 (.00159) | .000761 (.0962) |
| Other | 2.810** (.858) | -.0400** (.0145) | .00758* (.00301) | .0865 (.190) |
| Booker | -1.627 (1.043) | .115** (.0112) | .00719 (.00384) | -.187 (.241) |
| <i>Booker, Gall/Kimbrough:</i> | | | | |
| Booker × Black | 1.623* (.645) | -.00810 (.00616) | .00628** (.00217) | .735** (.229) |
| Booker × Hispanic | -.818 (.745) | -.0213 (.0153) | .00220 (.00173) | .151 (.103) |

| | | | | |
|---------------------------|--------------------|----------------------|---------------------|-------------------|
| Booker × Other | -1.092 (.909) | .00519 (.0117) | -.00219 (.00329) | .126 (.296) |
| Gall/Kimbrough × Black | 3.176** (.896) | -.0251** (.00828) | .0108** (.00286) | 1.187** (.350) |
| Gall/Kimbrough × Hispanic | -.247 (1.100) | -.0633** (.0214) | .000419 (.00267) | .313* (.148) |
| Gall/Kimbrough × Other | -1.370 (1.815) | .00452 (.0166) | -.00966 (.00570) | -.418 (.309) |
| Booker | -1.503 (1.023) | .112** (.0115) | .00725 (.00387) | -.149 (.243) |
| Gall/Kimbrough | -2.647* (1.243) | .153** (.0114) | .00129 (.00422) | -.127 (.334) |
| N | 901,000 | 735,032 | 901,000 | 475,810 |
| R ² | .560 | .273 | .040 | .980 |

Source. For 2002–10 data, see US Sentencing Commission, Commission Datafiles (<http://www.uscc.gov/research-and-publications/commission-datafiles>); data for 1994–2001 are on file with the author.

Note. All regressions contain controls for offense type, dummies for each offense level and criminal history combination, and district, sentencing-year, and sentencing-month fixed effects. Standard errors, in parentheses, are clustered at the district level.

* Significant at the 5 percent level.

** Significant at the 1 percent level.

For the results for below-range departures, which occur approximately 40 percent of the time, I separate the effect of judicial departures from government motions by excluding downward departures that are the result of a government-sponsored substantial assistance motion. Below-range departures exhibit a discontinuous increase at the timing of *Booker* by over 11 percent for white defendants. However, black defendants were 1.2 percent less likely than similar white offenders to be sentenced below range after *Booker* ($p = .06$). Racial disparities also increased in the rate of above-range departures, which occur approximately 2 percent of the time. Column 3 indicates that after *Booker*, black defendants were .7 percent more likely than white defendants to receive an above-range departure, an increase of more than 30 percent from the mean rate.

Column 4 of Table 5 suggests that, conditional on being sentenced within range, black offenders received a .8-month-longer sentence than their white counterparts after *Booker*. Similarly, Hispanic defendants received a .2-month-longer sentence than did similar white offenders after *Booker*. Recall that prior to *Booker*, judges were generally not allowed to consider factors such as the defendant's age, education, physical or mental problems, or family in making sentencing decisions, except for within-range sentences. The finding that disparities increased after *Booker* even for the subset of within-range sentences suggests that disparities are not driven solely by the ability of judges to consider various unobservable factors in the aftermath of *Booker*. Thus, it appears that increased racial disparities in sentencing after *Booker* occurred in the differential application of upward and downward departures as well as disparate sentence lengths for within-range sentences.¹⁹

5.3. Increasing Disparities in Other Characteristics of Defendants

The previous results identify growing racial disparities in sentence length and departures from the guidelines after *Booker*. However, the increase in racial disparities after *Booker* may have been driven by harsher treatment of other characteristics that are associated with black defendants.

19. An analysis of other sentence outcomes is presented in Table A5 in the online appendix. Black offenders are generally more likely than white offenders to be incarcerated, and even more likely after *Booker*. Probation lengths by defendant race did not change significantly after *Booker*. However, length of supervised release (served after imprisonment) changed substantially. Black defendants generally receive almost 2 more months of supervised release than similar white defendants. After *Booker*, black defendants received about 1.7 fewer months of supervised release than white defendants. The divergent changes in racial disparities in sentence length and supervised release length after *Booker* may be a result of judges replacing sentences for supervised release time for black defendants.

Table 6. Disparities in Sentence Length by Other Characteristics

| | (1) | (2) | (3) |
|---------------------------|--------------------|--------------------|--------------------|
| Booker × Black | 1.919** (.652) | 1.466* (.671) | 1.261 (.669) |
| Black | 2.477** (.607) | 2.605** (.618) | 2.687** (.622) |
| Booker × Non-US Citizen | | -2.814** (.619) | -2.719** (.683) |
| Non-US Citizen | 2.747** (.531) | 3.849** (.717) | 3.813** (.714) |
| Booker × Some College | | -1.775** (.520) | -1.101* (.497) |
| Some College | -1.432** (.273) | -.762 (.412) | -1.018* (.408) |
| Booker × College Graduate | | -3.282** (.767) | -2.302** (.732) |
| College Graduate | -1.437** (.283) | -.212 (.430) | -.610 (.441) |
| Booker × Female | | -2.679** (.411) | -2.023** (.395) |
| Female | -4.143** (.323) | -3.112** (.360) | -3.336** (.361) |
| R ² | .560 | .560 | .561 |

Source. For 2002–10 data, see US Sentencing Commission, Commission Datafiles (<http://www.ussc.gov/research-and-publications/commission-datafiles>); data for 1994–2001 are on file with the author.

Note. Column 1 shows the baseline results. Column 2 includes interactions between a *Booker* indicator and citizenship status, educational attainment, number of dependents, gender, and age. Column 3 adds interactions between a *Booker* indicator and offense type. All regressions contain controls for offense type, dummies for each offense level and criminal history combination, and district, sentencing-year, and sentencing-month fixed effects. Standard errors, in parentheses, are clustered at the district level. $N = 901,000$.

* Significant at the 5 percent level.

** Significant at the 1 percent level.

For instance, if black defendants disproportionately have lower levels of educational attainment, and judges took a harsher sentencing stance on less-educated defendants after *Booker*, racial disparities may mechanically increase. To account for possible disparities driven by other demographic and crime characteristics, I include full interactions between the *Booker* indicator and a variety of relevant observables in Table 6.

Column 1 of Table 6 replicates the preferred specification from Table 3 to show the baseline results. Significant racial disparities are robust to the inclusion of controls for citizenship status, educational attainment, number of dependents, gender, and age. However, racial differences in

sentencing are not the only disparities that emerged after *Booker*. Column 2 reveals decreasing disparities among defendants based on citizenship status, with non-US citizens receiving relatively shorter sentences after *Booker* than similar US citizens. The results also suggest growing disparities by educational attainment. After *Booker*, defendants with some college and those with a college degree received sentences that were approximately 2 months shorter than those of their less-educated counterparts. Disparities also increased by gender, with female defendants receiving even shorter sentences than similar male defendants after *Booker*.

The results in column 3 account for additional disparities attributable to differential treatment of offense type after *Booker*. Accounting for these additional controls eliminates the statistical significance of the coefficient on *Booker* × Black at the 5 percent level ($p = .06$), but the coefficient remains economically large, and the magnitude of racial disparities is not statistically different across specifications. None of the coefficients on offense type interacted with the *Booker* indicator are significant and are thus not reported, which suggests that judges were not sentencing differentially across offenses in the aftermath of *Booker*.²⁰ Overall, these results reveal growing disparities among a variety of demographic characteristics after *Booker*, but racial disparities are robust to differential treatment of defendants by other factors. Despite increasing disparities by citizenship, educational attainment, and gender, racial disparities persist.

5.4. How Constraining Is Appellate Review? Evidence from *Rita*, *Gall*, and *Kimbrough*

Booker changed the legal landscape by invalidating the mandatory nature of the guidelines, but the series of Supreme Court decisions that followed also changed the standard of appellate review. In the first 2.5 years after *Booker*, judges were no longer bound to the guidelines yet still faced a potentially high level of appellate scrutiny. Beginning in late 2007, the *Rita* presumption of reasonableness for within-range sentences provided judges with a safe harbor from appellate scrutiny. By *Gall* and *Kimbrough*, the Court made clear that the advisory guidelines were truly advisory by removing the presumption of unreasonableness for sentences

20. To further test the robustness of the results, I explore whether increasing racial disparities may be mechanically driven by black defendants being less likely to show remorse for their crimes. I capture this through courts' decisions to reduce a defendant's offense level by either 2 or 3 points through the acceptance-of-responsibility provision. In unreported results, I find that lack of remorse as proxied by acceptance of responsibility does not explain the growing racial disparities in the aftermath of *Booker*.

outside the guidelines range. Accordingly, judges constrained by appellate review would be most free to deviate in the aftermath of *Gall* and *Kimbrough*.

Moreover, the timing of *Gall* and *Kimbrough* was much more of a surprise than that of *Booker* and thus is a more convincing quasi experiment. Given the exogenous timing of *Gall* and *Kimbrough*, I also separately test for the impact of *Booker* by splitting the post-*Booker* time period into the period between *Booker* and *Gall/Kimbrough* and a post-*Gall/Kimbrough* period. Table 5 reveals that while racial disparities generally increased in the period from *Booker* to *Gall/Kimbrough*, the magnitudes of disparities increased further after *Gall/Kimbrough* in all outcomes. In particular, racial disparities in below-range departures appeared primarily after *Gall/Kimbrough* further encouraged judges to depart.

5.5. Free at Last? Effects of Judicial Sentencing Philosophies and Experience

While disparities in sentencing outcomes increased in the wake of *Booker*, the response to increased judicial discretion may differ by judges' sentencing philosophies and experience. In particular, judges appointed before *Booker* may sentence differently than judges appointed after *Booker*. Judges with substantial experience sentencing under the mandatory-guidelines regime may have become acculturated to the guidelines and less likely to change their sentencing practices in the aftermath of *Booker*.

Since *Booker*, there have been 123 confirmed judicial appointments to US district courts through the end of the fiscal sentencing year 2010. The judges appointed prior to 2009 were appointed by George W. Bush, and the remaining judges were appointed by Barack Obama. However, all Obama appointees began active service following the end of fiscal year 2009, so I cannot identify the sentencing patterns of a large enough sample of new Democratic-appointed judges. In the matched data from 2000–2010, post-*Booker*-appointed judges have sentenced a growing share of criminal defendants, to over 10 percent of cases in fiscal year 2010.

Recall that random assignment of cases to judges is necessary to compare sentencing practices of judges in a district court. According to the Administrative Office of the US Courts, “[t]he majority of courts use some variation of a random drawing” as prescribed by local court orders.²¹

21. See Administrative Office of the US Courts, Frequently Asked Questions: Federal Judges (<http://www.uscourts.gov/Common/FAQS.aspx>).

However, random assignment may be violated in some instances. For example, senior-status judges with reduced caseloads may select the types of cases they hear during the year.

To exclude senior-status judges who may not obtain cases through a random-assignment process, I exclude judges who were formally retired prior to 2000, cases heard by senior-status judges, and judges and district courthouses with annual caseloads of fewer than 25 cases. To ensure that I include only courthouses with random assignment of cases, I then test for random assignment by district courthouse using the matched USSC, TRAC, and Federal Judicial Center data from 2000 to 2010 for a set of five predetermined defendant characteristics: gender, age, an indicator for black defendants, number of dependents, and an indicator for less than a high school degree. I regress each characteristic on district courthouse by sentencing-year fixed effects, sentencing-month fixed effects, and judge fixed effects. I test the hypothesis of no judge effects (the null hypothesis) using an *F*-test for whether the judge fixed effects are equal to 0 using seemingly unrelated regression (Autor and Houseman 2010). The *p*-values for these tests by district courthouse are presented in Table A6 in the online appendix. I exclude all courthouses with *F*-test *p*-values less than .05, but the results are robust to other cutoffs. The subsample of district courts with random case assignment includes 67 courts representing about 50 percent of the matched cases from 2000 to 2010.²²

Table 7 presents the results for sentence length using this subsample of district courts. Column 1 controls for an interaction between defendant race, the Booker indicator, and an indicator variable equal to one for judges appointed after *Booker* in addition to the interaction between defendant race and the Booker indicator.²³ The double interaction term measures the different sentencing practices of post-*Booker*-appointed judges on disparities in sentencing compared with those of pre-*Booker*-appointed judges in the aftermath of *Booker*. The results suggest that racial disparities after *Booker* were particularly driven by judges appointed after *Booker*. Relative to their colleagues, these new judges sentenced black defendants to an additional 5.5 months in prison compared with similar white defendants.

Moreover, greater racial disparities among post-*Booker*-appointed judges cannot be fully explained by the fact that these judges were ap-

22. Main results from Table 3 are robust to using the random or full matched sample.

23. Note that because all “new” judges were appointed after *Booker*, in this instance the double interaction is identical to an interaction between the defendant’s race and Post-Booker Judge.

Table 7. Sentencing Patterns for Judges Appointed after *Booker*: Subsample of Random District Courts

| | (1) | (2) | (3) | (4) |
|------------------------------|--------------------|--------------------|--------------------|--------------------|
| Post-Booker Judge | .796 (1.433) | .326 (1.394) | .134 (1.478) | .658 (1.441) |
| Post-Booker Judge × Black | 5.451* (2.430) | 4.381 (2.583) | 4.751* (2.429) | 5.007* (2.545) |
| Post-Booker Judge × Hispanic | -1.457 (1.762) | -1.810 (1.736) | -1.666 (1.790) | -1.045 (1.705) |
| Post-Booker Judge × Other | 4.545 (5.390) | 3.116 (5.407) | 5.016 (5.127) | 2.947 (5.639) |
| Booker × Black | .817 (.983) | 1.336 (1.039) | 1.654 (1.217) | -.325 (1.559) |
| Booker × Hispanic | 1.472 (.827) | 1.664* (.820) | 1.719 (.958) | 1.283 (1.313) |
| Booker × Other | -1.146 (1.705) | -3.342* (1.640) | .0603 (1.835) | 2.051 (2.900) |
| Black | 3.974** (.839) | 3.002** (.934) | 4.291** (.897) | 5.665** (1.306) |
| Hispanic | .988 (.819) | .700 (.838) | .914 (.843) | .726 (1.087) |
| Other | 4.642** (1.289) | 5.716** (1.330) | 3.614** (1.267) | 3.390 (1.861) |
| Booker | -4.988* (2.136) | -4.863* (2.176) | -4.277 (2.194) | -5.225* (2.211) |
| R ² | .760 | .761 | .761 | .760 |

Source. Data are from the matched 2000–2010 US Sentencing Commission, Transactional Records Access Clearinghouse, and Federal Judicial Center data for courts with random assignment, excluding judges who formally retired prior to 2000.

Note. All regressions contain controls for offense type, dummies for each offense level and criminal history combination, and district, sentencing-year, and sentencing-month fixed effects. Standard errors, in parentheses, are clustered at the district level. $N = 180,789$.

* Significant at the 5 percent level.

** Significant at the 1 percent level.

pointed by Bush. In column 2, I include all interactions between the defendant's race, the Booker indicator, and an indicator variable for all Bush appointees, those appointed before and after *Booker*. The coefficients on Post-Booker Judge and its interactions with defendant race fall slightly in magnitude ($p = .09$) but are not statistically different from those presented in column 1.²⁴

Different sentencing philosophies and practices between judges may be driven not only by experience under a mandatory-guidelines regime

24. I cannot rule out the possibility that Bush appointed more conservative judges after *Booker*.

but also by other personal preferences. To proxy for sentencing philosophies, column 3 adds additional controls for the judge's gender, race, and political affiliation interacted with the defendant's race and the Booker indicator. However, the coefficients on Post-Booker Judge interacted with defendant race remain largely unchanged, which suggests that post-*Booker*-appointed judges were more responsive than their peers to a shift to advisory guidelines.

Furthermore, new judges did not sentence differently from their more experienced colleagues, either before or after *Booker*. In column 4, I present main results including a tenure variable that captures the number of years on the bench and an interaction between tenure and defendant race.²⁵ Tenure is not predictive of racial disparities and thus is unreported. The results suggest that a judge's experience alone cannot explain interjudge differences in sentencing and that the results are not driven by a new-judge effect. Instead, the results suggest that exposure to sentencing under a mandatory-guidelines regime may drive the differential sentencing patterns between judges appointed before and after *Booker*.²⁶

Given that cases are randomly assigned in a district, it is unlikely that these post-*Booker*-appointed judges were assigned cases in which black defendants deserved longer sentences than their observably similar white counterparts. The findings are also not the result of differential fact finding, as I find no difference in defendants' offense levels between post-*Booker*-appointed judges and pre-*Booker*-appointed judges. Furthermore, the results are not driven by prosecutors being more likely to charge mandatory minimum sentences when cases were assigned to post-*Booker*-appointed judges (see Table A7 in the online appendix).

5.6. Response of Prosecutors to Increased Judicial Discretion

While the disparities I estimate do not capture the compounded disparities that can result at each stage of the criminal process, I next explore the impact of increased judicial discretion on changes in prosecutorial decisions to charge mandatory minimum sentences. Given that *Booker* left congressionally enacted statutory minimums intact, one would not necessarily expect judicial treatment of mandatory minimum sentences to change in the aftermath of *Booker*. However, prosecutors may strategically respond to increased judicial discretion after *Booker* if they want to

25. In unreported results, I compare Bush's post-*Booker* appointees to the only other cohort of recent judges in the sample—new judges appointed by William J. Clinton after 2000. These new judges do not exhibit racial disparities in their sentencing practices either before or after *Booker*.

26. Results are also robust to using the full matched sample.

bind judges from departing downward. After *Booker*, prosecutors commented that they were far less willing to forgo charging mandatory minimum sentences because judges may ultimately sentence defendants below the guidelines minimum.

Table 8 presents results suggesting that prosecutorial discretion after *Booker* did not differentially affect black and white defendants in terms of charging offenses that carry mandatory minimum sentences, although black and Hispanic defendants are far more likely to receive them. However, column 2 reveals that black defendants were significantly more likely than white defendants to face a binding mandatory minimum sentence after *Booker*.²⁷ The greater prevalence of binding mandatory minimum sentences for black defendants in the aftermath of *Booker* reflects that more statutory minimums exceeding the guidelines-recommended sentences were applied to black defendants than to similar white offenders. Accordingly, black defendants were also more likely to be sentenced at the mandatory minimum after *Booker* (column 3). See Figure 3 for trends in the rate of defendants sentenced at the statutory minimum after *Booker*, which reveals that large racial disparities expanded after *Booker* and coincided with the timing of *Kimbrough* and *Gall*.

However, when a defendant is convicted of a charge that carries a mandatory minimum sentence, prosecutors can reduce sentences below the mandatory minimum if the defendant offered “substantial assistance” during another investigation or prosecution under federal sentencing guidelines section 5K1.1 and 18 U.S.C. sec. 3553(e). Column 4 indicates that government-sponsored substantial assistance motions for cases with mandatory minimum sentences did not change differentially between offenders after *Booker*, although nonwhite defendants are significantly less likely to receive substantial assistance motions in general.

While prosecutorial charging decisions likely contributed to increasing racial disparities after *Booker*, judicially induced disparities remain. Column 5 of Table 8 replicates the main results from Table 3 for the subset of cases in which mandatory minimum sentences are not applied. Racial disparities increased after *Booker* in this subsample, despite these cases being less subject to prosecutorial discretion, at least regarding the decision to charge a mandatory minimum sentence.²⁸ These results indicate that while prosecutorial charging is a large contributor to increases

27. This finding is robust to examining only statutory minimum sentences for drug offenses (the majority of statutory minimum cases) and controlling for specific drug type.

28. Table A8 in the online appendix reveals racial disparities even among offenders in the lowest criminal history category and those with no criminal history points.

Table 8. Treatment of Mandatory Minimum Sentences

| | Mandatory Minimum (1) | Binding Minimum (2) | Sentenced at Minimum (3) | Substantial Assistance (4) | Sentence No Mandatory Minimum (5) |
|-------------------|-----------------------------|---------------------------|--------------------------------|----------------------------------|---|
| Booker × Black | -.00610 (.00689) | .00815* (.00364) | .0115** (.00221) | .00251 (.00703) | 1.536** (.437) |
| Booker × Hispanic | -.0199** (.00758) | -.00621* (.00251) | .00278 (.00259) | .0276** (.00718) | .933* (.467) |
| Booker × Other | -.0393** (.0102) | .00472 (.00404) | -.00506 (.00411) | .0127 (.0116) | 1.509 (.825) |
| Black | .0492** (.00730) | -.00259 (.00320) | .00388 (.00207) | -.0525** (.00719) | .621 (.356) |
| Hispanic | .0464** (.00840) | .0112** (.00281) | .000881 (.00204) | -.0790** (.00770) | .468 (.304) |
| Other | .00196 (.00886) | -.00343 (.00293) | .00191 (.00247) | -.00654 (.0129) | -.0221 (.518) |
| Booker | .00702 (.00913) | -.0113 (.00642) | -.00232 (.00629) | .000930 (.00978) | -1.684* (.665) |
| Observations | 907,388 | 907,381 | 905,449 | 892,387 | 633,596 |
| R ² | .644 | .693 | .193 | .167 | .352 |

Source. For 2002–10 data, see US Sentencing Commission, Commission Datafiles (<http://www.usssc.gov/research-and-publications/commission-datafiles>); data for 1994–2001 are on file with the author.

Note. All regressions contain controls for offense type, dummies for each offense level and criminal history combination, and district, sentencing-year, and sentencing-month fixed effects. Standard errors, in parentheses, are clustered at the district level.

* Significant at the 5 percent level.

** Significant at the 1 percent level.

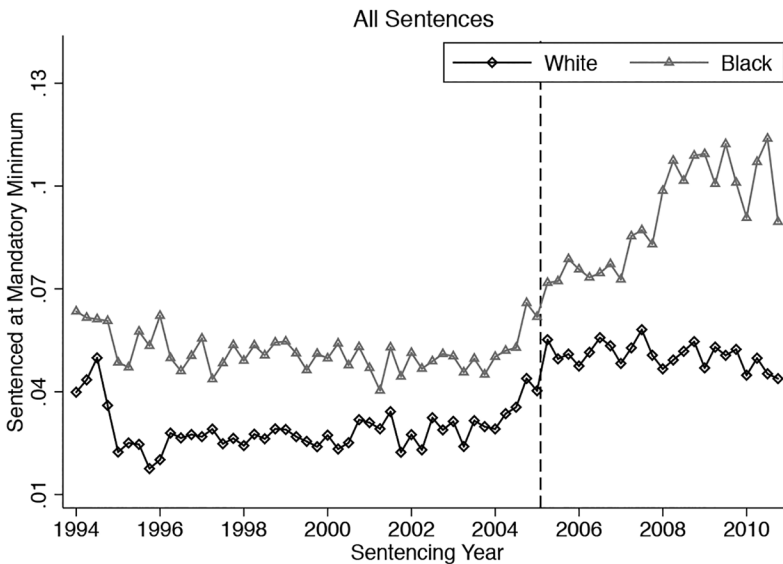


Figure 3. Defendants sentenced at the mandatory minimum by race

in racial disparities, it is unlikely able to fully explain recent increases in racial disparities.

5.7. Consistency with Prior Findings

Finally, I explore the alternative specifications used by previous researchers who find no increase in racial disparities after *Booker* (see Table A9 in the online appendix). Ulmer, Light, and Kramer (2011) condition on the presumptive guidelines minimum sentence and departures from the guidelines (below range, above range, and substantial assistance). Unsurprisingly, given the endogenous changes in departures by the defendant's race after *Booker* (Table 5), this specification yields the appearance of no increase in racial disparities after *Booker*.

I also explore the event-study approach taken by Starr and Rehavi (2013). I replicate their event study using 18-month and 12-month windows around *Booker* to estimate the immediate impacts of the decision on racial disparities. Using the same linked arrest-to-sentencing data and controlling for arrest offense, I find results consistent with theirs. However, this short-term estimate cannot account for racial disparities that emerge after later important changes in appellate review in 2007 and the entrance of new judges to the federal bench, two effects discussed in Sections 5.4 and 5.5, respectively.

6. CONCLUSION

After almost 2 decades of mandatory-guidelines sentencing, the Supreme Court struck down the guidelines in *United States v. Booker*, which greatly increased the degree of judicial discretion. In subsequent decisions, the Court further increased judicial discretion by reducing the degree of appellate review and granting judges explicit permission to reject the policies of the USSC.

Using comprehensive data on federal defendants sentenced from 1994 to 2010, I find evidence that increased judicial discretion after *Booker* has led to large and robust increases in racial disparities in sentencing, particularly after periods of reduced appellate scrutiny. After changes in appellate review, the racial sentencing gap increased to over 3 months, a 6 percent increase in the average sentence length. I also find that recent increases in racial disparities in sentencing were larger among judges appointed after *Booker*, which is consistent with a story in which judges who are experienced with sentencing under rule-based sentencing continued to follow the guidelines even when given more discretion. These findings should, however, be interpreted cautiously, as they apply predominantly to new Bush appointees. Obama-appointed judges may exhibit different sentencing patterns. Finally, my results suggest that prosecutors charged black defendants with higher rates of binding mandatory minimum sentences compared with white defendants after *Booker*, consistent with prosecutors attempting to bind judges to prevent them from departing downward from the guidelines in response to increased judicial discretion.

Despite the increase in racial disparities in federal sentencing after *Booker*, 75 percent of federal district judges believe that the current advisory regime better achieves the purposes of sentencing than did the mandatory-guidelines regime prior to *Booker* (3 percent) or the indeterminate-sentencing regime before the implementation of the guidelines (8 percent). Only 14 percent of judges believe that a new mandatory-guidelines regime that complies with the Sixth Amendment would best achieve sentencing goals (USSC 2010b).

The findings in this paper suggest that while most federal district judges prefer the expanded judicial discretion under the current advisory system to the mandatory-guidelines regime, discretion comes with potentially undesirable consequences. An increase in disparities in the wake of increased judicial discretion can reflect unwarranted disparities if judicial bias enters into decision making. On the other hand, disparities may be

warranted if expanded discretion allows judges to tailor a sentence to the unique circumstances of an offender. For instance, warranted disparities may emerge if judges are sentencing according to defendants' characteristics, both observed and unobserved, that are correlated with recidivism risk.

In fact, recidivism rates are higher among nonwhite offenders, offenders with more extensive criminal histories, and offenders with lower levels of educational attainment, and I find that judges sentenced these defendants to longer prison terms after *Booker*. Not conditional on other characteristics, black offenders are more likely to recidivate (32.8 percent) than Hispanic offenders (24.3 percent) and white offenders (16.0 percent) (USSC 2004). Even controlling for basic demographics, criminal history, and severity of offense, blacks are about 3.2 percentage points more likely to recidivate than white offenders (Kuziemko 2013). If the mandatory guidelines constrained judges' ability to equalize recidivism risk across defendants, a shift to advisory guidelines may have allowed judges to tailor sentences more accurately to recidivism risk.

Future work could analyze the extent to which disparities in sentencing are warranted by looking at rates of recidivism in the federal criminal justice system. More generally, the framework in this paper can be applied to analyzing the impact of increased discretion on many other actors in the criminal justice system. Further work on the interactions of actors at various stages in the criminal process is critical to a thorough exploration of disparities in the criminal justice system.

REFERENCES

- Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. Do Judges Vary in Their Treatment of Race? *Journal of Legal Studies* 41:347–83.
- Alschuler, Albert W. 1978. Sentencing Reform and Prosecutorial Power: A Critique of Recent Proposals for “Fixed” and “Presumptive” Sentencing. *University of Pennsylvania Law Review* 126:550–77.
- American Friends Service Committee. 1971. *Struggle for Justice: A Report on Crime and Punishment in America*. New York: Hill & Wang.
- Anderson, James M., Jeffrey R. Kling, and Kate Stith. 1999. Measuring Interjudge Sentencing Disparity: Before and after the Federal Sentencing Guidelines. *Journal of Law and Economics* 42:271–307.
- Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. The Impact of Jury Race in Criminal Trials. *Quarterly Journal of Economics* 127:1017–55.
- Austin, William, and Thomas A. Williams III. 1977. A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity.

- Journal of Criminal Law and Criminology* 69:306–10.
- Autor, David H., and Susan N. Houseman. 2010. Do Temporary-Help Jobs Improve Labor Market Outcomes for Low-Skilled Workers? Evidence from “Work First.” *American Economic Journal: Applied Economics* 2:96–128.
- Berman, Douglas A. 2005. Foreword: Beyond *Blakely* and *Booker*: Pondering Modern Sentencing Process. *Journal of Criminal Law and Criminology* 95:653–88.
- Fischman, Joshua B., and Max M. Schanzenbach. 2011. Do Standards of Review Matter? The Case of Federal Criminal Sentencing. *Journal of Legal Studies* 40:405–37.
- . 2012. Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums. *Journal of Empirical Legal Studies* 9:729–64.
- Frankel, Marvin E. 1973. *Criminal Sentences: Law without Order*. New York: Hill & Wang.
- Freed, Daniel J. 1992. Federal Sentencing in the Wake of the Guidelines: Unacceptable Limits on the Discretion of Sentencers. *Yale Law Journal* 101:1681–1754.
- Hofer, Paul J., Kevin R. Blackwell, and R. Barry Ruback. 1999. The Effect of the Federal Sentencing Guidelines on Inter-judge Sentencing Disparity. *Journal of Criminal Law and Criminology* 90:239–321.
- Kuziemko, Ilyana. 2013. How Should Inmates Be Released from Prison? An Assessment of Parole versus Fixed Sentence Regimes. *Quarterly Journal of Economics* 128:371–424.
- Mustard, David B. 2001. Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts. *Journal of Law and Economics* 44:285–314.
- Nagel, Ilene H., and Steven J. Schulhofer. 1992. A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices under the Sentencing Guidelines. *Southern California Law Review* 66:501–66.
- Posner, Richard A. 2005. Judicial Behavior and Performance: An Economic Approach. *Florida State University Law Review* 32:1259–79.
- Rehavi, M. Marit, and Sonja B. Starr. 2014. Racial Disparity in Federal Criminal Sentences. *Journal of Political Economy* 122:1320–54.
- Schanzenbach, Max M. 2005. Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics. *Journal of Legal Studies* 34:57–92.
- Schanzenbach, Max M., and Emerson H. Tiller. 2007. Strategic Judging under the United States Sentencing Guidelines: Positive Political Theory and Evidence. *Journal of Law, Economics, and Organization* 23:24–56.
- . 2008. Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform. *University of Chicago Law Review* 75:715–60.
- Scott, Ryan W. 2010. Inter-judge Sentencing Disparity after *Booker*: A First Look. *Stanford Law Review* 63:1–66.

- Sessions, William K., III. 2011. At the Crossroads of the Three Branches: The U.S. Sentencing Commission's Attempts to Achieve Sentencing Reform in the Midst of Inter-branch Power Struggles. *Journal of Law and Politics* 26:305–57.
- Starr, Sonja B., and M. Marit Rehavi. 2013. Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of *Booker*. *Yale Law Journal* 123:2–80.
- Stith, Kate. 2008. The Arc of the Pendulum: Judges, Prosecutors, and the Exercise of Discretion. *Yale Law Journal* 117:1420–97.
- Stith, Kate, and José A. Cabranes. 1998. *Fear of Judging: Sentencing Guidelines in the Federal Courts*. Chicago: University of Chicago Press.
- Tonry, Michael. 2005. Obsolescence and Immanence in Penal Theory and Policy. *Columbia Law Review* 105:1233–75.
- Ulmer, Jeffery T., Michael T. Light, and John H. Kramer. 2011. Racial Disparity in the Wake of the *Booker/Fanfan* Decision: An Alternative Analysis to the USSC's 2010 Report. *Criminology and Public Policy* 10:1077–1118.
- US Department of Justice. Bureau of Justice Statistics. 1994–2009. Federal Justice Statistics Program: Arrests and Bookings for Federal Offenses [United States] (computer file). Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor).
- . 2012. Federal Justice Statistics Program: Paired-Agency Linked Files, 2012 (computer file). Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor).
- USSC (US Sentencing Commission). 2004. *Measuring Recidivism: The Criminal History Computation of the Federal Sentencing Guidelines*. Washington, DC: US Sentencing Commission.
- . 2006. *Final Report on the Impact of “United States v. Booker” on Federal Sentencing*. Washington, DC: US Sentencing Commission.
- . 2010a. *Demographic Differences in Federal Sentencing Practices: An Update of the “Booker” Report’s Multivariate Regression Analysis*. Washington, DC: US Sentencing Commission.
- . 2010b. *Results of Survey of United States District Judges January 2010 through March 2010*. Washington, DC: US Sentencing Commission.
- US Senate Committee on the Judiciary. 1983. *Legislative History of the Comprehensive Crime Control Act of 1984*. Senate Report No. 98-225. 98th Cong., 1st sess., September 14.
- Welch, Susan, Michael Combs, and John Gruhl. 1988. Do Black Judges Make a Difference? *American Journal of Political Science* 32:126–36.
- Wheeler, Stanton, Kenneth Mann, and Austin Sarat. 1988. *Sitting in Judgment: The Sentencing of White-Collar Criminals*. New Haven, CT: Yale University Press.
- Yang, Crystal S. 2014. Have Inter-judge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from *Booker*. *New York University Law Review* 89:1268–1342.