HAVE INTERJUDGE SENTENCING DISPARITIES INCREASED IN AN ADVISORY GUIDELINES REGIME? EVIDENCE FROM *BOOKER*

CRYSTAL S. YANG*

The Federal Sentencing Guidelines were promulgated in response to concerns of widespread disparities in sentencing. After almost two decades of determinate sentencing, the Guidelines were rendered advisory in United States v. Booker. *How has greater judicial discretion affected interjudge disparities, or differences in sentencing outcomes that are attributable to the mere happenstance of the sentencing judge assigned? This Article utilizes new data covering almost 400,000 criminal defendants linked to sentencing judges to undertake the first national empirical analysis of interjudge disparities after* Booker.

The results are striking: Interjudge sentencing disparities have doubled since the Guidelines became advisory. Some of the recent increase in disparities can be attributed to differential sentencing behavior associated with judge demographic characteristics, with Democratic and female judges being more likely to exercise their enhanced discretion after Booker. *Newer judges appointed post-*Booker *also appear less anchored to the Guidelines than judges with experience sentencing under the mandatory Guidelines regime.*

Disentangling the effects of various actors on sentencing disparities, I find that prosecutorial charging is likely a prominent source of disparities. Rather than charging mandatory minimums uniformly across eligible cases, prosecutors appear to selectively apply mandatory minimums in response to the identity of the sentencing judge, potentially through superseding indictments. Drawing on this empirical evidence, this Article suggests that recent sentencing proposals calling for a reduction in judicial discretion in order to reduce disparities may overlook the substantial contribution of prosecutors.

 Assistant Professor of Law, Harvard Law School. Support was provided by the John M. Olin Center for Law, Economics, and Business at Harvard Law School and the Multidisciplinary Program on Inequality and Social Policy at the Harvard Kennedy School. I also thank Kevin Blackwell, Emily Cauble, Gary Chamberlain, Raj Chetty, Adam Chilton, Zachary Clopton, Will Dobbie, Judge Frank Easterbrook, Judge Harry T. Edwards, Lee Epstein, Joshua Fischman, Stuart Ford, Judge Nancy Gertner, Claudia Goldin, Geoffrey Heeren, Paul Hofer, William H.J. Hubbard, Dennis Hutchinson, Louis Kaplow, Lawrence Katz, Joshua Kleinfeld, Ilyana Kuziemko, William Landes, Thomas Miles, Judge Richard Posner, J.J. Prescott, John Rappaport, Ryan Sakoda, Shayak Sarkar, Max Schanzenbach, Steven Shavell, Kate Stith, Susan Yeh, and Judge William G. Young for helpful comments and suggestions, and participants at the Harvard Labor Economics/Public Finance Seminars, Harvard Law School Law and Economics Seminar, Chicago Junior Faculty Workshop, Conference on Empirical Legal Studies, University of Chicago Judicial Behavior Workshop, and faculties at Chicago, Columbia, Harvard, Northwestern, NYU, Stanford, and Virginia law schools. The Transactional Records Access Clearinghouse (TRAC) at Syracuse University, in particular Susan Long, generously provided sentencing data for use in this project in my role as a TRAC Fellow of the Center. All errors are my own. E-mail: cyang@law.harvard.edu.

INTRODUCTION

The Federal Sentencing Guidelines were adopted to counter widespread disparities in federal sentencing. By the 1970s, the federal system exhibited "an unjustifiably wide range of sentences to offenders convicted of similar crimes" because each judge was "left to apply his own notions of the purposes of sentencing."[1] Disparities were so pronounced that a defendant sentenced to three years by one

[1] S. REP. NO. 98-225, at 38 (1983).

judge would have been sentenced to twenty years had he been assigned to another judge.[2] Decrying these disparities and championing sentencing reform, Judge Marvin Frankel of the Southern District of New York claimed that "the almost wholly unchecked and sweeping powers we give to judges in the fashioning of sentences are terrifying and intolerable for a society that professes devotion to the rule of law."[3]

In response, policymakers sought to limit the "unfettered discretion the law confers on those judges and parole authorities [that implement] the sentence."[4] Under the Sentencing Reform Act of 1984 (SRA), Congress created the United States Sentencing Commission to promulgate the Guidelines,[5] a new regime intended to reduce disparities stemming from judicial preferences and biases rather than offense and offender characteristics.[6] Congress directed the Commission, an independent agency within the judicial branch,[7] to fashion sentencing

---

[2] ANTHONY PARTRIDGE & WILLIAM B. ELDRIDGE, FED. JUDICIAL CTR., THE SECOND CIRCUIT SENTENCING STUDY: A REPORT TO THE JUDGES OF THE SECOND CIRCUIT 6 tbl.1 (1974). A unique experiment, the Second Circuit study presented the district court judges in the Second Circuit with approximately twenty presentence reports of actual defendants in representative cases and asked each of the judges to sentence the defendant described in the report. *Id.* at 1. The researchers were thus able to observe how judges render sentences in identical cases. *Id.* at 30 tbl.7. Large differences of several years in sentence lengths were common, and in sixteen of the twenty cases, there was no unanimity about whether incarceration was even appropriate. *Id.* at 6–7. In a representative extortion case, sentences ranged from three years in prison with no fine to twenty years in prison with a $65,000 fine. *Id.* at 6 tbl.1. In a representative bank robbery case, nearly half the judges imposed a sentence length of fifteen years or more, while at least six judges imposed sentence lengths of five years or less. *Id.*

[3] MARVIN E. FRANKEL, CRIMINAL SENTENCES: LAW WITHOUT ORDER 5 (1973).

[4] S. REP. NO. 98-225, at 38; *see also id.* at 49 ("[T]he present practices of the federal courts and of the parole commission clearly indicate that sentencing in the federal courts is characterized by unwarranted disparity and by uncertainty about the length of time offenders will serve in prison.").

[5] Pub. L. No. 98-473, 98 Stat. 1837, 1987 (1984) (codified as amended at 18 U.S.C. § 3551 (2012) and 28 U.S.C. § 991 (2012)). For a discussion of the efforts leading up to the passage of the SRA, see Kate Stith & Steve Y. Koh, *The Politics of Sentencing Reform: The Legislative History of the Federal Sentencing Guidelines*, 28 WAKE FOREST L. REV. 223, 230–60 (1993).

[6] *See* U.S. SENTENCING GUIDELINES MANUAL ch. 1, pt. A, at 2 (2013) ("Congress sought reasonable uniformity in sentencing by narrowing the wide disparity in sentences imposed for similar criminal offenses committed by similar offenders."); S. REP. NO. 98-225, at 45 ("Sentencing disparities that are not justified by differences among offenses or offenders are unfair both to offenders and to the public.").

[7] *See* 28 U.S.C. § 991(a) (2012) (placing the Commission in the judicial branch); *see also* S. REP. NO. 98-225, at 159 (explaining that the Commission was placed within the judicial branch as an effort to keep sentencing "primarily a judicial function" and was also done in part because sitting judges would serve on the Commission). The Commission is comprised of seven voting members. 28 U.S.C. § 991(a). The SRA provides that "[a]t least three of the [Commission] members shall be Federal judges selected after considering a list of six judges recommended to the President by the Judicial Conference of the United

guidelines aimed primarily at avoiding unwarranted sentencing disparity.[8]

After the implementation of the Guidelines, researchers began to investigate the extent to which the Guidelines reduced disparities.[9] Initial work by James M. Anderson, Jeffrey R. Kling, and Kate Stith revealed that the Guidelines were somewhat successful in reducing interjudge sentencing disparity.[10] The authors estimated that the expected difference in sentence length between judges fell from 17% (4.9 months) in 1986–1987 to 11% (3.9 months) in 1988–1993 in twenty-six cities where case assignment was found sufficiently random.[11] Another study by Paul T. Hofer, Kevin R. Blackwell, and R. Barry Ruback also found evidence of reduced interjudge sentencing disparities after the promulgation of the Guidelines.[12] The study concludes that the Guidelines achieved "modest but meaningful success"[13] in reducing interjudge disparities, documenting that the sentencing judge accounted for 2.32% of the variation in sentences in 1984–1985, but only 1.24% under the Guidelines in 1994–1995.[14] Convergence in findings by researchers and the Commission led the Commission to conclude that "the federal sentencing guidelines have made

States" and no more than four members of the Commission can be members of the same political party. *Id.* The Commission later withstood separation-of-powers challenges, as the Court rejected constitutional challenges to the Commission and its delegated authority. *See* Mistretta v. United States, 488 U.S. 361, 390 (1989) (holding that there is "no separation-of-powers impediment to the placement of the Sentencing Commission within the Judicial Branch").

[8] *See, e.g.*, Stephen Breyer, *The Federal Sentencing Guidelines and the Key Compromises upon Which They Rest*, 17 HOFSTRA L. REV. 1, 3 (1988) (explaining that one of Congress's primary purposes in enacting the SRA was to reduce sentencing disparities); Kenneth R. Feinberg, *Federal Criminal Sentencing Reform: Congress and the United States Sentencing Commission*, 28 WAKE FOREST L. REV. 291, 295 (1993) ("The first and foremost goal of the sentencing reform effort was to alleviate the perceived problem of federal criminal sentencing disparity.").

[9] While the Guidelines were mandatory, they provided a range of permissible sentence lengths. 18 U.S.C. § 3553(a)(4) (2012). Thus, it makes sense that the Guidelines would ameliorate, but not completely eliminate, interjudge disparities.

[10] James M. Anderson et al., *Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines*, 42 J.L. & ECON. 271, 303 (1999) ("The Guidelines have reduced the net variation in sentence attributable to the happenstance of the identity of the sentencing judge.").

[11] *Id.* at 294.

[12] *See* Paul J. Hofer et al., *The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity*, 90 J. CRIM. L. & CRIMINOLOGY 239, 241 (1999) ("[T]he guidelines have significantly reduced overall inter-judge disparity in sentences imposed . . . ."); *see also id.* at 284–86 (discussing the statistical techniques employed in the study).

[13] *Id.* at 241.

[14] *Id.* at 288 tbl.1.

significant progress toward reducing disparity caused by judicial discretion."[15]

Following nearly two decades of mandatory Guidelines sentencing, the Guidelines were struck down in *United States v. Booker*,[16] dramatically altering the sentencing landscape. In *Booker*, the Supreme Court found that the mandatory application of the Guidelines violated defendants' Sixth Amendment right to a jury trial because judges based sentences on "facts beyond those found by the jury."[17] The Court thus rendered the Guidelines advisory.[18]

Subsequent Supreme Court decisions further weakened the effect of the Guidelines on criminal sentencing by reducing the degree of appellate scrutiny applied to sentences both within and outside the Guidelines range. In *Rita v. United States*, the Court permitted courts of appeals to apply a presumption of reasonableness to within-Guidelines sentences.[19] In the same year, the Court held in *Gall v. United States* that appellate courts could not presume that a sentence outside the Guidelines range was unreasonable, reducing the degree of appellate review to the more deferential abuse-of-discretion standard.[20] Concurrently with *Gall*, the Court held in *Kimbrough v. United States* that a district court judge has the discretion to impose sentences outside the recommended Guidelines range if that judge has policy disagreements with the Sentencing Commission.[21]

After *Booker* and its progeny—*Rita*, *Gall*, and *Kimbrough*—the legal community expressed concern as to how these decisions would impact interjudge sentencing disparities. Congressman Tom Feeney wrote that "[t]he Supreme Court's decision [in *Booker*] to place this extraordinary power to sentence a person solely in the hands of a single federal judge—who is accountable to no one—flies in the face of the clear will of Congress."[22] Patrick Fitzgerald, the United States Attorney for the Northern District of Illinois, stated that *Booker* "reintroduced into federal sentencing both substantial district-to-

---

15 U.S. SENTENCING COMM'N, FIFTEEN YEARS OF GUIDELINES SENTENCING 99 (2004), *available at* http://cdn.ca9.uscourts.gov/datastore/library/2013/02/26/Henderson_15Year.pdf.

16 *See* 543 U.S. 220, 233 (2005) (noting that the Court's holding hinged on the fact that "the relevant sentencing rules are mandatory and impose binding requirements on all sentencing judges").

17 *Id*. at 236.

18 *Id*. at 247 (opinion of Breyer, J.).

19 551 U.S. 338, 347 (2007).

20 552 U.S. 38, 56 (2007).

21 552 U.S. 85, 101 (2007).

22 Carl Hulse & Adam Liptak, *New Fight over Controlling Punishments Is Widely Seen*, N.Y. TIMES, Jan. 13, 2005, at A29.

district variations and substantial judge-to-judge variations."[23] Similarly, scholars noted the dramatic increase in the discretion afforded to judges,[24] and predicted an increase in unwarranted sentencing disparities.[25]

Due to suggestive evidence of increasing disparities post-*Booker*, the United States Sentencing Commission and policymakers have commented on possible ways to constrain judicial discretion. Then–U.S. Attorney General Alberto Gonzales claimed that, in light of "increasing disparity in sentences" since *Booker*, the Guidelines needed to be fixed.[26] As a potential "fix," former Chair of the Sentencing Commission Judge William K. Sessions III has proposed resurrecting the mandatory Guidelines in order to give them greater weight during sentencing.[27] To comply with *Booker*, Judge Sessions suggests that aggravating facts on offense conduct be proven by a jury beyond a reasonable doubt or admitted by the defendant, resulting in a broad but mandatory sentencing range.[28]

On the other side of the debate, some scholars have suggested that *Booker* improved the "quality, transparency, and rationality" in federal sentencing, and thus *Booker* is the fix.[29] Indeed, the vast majority of federal district court judges also indicate that they prefer the current advisory Guidelines system to alternative regimes. In a

---

[23] Transcript at 243, U.S. Sentencing Comm'n Public Hearing (Sept. 10, 2009) (statement of Patrick J. Fitzgerald, U.S. Attorney, N. Dist. of Ill.).

[24] *See* Douglas A. Berman, *Foreword: Beyond* Blakely *and* Booker*: Pondering Modern Sentencing Process*, 95 J. CRIM. L. & CRIMINOLOGY 653, 676 (2005) ("Booker devised a new system of federal sentencing which granted judges more sentencing power than they had ever previously wielded."); Nancy Gertner, *A Short History of American Sentencing: Too Little Law, Too Much Law, or Just Right*, 100 J. CRIM. L. & CRIMINOLOGY 691, 706 (2010) ("It is clear that *Booker* has enhanced the position of the judge, whose sentencing expertise has been formally acknowledged again, at the cost of diminishing the position of the Sentencing Commission."); Luiza Ch. Savage, *Chaos Ahead After Sentencing Guidelines Decision*, N.Y. SUN, Jan. 13, 2005, at 1 (quoting Professor Frank Bowman, who argued that *Booker* resulted in "the most amount of judicial discretion ever afforded to sentencing judges").

[25] *See* Frank O. Bowman, III, *Debacle: How the Supreme Court Has Mangled American Sentencing Law and How It Might Yet Be Mended*, 77 U. CHI. L. REV. 367, 470 (2010) (claiming that *Kimbrough* and *Gall* were in a line of cases that "have so thoroughly denatured appellate review that the federal system's ability to control regional and judge-to-judge sentencing disparity has been effectively eliminated").

[26] Alberto Gonzales, U.S. Attorney Gen., Is a *Booker* Fix Needed?, Federal Sentencing Guidelines Speech Before National Center for Victims of Crime (June 21, 2005), *in* 17 FED. SENT'G REP. 324, 325–26 (2005).

[27] William K. Sessions III, *At the Crossroads of the Three Branches: The U.S. Sentencing Commission's Attempts to Achieve Sentencing Reform in the Midst of Inter-Branch Power Struggles*, 26 J.L. & POL. 305, 346 (2011).

[28] *Id.*

[29] Amy Baron-Evans & Kate Stith, Booker *Rules*, 160 U. PA. L. REV. 1631, 1633 (2012).

2010 Sentencing Commission survey of district court judges, seventy-five percent of judges indicated that the current advisory Guidelines system "best achieves the purposes of sentencing," while only three percent preferred the mandatory Guidelines regime in place before *Booker*.[30]

This Article asks how reintroducing greater judicial discretion after *Booker* has impacted interjudge sentencing disparities. The question is critical given that past research suggested that the Guidelines themselves had reduced interjudge sentencing disparities. Thus, it is imperative to understand whether weakening the effect of those Guidelines has reversed this trend. A finding of increased interjudge sentencing disparities implicates equity concerns because it suggests that similar defendants convicted of similar crimes may receive different sentences due to the mere happenstance of the sentencing judge assigned.

The primary empirical work on *Booker*'s impact on sentencing disparities suggests there have been increases in interjudge disparity. Using sentencing data from the District of Massachusetts, Ryan W. Scott observed an increase in judge differences with respect to sentencing.[31] While an important first step, Scott's study was limited to

---

[30] U.S. SENTENCING COMM'N, RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES JANUARY 2010 THROUGH MARCH 2010 app. at 19 (2010).

[31] Ryan W. Scott, *Inter-Judge Sentencing Disparity After* Booker*: A First Look*, 63 STAN. L. REV. 1, 4–5 (2010). Scott finds an almost doubling of the effect of judge assignment on sentence length following *Booker*, resulting in an average difference of over two years between lenient and harsh judges for cases not subject to a mandatory minimum. *Id.* at 40–41. Scott also finds significant variation in the rate of below-range sentencing among judges. Some judges sentenced below range at the same rate prior to *Booker* (around sixteen percent), while others increased their rate of sentencing below range to as high as fifty-three percent. *Id.* The Transactional Records Access Clearinghouse (TRAC) recently compiled a data set of the sentencing records of over 800 federal judges from fiscal years 2007 to 2011. Susan B. Long & David Burnham, *TRAC Report: Examining Current Federal Sentencing Practices: A National Study of Differences Among Judges*, 25 FED. SENT'G REP. 6, 7 (2012); *see also* Mosi Secret, *Big Sentencing Disparity Seen for Judges*, N.Y. TIMES, Mar. 6, 2012, at A23 (describing the release of the TRAC study). Relying on the random assignment of cases to judges within district courthouses, the TRAC study found unexplained differences in the typical sentences of judges in over half of the courthouses studied statistically significant. Long & Burnham, *supra*, at 15. The most recent Commission report also finds evidence suggesting that variation among judges within the same district, in particular the rates of non-government-sponsored below-range sentences, has increased after *Booker* and *Gall*. U.S. SENTENCING COMM'N, REPORT ON THE CONTINUING IMPACT OF *UNITED STATES V. BOOKER* ON FEDERAL SENTENCING 69 (2012). The Commission concludes that "sentencing outcomes increasingly depend upon the judge to whom the case is assigned." *Id.* at 8. However, the Commission does not account for caseload composition differences across judges within the same district and analyzes all ninety-four districts, despite evidence from previous researchers that random assignment of cases is not universal. *See id.* at 98 ("The Commission recognizes, however, that caseload composition may differ substantially across divisions within the same

ten judges in the Boston courthouse. Therefore, a comprehensive analysis of disparities post-*Booker* is essential to informing ongoing policy debates.[32]

This Article fills this gap by undertaking the first national, multi-district empirical analysis of interjudge disparities in federal sentencing after *Booker*, utilizing a new and comprehensive data set constructed for this study. The Article proceeds in four parts. Part I provides a brief background of the legal landscape. Part II describes the study's framework, data set, and empirical methods. Merging three data sources, I construct a data set of approximately 400,000 criminal defendants linked to sentencing judges from 2000–2009.

Part III presents empirical results. Relying on the random assignment of cases to judges in district courthouses, I find evidence of significant increases in interjudge disparities. A defendant who is randomly assigned a one-standard-deviation "harsher" judge in the district court received a 2.8-month longer prison sentence compared to the average judge before *Booker*, but received a 5.9-month longer sentence following *Kimbrough*/*Gall*, a doubling of interjudge disparities. Similarly, a defendant randomly assigned to a one-standard-deviation more "lenient" judge faced a 4.7% chance of receiving a below-range departure before *Booker*, but over a 6.9% chance after *Kimbrough*/*Gall*.

---

district."); Anderson et al., *supra* note 10, at 288–89 (noting that the authors excluded some districts from their analysis because they did not assign caseloads in a sufficiently random way). Thus, the Commission's findings are only suggestive.

[32] Another strand of empirical research analyzes the impact of *Booker* on racial disparities in sentencing. The United States Sentencing Commission has found evidence of large racial disparities in sentencing outcomes after *Booker*. *See* U.S. SENTENCING COMM'N, DEMOGRAPHIC DIFFERENCES IN FEDERAL SENTENCING PRACTICES: AN UPDATE OF THE *BOOKER* REPORT'S MULTIVARIATE REGRESSION ANALYSIS 23 (2010) (providing evidence that variations in sentencing were associated with demographic differences to a significant extent during post-*Booker* periods); U.S. SENTENCING COMM'N, FINAL REPORT ON THE IMPACT OF *UNITED STATES V. BOOKER* ON FEDERAL SENTENCING 106 (2006) (concluding, contrary to a report issued four years later, that "demographic factors are associated with sentence length" but "their contribution to sentence lengths before and after Booker are identical"). However, other scholars have found no significant change in racial disparities, at least in sentence length. *See* Jeffery 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, 1081 (2011) ("Multivariate analyses showed that social status factors were associated moderately with sentence length but that their effects pre- and post-*Booker* were similar . . . ."). Some scholars argue that judicial discretion may actually mitigate recent increases in racial disparities. *See* Joshua B. Fischman & Max M. Schanzenbach, *Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, 9 J. EMPIRICAL LEGAL STUD. 729, 730 (2012) (arguing that recent increases in racial disparities are mainly due to the increased relevance of mandatory minimums).

Part III.D analyzes the sources of increases in interjudge disparities. Many scholars have suggested that judges have different sentencing philosophies,[33] which may be affected by the standard of appellate review.[34] Sentencing practices are correlated with judge demographic characteristics such as race,[35] gender,[36] and political affiliation.[37] In particular, the inevitable shift in the composition of the federal district courts may have profound consequences on unwar-

[33] *See, e.g.*, John S. Carroll et al., *Sentencing Goals, Causal Attributions, Ideology, and Personality*, 52 J. PERSONALITY & SOC. PSYCHOL. 107, 107 (1987) (arguing that differences in judges' sentencing goals, penal philosophies, and beliefs about the causes of crimes generate sentencing disparities); Shari Seidman Diamond & Hans Zeisel, *Sentencing Councils: A Study of Sentence Disparity and Its Reduction*, 43 U. CHI. L. REV. 109, 114 (1975) ("[I]t is reasonable to infer that the judges' differing sentencing philosophies are a primary cause of the disparity."); *see also* Hofer et al., *supra* note 12, at 250 (claiming that there are differences between how liberals and conservatives view the goals of sentencing, which can drive different sentencing practices).

[34] In Joshua B. Fischman & Max M. Schanzenbach, *Do Standards of Review Matter? The Case of Federal Criminal Sentencing*, 40 J. LEG. STUD. 405, 406 (2011), the authors find that Democratic appointees are more lenient than Republican appointees and differences in sentencing practices increase when appellate review is more deferential, suggesting that judges are constrained by the fear of reversal. The authors also find evidence that pre-Guidelines-appointed judges are more likely to depart from the Guidelines than post-Guidelines appointees. *Id.*

[35] *See* THOMAS M. UHLMAN, RACIAL JUSTICE: BLACK JUDGES AND DEFENDANTS IN AN URBAN TRIAL COURT 78 (1979) (claiming that both Black and White judges sentenced Black defendants more harshly compared to White defendants); Susan Welch et al., *Do Black Judges Make a Difference?*, 32 AM. J. POL. SCI. 126, 134 (1988) (finding "significant differences, though modest in magnitude," in the way Black and White judges sentence defendants).

[36] *See, e.g.*, Max Schanzenbach, *Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics*, 34 J. LEGAL STUD. 57, 60–85 (2005) (presenting some evidence that minority and female judges sentence differently using district-level variation in judicial characteristics); Darrell Steffensmeier & Chris Hebert, *Women and Men Policymakers: Does the Judge's Gender Affect the Sentencing of Criminal Defendants?*, 77 SOC. FORCES 1163, 1175, 1177 (1999) (finding that female judges in Pennsylvania criminal cases sentence defendants for longer terms, are more likely to incarcerate minorities, and are less likely to incarcerate women).

[37] For instance, Max M. Schanzenbach and Emerson H. Tiller explore the impact of ideology on federal criminal sentencing decisions from 1992 through 2001. Max M. Schanzenbach & Emerson H. Tiller, *Strategic Judging Under the U.S. Sentencing Guidelines: Positive Political Theory and Evidence*, 23 J.L. ECON. & ORG. 24 (2007) [hereinafter Schanzenbach & Tiller, *Strategic Judging*]. They find that sentences for serious crimes in districts comprised of more Democratic-appointed judges are lower than sentences in districts with more Republican-appointed judges. *Id.* at 52–53. The ideology of the reviewing court also increased departures from the Guidelines. *Id.* More recent work in Max M. Schanzenbach & Emerson H. Tiller, *Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform*, 75 U. CHI. L. REV. 715, 734 (2008) [hereinafter Schanzenbach & Tiller, *Reviewing the Sentencing Guidelines*] reveals that Republican-appointed judges tend to give longer sentences for the same crime compared to their Democratic-appointed counterparts. Moreover, Democratic-appointed judges are more likely to depart downward from the Guidelines when the reviewing circuit court is majority Democratic appointed. *Id.* at 735.

ranted disparities as judges who have no experience sentencing under a presumptive Guidelines regime take the federal bench.[38] Federal defense lawyer James Felman predicted that, following the introduction of the advisory Guidelines, "unwarranted disparity in the near term would be considerably less than that which existed prior to 1987," but "there will be a minority of judges who will generate unwarranted disparity, and this number seems likely to increase as the years go by and the bench is filled with individuals who have no history with binding guidelines."[39]

I find that female judges and Democratic-appointed judges issue shorter sentences and are more likely to depart downward from the Guidelines after *Booker*, compared to their male and Republican-appointed peers, respectively. Also striking are the differential sentencing practices of pre- and post-*Booker* judicial appointees. Judges who have no prior experience sentencing under the mandatory Guidelines regime are more likely to depart from the Guidelines-recommended range than their pre-*Booker* counterparts, suggesting that newer judges are less anchored to the Guidelines.

This Article also contributes to the literature on geographical variations in sentencing patterns. Regional variations in sentencing can exist through both exercise of judicial discretion and policy differences across prosecutors' offices.[40] In Part III.E, I present evidence of substantial interdistrict differences in sentencing outcomes. Significant differences exist in sentence length, rates of below-range departures, mandatory minimums, and substantial-assistance motions, with interdistrict differences expanding after *Booker*.

---

[38] Prior studies have been unable to identify the impact of post-*Booker*-appointed judges on interjudge sentencing disparities. The Scott study, which looks only at a single courthouse, is unable to account for changes in judicial composition because the study examined ten judges who served continuously from 2001 through 2008 (both pre- and post-*Booker*). Scott, *supra* note 31, at 25. Recent work suggests that racial disparities in sentencing are greater among judges appointed after *Booker* than among those who were appointed pre-*Booker*. *See, e.g.*, Crystal S. Yang, *Free At Last? Judicial Discretion and Racial Disparities in Federal Sentencing* 3 (Harvard John M. Olin Center for Law, Economics, and Business Fellows' Discussion Paper Series No. 47, 2012), *available at* http://www.law.harvard.edu/programs/olin_center/Prizes/2013-2.pdf ("[M]uch of the increases in racial disparities post *Booker* are driven by post *Booker* appointed judges.").

[39] James Felman, *How Should Congress Respond if the Supreme Court Strikes Down the Federal Sentencing Guidelines?*, 17 FED. SENT'G REP. 97, 98–99 (2004).

[40] *See, e.g.*, L. PAUL SUTTON, FEDERAL SENTENCING PATTERNS: A STUDY OF GEOGRAPHICAL VARIATIONS 7 (1978) (confirming "that sentencing policy varies across jurisdiction" through statistical analysis of sentencing disparity); Charles D. Weisselberg & Terrence Dunworth, *Inter-District Variation Under the Guidelines: The Trees May Be More Significant than the Forest*, 6 FED. SENT'G REP. 25, 26–27 (1993) (finding that the Guidelines do not impact all cases and all districts equally and that the Guidelines mean different things in different contexts).

In Part III.F, I present evidence on how prosecutorial decisions contribute to interjudge disparities. Undoubtedly, a defendant's sentence is determined by the discretionary actions of multiple actors in the criminal justice process, culminating in sentencing. Therefore, any study of interjudge sentencing disparities is only a partial portrayal of the disparities that can arise in the criminal justice system. Previous scholars rightly suggested that arrest, charge, and plea-bargaining decisions made earlier in the process are all ripe avenues for unwarranted bias.[41]

In particular, I analyze the impact of mandatory minimums on interjudge disparities, which has been largely unexplored by previous work. If charged prior to judge assignment, one would expect mandatory minimums to reduce interjudge disparities.[42] However, prosecutorial discretion can lead to disparate application of mandatory minimums across judges, potentially through the use of superseding indictments filed after judge assignment to the case.[43] Indeed, I find evidence that the application of a mandatory minimum

---

[41] *See, e.g.*, Ilene H. Nagel & Stephen J. Schulhofer, *A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices Under the Federal Sentencing Guidelines*, 66 S. CAL. L. REV. 501, 502 (1992) ("[B]oth Congress and the U.S. Sentencing Commission were well aware that plea bargaining posed a potential threat to the success of guidelines sentencing."); Franklin E. Zimring, *Making the Punishment Fit the Crime*, 6 HASTINGS CENTER REP. 13, 13 (1976) (arguing that there is "multiple discretion" in the criminal justice system from the legislature, prosecutor, judge, and parole board). As a result, the Guidelines system has been attacked by many for its rigidity and for shifting power to prosecutors in their charging and plea-bargaining decisions. *See, e.g.*, KATE STITH & JOSE A. CABRANES, FEAR OF JUDGING: SENTENCING GUIDELINES IN THE FEDERAL COURTS 6 (1998) (discussing the increase in power to prosecutors caused by the Sentencing Guidelines); Albert W. Alschuler, *Sentencing Reform and Prosecutorial Power: A Critique of Recent Proposals for "Fixed" and "Presumptive" Sentencing*, 126 U. PA. L. REV. 550, 551–52 (1978) (arguing that fixed and presumptive sentencing schemes "are unlikely to achieve their objectives so long as they leave the prosecutor's power to formulate charges and to bargain for guilty pleas unchecked"); Daniel J. Freed, *Federal Sentencing in the Wake of Guidelines: Unacceptable Limits on the Discretion of Sentencers*, 101 YALE L.J. 1681, 1723–24 (1992) (discussing the role prosecutorial discretion has on sentencing disparity and attributing differences in such discretion to regional policy differences); Kate Stith, *The Arc of the Pendulum: Judges, Prosecutors, and the Exercise of Discretion*, 117 YALE L.J. 1420, 1430 (2008) (arguing that reducing judicial discretion "threatens to enhance prosecutorial authority over sentencing").

[42] *See infra* Part III.F (discussing the effects on sentencing disparity of prosecutorial discretion over which charges to bring against a defendant).

[43] For instance, prosecutors often bring superseding indictments under various statutes, such as 18 U.S.C. § 924(c)(1)(A) (2012) (making it a separate crime to carry a deadly or dangerous weapon during the commission of a violent crime). Scholarship documents the large degree of prosecutorial discretion in bringing superseding indictments, as well as the associated due process concerns. *See, e.g.*, Donald C. Smaltz, *Due Process Limitations on Prosecutorial Discretion in Re-charging Defendants:* Pearce *to* Blackledge *to* Bordenkircher, 36 WASH. & LEE L. REV. 347, 353 (1979) ("Absent an adequate justification for the superseding or additional charges, vindictiveness will be presumed.").

is a large contributor to interjudge and interdistrict disparities, such that measures of disparity are reduced by almost a factor of two when I exclude cases in which mandatory minimums are charged. The results also indicate substantial, unequal application of mandatory minimums to similar cases within district courthouses, and different mandatory minimum policies by prosecutors across district courts. There are also large differences in the rates of substantial-assistance motions filed by prosecutors across judges and district courts. Such results indicate that prosecutorial charging is likely a measurable contributor to disparities in sentencing.

In Part IV, I describe recent proposals to reform federal sentencing and apply this Article's empirical findings to shed light on the soundness of these suggestions. Proposals include introducing jury fact-finding into a mandatory Guidelines system and a return to a simplified-but-presumptive Guidelines regime, both of which would comport with the Court's holding in *Booker*. While such proposals may mechanically reduce the presence of interjudge disparities, they would effectively bind judges to the charges brought by prosecutors. In light of the finding in this Article that a substantial portion of interjudge disparities and regional disparity may be attributable to the application of mandatory minimums, any proposal that contemplates shifting power to prosecutors will likely exacerbate unwarranted disparities. Indeed, many judges and scholars have suggested that mandatory minimums are "fundamentally inconsistent with the sentencing guidelines system."[44] Instead, I argue that strengthened appellate review and elimination of mandatory minimums are potential steps in the direction of reducing unwarranted disparities in sentencing.

[44] *See* Edward M. Kennedy, *Sentencing Reform—An Evolutionary Process*, 3 FED. SENT'G REP. 271, 272 (1991) ("Mandatory minimum sentencing statutes have also hampered the guideline system and are becoming an increasingly serious obstacle to its success. . . . Mandatory minimums inevitably lead to sentencing disparity because defendants with different degrees of guilt and different criminal records receive the same sentence."); Stephen Breyer, *Federal Sentencing Guidelines Revisited*, 11 FED. SENT'G REP. 180, 184–85 (1999) ("Congress, in simultaneously requiring Guideline sentencing and mandatory minimum sentences, is riding two different horses. And those two horses, in terms of coherence, fairness, and effectiveness, are traveling in opposite directions. . . . [Congress needs to] abolish mandatory minimums altogether.").

I
BRIEF LEGAL BACKGROUND ON FEDERAL SENTENCING

A. Adoption of the Federal Sentencing Guidelines

In the early twentieth century, criminal justice was premised on the notion of rehabilitation.[45] This goal of rehabilitation manifested itself in the form of indeterminate sentencing, which allowed prison sentences and probation to be tailored to each offender's progress toward reform. As a result, judges and parole boards possessed substantial discretion in their sentencing determinations.[46] In this regime of "free at last" sentencing,[47] federal judges had essentially unlimited authority when imposing sentences, restrained only by statutorily prescribed minimum and maximum sentences.[48] Lack of appellate review of sentences meant that judges faced no meaningful check to ensure uniformity and consistency in sentencing.[49]

By the 1970s, faith in the rehabilitative model of sentencing declined due to a confluence of changing social norms, escalating public anxiety over rising crime rates, and public skepticism of the ability to rehabilitate criminal offenders.[50] The legal community and

---

[45] *See* David J. Rothman, *Sentencing Reforms in Historical Perspective*, 29 CRIME & DELINQUENCY 631, 638 (1983) (explaining that reformers "pursue[d] rehabilitation, which meant treating the criminal not the crime, calculating the sentence to fit the individual needs and characteristics of the offender"); *see also* Frank O. Bowman, III, *The Quality of Mercy Must Be Restrained, and Other Lessons in Learning to Love the Federal Sentencing Guidelines*, 1996 WIS. L. REV. 679, 680–89 (describing the rehabilitative sentencing model and other factors that led to the SRA).

[46] *See* Rothman, *supra* note 45 ("[T]he judge would make his decision (probation or such a minimum-maximum term); eventually the prison classification committee experts would make their decision (this program or that program), and the parole board experts would make theirs (release at the minimum, or later).").

[47] "Free at last" is a term coined by Judge Nancy Gertner to describe the state of indeterminate sentencing prior to 1984. *See* United States v. Jaber, 362 F. Supp. 2d 365, 370 (D. Mass. 2005) (referring again to "a return to pre-1984 indeterminate sentencing" as the "free at last" period); United States v. Mueffelman, 400 F. Supp. 2d 368, 372 (D. Mass. 2005) (describing the nonbinding approach to sentencing as one in which judges are "free at last").

[48] *See* U.S. SENTENCING COMM'N, THE FEDERAL SENTENCING GUIDELINES: A REPORT ON THE OPERATION OF THE GUIDELINES SYSTEM AND SHORT-TERM IMPACTS ON DISPARITY IN SENTENCING, USE OF INCARCERATION, AND PROSECUTORIAL DISCRETION AND PLEA BARGAINING 9 (1991) ("[Judges] decided the various goals of sentencing, the relevant aggravating and mitigating circumstances, and the way in which these factors would be combined in determining a specific sentence."); *see also* Koon v. United States, 518 U.S. 81, 96 (1996) ("Before the Guidelines system, a federal criminal sentence within statutory limits was, for all practical purposes, not reviewable on appeal.").

[49] *See, e.g.*, Dorszynski v. United States, 418 U.S. 424, 431 (1974) (stating the general proposition that appellate review ends if a sentence is within the limitations set forth in the statute).

[50] *See, e.g.*, FRANCIS A. ALLEN, THE DECLINE OF THE REHABILITATIVE IDEAL: PENAL POLICY AND SOCIAL PURPOSE 24–25, 28–32 (1981) (discussing the decline of the

public expressed alarm at the widespread disparities in criminal sentencing. Some argued that judges and parole boards endangered public safety with lenient sentencing of criminals.[51] Others were distressed by inequitable and arbitrary treatment within the criminal justice system, as studies showed that similar offenders were often punished very differently.[52] A 1977 study of forty-seven Virginia state district court judges revealed that while judges generally agreed on the verdict in legal cases, they applied drastically different sentences.[53] A Federal Judicial Center study in the Second Circuit found large interjudge differences in the sentences imposed based on identical presentence reports of defendants.[54] The same defendant was sentenced to three years in prison by one judge and twenty years by another.[55]

Some concluded that this disparate treatment of defendants by judges produced racial inequities in sentencing, with judicial discretion resulting in higher sentences for minorities and poor defendants.[56] The American Friends Service Committee claimed that decreasing discretion among judges and parole boards was needed to eliminate racial discrimination and sentencing disparities in the criminal justice system.[57]

Other studies identified large intercourt differences. A 1988 study of federal courts found that white-collar offenders who committed similar offenses received very different sentences depending on the

---

rehabilitative model in the 1970s); *see also* Bowman, *supra* note 25, at 374 (attributing demand for social controls to rising crime rates and social upheaval).

[51] *See* Michael Tonry, *Obsolescence and Immanence in Penal Theory and Policy*, 105 COLUM. L. REV. 1233, 1247 (2005) (explaining that conservatives criticized indeterminate sentencing for being uncertain, lenient, and increasing crime rates).

[52] *See* AM. FRIENDS SERV. COMM., STRUGGLE FOR JUSTICE: A REPORT ON CRIME AND PUNISHMENT IN AMERICA 125–30 (1971) (demonstrating disparity in sentences for similar crimes and similarity in sentences for wildly differing crimes); Bowman, *supra* note 45, at 686–88 (arguing that judicial discretion, plea bargaining, and parole systems all contributed to sentencing disparities).

[53] William Austin & Thomas A. Williams III, *A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity*, 69 J. CRIM. L. & CRIMINOLOGY 306, 307 (1977). In a marijuana possession case, for example, some judges imposed prison terms while others only recommended probation. *Id.* at 308. In a burglary case, recommended prison sentences ranged from thirty days and a $100 fine, to five years in prison. *Id.* at 309.

[54] PARTRIDGE & ELDRIDGE, *supra* note 2, at 36.

[55] *Id.*

[56] *See* AM. FRIENDS SERV. COMM., *supra* note 52, at 125 (arguing that discretion anywhere in the criminal justice system, including sentencing, leads to discrimination).

[57] *Id.*; *see also* Bowman, *supra* note 25, at 686–88 (explaining that critics argued that discretion produced unjustifiable disparities and discriminatory treatment).

court in which they were sentenced,[58] with one study observing that "a defendant sentenced by a Southern [California] judge was likely to serve six months more than average, while a defendant sentenced in Central California was likely to serve twelve months less."[59]

These large disparities in sentencing prompted calls for sentencing reform. Championing the call for reform, Federal District Judge Marvin Frankel expressed grave concern over the indeterminate and individualized sentencing regime of the period.[60] Judge Frankel called for the creation of an independent sentencing commission that would replace judicial and parole board discretion.[61]

In response, Congress created the United States Sentencing Commission to adopt and administer the Sentencing Guidelines, which were aimed at eliminating unwarranted sentencing disparities "among defendants with similar records who have been found guilty of similar criminal conduct."[62] As part of the SRA, the Guidelines were applied to all federal offenses committed after November 1, 1987.[63] Some have argued that the SRA itself preserved judicial discretion,[64] while most view the Commission as having substantially increased the severity of punishment and dramatically reduced judges' discretion to consider the particular circumstances of each offense and offender.[65]

Notwithstanding disagreement about the degree to which sentencing reform changed the legal landscape, the new SRA introduced a shift from a regime of nearly unfettered judicial discretion to more restricted discretion within a system of determinate sentencing.[66] By

---

[58] STANTON WHEELER ET AL., SITTING IN JUDGMENT: THE SENTENCING OF WHITE-COLLAR CRIMINALS 139 (1988).

[59] *See* Breyer, *supra* note 44, at 180.

[60] FRANKEL, *supra* note 3, at 5. Frankel also argued that individualized sentencing was "out of hand," and criticized the state of indeterminate sentencing. *Id.* at 10, 26–49; *see also* Marvin E. Frankel, *Lawlessness in Sentencing*, 41 U. CIN. L. REV. 1, 31 (1972) ("[T]he excessive extension of indeterminacy [of sentences] has probably resulted in much cruelty and injustice, rather than the great goods its proponents envisage.").

[61] Frankel, *supra* note 60, at 50–54.

[62] 28 U.S.C. § 991(b)(1)(B) (2012).

[63] Sentencing Reform Amendments Act of 1985, Pub. L. No. 99-217, 99 Stat. 1728 (1985).

[64] *See* Marc L. Miller & Ronald F. Wright, *Your Cheatin' Heart(land): The Long Search for Administrative Sentencing Justice*, 2 BUFF. CRIM. L. REV. 725, 750 (1999) (stating that the SRA allows the Commission to decide how much discretion to grant sentencing judges).

[65] *See* Stith & Koh, *supra* note 5, at 284 ("[I]n those areas where the statute is ambiguous or otherwise deliberately leaves important policy issues to the Commission, the Commission has generally chosen to increase the rigidity and complexity of its guidelines . . . . [J]udges have found it difficult to depart . . . [which] is precisely what Congress intended.").

[66] In addition to creating the Guidelines, the SRA also abolished federal parole and instituted supervised release in its place. *See* Comprehensive Crime Control Act of 1984,

requiring judges to sentence within the recommended Guidelines range unless the court found aggravating or mitigating circumstances,[67] the Guidelines were presumptively mandatory, although the particular standards for departure were ambiguous.[68] Later in *Koon v. United States*, the Supreme Court clarified that a district court judge's decision to depart from the Guidelines range would be subject to an abuse-of-discretion standard of appellate review.[69]

Under the Guidelines, federal district court judges assign each federal crime to one of forty-three offense levels, and assign each federal defendant to one of six criminal history categories. The more serious the offense and the greater the harm associated with the offense, the higher the base offense level assigned under Chapter Two of the Guidelines.[70] For example, trespass offenses are assigned a base offense level of four,[71] while kidnapping is assigned a base offense level of thirty-two.[72] From a base offense level, the final offense level is calculated by adjusting for applicable offense characteristics and other prescribed adjustments.[73] Relevant adjustments include the amount of loss involved in the offense, use of a firearm, and the age or condition of the victim.[74] Chapter Three of the Guidelines allows for further adjustments based on aggravating or mitigating factors, such as a defendant's acceptance of responsibility.[75]

The criminal history category reflects the frequency and severity of a defendant's prior criminal convictions. To determine a defendant's criminal history category, a judge adds points for prior sentences in the federal system, states, territories, and foreign or mili-

---

Pub. L. No. 98-473, § 218(a)(5), 98 Stat. 1837, 2027 (1984). Supervised release is meant "to assist individuals in their transition to community life," and "fulfills rehabilitative ends, distinct from those served by incarceration." United States v. Johnson, 529 U.S. 53, 59 (2000). The term of supervised release is imposed along with a prison term at the time of sentencing. U.S. SENTENCING GUIDELINES MANUAL § 5D1.1(a) (2013).

67 S. REP. NO. 98-225, at 51–52 (1983); Stith & Koh, *supra* note 5, at 245.

68 *See* Miller & Wright, *supra* note 64, at 728 (explaining that while the Commission allowed judges to depart from the Guidelines when the case fell outside the "heartland," the "heartland" concept was too general to provide effective guidance).

69 518 U.S. 81, 99 (1996). The *Koon* Court stated that Congress "did not intend, by establishing limited appellate review, to vest in appellate courts wide-ranging authority over district court sentencing decisions." *Id.* at 97.

70 U.S. SENTENCING GUIDELINES MANUAL ch. 2.

71 *Id.* § 2B2.3.

72 *Id.* § 2A4.1.

73 *Id.*

74 *Id.*

75 For instance, the Guidelines allow for a decrease in base offense level for a defendant's acceptance of responsibility under § 3E1.1 or for minimal participation in the offense under § 3B1.2. Base offense level is increased for defendants who obstruct or impede the administration of justice under § 3C1.1.

tary courts.[76] For example, three points are added for each prior sentence of imprisonment exceeding thirteen months, and two points are added for each prior sentence of imprisonment exceeding sixty days but less than thirteen months. Two points are also added if the defendant committed the offense while already under a criminal justice sentence.[77]

The intersection of the final offense level and criminal history category yields a fairly narrow recommended sentencing range, where the top of the range exceeds the bottom by the greater of either six months or twenty-five percent. If a judge determines that aggravating or mitigating circumstances warrant a departure from the Guidelines, she would have to justify her reasons for departure to the appellate court,[78] but in general the Guidelines were treated as effectively mandatory prior to *Booker*.[79] After the imposition of a sentence, the government is permitted to appeal any sentence resulting in a departure below the Guidelines range, and the defendant can appeal an upward departure.[80]

There are numerous other ways in which Congress has attempted to limit unwarranted disparities in sentencing. Beginning in 1984, and subsequently in 1986 and 1988, Congress enacted a series of mandatory minimum statutes directed at drug and firearms offenses.[81] Mandatory minimums were also applied to recidivist offenders, through the Armed Career Criminal Act,[82] enhancements for career

---

[76] *Id.* ch. 4, pt. A, at 369 ("A defendant with a record of prior criminal behavior is more culpable than a first offender and thus deserving of greater punishment. General deterrence of criminal conduct dictates that a clear message be sent to society that repeated criminal behavior will aggravate the need for punishment with each recurrence.").

[77] *Id.* § 4A1.1.

[78] 18 U.S.C. § 3553(b) (2000) ("[T]he court shall impose a sentence of the kind, and within the range, referred to in subsection (a)(4) unless the court finds that there exists an aggravating or mitigating circumstance of a kind, or to a degree, not adequately taken into consideration by the Sentencing Commission . . . .").

[79] The Court in *Booker* noted that "[t]he Guidelines as written . . . are not advisory; they are mandatory and binding on all judges" and therefore "have the force and effect of laws." United States v. Booker, 543 U.S. 220, 233–34 (2005).

[80] 18 U.S.C. § 3742(a)–(b) (2012).

[81] Anti-Drug Abuse Act of 1988, Pub. L. No. 100-690, 102 Stat. 4181; Anti-Drug Abuse Act of 1986, Pub. L. No. 99-570, 100 Stat. 3207; Pub. L. No. 98-473, §1005(a), 98 Stat. 1837, 2138–39 (1984).

[82] The Armed Career Criminal Act (ACCA) imposes a minimum fifteen-year term of imprisonment for defendants convicted of unlawful possession of a firearm under 18 U.S.C. § 922(g) (2012) with three prior state or federal convictions for violent felonies or serious drug offenses. 18 U.S.C. § 924(e)(1) (2012).

offenders,[83] and enhancements for repeat and dangerous sex offenders.[84]

In 2003, Congress passed the PROTECT Act to curtail judicial departures due to a concern that the standard for appellate review of departures had led to undesirably high rates of downward departures for child sex offenses.[85] Under the Feeney Amendment to the PROTECT Act, judicial departures were only allowed for certain reasons outlined in the Guidelines Manual.[86] Additionally, the Feeney Amendment replaced the prior abuse-of-discretion standard of review for downward departures with de novo review, thus overturning the Supreme Court's holding in *Koon*.[87]

B. Challenges to the Mandatory Guidelines Regime

The initial challenge to the federal sentencing regime began with the "watershed" ruling in *Apprendi v. New Jersey*.[88] In *Apprendi*, the Supreme Court found a New Jersey hate-crime statute unconstitutional because it authorized judges to impose higher sentences based on facts other than those submitted to a jury and proved beyond a reasonable doubt.[89]

These principles were first applied to the constitutionality of mandatory sentencing guidelines in a case involving the Washington State Sentencing Guidelines. In *Blakely v. Washington*, 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 based on facts other than those decided by the jury beyond a reasonable doubt.[90] As a result, Washington's mandatory sentencing

---

[83] *See* 28 U.S.C. § 994(h) (2012) (mandating that the Commission impose imprisonment "at or near the maximum term authorized" for defined "career" offenders); U.S. SENTENCING GUIDELINES MANUAL § 4B1.1 (2013) (defining a "career" offender).

[84] 18 U.S.C. §§ 2247, 2426; U.S. SENTENCING GUIDELINES MANUAL § 4B1.5.

[85] Prosecutorial Remedies and Other Tools to End the Exploitation of Children Today (PROTECT) Act, Pub. L. No. 108-21, 117 Stat. 650 (2003) (codified as amended in scattered sections of 18, 21, 28, 42, and 47 U.S.C.).

[86] For certain offenses, such as child abduction and child sex offenses, the PROTECT Act amended 18 U.S.C. § 3553(b) in 2003 to allow the sentencing court to depart downward only if there are mitigating circumstances of a kind or to a degree that has been "affirmatively and specifically identified" as permissible grounds for downward departure. 18 U.S.C. § 3553(b)(2)(ii) (2012). The PROTECT Act also amended the Guidelines Manual to state that "the grounds enumerated in this Part K of Chapter Five are the sole grounds that have been affirmatively and specifically identified as a permissible ground of downward departure." U.S. SENTENCING GUIDELINES MANUAL § 5K2.0(b).

[87] PROTECT Act, Pub. L. No. 108-21, § 401(d), 117 Stat. 650, 670–71 (2003).

[88] 530 U.S. 466, 524 (2000) (O'Connor, J., dissenting) (calling the *Apprendi* decision a "watershed change in constitutional law").

[89] *Id.* at 468–69, 490.

[90] 542 U.S. 296 (2004).

guidelines were struck down.[91] While the majority opinion in *Blakely* emphasized that the decision was not passing judgment on the constitutionality of the Federal Sentencing Guidelines,[92] the parallels were apparent, and shortly after, the Court applied its reasoning in *Blakely* to the Federal Guidelines.

In *United States v. Booker*, the Court held that the mandatory Federal Sentencing Guidelines violated the Sixth Amendment by mandating judicial fact-finding in determining sentencing ranges.[93] The *Booker* ruling, however, did not apply to congressionally enacted mandatory minimum sentences.[94] Rather than invalidate the Guidelines wholly, or prescribe an enhanced role for jury fact-finding, the Court held in a separate remedial decision by Justice Breyer that the remedy for the Sixth Amendment violation was to declare the Guidelines no longer mandatory but "effectively advisory."[95] The Court explained that "district courts, while not bound to apply the Guidelines, must consult those Guidelines and take them into account when sentencing."[96]

In the immediate aftermath of the Court's decision, district courts took varied approaches in applying *Booker*.[97] Some courts sentenced with minimal consideration of the applicable Guidelines range, while

---

[91] *Id.* at 305.

[92] *See id.* at 305 n.9 ("The Federal Guidelines are not before us, and we express no opinion on them.").

[93] 543 U.S. 220, 226–27, 243–44 (2005). In *Booker*, a federal district court judge enhanced Booker's sentence under the Guidelines based on facts determined by the judge. The judge found by a preponderance of the evidence that Booker had distributed additional amounts of crack cocaine and that the defendant obstructed justice. Booker appealed and the Seventh Circuit ruled that the Guidelines violated the Sixth Amendment in light of *Blakely*. United States v. Booker, 375 F.3d 508, 515 (7th Cir. 2004). A few months later, the Supreme Court, in a 5–4 opinion delivered by Justice Stevens, held that the Guidelines violated the Sixth Amendment by allowing judges to unilaterally enhance sentences using facts not determined by juries. *Booker*, 543 U.S. at 243–44.

[94] *See* U.S. SENTENCING COMM'N, REPORT TO THE CONGRESS: MANDATORY MINIMUM PENALTIES IN THE FEDERAL CRIMINAL JUSTICE SYSTEM 37 (2011) (discussing the compatibility of mandatory minimums and the Guidelines after *Booker*).

[95] *Booker*, 543 U.S. at 245. Similarly, the provisions on supervised release also became advisory, although the United States Sentencing Commission (USSC) states that the majority of courts continue to impose at least the minimum terms in U.S. SENTENCING GUIDELINES MANUAL § 5D1.2 (2013). U.S. SENTENCING COMM'N, FEDERAL OFFENDERS SENTENCED TO SUPERVISED RELEASE 7 (2010), *available at* http://www.ussc.gov/sites/default/files/pdf/training/annual-national-training-seminar/2012/2_Federal_Offenders_Sentenced_to_Supervised_Release.pdf.

[96] *Booker*, 543 U.S. at 264.

[97] *See* Gilles R. Bissonnette, *"Consulting" the Federal Sentencing Guidelines After* Booker, 53 UCLA L. REV. 1497, 1520–21 (2006) (arguing that *Booker* left open how much sentencing judges could deviate from the Guidelines).

others treated the Guidelines as a dominant factor.[98] Circuit courts later reached a consensus that sentencing must begin with the calculation of the applicable Guidelines range.[99] Today, after a sentencing judge has calculated the applicable Guidelines range, he or she must consider seven factors before imposition of punishment: (1) the nature and circumstances of the offense and the history and characteristics of the defendant, (2) the need for the sentence imposed, (3) the kinds of sentences available, (4) the kinds of sentence and the sentencing range established, (5) any pertinent policy statement issued by the Sentencing Commission, (6) the need to avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct, and (7) the need to provide restitution to any victims of the offense.[100] After considering the above factors, the sentencing judge is instructed to "impose a sentence sufficient, but not greater than necessary, to comply with" the basic goals of sentencing.[101]

Subsequent Supreme Court decisions further weakened the effect of the Guidelines on criminal sentencing. While the Court in *Booker* declared that departures from the Guidelines should be reviewed under a "reasonableness" standard,[102] it did not clarify the meaning of this standard until a sequence of 2007 decisions. In the first of these decisions, *Rita v. United States*, 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 consis-

[98] *See id.* at 1522–33 (claiming that district courts have taken two approaches in applying *Booker*: the "substantial-weight" approach and the "consultative" approach). For examples of the two approaches taken by district courts, see, for example, United States v. Ranum, 353 F. Supp. 2d 984, 987 (E.D. Wis. 2005) (noting that "*Booker* is not . . . an invitation to do business as usual," and urging instead that courts consider all the factors in § 3553(a)); United States v. Wilson, 350 F. Supp. 2d 910, 925 (D. Utah 2005) (suggesting courts give "heavy weight" to the Guidelines after *Booker*).

[99] *See, e.g.*, United States v. Howard, 454 F.3d 700, 703 (7th Cir. 2006) ("[After *Booker*,] the district court must first calculate the proper Guidelines range and then, by reference to the factors specified in 18 U.S.C. § 3553(a), select an appropriate sentence."); United States v. Crosby, 397 F.3d 103, 113–14 (2d Cir. 2005) ("[C]onsideration of the Guidelines will normally require determination of the applicable Guidelines range, or at least identification of the arguably applicable ranges . . . it would be a mistake to think that, after *Booker*/*Fanfan*, district judges may return to the sentencing regime that existed before 1987 and exercise unfettered discretion . . . .").

[100] *See* 18 U.S.C. § 3553(a) (2012) (laying out factors).

[101] *Id.*

[102] *Booker*, 543 U.S. at 260–63.

tent with the Commission's judgment in general."[103] In *Gall v. United States*, the Court further held that federal appeals courts could not presume that a sentence outside the range recommended by the Guidelines was unreasonable, reducing the degree of appellate review.[104] The Court in *Gall* concluded that in reviewing a sentence outside the Guidelines range, an appellate court could consider the extent of deviation from the Guidelines, but must give "due deference to the district court's decision that the § 3553(a) factors, on a whole, justify the extent of the variance."[105] In the aftermath of *Gall*, appellate courts could only review sentencing decisions under the more deferential abuse-of-discretion standard. Concurrent with the *Gall* decision, in *Kimbrough v. United States*, the Supreme Court confirmed the holding in *Booker* as applied to cases involving possession, distribution, and manufacture of crack cocaine.[106] The *Kimbrough* Court held that federal district court judges have the discretion to impose sentences outside the recommended Guidelines range on the basis of policy disagreements with the Sentencing Commission.[107]

Ultimately, federal sentencing as it stands today is a mere remnant of the pre-*Apprendi* era. Fundamentally altering the state of sentencing, the Court in *Booker* struck down the long-standing mandatory Guidelines as violative of the Sixth Amendment right to a jury trial. Today, a district court judge calculates the applicable Guidelines sentence, but has the discretion to deviate from the Guidelines recommendation, with departures from the Guidelines reviewed under an abuse-of-discretion standard. The reduced role of the Guidelines has led many to speculate on whether federal judges are indeed "free at last," returning to the state of indeterminate sentencing that spurred the adoption of the Guidelines over two decades ago.

---

[103] 551 U.S. 338, 347–50 (2007) ("[A] court of appeals may apply a presumption of reasonableness to a district court sentence that reflects a proper application of the Sentencing Guidelines.").

[104] 552 U.S. 38, 51 (2007).

[105] *Id.* at 51 (arguing that an appellate court's disagreement with the appropriateness of a sentence is "insufficient to justify reversal").

[106] Kimbrough v. United States, 552 U.S. 85, 91 (2007).

[107] *See id.* at 91, 105 (granting sentencing judges permission to deviate from the Guidelines based on disagreement with the disparate treatment of crack and powder cocaine offenses—the "100-to-1 ratio").

II
Framework, Data, and Methods

A. Judicial Behavior in Criminal Sentencing

While judges have an obligation to "follow the law,"[108] they have additional motivations that affect their decisionmaking.[109] Scholars have suggested that judges care about a variety of factors, such as public recognition, leisure, and reputation.[110] In addition, judges have preferences for sentencing according to their personal tastes.[111] In the context of criminal sentencing, a judge may prefer to sentence a defendant based on personal, political, or ideological goals, rather than according to the mandated Guidelines sentence.[112]

How might the choice of sentencing regime affect judicial behavior? Given judges' individual preferences for sentencing, one likely would observe large interjudge disparities if judges were left unconstrained in the exercise of discretion. Consistent with this prediction, scholars have suggested that the large variances in federal sentences prior to the adoption of the Guidelines likely were due to differing judicial attitudes towards rehabilitation and deterrence.[113]

---

[108] *See* Lewis A. Kornhauser, *Judicial Organization and Administration*, *in* 7 Encyclopedia of Law and Economics 27, 29 (Boudewijn Bouckaert & Gerrit De Geest eds., 2000) (arguing that although judges' motivations may differ, they all maintain an obligation to follow the law).

[109] The economic analysis of judicial behavior builds on work by Judge Richard A. Posner. *See* Richard A. Posner, *Judicial Behavior and Performance: An Economic Approach*, 32 Fla. St. U. L. Rev. 1259, 1259 (2005) ("[J]udicial behavior is best understood as a function of the incentives and constraints that particular legal systems place on their judges."); *see also* Richard A. Posner, How Judges Think (2008) (attempting to define factors in the judicial thought process and their consequences); Lee Epstein, William M. Landes & Richard A. Posner, The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice (2013) (attempting to create an empirically testable model of judicial behavior). For a theoretical economic model of judicial discretion in fact determination, see Nicola Gennaioli & Andrei Shleifer, *Judicial Fact Discretion*, 37 J. Legal Stud. 1 (2008).

[110] Federal district judges occupy a unique position because most district judges retain their positions indefinitely, thus "insulat[ing] the judges from the normal incentives and constraints that determine the behavior of rational actors, except for the relative handful of judges who are ambitious for promotion to the court of appeals . . . ." Posner, *supra* note 109, at 1269.

[111] *See id.* at 1269–70 ("Deciding a particular case in a particular way might increase the judge's utility just by the satisfaction that doing a good job produces . . . [or] by advancing a political or ideological goal . . . .").

[112] Indeed, federal district court judges have expressed a great degree of dissatisfaction with the Guidelines. In a 2010 survey of federal district judges, sixty-five percent of judges indicated that they thought the departure policy statements in the Guidelines Manual were too restrictive, suggesting that many judges would prefer to deviate from the Guidelines. U.S. Sentencing Comm'n, *supra* note 30, tbl.14.

[113] *See* Posner, *supra* note 109, at 1270 (inferring from the "extraordinary variance" in federal sentences prior to the promulgation of the Guidelines that differences in sentence

Conversely, judges who are restricted in the exercise of discretion would be unable to fully sentence according to their preferences. For instance, the adoption of determinate sentencing under the Guidelines introduced a mechanism to constrain judges, which likely explains the finding of reduced interjudge sentencing disparity after the promulgation of the Guidelines.[114]

The prospect of appellate review further restricts judges' sentencing discretion. Judges who depart from the Guidelines incur economic and social costs from deviating. A high reversal rate not only generates administrative burdens, but also potentially harms a trial judge's reputation and prospects for promotion to a court of appeals.[115] Indeed, under the mandatory sentencing regime, departures from the Guidelines were relatively rare.[116] After the Feeney Amendment of the PROTECT Act, which subjected departures from the Guidelines to de novo review, departures decreased even further, suggesting that judges respond to changes in the level of appellate scrutiny.[117]

Given the countervailing forces of judge sentencing preferences versus costs of exercising discretion, how—theoretically—should *Booker*, *Rita*, *Gall*, and *Kimbrough* affect interjudge disparities? Immediately following *Booker*, the total cost of exercising discretion

---

lengths were due to judicial attitudes on responsibility and deterrence); *see also* Brian Forst & Charles Wellford, *Punishment and Sentencing: Developing Sentencing Guidelines Empirically from Principles of Punishment*, 33 RUTGERS L. REV. 799, 804–06 (1981) (documenting disagreement between judges regarding five goals of sentencing: general deterrence, special deterrence, incapacitation, rehabilitation, and just deserts).

114 *See supra* notes 9–15 and accompanying text (discussing how imposition of the Guidelines forced judicial restraint).

115 *See* Evan H. Caminker, *Precedent and Prediction: The Forward-Looking Aspects of Inferior Court Decisionmaking*, 73 TEX. L. REV. 1, 77–78 (1994) (describing anecdotal evidence that lower court judges dislike being reversed on appeal because reversals affect their professional reputation, chances of advancement, and judicial power); Richard S. Higgins & Paul H. Rubin, *Judicial Discretion*, 9 J. LEGAL STUD. 129, 129–30 (1980) (discussing reasons why the possibility of appellate reversal constrains judges); Posner, *supra* note 109, at 1270–71 (noting the limiting effect of judges' desire to avoid reversal on their discretion in handling their dockets); Stephen J. Choi et al., *What Do Federal District Judges Want?: An Analysis of Publications, Citations, and Reversals* 3–4 (Univ. of Chi. Law & Econ., Olin Working Paper No. 508, 2010) ("[D]istrict judges care about minimizing their workload and maximizing their reputation . . . by avoiding appellate reversal.").

116 The rate of departure from the Guidelines was less than 15% in the early 1990s. *See* U.S. SENTENCING COMM'N, DOWNWARD DEPARTURES FROM THE FEDERAL SENTENCING GUIDELINES 55 (2003), *available at* http://www.ussc.gov/Legislative_and_Public_Affairs/Congressional_Testimony_and_Reports/Departures/200310_RtC_Downward_Departures/ch4fnl.pdf.

117 Recall that the Commission found that demographic differences under the mandatory Guidelines regime were lower during the PROTECT Act. *See supra* note 32 (surveying empirical research analyzing racial disparities in sentencing under *Booker* as compared to the PROTECT Act era).

fell substantially for a judge who wanted to depart from the Guidelines, because the Guidelines were rendered advisory. This major shift in sentencing may have been accompanied by increases in interjudge disparity.

However, to the extent that the relative cost associated with de novo appellate review was still binding, judges may have been hesitant to alter their practices without further clarification of appellate review standards. Indeed, not until *Rita*, *Gall*, and *Kimbrough* did the Court explicitly reduce the level of appellate review to abuse of discretion, lowering the probability of appellate reversal.[118] Thus, one might expect further increases in interjudge disparities following *Rita*, *Gall*, and *Kimbrough*. Nevertheless, given the *Rita* presumption of reasonableness attached to within-range sentences, the Guidelines still provide a safe harbor from appellate scrutiny, suggesting that judges may have continued adhering to the Guidelines in making sentencing decisions.

There are other reasons why judicial behavior and interjudge disparities may not have changed much after *Booker* and its progeny. First, judges may have become acculturated to the Guidelines if they had substantial experience sentencing under the previous Guidelines regime.[119] Acculturation would suggest that judges with greater exposure to Guidelines sentencing would be less likely to depart from the Guidelines in the aftermath of *Booker*.

---

[118] The probability of reversal on sentencing matters fell from thirty-six percent in 2006 (under "reasonableness" review), to twenty-six percent in 2008 (under abuse-of-discretion review). I calculated this statistic using yearly data on the universe of criminal appeals from the United States Sentencing Commission, and results are on file with the *New York University Law Review*. I define reversal as all reversals and remands on appeals arising out of *Booker* sentencing issues. According to former District Judge Nancy Gertner, in 2012, sentences that had been appealed on the grounds of unreasonableness were affirmed 95% of the time. *See* Douglas A. Berman, *Advice for the US Sentencing Commission from Former USDJ Nancy Gertner*, SENT'G L. & POL'Y (Sept. 30, 2013, 10:15 AM), http://sentencing.typepad.com/sentencing_law_and_policy/2013/09/advice-for-the-us-sentencing-commission-from-former-usdj-nancy-gertner.html (arguing that the Commission should disseminate district court decisions more widely because so few are overturned on appeal).

[119] *See* Frank O. Bowman, III, *Beyond Band-Aids: A Proposal for Reconfiguring Federal Sentencing After* Booker, 2005 U. CHI. LEGAL F. 149, 187 (2005) (arguing that advisory guidelines might still constrain judicial discretion in the short term "if for no other reason than that the federal bench has become acculturated to the Guidelines over the last seventeen years"); Nancy Gertner, *Supporting Advisory Guidelines*, 3 HARV. L. & POL'Y REV. 261, 262 (2009) ("[A]fter twenty years of strict enforcement, the Federal Sentencing Guidelines have a gravitational pull on sentencing . . . even if they are now only advisory. Indeed, the greatest danger is not that judges *will* exercise their new discretion, but that they *will not*."); STITH, *supra* note 41, at 1497 ("[I]ncumbent sentencing decision makers may be reluctant to regard as unreasonable the sentences they were obliged to seek and impose for two decades . . . .").

Another potential mechanism is anchoring, a type of cognitive bias in which decisionmakers rely heavily on one piece of information and fail to make rational adjustments.[120] Judge Nancy Gertner of the District of Massachusetts predicted that the Guidelines would still play a predominant role for all judges post-*Booker* because "[a]ppellate courts have insisted that district court judges begin with—effectively, 'anchor' their decisions in the Guidelines before considering anything else."[121] Thus, to the extent that federal district judges are effectively anchored to the Guidelines, one may not observe much deviation in sentencing practices after *Booker*. Indeed, because district courts continued to calculate the applicable Guidelines range in the aftermath of *Booker*, scholars commented in the year following *Booker* that the federal sentencing system remained virtually unchanged.[122]

B. Sentencing Data

This Article provides the first comprehensive empirical evidence on the impact of *Booker* and its progeny on interjudge sentencing disparities. As noted above,[123] the Scott study presents the only empirical evidence on these trends thus far, but it is limited to the 2262 sentences imposed by ten judges who served continuously from 2001 to 2008 in the Boston courthouse of the District of Massachusetts.[124] While Scott's study represents a first step in characterizing the extent to which interjudge sentencing practices have changed in the aftermath of *Booker*, a single courthouse is likely unrepresentative of other courthouses across the United States.[125] The presence of growing interjudge sentencing disparities after *Booker* in the Boston

---

[120] *See, e.g.*, Birte Englich et al., *Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making*, 32 PERSONALITY & SOC. PSYCHOL. BULL. 188, 190–92 (2006) (presenting experimental results showing that experienced legal professionals chose to issue significantly higher criminal sentences when previously confronted with a randomly high rather than a low anchor); Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1128–30 (1974) (showing results of psychological experiments demonstrating the existence of the anchoring effect).

[121] Nancy Gertner, *What Yogi Berra Teaches About Post-*Booker *Sentencing*, 115 YALE L.J. POCKET PART 137, 138–40 (2006).

[122] *See* Douglas A. Berman, *Tweaking* Booker*: Advisory Guidelines in the Federal System*, 43 HOUS. L. REV. 341, 349 (2006) ("[A] culture of guideline compliance has persisted after *Booker*."). Berman also observes that Commission data from the year after *Booker* suggest that "federal sentencing judges are exercising their new discretion relatively sparingly." *Id.* at 351.

[123] *See supra* note 31 (discussing the Scott study).

[124] Scott, *supra* note 31, at 25.

[125] *See supra* note 40 and accompanying text for a discussion of the large geographical differences in sentencing and Guidelines treatment.

courthouse may be the result of the particular caseload and judicial composition of that court. Therefore, the conclusion that *Booker* has increased interjudge sentencing disparities is difficult to generalize across other courts. In fact, a comparison of the Boston courthouse to other district courthouses reveals that the former experienced a greater increase in interjudge disparities following *Booker* than the average court in the nation.[126]

Prior empirical research on interjudge disparity and the impact of judicial demographics on sentencing practices has been hampered by the lack of judge identifiers in available data.[127] Because cases are generally randomly assigned to judges within a district courthouse,[128] judge identifiers allow one to compare judges within the same court and in the same time period, capturing judge differences in sentencing rather than differences in caseloads. However, the Sentencing Commission data do not identify the sentencing judge.[129] In response, most researchers have resorted to using aggregate district-level variation in judicial demographics to control for judge sentencing preferences,[130] but this methodology is flawed if districts with different judicial compositions differ in ways that affect all judges within the district.[131] A few researchers have resorted to manually matching sentencing data from the Sentencing Commission with the Public Access to Court Electronic Records (PACER) system, but due to the intensive

[126] The results are on file with the *New York University Law Review*.

[127] The Anderson et al. study is the only empirical work in the past twenty-five years with comprehensive sentencing data including judge identifiers. *See* Anderson et al., *supra* note 10, at 287 (describing the data used and noting that the authors were able to group together cases heard by the same judge only by obtaining previously unavailable data from the Statistics Division of the Administrative Office of the U.S. Courts).

[128] "The majority of courts use some variation of a random drawing." *Frequently Asked Questions*, U.S. COURTS, http://www.uscourts.gov/Common/FAQS.aspx (last visited Aug. 26, 2014).

[129] U.S. SENTENCING COMM'N, 2010/2011 GUIDE TO PUBLICATIONS & RESOURCES 28 (2010), *available at* http://www.ussc.gov/Publications/2010_Guide_to_Publications_and_Resources.pdf ("Pursuant to the policy on public access to Commission documents and data, all case and defendant identifiers have been removed from the data." (citation omitted)).

[130] *See* Schanzenbach & Tiller, *Strategic Judging*, *supra* note 37, at 35 (relying only on district-level variation in observable characteristics of judges).

[131] For instance, a district with a greater percentage of judges appointed by Democratic presidents could differ from other districts in other ways as well. It may be that both Democratic- and Republican-appointed judges within the district sentence differently from judges in other districts, so any effect cannot be attributed solely to Democratic-appointed judges.

matching process, sample sizes in studies using this method have been severely limited.[132]

This Article is the first in over twenty-five years to match sentencing data to judge identifiers in all ninety-four district courts, allowing for a comprehensive look at interjudge sentencing disparities after *Booker*. In order to overcome the lack of judge identifiers in sentencing data, I use data from three sources: the United States Sentencing Commission, the Transactional Records Access Clearinghouse, and the Federal Judicial Center. I describe each data set in turn.

1. United States Sentencing Commission

I use data from the United States Sentencing Commission (USSC) on records of all federal offenders sentenced pursuant to the Sentencing Guidelines and Policy Statements of the SRA in fiscal years 2000 through 2009 (October 1, 1999 to September 30, 2009).[133] The USSC provides detailed sentencing data on federal defendants, but defendant and judge identifiers are redacted. The data set contains information from numerous documents on every offender: Indictment, Presentence Report, Report on the Sentencing Hearing, Written Plea Agreement (if applicable), and Judgment of Conviction.

Court characteristics include the district court and circuit in which sentencing occurred, in addition to the sentencing month and year.[134] Demographic variables include defendant's race, gender, age, citizenship status, educational attainment, and number of dependents. The primary offense variable is the primary offense type for the case generated from the count of conviction with the highest statutory maximum.[135] Data are also available on whether the offense carries a

---

[132] *See, e.g.*, Schanzenbach & Tiller, *Reviewing the Sentencing Guidelines*, *supra* note 37, at 729–30 (describing the PACER method used to match records to sentencing data); Scott, *supra* note 31, at 54–55 (same).

[133] By 1999, it is estimated that over ninety percent of felony defendants in the federal criminal justice system were sentenced pursuant to the SRA of 1984. BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, COMPENDIUM OF FEDERAL JUSTICE STATISTICS, 1999, at 111 (2001), *available at* http://www.bjs.gov/content/pub/pdf/cfjs99.pdf.

[134] USSC data prior to 2004 include information on the exact sentencing date, but this variable is not available in later years. U.S. SENTENCING COMM'N, VARIABLE CODEBOOK FOR INDIVIDUAL OFFENDERS 40–41 (2013), *available at* http://www.ussc.gov/Research_and_Statistics/Datafiles/Variable_Codebook_for_Individual_Offenders.pdf. Information is also collected on the Guidelines amendment year used in calculations. All results are robust to controlling for amendment year, although the results presented in this paper do not include this control. Here, "robust" means that the results are the same even if control variables are changed slightly.

[135] There are a total of thirty-five offense categories in the data set. The most common offense is drug trafficking, followed by immigration, fraud, firearms, and larceny. Yang, *supra* note 38, at 8, 27 tbl.1.

mandatory minimum sentence under various statutes, and whether departures from the statutory minimum are granted, either under a substantial-assistance motion or application of the safety valve. Offense level variables include the base offense level, the base offense level after Chapter Two adjustments and the final offense level after Chapter Three adjustments. Criminal history variables include whether the defendant has a prior criminal record (first time offender or prior offender) and whether armed career criminal status, career offender status, or repeat and dangerous sex offender status is applied.[136]

For each offender, there is a computed Guidelines range and a Guidelines range adjusted for applicable mandatory minimums. Sentencing outcomes include imprisonment or probation, sentence length, and length of supervised release. Based on departure status variables, I construct indicator variables for above-range and below-range departures from the Guidelines.[137]

2. Transactional Records Access Clearinghouse

The Transactional Records Access Clearinghouse (TRAC) provides less comprehensive sentencing records obtained from detailed federal records and information from the Justice Department and the Office of Personnel Management. TRAC provides an especially useful set of data: While defendant, offense characteristics, and Guidelines application information are not included, defendant sentences are linked to sentencing judges.[138]

3. Federal Judicial Center

Finally, I use demographic information on federal district court judges from the Federal Judicial Center. As of January 11, 2014, there

---

[136] Data are also collected on the total number of criminal history points applied and the final criminal history category. *See* BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, FEDERAL JUSTICE STATISTICS PROGRAM: DEFENDANTS SENTENCED UNDER THE SENTENCING REFORM ACT, 2009 CODEBOOK 8–10, 12, 15, 81 (2011), *available at* http://www.icpsr.umich.edu/cgi-bin/file?comp=none&study=30795&ds=1&file_id=1055605 (detailing the information collected for each sentence in the data set).

[137] An above-range departure is coded as 1 for a defendant who received an upward departure from the maximum Guidelines-recommended range. Similarly, a below-range departure is coded as 1 for a defendant who received a downward departure from the minimum Guidelines-recommended range.

[138] TRAC has compiled records on the criminal cases and the civil matters handled by federal district court judges in each of the ninety-four federal judicial districts through over twenty years of FOIA requests. *See About Us*, TRANSACTIONAL RECORDS ACCESS CLEARINGHOUSE, http://trac.syr.edu/aboutTRACgeneral.html (last visited Aug. 26, 2014) (providing a general description of TRAC).

are a total of 677 authorized Article III district judgeships.[139] The number of federal district judgeships in each district court varies. The largest district court is the Southern District of New York, with twenty-eight authorized judgeships.[140] The majority of other district courts have between two to seven district court judgeships.[141] I collect information on judge race, gender, affiliation of appointing president, tenure, whether the judge was appointed prior to the adoption of the Guidelines in 1987, and whether the judge was appointed after *Booker* was decided in 2005.

C. Merging

In order to connect defendants to judges, I merge observations across these three data sets. First, I match sentencing records from the USSC to TRAC data. By district court, matching is conducted on several key variables that can uniquely identify records: sentencing year, sentencing month, offense type,[142] 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.[143] Then, I match the USSC and TRAC combined data to judge biographical data from the Federal Judicial Center.[144] I successfully match approximately sixty percent of all USSC cases from fiscal years 2000 through 2009, resulting in a merged data set of 373,340 cases that represent ninety-one district courts (what I call the "full sample").

---

139 Authorized judgeships only refer to full-time, non-senior-status judges. As of August 26, 2014, fifty-nine positions were vacant. For statistics on judicial vacancies, see *Judicial Vacancies*, U.S. COURTS, http://www.uscourts.gov/JudgesAndJudgeships/JudicialVacancies.aspx (last visited Aug. 26, 2014).

140 *See U.S. District Courts: Additional Authorized Judgeships*, U.S. COURTS, http://www.uscourts.gov/JudgesAndJudgeships/Viewer.aspx?doc=/uscourts/JudgesJudgeships/docs/district-judgeships.pdf (last visited Aug. 26, 2014) (compiling statistics on authorized judgeships by district court).

141 *Id.*

142 Data from USSC are coded to correspond with the offense categories in the TRAC data.

143 These match variables are comparable to those used by previous scholars under the PACER method. *See* Schanzenbach & Tiller, *Reviewing the Sentencing Guidelines*, *supra* note 37, at 729 (matching USSC data with PACER records based on the date and length of the sentence, and when necessary, the amount of any fine, the offense type, and the Hispanic ethnicity of the defendant).

144 The Federal Judicial Center does not collect demographic information on judges in three districts: Guam, Virgin Islands, and Northern Mariana Islands. *About the Biographical Directory of Federal Judges, 1789–Present*, FED. JUDICIAL CENTER, http://www.fjc.gov/history/home.nsf/page/about_judges.html (last visited Aug. 26, 2014).

D. Testing for Random Assignment

In an ideal experiment to test the impact of *Booker*, one would randomly allocate the *Booker* treatment—sentencing under advisory Guidelines—to certain groups of judges. In this hypothetical, a group of judges within each district court would be randomly assigned to the treatment group, while the others would comprise the control group (sentencing under a mandatory Guidelines regime). Because of the random assignment of the *Booker* treatment, any differences in caseload composition or judge characteristics would be on average the same across both treatment and control groups. As a result, a straightforward comparison of the sentencing practices between judges in the treatment group (those who sentence under advisory Guidelines), and the judges in the control group (those who sentence under mandatory Guidelines), would capture the causal effect of greater judicial discretion via *Booker* on interjudge differences in sentencing.

Unfortunately, this hypothetical experiment does not exist because the Supreme Court's decision in *Booker* applied to all judges. However, one can utilize the quasi-experiment created by the timing of the *Booker* decision. Assuming that judges within the same district courthouse are randomly assigned cases from the same underlying caseload, one can compare interjudge disparities before *Booker* to interjudge disparities after *Booker*. If there are no other contemporaneous changes that affect interjudge sentencing, an increase in interjudge disparities is attributable to changes in judicial behavior, rather than underlying differences in case composition.[145] Moreover, random assignment of cases can also lead to estimates of the relationships between judicial demographics (such as gender and experience) and sentencing practices.

The crucial assumption underlying the validity of the quasi-experiment is the random assignment of cases to judges.[146] According to the Administrative Office of the United States Courts, "[t]he majority of courts use some variation of a random drawing" as prescribed by local court orders.[147] While each district court is authorized to specify its own methods of case assignment,[148] "[m]ost district and

[145] Previous research indicates that, after controlling for unobservable annual and monthly changes that might affect sentencing, *Booker*—and to an even greater extent *Rita*, *Gall*, and *Kimbrough*—did have a causal impact on judicial behavior. *See* Yang, *supra* note 38, at 17–18 (finding increasing racial disparities in sentencing after these decisions).

[146] As mentioned previously, the Anderson et al., Hofer et al., Scott, and TRAC studies all rely on random case assignment. *See supra* notes 10–14 and accompanying text (discussing the history and results of these studies).

[147] *Frequently Asked Questions*, *supra* note 128.

[148] Under 28 U.S.C. § 137 (2012), the rules and orders of a district court largely determine the assignment of cases among the judges on that court, and the chief judge of

bankruptcy courts use random assignment, which helps to ensure a fair distribution of cases and also prevents 'judge shopping,' or parties' attempts to have their cases heard by the judge who they believe will act most favorably."[149]

However, random assignment may be violated in some instances. For example, senior-status judges with reduced caseloads may select the type of cases they hear during the year,[150] and some courts assign certain types of cases to particular judges.[151] Moreover, even if a court has local rules and orders that specify the use of random assignment, empirically testing for random assignment is important because random assignment can be violated if individuals—either parties or judges—seek to game the system.[152] For instance, prosecutors might seek to tilt the scales by indirectly steering criminal cases towards particular judges viewed as favorably disposed toward the government.[153]

---

the court is responsible for assigning cases in a manner consistent with these rules and orders. For example, in the District of Arizona, "the Clerk must assign criminal cases to District Judges within each division by automated random selection and in a manner so that neither the Clerk nor any parties or their attorneys will be able to make a deliberate choice of a particular Judge." D. ARIZ. R. CRIM. P. 5.1(a). Similarly, in the Northern District of California, "[c]ases shall be assigned blindly and at random by the Clerk by means of an automated system approved by the judges of the court." U.S. DIST. COURT, N. DIST. CAL., GENERAL ORDER NO. 44: ASSIGNMENT PLAN (2013), *available at* http://www.cand.uscourts.gov/filelibrary/132/G.O.%2044_Rev_12.17.13.pdf. In Colorado, "[t]he clerk shall maintain a computerized program to achieve work parity among judicial officers through random and public assignment of new cases." D. COLO. CIV. R. 40.1.

149 *Federal Judges and How They Get Appointed*, FED. JUDICIAL CENTER, http://www.fjc.gov/federal/courts.nsf/page/183 (last visited Aug. 26, 2014).

150 28 U.S.C. § 294 (2012) governs the assignment of cases to senior status judges, and the district courts have adopted more specific rules governing such assignments. *See, e.g.*, District Guidelines Regarding Senior Status in the District of Massachusetts § 4.2, General Order No. 10-04 (D. Mass. Mar. 2, 2010) (''A senior judge may limit the category of cases assigned to him or her or may select a special category of cases for assignment. For example, a senior judge may elect not to be assigned criminal cases or may elect to be assigned only patent cases.'').

151 For instance, the Southern District of New York assigns civil and criminal cases such that all active judges except the chief judge receive "substantially an equal share of the categories of cases of the court," but the chief judge is assigned only "such matters as the chief judge is willing and able to undertake, consistent with the chief judge's administrative duties." U.S. DIST. COURT, S. DIST. N.Y., RULES FOR THE DIVISION OF BUSINESS AMONG DISTRICT JUDGES (2013), *available at* http://www.nysd.uscourts.gov/rules/rules.pdf. Thus, assignment is based on equal caseload, rather than pure random assignment, and the chief judge is exempt from the rules.

152 *See* J. Robert Brown, Jr. & Allison Herren Lee, *Neutral Assignment of Judges at the Court of Appeals*, 78 TEX. L. REV. 1037, 1041–43 (2000) (describing opportunities to manipulate case assignment in federal courts of appeals).

153 *See* United States v. Pearson, 203 F.3d 1243, 1263–64 (10th Cir. 2000) (expressing concern over a defendant's claim that a prosecutor attempted to subvert neutral case assignment). For recent allegations of a court gaming the random assignment system, see William Safire, *Norma the Plumber*, N.Y. TIMES, July 31, 2000, at A19 (alleging that the chief judge of the U.S. District Court for the District of Columbia intentionally assigned

In order to dispose of potential violations to random case assignment, I empirically test for random assignment. To begin, I employ several sample restrictions. First, I drop judges who were formally retired prior to the start of the data set in 2000 to reduce the possibility of nonrandom assignment to senior status judges who continued to hear cases during the sample period. Second, I drop judges and district courthouses with annual caseloads of fewer than twenty-five cases in order to obtain a sufficient number of cases per judge for statistical power.[154] Finally, I drop district courthouses with only one active judge.

With these sample restrictions, I test for random assignment using the merged USSC, TRAC, and Federal Judicial Center data from 2000–2009.[155] If cases are randomly assigned to judges, then judges should see, on average, cases with the same distribution of predetermined defendant characteristics. I test for random assignment based on five fixed defendant characteristics: gender, age, a black race indicator, number of dependents, and an indicator for education less than a high school degree.[156] Intuitively, if cases are randomly assigned, there should be no significant correlation between a particular judge and defendant characteristics.[157] I drop all courthouses that violate random assignment,[158] resulting in a subsample of 156 courthouses from seventy-four district courts—representing about fifty percent of

politically sensitive cases in a nonrandom fashion to avoid embarrassing Democrats). For another example, see Dan Fitzpatrick, *Bank of America Manages to Avoid Rakoff*, WALL ST. J., May 17, 2010, at C1 (describing the nonrandom assignment of a case in the Southern District of New York).

154 Results are robust to choice of caseload minimums, but follow the convention in prior literature. The Scott study excludes judges who imposed fewer than twenty-five sentences in a two-year period. Scott, *supra* note 31, at 24 & n.143. Similarly, Anderson et al. exclude judges who handled fewer than thirty cases during a two-year period. Anderson et al., *supra* note 10, at 288.

155 See *infra* Appendix A for details on the statistical methods I used to test for random assignment.

156 For each of the five defendant characteristics, I regress the 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 zero. I employ seemingly unrelated regression (SUR). For a discussion of the SUR technique, see David H. Autor & Susan N. Houseman, *Do Temporary-Help Jobs Improve Labor Market Outcomes for Low-Skilled Workers? Evidence from "Work First,"* 2 AM. ECON. J.: APPLIED ECON. 96, 106–07 (2010). See *infra* Table A1 for the results of this analysis for each district courthouse.

157 Randomization results are robust to nonparametric permutation tests.

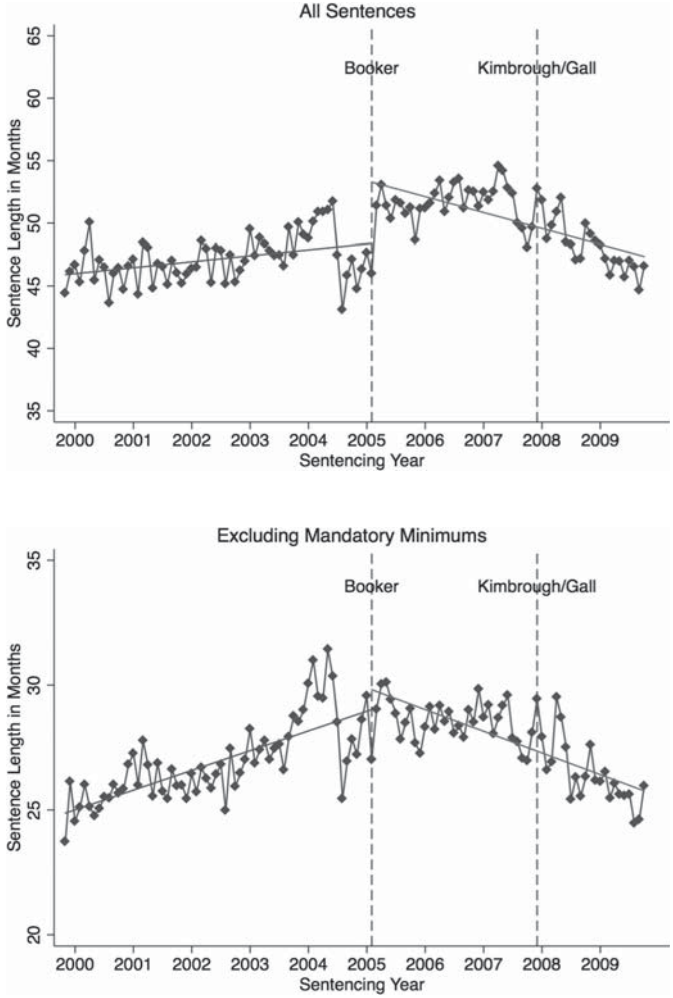
158 It is important to note that my tests of randomization may result in exclusion of courthouses that do not actually violate random assignment. Due to the nature of the USSC data, each defendant is listed as a separate observation, rather than each case, the unit of randomization. Therefore, district courthouses with large numbers of related, or multidefendant cases, may look nonrandom, although they in fact use random case assignment.

all merged cases in district courts from 2000–2009—for a total of 158,126 cases (what I call the "random sample").[159]

E. Trends in Sentencing

I first present graphical evidence of trends in sentence lengths and rates of below-range departures over the time period 2000–2009. Graphical analyses confirm that *Booker* did significantly alter the sentencing practices of judges. Figure 1 below presents a graph of average sentence lengths in months over time, with the timing of *Booker*

FIGURE 1. SENTENCE LENGTHS IN MONTHS



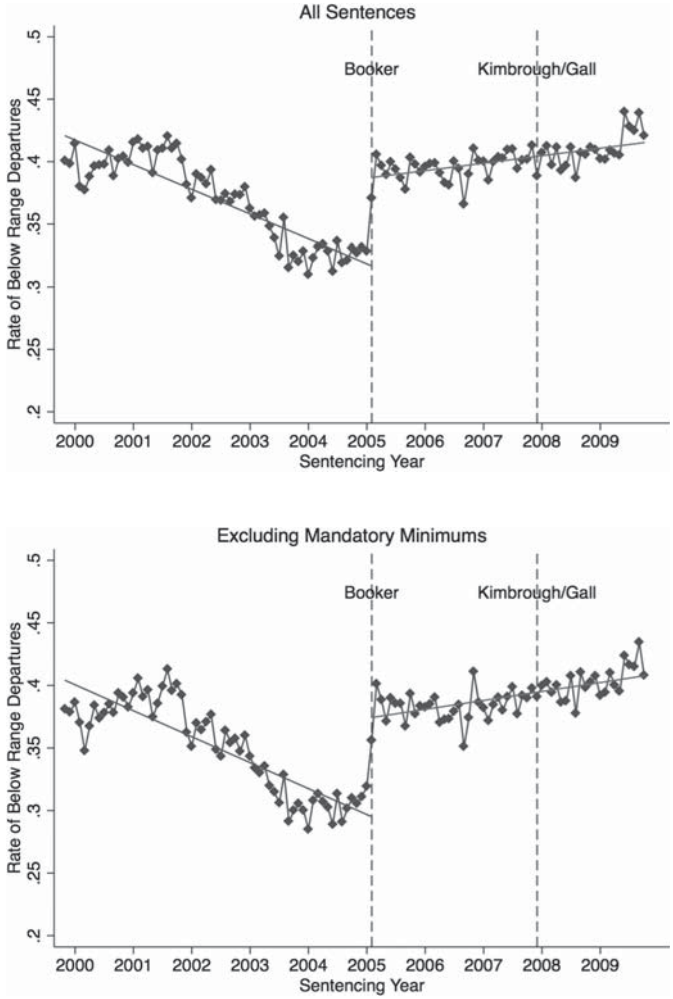
Notes: Data is from the USSC 2000–2009 on all defendants. Data points are monthly averages.

[159] All results presented are robust to using the "full sample." The results using the full sample are on file with the *New York University Law Review*.

delineated by the first vertical dash and *Kimbrough*/*Gall* by the second vertical dash, along with predicted trend lines before and after *Booker*. Figure 1 shows that overall sentence lengths were relatively stable in the five years prior to *Booker*, but began to decrease afterwards, particularly for cases in which a mandatory minimum was not charged.[160]

Figure 2 presents a graph of the average rate of below-range departures from 2000–2009, that is, the percentage of cases in which the defendant receives a sentence below the Guidelines-recom-

FIGURE 2. BELOW RANGE DEPARTURE RATES



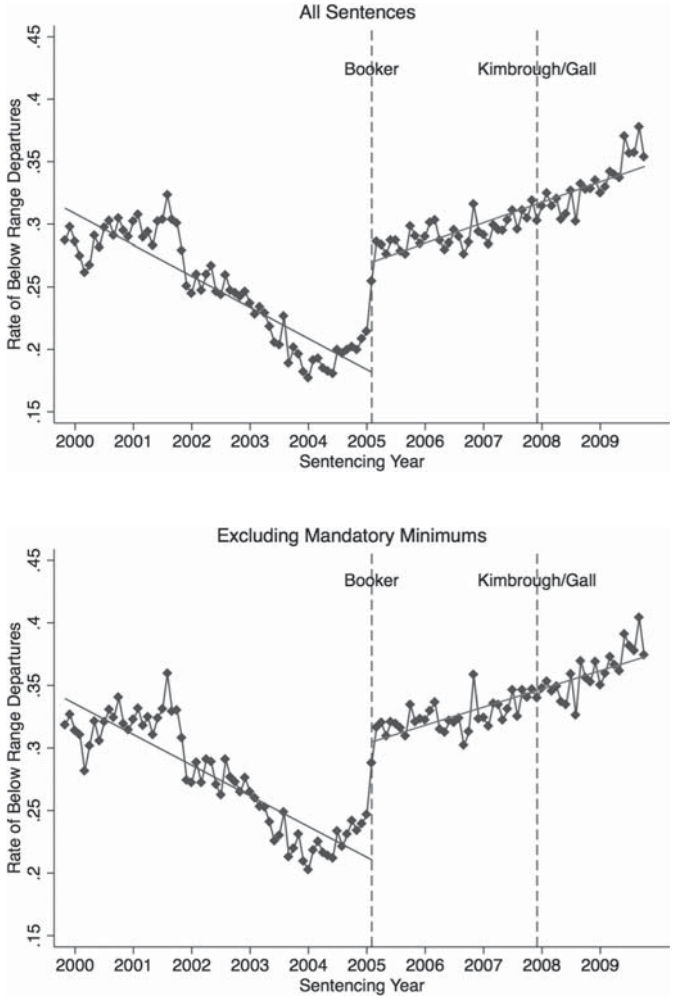
Notes: Data is from the USSC 2000–2009 on all defendants. Data points are monthly averages.

[160] Cases defined as excluding mandatory minimums are those in which a statutory mandatory minimum was not charged, which represent over two-thirds of all cases.

mended minimum sentence. Figure 2 reveals a trend of decreasing rates of below-range departures before *Booker*, characterized by particularly low relative rates of departures in the PROTECT Act era. The decreasing trend in below-range departures reversed after *Booker*, which induced a sudden jump in the rate of departure, and the new increasing trend continued throughout *Rita*, *Gall*, and *Kimbrough* until the rate of below-range departures returned to pre–PROTECT Act levels.

Below-range departures can be the product of both prosecutorial and judicial action. To disentangle these two factors, Figure 3 presents trends in the average rate of below-range departures that are not the

FIGURE 3. NON–GOVERNMENT SPONSORED BELOW RANGE DEPARTURES



Notes: Data is from the USSC 2000–2009 on all defendants. Data points are monthly averages.

result of a government-sponsored substantial-assistance motion.[161] Figure 3 confirms a similar trend with respect to rates of non-government-sponsored below-range departures, suggesting that judicial behavior has truly changed following the shift to an advisory Guidelines regime. However, while overall trends in sentencing have changed in the aftermath of *Booker* and its progeny—*Rita*, *Gall*, and *Kimbrough*—aggregate trends mask whether interjudge variation has increased. For this reason, I next examine interjudge disparities in these outcomes.

F. Measuring Interjudge Disparity: Analysis of Variance

To identify changes in interjudge disparity, I employ an analysis of variance methodology. Variants of this methodology have been used in the federal sentencing literature,[162] as well as in the economics literature on teacher value-added.[163] The analysis-of-variance technique measures interjudge dispersion in sentencing outcomes based on the variance of a judge-specific random variable.[164]

---

[161] As noted previously, a substantial-assistance motion by the government permits a departure from the Guidelines for a defendant who provides substantial assistance in the prosecution or investigation of another person. *See* 18 U.S.C. § 3553(e) (2012) (permitting sentences below a statutory minimum on motion by the prosecution indicating the defendant's substantial assistance in investigating another); U.S. SENTENCING GUIDELINES MANUAL § 5K1.1 (2013) (listing factors a judge should consider when deciding whether and how much to reduce a sentence for substantial assistance).

[162] Studies using a similar methodology include Anderson et al., *supra* note 10, at 280–87 (comparing alternative statistical methods and electing to use a modified negative binominal random effects model of sentences and a Gini coefficient to measure sentencing disparities), A. Abigail Payne, *Does Inter-Judge Disparity Really Matter? An Analysis of the Effects of Sentencing Reforms in Three Federal District Courts*, 17 INT'L REV. L. & ECON. 337, 344 (1997) (using a fixed-effects regression model to isolate the proportion of the variance in prison sentences attributable to disparities between judges), and Joel Waldfogel, *Aggregate Inter-Judge Disparity in Federal Sentencing: Evidence from Three Districts (D.Ct., S.D.N.Y., N.D.Cal.)*, 4 FED. SENT'G REP. 151, 152 & 153 n.4 (1991) (applying a likelihood ratio test to determine whether interjudge disparity exists).

[163] *See* Raj Chetty et al., *How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR*, 126 Q.J. ECON. 1593, 1595–96 (2011) (testing whether future student earnings are clustered by kindergarten classroom using an analysis-of-variance approach).

[164] This Article does not employ the statistical technique used by Scott, who regresses sentencing outcomes on dummy indicators for each sentencing judge, so that the corresponding *R*-squared measures the percentage of variance in the dependent variable for which that sentencing judge identity accounts. Scott, *supra* note 31, at 58. Scott interprets an increase in the *R*-squared in time periods following *Booker* as indicative of growing interjudge disparities. *Id.* at 40–41. For instance, in regard to sentence lengths for cases excluding mandatory minimums, Scott finds an increase in *R*-squared from 0.014 pre-*Booker* to 0.080 post-*Kimbrough*/*Gall*. *Id.* at 34 tbl.2. However, Scott's reliance on the *R*-squared measure is problematic for two main reasons. First, the measure of *R*-squared does not necessarily correspond directly to actual interjudge variation, in contrast to a measure of the variance in a judge-specific random variable. Second, the magnitude of an

The analysis-of-variance technique assumes that the impact of a judge on sentencing outcomes is randomly and normally distributed within each district courthouse such that the judge effect has mean of 0 and variance of σ^2.[165] For instance, suppose that there are n judges in a district courthouse. If the judges were identical in their sentencing preferences, and cases are randomly assigned to judges, there would be no impact of a particular judge on sentencing outcomes. Each judge would sentence in the exact same way and variation in the judge effect, as measured by σ^2, would equal zero. To the extent that judges do differ in their sentencing practices based on personal ideologies or goals, one would observe a distribution of judge effects, as measured by the standard deviation in judge effects, σ. The greater the interjudge variation in outcomes, the larger σ.

Analysis of variance allows one to estimate the standard deviation of judge effects on sentence length, σ, after controlling for case and defendant characteristics. For example, a finding that $\sigma = 5$ implies that a defendant assigned to a judge that is one standard deviation "harsher"[166] than the average judge receives a sentence that is five months longer. In order to capture changes in interjudge disparity, this paper measures σ in periods before and after *Booker*. Specifically, I separate the sample timeframe of 2000–2009 into four main periods: (1) *Koon* (October 1999 to April 2003), (2) the PROTECT Act (May 2003 to January 2005),[167] (3) *Booker* (January 2005 to November 2007), and (4) *Kimbrough*/*Gall* (December 2007 to

R-squared cannot be taken literally without some discussion of its statistical significance, for which the significance of the underlying linear regression model serves as a proxy. Scott's linear regression models are often statistically insignificant, *see, e.g.*, *id.* (presenting data on linear regression models for sentence length excluding mandatory minimums); *id.* at 40 tbl.3. (presenting data on linear regression models for the distance of sentences from the guideline range), suggesting that judge fixed effects are a poor predictor of sentencing outcomes. Despite this insignificance, Scott does not qualify the magnitudes of the *R*-squared measures. For instance, the model is statistically insignificant in two out of three of the studied periods in Table 2, *id.* at 34 tbl.2, and four out of five of the studied periods in Table 3, *id.* at 40 tbl.3. In contrast, measures of interjudge variance in an analysis of variance can be rigorously tested for statistical significance. Scott acknowledges this issue, noting that "[t]he fact that the model for the *Kimbrough*/*Gall* period is not significant reinforces the need for caution in interpreting the results for cases not governed by a mandatory minimum. . . . Although the relationship in the *Kimbrough*/*Gall* period is strongly positive, the model falls well short of statistical significance." *Id.* at 34 n.177.

165 See *infra* Appendix B for details on the empirical methodology used here.

166 "Harshness" is used solely to connote a tendency to impose longer sentence lengths, and I call any judge who sentences above range more than the court average as "harsher." Conversely, any judge who sentences below range more often than the court average can be called "lenient."

167 The PROTECT Act took effect on April 30, 2003. Pub. L. No. 108-21, 117 Stat. 650 (2003) (codified as amended in scattered sections of 18 U.S.C.).

September 2009).[168] An increase in σ after *Booker* would imply an increase in interjudge sentencing disparity after the Guidelines were rendered advisory.

III
Results on Interjudge and Regional Disparities

This Part presents empirical findings on interjudge sentencing disparities in different time periods from 2000–2009. In Parts III.A–C, I explore disparities in three main sentencing outcomes: sentence length, below-range departures, and above-range departures, respectively. Then, in Part III.D, I explore potential sources of increasing disparities by analyzing changes in sentencing practices by demographic characteristics of judges. In Part III.E, I analyze whether disparities have also increased regionally across district courts. Finally, in Part III.F, I consider to what extent prosecutors contribute to measures of interjudge disparities.

A. Sentence Length

The following graphs present boxplots of the average sentence length imposed by each judge relative to the average sentence length of the district courthouse in which the judge sits.[169] A boxplot captures the distribution of judge sentencing practices, with a narrower spacing of the boxplot indicating less interjudge disparity. In particular, the top and bottom of the box capture the spread between the twenty-fifth percentile and seventy-fifth percentile mean judge effect, also known as the interquartile range (IQR). More extreme judge effects are represented in the whiskers of the boxplot. The greater the IQR, the larger the interjudge disparity within district courts.

---

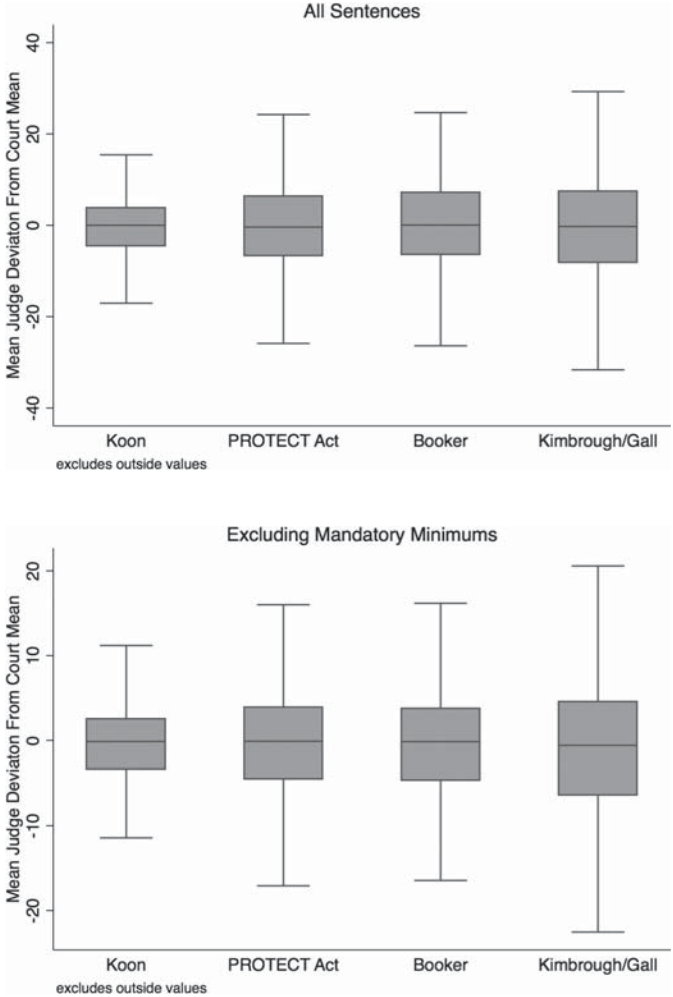
[168] *Booker* was decided on January 12, 2005. United States v. Booker, 543 U.S. 220 (2005). *Kimbrough* and *Gall* were decided on December 10, 2007. Kimbrough v. United States, 552 U.S. 85 (2007); Gall v. United States, 552 U.S. 38 (2007). Although the USSC data only contain information on sentencing month and year, the data are coded to denote which January 2005 cases were pre- and post-*Booker* and which December 2007 cases were pre- and post-*Kimbrough/Gall*. *See* Bureau of Justice Statistics, U.S. Dep't of Justice, Federal Justice Statistics Program: Defendants Sentenced Under the Sentencing Reform Act, 2005 Codebook 80 (2009), *available at* http://www.icpsr.umich.edu/cgi-bin/file?comp=none&study=24200&ds=1&file_id=1043712; Bureau of Justice Statistics, U.S. Dep't of Justice, Federal Justice Statistics Program: Defendants Sentenced Under the Sentencing Reform Act, 2007 Codebook (2009), *available at* http://www.icpsr.umich.edu/cgi-bin/file?comp=none&study=24148&ds=1&file_id=1043559.

[169] Case composition likely varies across district courts. Thus, I use a measure of judge sentence length that is comparable across all districts, which I obtain by calculating average sentence length by judge relative to the mean district courthouse sentence. All boxplots exclude outside values, defined as values that fall outside 1.5 times the interquartile range.

The first panel of Figure 4 indicates that more than half of judges are sentencing within a few months of the average courthouse mean, with some outliers in both directions. However, the spread of the distributions over the four time periods indicates an increase in the distribution of judge average sentence lengths relative to the court mean following *Kimbrough*/*Gall*. The spread between the twenty-fifth and seventy-fifth percentile (IQR) expands modestly but noticeably across the time periods.

If anything, the PROTECT Act period appears to be characterized by larger interjudge difference than during *Koon*. Recall that during the PROTECT Act period, the Guidelines were still mandatory and judges were subject to de novo appellate review. One

FIGURE 4. AVERAGE JUDGE SENTENCE LENGTHS IN MONTHS



Notes: Data is from the random sample 2000–2009.

might therefore expect measures of disparities to be lowest during this period. However, the direct effect of the PROTECT Act was to reduce judge-initiated below-range departures, rather than differences in overall sentence length. Moreover, as I show in Part III.F, measures of judge disparities necessarily capture both judge-induced differences as well as prosecutor-induced differences across cases, which are difficult to disentangle.

Note, however, that some of the interjudge disparities may be attributable to prosecutors' uneven application of mandatory minimums. The second panel of Figure 4 presents distributions of average judge sentences for those cases in which a mandatory minimum was not charged, which constitute approximately two-thirds of all cases. These cases better represent disparities more likely attributable primarily to judicial behavior. Judge differences from the courthouse mean are smaller in this subset of cases, but disparities appear to grow larger following *Booker*. As the figure shows, after *Kimbrough* and *Gall*, judge sentence lengths begin to depart more radically from court averages compared to the *Koon* period.

Statistical analysis of variance confirms the graphical patterns. Table 1 presents a measure of σ for sentence lengths, the causal impact of being randomly assigned a one-standard-deviation harsher judge in the courthouse.[170] Each measure of σ is also accompanied by a ninety-five percent confidence interval, which indicates the statistical probability that the true measure of σ lies within the interval range. During the *Koon* period, in which the Guidelines were binding and judges were governed by an abuse-of-discretion standard of appellate review, a defendant assigned to a harsher judge received a sentence 2.8 months longer than similar defendants sentenced by an average judge in the courthouse. By the time of the PROTECT Act, a defendant randomly assigned to a harsher judge received a 2.5 month longer sentence. Interjudge disparities increased following *Booker*, such that *Booker* and *Kimbrough*/*Gall* induced a doubling of interjudge disparity compared to the *Koon* period. A harsher judge sentenced a defendant 4.8 months longer than the court average in the immediate aftermath of *Booker* and over 5.9 months longer after *Kimbrough* and *Gall*.

The second panel of Table 1 excludes from the analysis those cases in which a mandatory minimum was charged. A one-standard-deviation harsher judge sentenced a defendant to 1.6 months more than the court average during *Koon* and 2.5 months longer during the

[170] Table 1 includes all defendants who are sentenced, including those who received probation (or sentence lengths of zero).

PROTECT Act period. Interjudge disparity for non–mandatory minimum cases remains at 2.5 months during *Booker*, rising to 3.2 months after *Kimbrough* and *Gall*. Changes in σ are not statistically significant from the PROTECT Act to *Kimbrough* and *Gall*. However, interjudge disparities are significantly larger following *Gall*/*Kimbrough* compared to the *Koon* period, more than doubling.[171] This evidence suggests that even in cases in which mandatory minimums were not charged, interjudge disparities have increased significantly.

Interestingly, the estimates of σ in the bottom panel are almost halved compared to those presented in the top panel, where all sentences are included. These results indicate that interjudge disparities are much larger in the subset of cases in which mandatory minimums are charged. This finding could be attributable to larger interjudge differences among types of cases with eligible mandatory minimums (such as drug and firearms offenses), disparate application of mandatory minimums by prosecutors, or both. Results presented in Part III.F below suggest the latter explanation: Both judges and prosecutors likely contribute to disparities in cases involving mandatory minimums.

TABLE 1. INTERJUDGE VARIATION IN SENTENCE LENGTHS

		ALL SENTENCES		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	2.818	2.262	3.511	66150
PROTECT Act	2.505	1.349	4.651	22469
Booker	4.833	3.951	5.914	35132
Kimbrough/Gall	5.952	4.853	7.299	22431
		EXCLUDING MANDATORY MINIMUMS		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	1.560	1.191	2.043	44403
PROTECT Act	2.474	1.704	3.592	15400
Booker	2.546	1.974	3.283	22056
Kimbrough/Gall	3.240	2.496	4.207	13571

Notes: Data is from the random sample from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

[171] A formal statistical analysis of whether σ is statistically significant from one period to another relies on the chi-squared distribution. However, an easier and more informal (albeit conservative) way of assessing significance is to look at whether the confidence intervals of σ overlap. That is, if the lower bound of σ in one period is greater than the upper bound of σ in another period, the estimates of σ are statistically significant from one another, suggesting a measurable change in disparity.

Table 2 presents evidence of interjudge variation in the decision to incarcerate, disentangling the decision to incarcerate from the sentence-length decision.[172] Given that *Booker* rendered the Guidelines advisory, a judge could potentially impose no prison sentence on a defendant, even if the Guidelines-recommended minimum was nonzero. Interestingly, Table 2 reveals that interjudge disparity in incarceration has not significantly increased throughout the four time periods. During *Koon*, a one-standard-deviation harsher judge was 2.7% more likely to incarcerate than the courthouse average. The effect fell to 2.2% during the PROTECT Act, to 1.8% during *Booker*, and rose back to 2.7% following *Kimbrough*/*Gall*.

TABLE 2. INTERJUDGE VARIATION IN INCARCERATION RATE

		ALL SENTENCES		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0275	.0232	.0326	66150
PROTECT Act	.0225	.0166	.0307	22469
Booker	.0178	.0131	.0243	35132
Kimbrough/Gall	.0267	.0209	.0339	22431

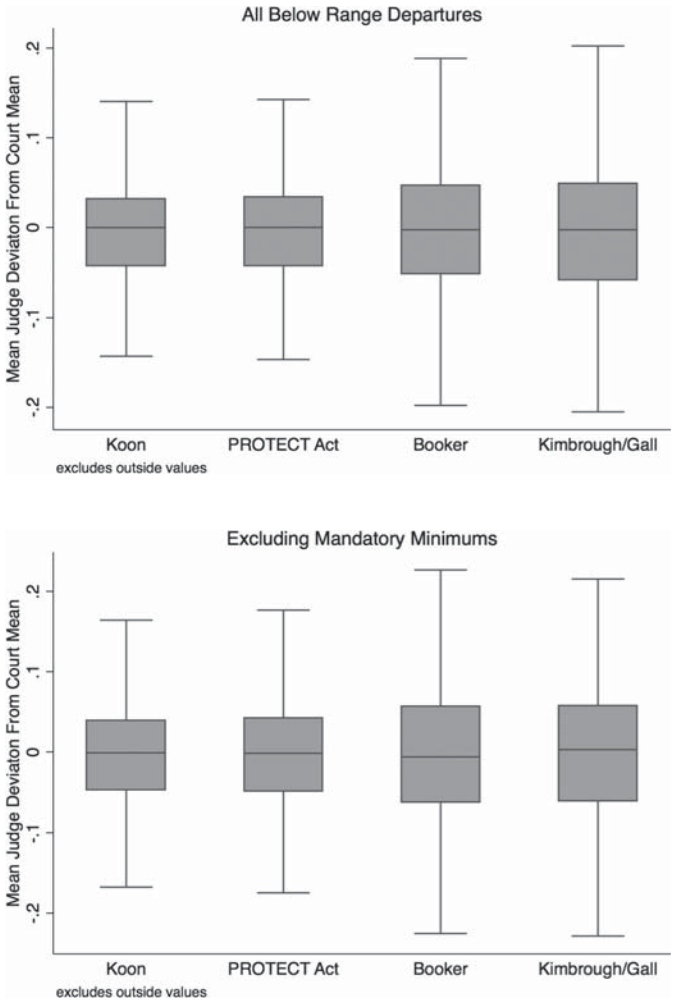
Notes: Data is from the random sample from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

B. Below-Range Departures

Interjudge disparities have also increased in the rate of below-range departures, which occur in approximately forty percent of cases. Figure 5 shows boxplots of average rates of below-range departures by judge, relative to the district courthouse mean. The rate of below-range departures in the aggregate was lowest during the PROTECT Act period (Figure 2), and the top panel of Figure 5 suggests that the distribution of judge below-range departures was also more compressed during the PROTECT Act compared to *Koon*. These results are consistent with the intended effect of the PROTECT Act, which was targeted at reducing below-range departures. However, interjudge deviations from the court mean expand visibly following *Booker* and *Kimbrough*/*Gall*. Figure 5 suggests that increasing interjudge disparities in sentence length, as described in Part III.A below, are partly attributable to growing interjudge disparities in the rate of below-range departures.

[172] I define incarceration as a binary indicator, where 1 indicates that the defendant has received a prison sentence and 0 indicates no prison sentence imposed. Defendants who do not receive a prison sentence often pay fines and serve probationary periods.

FIGURE 5. AVERAGE JUDGE RATES OF BELOW RANGE DEPARTURES



Notes: Data is from the random sample 2000–2009.

Table 3 confirms these graphical trends. The top panel of Table 3 presents results including all sentences. During the *Koon* period, a defendant who was assigned to a judge one standard deviation more "lenient" than the average judge was 4.7% more likely to be sentenced below the Guidelines-recommended minimum.[173] During the PROTECT Act, a similar judge was 4.1% more likely to sentence below range. Following *Booker*, the "lenient" judge's practices greater deviated from the courthouse average, with a 6.9% rate immediately

[173] As noted above, a judge is more "lenient" if she more frequently sentences defendants below the Guidelines range. *Supra* note 166.

following *Booker* and a 6.9% rate after *Kimbrough/Gall*. The increased likelihood of below-range departures following *Booker* and *Kimbrough*/*Gall* is statistically significant from the *Koon*-era rate and PROTECT Act rate, revealing markedly higher interjudge disparities.[174]

Excluding cases with mandatory minimums reveals a very similar trend, with the one-standard-deviation more lenient judge being 4.5% more likely to sentence below range during *Koon*, rising to 7.4% following *Booker*. Note that the magnitudes of σ when all sentences are included (top panel), and when mandatory minimums are excluded (bottom panel), are very similar. This finding suggests that interjudge disparities in below-range departures are real and substantial, and not merely the product of mandatory minimums. If anything, measures of interjudge disparity are lower in most periods when mandatory minimums are included. Recall that a mandatory minimum that exceeds the Guidelines-recommended minimum trumps the Guidelines and becomes the binding statutory minimum, potentially reducing interjudge disparity. The findings suggest that the application of

TABLE 3. INTERJUDGE VARIATION IN BELOW RANGE DEPARTURES

		ALL SENTENCES		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0470	.0402	.0549	66150
PROTECT Act	.0413	.0322	.0530	22469
Booker	.0689	.0595	.0799	35132
Kimbrough/Gall	.0689	.0578	.0821	22431
		EXCLUDING MANDATORY MINIMUMS		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0451	.0375	.0544	44403
PROTECT Act	.0436	.0325	.0585	15400
Booker	.0739	.0627	.0870	22056
Kimbrough/Gall	.0627	.0495	.0793	13571

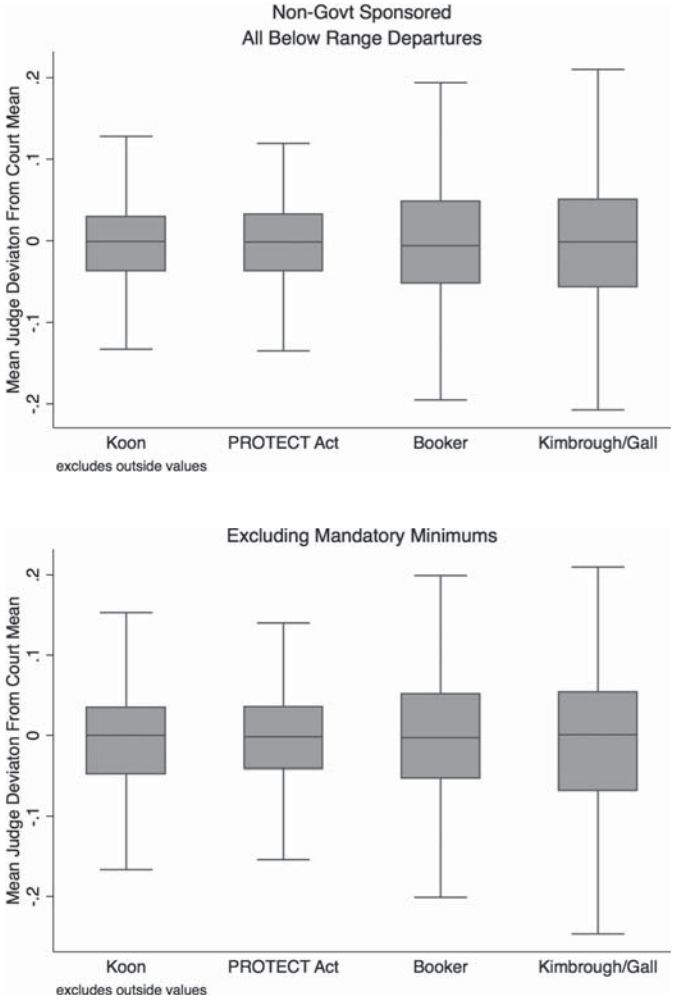
Notes: Data is from the random sample from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

[174] Looking at the top panel of Table 3, the ninety-five percent confidence interval of σ during *Koon* is (.0402, .0549) and the confidence interval of σ during *Kimbrough*/*Gall* is (.0578, .0821). Note that the two confidence intervals do not overlap, indicating that interjudge disparities are significantly larger following *Kimbrough*/*Gall* compared to *Koon*.

mandatory minimums may yield the appearance of interjudge consistency.[175]

Not only do mandatory minimums confound the accurate determination of interjudge disparities, so too do below-range departures that are government sponsored, which occur in approximately fifteen percent of cases.[176] Figure 6 thus presents boxplots of average rates of

FIGURE 6. AVERAGE JUDGE RATES OF BELOW RANGE DEPARTURES—NON–GOVERNMENT SPONSORED



Notes: Data is from the random sample 2000–2009.

175 *See* Scott, *supra* note 31, at 26 ("[M]andatory minimums may interfere with accurate assessment of inter-judge sentencing disparity by creating the illusion of inter-judge consistency.").

176 I define government-sponsored below-range departures as those that arise from a substantial-assistance motion under 18 U.S.C. § 3553(e) (2012) (permitting sentences

below-range departures by judge, relative to the district courthouse mean, excluding below-range departures that are government sponsored. Here, the changes in judge distributions follow the expected pattern. Judge deviations from the court mean were lowest during the PROTECT Act, which specifically focused on restricting non-government-sponsored below-range departures. After *Booker*, judge deviations in non-government-sponsored below-range departures expand, and continue to do so through *Kimbrough*/*Gall*.

Table 4 analyzes interjudge variation in this subset of below-range departures that are judicially initiated rather than the result of a government substantial-assistance motion. Table 4 indicates that the increasing interjudge disparities in below-range departures evidenced in Table 3 persist in this subset of departures. Interjudge disparities increased from 4.4% during *Koon* to 7.6% after *Booker* and *Kimbrough*/*Gall*, with the lowest interjudge disparities during the PROTECT Act. Interjudge disparities similarly increased throughout the period for the subset of cases not subject to a mandatory minimum, from 4.4% during *Koon* to over 7.4% after *Booker*. These results indicate that in the subset of departures that are most likely attributable to judicial behavior, the PROTECT Act was not only

TABLE 4. INTERJUDGE VARIATION IN BELOW RANGE NON–GOVERNMENT SPONSORED DEPARTURES

		ALL SENTENCES		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0438	.0372	.0515	52736
PROTECT Act	.0371	.0290	.0473	18552
Booker	.0759	.0661	.0871	28723
Kimbrough/Gall	.0755	.0638	.0894	18331
		EXCLUDING MANDATORY MINIMUMS		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0441	.0365	.0533	38019
PROTECT Act	.0374	.0273	.0514	13548
Booker	.0743	.0632	.0873	19540
Kimbrough/Gall	.0676	.0541	.0845	12131

Notes: Data is from the random sample from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

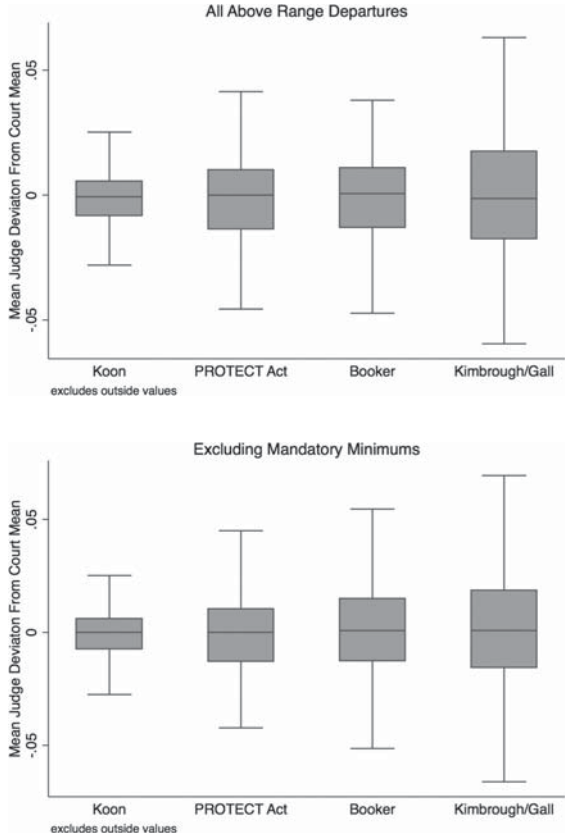
below a statutory minimum on motion by the prosecution indicating the defendant's substantial assistance in investigating another case) or U.S. SENTENCING GUIDELINES MANUAL § 5K1.1 (2013) (listing factors a judge should consider when deciding whether and how much to reduce a sentence for substantial assistance).

associated with the lowest aggregate rates of downward departures, but also the lowest interjudge disparities.

C. *Above-Range Departures*

Interjudge disparities have also increased in the rate of above-range departures, which comprise approximately two percent of cases. Figure 7 presents the distribution of average rates of above-range departures by judge—relative to their district courthouse mean—for all defendants. The graphs reveal an expansion in the distribution of above-range departure rates within district courts, particularly between the twenty-fifth and seventy-fifth percentile of the distribution. Increased interjudge deviations are also reflected in the rate of above-range departures for cases with no mandatory minimums charged. The spread between the twenty-fifth and seventy-fifth

FIGURE 7. AVERAGE JUDGE RATES OF ABOVE RANGE DEPARTURES



Notes: Data is from the random sample 2000–2009.

percentile following *Kimbrough*/*Gall* is visibly larger compared to pre-*Booker* spreads.

Table 5 presents measures of interjudge variation from the analysis of variance and reveals significant and substantial increases in interjudge disparities in above-range departures. In the top panel where all sentences are analyzed, a one-standard-deviation harsher judge was 0.65% more likely to sentence a defendant above range compared to the average judge in the courthouse during *Koon*. While interjudge variation did not change substantially from *Koon* to the PROTECT Act period, interjudge disparity is significantly higher after *Booker* and *Kimbrough*/*Gall*, with this harsher judge 1.7% more likely to sentence a defendant above range after *Kimbrough*/*Gall*, a statistically significant increase from *Koon*. Interjudge disparities in above-range departures more than doubled from the beginning to the end of the time period.

TABLE 5. INTERJUDGE VARIATION IN ABOVE RANGE DEPARTURES

	ALL SENTENCES			
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0065	.0047	.0089	66150
PROTECT Act	.0114	.0078	.0167	22469
Booker	.0123	.0091	.0165	35132
Kimbrough/Gall	.0170	.0125	.0230	22431
	EXCLUDING MANDATORY MINIMUMS			
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0072	.0053	.0100	44403
PROTECT Act	.0121	.0083	.0177	15400
Booker	.0129	.0091	.0184	22056
Kimbrough/Gall	.0154	.0102	.0234	13571

Notes: Data is from the random sample from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

Of course, mandatory minimums may explain a sizable fraction of defendants who are sentenced above range if the mandatory minimum trumps the maximum Guidelines-recommended sentence. When cases with mandatory minimums are excluded, patterns in interjudge disparities are similar. During *Koon*, interjudge disparities in above-range departures were minimal, with a one-standard-deviation harsher judge only 0.72% more likely to sentence above range. However, interjudge disparities doubled by the end of the time period to 1.5% after *Kimbrough*/*Gall*, a statistically significant increase from *Koon*. These results suggest that increases in above-range interjudge disparities are not the mere byproduct of mandatory minimums.

D. Sentencing Practices by Judge Demographics

The previous section finds that interjudge disparities in sentence length, below-range departures, and above-range departures have increased significantly following *Booker* and in particular after *Kimbrough*/*Gall.* In this section, I analyze whether increases in interjudge disparities are idiosyncratic, resulting from all judges changing their behavior in similar ways, or if judges are systematically differing from their colleagues based on observable traits.[177] Recall that due to the random assignment of cases to judges within a district courthouse, any difference in judge sentencing practices can be exclusively attributed to a judge effect.[178]

Consistent with previous research,[179] I find significant and systematic differences in the sentencing practices of both Democratic judicial appointees compared to their Republican-appointed peers and female judges compared to male judges. These differences are magnified in the aftermath of *Booker* and *Kimbrough*/*Gall*, suggesting that they are some of the sources of the growing interjudge disparities identified earlier. The coefficients presented in Table 6 represent the sentencing tendency of a particular type of judge compared to his or her colleagues within the same district courthouse, for an observably identical defendant and case, sentenced in the same month and year.[180]

---

[177] See *infra* Appendix C for the empirical methodology used to analyze the effects of judges' observable demographic characteristics on their treatment of similar defendants.

[178] *See supra* notes 128, 146 and accompanying text (noting that random assignment of cases to judges allows analysis of judges' sentencing behavior).

[179] *See supra* notes 35–37 (citing previous research linking judges' demographic characteristics to their sentencing habits).

[180] All results discussed are statistically significant at the one- to ten-percent level as indicated in Table 6. A finding that a coefficient is statistically significant at the one-percent level suggests very strong evidence for the result; with ninety-nine percent certainty, the real value of the coefficient is different from zero, the null hypothesis.

TABLE 6. SENTENCING PRACTICES BY JUDGE CHARACTERISTICS
ALL SENTENCES

	(1) Sentence	(2) Below Range Non-Govt	(3) Above Range
Post Booker Judge	2.503*	0.0489**	0.0106**
	(1.446)	(0.0241)	(0.00518)
Booker	-2.153	0.0748***	0.00517
	(2.160)	(0.0206)	(0.0102)
Democratic	-1.812***	0.0268***	-0.00148
	(0.550)	(0.00704)	(0.00168)
Democratic*Booker	-0.484	0.0172*	-0.00235
	(0.711)	(0.0100)	(0.00292)
Female	1.006**	-0.0215	0.000180
	(0.478)	(0.0137)	(0.00164)
Female*Booker	-1.396*	0.00636	-0.00297
	(0.765)	(0.0137)	(0.00243)
Black	-0.865	0.0141	-0.00476**
	(0.688)	(0.0178)	(0.00229)
Black*Booker	-0.608	-0.0148	0.00524
	(0.968)	(0.0390)	(0.00399)
Tenure	-0.140***	-0.000157	0.000207
	(0.0505)	(0.00125)	(0.000200)
Tenure*Booker	-0.0114	0.00316*	3.16e-05
	(0.0714)	(0.00186)	(0.000333)
Pre Guidelines	0.460	-0.00384	-0.00272
	(0.648)	(0.00939)	(0.00276)
Pre Guidelines*Booker	0.251	0.0138	0.00200
	(1.050)	(0.0208)	(0.00543)
Observations	145,975	145,975	145,975
R-squared	0.805	0.157	0.041

Notes: Data is from the random sample from 2000–2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district courthouse fixed effects, sentencing year and sentencing month fixed effects, and standard errors are clustered at the district courthouse level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Column 1 presents results for sentence length. Post-*Booker* judicial appointees sentence observably similar defendants to 2.5 months longer in prison compared to their pre-*Booker* appointed peers, significant at the ten percent level. In general, Democratic judicial appointees issue significantly shorter sentences compared to their Republican peers. Female judges significantly altered their practices from their male counterparts within the same courthouse. Prior to *Booker*, female judges sentenced observably similar defendants to an additional month in prison compared to their male colleagues. However, after *Booker*, female judges sentenced observably similar defendants to approximately 1.4 months less than their male colleagues,

compared to pre-*Booker* differences. Finally, judges with greater experience under the mandatory Guidelines regime issue slightly shorter sentences than judges with less experience.

Column 2 presents results for the rate of below-range departures not sponsored by the government. Interjudge disparities in rates of below-range departures appear to be somewhat attributable to differences by judge political affiliation. On average, Democratic judicial appointees are 2.7% more likely to depart downward. Following *Booker*, Democratic judicial appointees are even more likely to depart downward from the Guidelines-recommended range, compared to their Republican appointed colleagues. For a similar defendant and crime, Democratic judges were an additional 1.7% more likely to depart downward. Also striking are the interjudge differences between judges appointed pre-*Booker* and judges appointed post-*Booker*. Post-*Booker* judicial appointees are 4.9% more likely to sentence below range compared to their pre-*Booker* appointed colleagues.

Finally, column 3 presents results for above-range departures. Black judges are in general slightly less likely to sentence above range compared to White judges. Again, interjudge differences persist between judges appointed pre-*Booker* and judges appointed post-*Booker*. In general, post-*Booker* judicial appointees are 1.1% more likely to sentence above range than their pre-*Booker* appointed peers.

Overall, these results suggest that sentencing differences associated with judge gender and political affiliation are significant after *Booker* and *Kimbrough*/*Gall.* Such differences are likely sources of growing interjudge disparities. Given these large changes in interjudge disparities following *Booker*, judges do not appear to be completely anchored to the Guidelines.[181]

However, the finding that post-*Booker* judicial appointees are more likely to depart from the Guidelines than pre-*Booker* appointees is consistent with a story in which judges with *no* prior experience sentencing under the Guidelines regime are less anchored to the regime.[182] The "anchor" effect of the Guidelines sentence may be more prominent for pre-*Booker* appointees because these judges are

[181] Of course, some degree of anchoring is likely occurring, which indicates that these results are only lower-bound estimates of increases in interjudge disparities in a hypothetical system in which sentencing does not begin with the Guidelines calculation.

[182] In robustness checks, I find that the behavior of post-*Booker* appointees in my data is not because they are George W. Bush appointees, a conclusion based on comparisons with pre-*Booker* George W. Bush appointees. Rather, sentencing behavior seems to be associated with lack of experience under the binding Guidelines. Yang, *supra* note 38, at 20–21.

more acculturated to and experienced with constraining their sentences to the dictates of the Guidelines. In contrast, the "anchor" effect is less prominent for post-*Booker* appointees. These potential anchoring differences between pre- and post-*Booker* appointees suggest that defense lawyer James Felman's predictions may be true: Disparities may "increase as the years go by and the bench is filled with individuals who have no history with binding guidelines."[183] Yet interjudge disparities due to the entrance of new judges to the federal bench might only reflect a short-term surge in disparity. Eventually, all sitting judges will have no history with the binding Guidelines and interjudge disparities attributable to different levels of "anchoring" may fall.

E. Regional Disparity: Interdistrict Variation

Commentators have suggested that different political climates across districts and circuits can affect sentencing practices,[184] yielding empirical findings that jurisdictional effects are prominent in federal sentencing.[185] The 2012 Commission report finds that rates of non-government-sponsored below-range sentences increasingly depend upon the district court in which the defendant is sentenced and that the influence of the Guidelines on sentence length varies significantly

---

[183] Felman, *supra* note 39, at 98–99.

[184] *See, e.g.*, JEFFREY T. ULMER, SOCIAL WORLDS OF SENTENCING: COURT COMMUNITIES UNDER SENTENCING GUIDELINES 168–71 (1997) (comparing the political orientations and sentencing practices of three county court communities); Nora Demleitner, *The Nonuniform Development of Guideline Law in the Circuits*, 6 FED. SENT'G REP. 239, 239–40 (1994) (describing district and circuit specific "personas" in sentencing case law); Jeffery T. Ulmer & John H. Kramer, *Court Communities Under Sentencing Guidelines: Dilemmas of Formal Rationality and Sentencing Disparity*, 34 CRIMINOLOGY 383, 402–03 (1996) (arguing, based on an analysis of three county courts in Pennsylvania, that local courts operate under formal sentencing standards articulated by a guidelines regime as well as under substantive, extralegal factors relevant to local courts, such as "perceptions of the defendant's characteristics, local concerns, and court actors' organizational and individual interests").

[185] *See* Celesta A. Albonetti, *Sentencing Under the Federal Sentencing Guidelines: Effects of Defendant Characteristics, Guilty Pleas, and Departures on Sentence Outcomes for Drug Offenses, 1991–1992*, 31 LAW & SOC'Y REV. 789, 815–16 (1997) (finding significant circuit-specific sentencing practices for black and white defendants); Ronald S. Everett & Roger A. Wojtkiewicz, *Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing*, 18 J. QUANTATATIVE CRIMINOLOGY 189, 197–200 (2002) (finding harsher sentencing in the southern circuits compared to other circuits); Paula M. Kautt, *Location, Location, Location: Interdistrict and Intercircuit Variation in Sentencing Outcomes for Federal Drug-Trafficking Offenses*, 19 JUST. Q. 633, 659 (2002) ("[D]espite the federal system's congressionally mandated return to determinate sentencing, extralegal factors (specifically jurisdictional effects) continue to influence the federal sentencing system and its outcomes directly and indirectly.").

by circuit court.[186] However, some researchers have found that between-district variation in the effects of extralegal factors on sentencing have not increased following *Booker*.[187]

Recall that the identification of the impact of *Booker* on interjudge disparity within a district courthouse relies on the random assignment of cases to judges.[188] Such random assignment does not exist between districts, such that differences in district sentencing practices are likely also due to differing caseloads. For instance, the Commission has noted that simple comparisons of regional variations might be attributable to different types of crimes within general offense categories, such that frauds sentenced in the Southern District of New York are substantially different from frauds sentenced in the District of North Dakota.[189]

While I cannot control for unobservable differences across districts, the empirical methodology in this Article does control comprehensively for observable offender and case characteristics.[190] For the interdistrict results, I utilize the full USSC data from 2000–2009 as random assignment is no longer a prerequisite.[191] In the context of interdistrict disparities, analysis of variance now yields an estimate of the standard deviation of district effects on sentence length, σ, after controlling for case and defendant characteristics. Thus, a finding of $\sigma = 5$ now suggests that a defendant sentenced in a one-standard-deviation harsher district is sentenced to five more months in prison than he would be if he were sentenced in an average district court.

Figure 8 presents raw distributions of sentence lengths by circuit court, excluding life sentences.[192] While uncontrolled differences cannot be treated as regional effects because districts have very different case compositions, the raw data reveal substantial differences in sentence length, both in the distribution between the twenty-fifth per-

---

186 U.S. SENTENCING COMM'N, *supra* note 31, at 87–91.

187 *See* Jeffery Ulmer et al., *The "Liberation" of Federal Judges' Discretion in the Wake of the* Booker/Fanfan *Decision: Is There Increased Disparity and Divergence Between Courts?*, 28 JUST. Q. 799, 829–30 (2011) (finding no greater variation in the effects of certain extralegal variables after *Booker* than before).

188 *See supra* note 146 and accompanying text (describing the use of random case assignment in previous research on interjudge disparities).

189 *See* U.S. SENTENCING COMM'N, *supra* note 15, at 99–100 ("Similarly, variations in the rates of a particular type of departure among different districts must be evaluated within a larger context of each district's distinctive adaptation to the guidelines system. Inferring unwarranted disparity from uncontrolled comparisons of average sentences or rates of departure may be erroneous.").

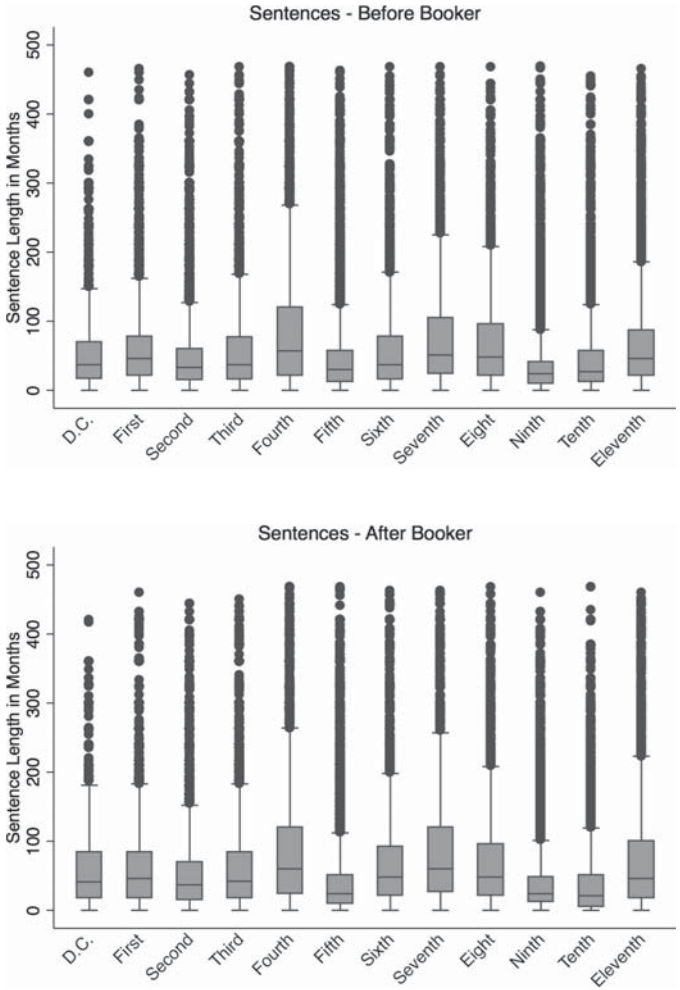
190 Nevertheless, the results on interdistrict variation should be interpreted with some unobserved differences across district courts that cannot be captured.

191 Random assignment of cases is no longer a prerequisite for statistical analysis because the unit of comparison is now district courts, rather than judges within courts.

192 Life sentences are top-coded as 470 months in the data set.

centile and seventy-fifth percentile and in the presence of outliers. For instance, prior to *Booker*, defendants sentenced in the Third Circuit received an average sentence of fifty-six months, compared to an average sentence of seventy-eight months for defendants sentenced in the neighboring Fourth Circuit. After *Booker*, the average Third Circuit defendant received sixty-two months in prison, while the average Fourth Circuit defendant received eighty-four months in prison.

FIGURE 8. DISTRIBUTION OF SENTENCE LENGTHS, BY CIRCUIT COURT



Notes: Data is from the USSC from 2000–2009.

Table 7 shows that after controlling for case and defendant characteristics, there is substantial variation in the sentence that a defen-

dant would receive depending on the district court in which he is sentenced. During the *Koon* period, a defendant sentenced in a one-standard-deviation harsher district court received a 7.4-month longer prison sentence. This interdistrict disparity increased to 7.8 months during the PROTECT Act, and then to 9.5 months immediately following *Booker*, reaching a 10.6-month difference after *Kimbrough*/*Gall*. By late 2007, interdistrict disparities were significantly larger than existed under *Koon*.

Analyzing the subset of cases in which a mandatory minimum was not charged more than halves the magnitude of σ, the measure of interdistrict differences. The lower panel of Table 7 indicates that a one-standard-deviation harsher district court sentenced a defendant to 3.5 months longer than the average district court under *Koon*, 4.1 months longer after the PROTECT Act, 4.5 months longer after *Booker*, and 4.6 months longer after *Kimbrough*/*Gall*. Again, the finding that interdistrict variation is more than halved when a statutory minimum is not charged indicates that the application of mandatory minimums is potentially a large contributor to interdistrict disparities, particularly in light of the fact that mandatory minimums only represent about one-third of the cases.

TABLE 7. INTERDISTRICT VARIATION IN SENTENCE LENGTHS

	ALL SENTENCES			
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	7.358	6.339	8.542	194510
PROTECT Act	7.829	6.716	9.126	99971
Booker	9.525	8.218	11.039	171097
Kimbrough/Gall	10.632	9.124	12.388	119957
	EXCLUDING MANDATORY MINIMUMS			
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	3.545	3.028	4.149	136932
PROTECT Act	4.119	3.516	4.825	72276
Booker	4.556	3.918	5.297	117585
Kimbrough/Gall	4.611	3.956	5.375	83917

Notes: Data is from the USSC from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects.

Table 8 reveals that district courts also significantly differ in their rates of below-range departures. During *Koon*, a defendant sentenced in a one-standard-deviation more lenient district is 12.1% more likely to be sentenced below the Guidelines range compared to the average district court. This measure of interdistrict variation for below-range departures remains relatively constant throughout the entire sample,

both including and excluding mandatory minimums. *Booker* and *Kimbrough*/*Gall* do not appear to have dramatically increased interdistrict disparity with regards to downward departures.

TABLE 8. INTERDISTRICT VARIATION IN BELOW RANGE DEPARTURES

		ALL SENTENCES		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.1208	.1041	.1402	157103
PROTECT Act	.1198	.1031	.1392	83453
Booker	.1341	.1156	.1555	147774
Kimbrough/Gall	.1289	.1109	.1497	105535
		EXCLUDING MANDATORY MINIMUMS		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.1125	.0969	.1308	103720
PROTECT Act	.1113	.0956	.1296	58066
Booker	.1251	.1078	.1453	97568
Kimbrough/Gall	.1256	.1073	.1463	71994

Notes: Data is from the USSC from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects and sentencing month fixed effects.

F. Prosecutorial Contribution to Disparities

Prosecutors likely contribute to observed interjudge disparities through their charging decisions. One area of great prosecutorial discretion is the decision to charge an offense that carries a mandatory minimum. As Justice Breyer has noted, mandatory minimum statutes "transfer sentencing power to prosecutors, who can determine sentences through the charges they decide to bring."[193] Strategic charging of mandatory minimums is likely more prominent after *Booker*, as some prosecutors charge mandatory minimums in order to curb judges' discretion.[194]

In a 2011 congressional report on mandatory minimum penalties, the Sentencing Commission found significant variation in the extent to which prosecutors applied enhancements for mandatory minimum

[193] Harris v. United States, 536 U.S. 545, 571 (2002) (Breyer, J., concurring).

[194] *See* Patrick J. Fitzgerald, U.S. Attorney, N. Dist. of Ill., Testimony Before the United States Sentencing Commission Hearing on the State of Federal Sentencing 252 (Sept. 10, 2009) (transcript available at http://www.ussc.gov/Legislative_and_Public_Affairs/Public_Hearings_and_Meetings/20090909-10/Public_Hearing_Transcript.pdf) ("[A] prosecutor is far less willing to forgo charging a mandatory minimum sentence when prior experience shows that the defendant will ultimately be sentenced to a mere fraction of what the guidelines range is.").

penalties under drug-trafficking offenses.[195] The report documented over seventy-five percent of eligible defendants receiving the statutory mandatory minimum penalty in some districts, but none of the eligible defendants in other districts receiving the enhancement.[196] Furthermore, recent work by researchers shows evidence of significant racial disparities in prosecutorial charging.[197]

Prosecutors are also in charge of the decision to reduce sentences below the mandatory minimum if the defendant offers "substantial assistance" during another investigation or prosecution.[198] If the government files a motion for substantial assistance for a case involving a mandatory minimum sentence, the court has the power to impose a sentence as low as probation.[199] Scholars have commented that the substantial assistance departure provision affords prosecutors immense discretion over both plea-bargaining and sentencing outcomes under the Guidelines.[200]

I find that the application of mandatory minimums may be a large contributor to interjudge disparities. Given the random assignment of

---

195 U.S. SENTENCING COMM'N, *supra* note 94, at 252–61.

196 *See id.* at 111–13, 255 (explaining that prosecutors reported wide variations in the district practices on seeking statutory minimum penalties).

197 *See* M. Marit Rehavi & Sonja B. Starr, *Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences* 1–5 (Univ. of Mich. Law & Econ. Working Paper No. 12-002, 2012) (finding that Black offenders are on average more than two times as likely to be subjected to a mandatory minimum sentence compared to similar White offenders, and that a major part of the racial gap in sentence length can be attributed to the prosecutorial bias in initial charging).

198 *See* 18 U.S.C. § 3553(e) (2012) (stating that courts may impose a sentence below the statutory minimum upon motion of the Government when the defendant offers substantial assistance); U.S. SENTENCING GUIDELINES MANUAL § 5K1.1 (2013) (noting that courts may also depart from the Guidelines in these circumstances). A judge has some leeway in reducing sentence length for certain drug-trafficking offenses under the "safety valve" provision, which allows a judge to reduce the punishment for low-level, first-time offenders. *See* 18 U.S.C. § 3553(f) (limiting the applicability of mandatory minimums for qualified offenders). The U.S. Sentencing Commission also notes that in recent years, White defendants in drug cases are more frequently granted the "safety valve" exception than other defendants. U.S. SENTENCING COMM'N, *supra* note 94, at 188.

199 According to the U.S. General Accounting Office, substantial-assistance motions reduce the average defendant's sentence length by fifty percent. *See* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-04-105, FEDERAL DRUG OFFENSES: DEPARTURES FROM SENTENCING GUIDELINES AND MANDATORY MINIMUM SENTENCES, FISCAL YEARS 1999–2001, at 15 (2003) ("On average, prosecutors' substantial assistance motions reduced drug offenders' 5-year mandatory minimum sentence by 33 months.").

200 *See* MICHAEL TONRY, SENTENCING MATTERS 81–82, 90–91 (1996) (describing how to amend substantial-assistance provisions to give judges more discretion); Nagel & Schulhofer, *supra* note 41, at 550 ("[U]se of the . . . substantial-assistance motion varies from jurisdiction to jurisdiction. . . . There is no limit on the amount of reduction once the motion is submitted. [It] is also used to avoid guideline ranges or mandatory minimum sentences for sympathetic defendants—even when there has been no genuine substantial assistance.").

cases to judges within a district courthouse, equal application of mandatory minimums among eligible cases prior to assignment would result in no significant judge differences in the rate of mandatory minimums applied. However, mandatory minimums can also be charged after assignment through the use of superseding indictments, giving prosecutors even greater control in their charging decisions. The results in Table 9 reveal small but significant differences in the percentage of cases with applicable mandatory minimums across judges. A judge one standard deviation from the mean was 2.2% more likely to see a case with a mandatory minimum during *Koon*, but 3.2% more likely after *Kimbrough*/*Gall* to see such a case. The increase in the differential rates of mandatory minimums after *Kimbrough*/*Gall* coincides with substantial increases in interjudge disparities in below-range departures. These results are consistent with a story in which prosecutors are attempting to rein in judicially induced downward departures through the use of mandatory minimums.

TABLE 9. INTERJUDGE VARIATION IN MANDATORY MINIMUMS

		ALL SENTENCES		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0222	.0179	.0275	66132
PROTECT Act	.0164	.0089	.0240	22469
Booker	.0292	.0231	.0368	35140
Kimbrough/Gall	.0323	.0242	.0431	22426

Notes: Data is from the random sample from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

If prosecutors strategically apply mandatory minimums in cases before lenient judges, they may actually increase interjudge *consistency*. However, mandatory minimums far exceed even the maximum of the Guidelines range roughly forty percent of the time when applied. Thus, given the crudeness and severity of mandatory minimums, strategic charging could plausibly enlarge differences in sentencing outcomes across judges. Of course, the contribution of disparate treatment of mandatory minimums to disparities is only the tip of the iceberg. Unobserved in the empirical data, but just as disconcerting, are the strategic charging decisions made by prosecutors even if a mandatory minimum is not applicable. For instance, prosecutors may vary in whether they seek sentencing enhancements by proving relevant aggravating facts by a preponderance of the evidence. Accordingly, the finding that prosecutorial charging of mandatory minimums contributes to measures of interjudge dispari-

ties is likely an underestimate of the real magnitude of the phenomenon.

TABLE 10. INTERJUDGE VARIATION IN SUBSTANTIAL ASSISTANCE

		ALL SENTENCES		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0372	.0307	.0450	49812
PROTECT Act	.0314	.0248	.0399	41298
Booker	.0286	.0233	.0351	73592
Kimbrough/Gall	.0362	.0301	.0434	52408

Notes: Data is from the random sample from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

Tables A2 and A3 in Appendix A confirm that a large portion of interdistrict differences in the sentencing of observably similar defendants arises from district variation in both the charging of mandatory minimums and the application of a substantial-assistance motion. Table A2 reveals that a defendant sentenced in a one-standard-deviation harsher district is approximately six percent more likely to be charged with a mandatory minimum.[201] Table A3 also presents evidence of large interdistrict differences in the rates of substantial-assistance motions, with a defendant being approximately eight percent more likely to be granted this form of downward departure in more "lenient" districts. As previously noted, the application of a substantial-assistance motion is often applied "to avoid guideline ranges or mandatory minimum sentences for sympathetic defendants—even when there has been no genuine substantial assistance."[202] However, interdistrict differences in average rates of mandatory minimums and rates of substantial-assistance motions do not appear to have increased significantly following *Booker* and *Kimbrough*/*Gall*.

IV
POLICY RECOMMENDATIONS

Numerous proposals for reforming federal sentencing arose in the aftermath of *Booker* due to dissatisfaction with the state of federal sentencing. Indeed, the equity stakes are high as similar offenders receive increasingly disparate sentences based on the mere happenstance of the sentencing judge to whom they are assigned. While sev-

[201] As with judges, a "harsher" district is defined as one in which the judges give longer sentences on average.

[202] Nagel & Schulhofer, *supra* note 41, at 550.

eral of the major proposals for reform contemplate a reduction in judicial discretion, the proposals largely ignore the role of other institutional actors—in particular, prosecutors—who play a central and powerful role in sentencing decisions and disparities.

Recognizing the degree of prosecutorial power, both scholars and district court judges have expressed the view that the current advisory Guidelines best achieve the goals of sentencing because they reflect the right balance between various actors in federal sentencing.[203] Of district judges surveyed in 2010, over seventy-five percent preferred the current advisory Guidelines system to other alternatives.[204] Fourteen percent of judges favored "[a] system of mandatory guidelines that comply with the Sixth Amendment . . . and have broader sentencing ranges than currently exist, coupled with fewer statutory mandatory minimum sentencing provisions."[205] Only three percent of judges preferred a return to the pre-*Booker* Guidelines system, suggesting that the overwhelming majority of judges would be opposed to reforms that solely curb judicial discretion.

This Part describes three of the major proposals for reform of federal sentencing after *Booker*: (A) "Topless" Guidelines that control only sentence minimums, but not maximums, (B) "*Blakely*-ized" Guidelines that require aggravating facts triggering longer maximums to be proven by a jury beyond reasonable doubt, and (C) a return to a presumptive Guidelines regime that more closely—though constitutionally—resembles the pre-*Booker* regime. I describe each proposal in turn, and then apply my empirical findings to assess the desirability of the various proposals in light of the goal of reducing interjudge disparities.

A. "Topless" Guidelines

A few months after *Booker*, the Department of Justice recommended a new "topless" Guidelines system in which the top of existing Guidelines ranges would essentially be removed.[206] Echoing the regime first proposed by Professor Frank Bowman, this construction would still allow judicial finding of facts that raise the minimum

---

[203] *See* Sara Sun Beale, *Is Now the Time for Major Sentencing Reform?*, 24 FED. SENT'G REP. 382, 384–85 (2012) (concluding that the system that has evolved since *Booker* is better than any alternative); Michael Tonry, *The U.S. Sentencing Commission's Best Response to* Booker *Is to Do Nothing*, 24 FED. SENT'G REP. 387, 387–88 (2012) (explaining why the current federal sentencing system is as good as it can be, despite its imperfections).

[204] U.S. SENTENCING COMM'N, *supra* note 30, tbl.19.

[205] *Id.*

[206] *See* Gonzales, *supra* note 26, at 326 (favoring "the construction of a minimum guideline system").

applicable sentence while remaining constitutional under the principles espoused in *Blakely*.[207]

As explained above, *Blakely* applied the Sixth Amendment to challenge judicial fact-finding that raised a defendant's maximum sentence.[208] As a result, the recommended topless Guidelines system—which allows for judicial fact-finding of precisely these aggravating factors—appears to comport with both *Blakely* and *Harris v. United States*,[209] which held that facts triggering a mandatory minimum sentence could be found by a judge rather than a jury.[210] However, commentators were skeptical of whether the Court's holding in *Harris* would survive *Booker*.[211] Indeed, since *Booker*, the constitutional viability of a topless Guidelines system has now been firmly rejected by the Supreme Court's recent holding in *Alleyne v. United States*, in which the Court squarely overruled *Harris*.[212]

Moreover, even if constitutionally permissible under the Sixth Amendment, the topless regime takes the prior mandatory Guidelines as the baseline, which some argue "would constitute a step backwards in the development and evolution of the federal sentencing system by exacerbating some of the worst features of the pre-*Booker* federal sentencing."[213] By binding judges to the applicable minimum sentence, the topless proposal would likely reintroduce the pre-*Booker* concerns associated with prosecutorial power in charging and plea bargaining.[214]

Indeed, the evidence from Part III provides empirical support for this proposition. Table 1 and Table 7, which present evidence of interjudge disparities and interdistrict disparities, show that these dis-

---

[207] *See* Memorandum from Frank Bowman to the U.S. Sentencing Comm'n (June 27, 2004), *in* 16 FED. SENT'G REP. 364, 367 (2004) (suggesting this scheme as a legislative response in the wake of *Blakely*).

[208] Blakely v. Washington, 542 U.S. 296, 304–06, 309 (2004).

[209] 536 U.S. 545 (2002).

[210] *Id.* at 567–69.

[211] *See, e.g.*, Frank O. Bowman, III, *Mr. Madison Meets a Time Machine: The Political Science of Federal Sentencing Reform*, 58 STAN. L. REV. 235, 261 (2005) (noting the logical inconsistencies between *Harris* and *Booker*); Erwin Chemerinsky, *Making Sense of* Apprendi *and Its Progeny*, 37 MCGEORGE L. REV. 531, 541 (2006) (finding *Harris* and *Booker* irreconcilable).

[212] *See* Alleyne v. United States, 133 S. Ct. 2151, 2163 (2013) ("Because there is no basis in principle or logic to distinguish facts that raise the maximum from those that increase the minimum, *Harris* was inconsistent with *Apprendi*. It is, accordingly, overruled.").

[213] Berman, *supra* note 122, at 363. Berman also discusses potential constitutional challenges to a "topless" Guidelines system. *Id.* at 359–62.

[214] *See id.* at 364 ("Consequently, the most problematic facets and the most disconcerting consequences in terms of prosecutorial power, disparity, and evasion experienced in the pre-*Booker* federal sentencing system would likely be aggravated by the enactment of any sort of topless guideline *Booker* fix.").

parities are reduced by almost a factor of two when mandatory minimums are excluded from analyses. These results suggest that the decision to charge a mandatory minimum contributes substantially to interjudge differences, such that these decisions are not made equally across all eligible cases. The results also indicate that mandatory minimum practices differ largely across U.S. district courts. Accordingly, any proposal that binds judges to the applicable minimum sentence would give greater power to prosecutors, likely resulting in greater disparities. Furthermore, the results in Table 5 suggest that there have been substantial increases in interjudge disparities in above-range departures even when a mandatory minimum is not charged. As a result, to the extent that a topless regime seeks to limit judicial discretion, it does so in an asymmetric manner.

*B. "*Blakely*-ized" Guidelines*

In *Booker*, Justice Breyer's ultimate remedy for the Sixth Amendment issues facing the Federal Sentencing Guidelines was to declare the Guidelines "effectively advisory."[215] But one could have imagined another approach: to "*Blakely*-ize" the Guidelines. Indeed, the Justices dissenting from the Breyer remedial opinion in *Booker* suggested leaving the mandatory Guidelines intact, but requiring that aggravating facts triggering longer maximums be either proven by a jury beyond a reasonable doubt or admitted by the defendant.[216]

However, introducing jury fact-finding into a mandatory Guidelines system would likely prove particularly complex. Justice Breyer, in his *Booker* remedial opinion, mused over how jury fact-finding might work, asking, "Would the indictment have to allege, in addition to the elements of robbery, whether the defendant possessed a firearm, whether he brandished or discharged it, whether he threatened death . . . ?"[217] Other scholars have echoed the concern that *Blakely*-izing the current version of the Guidelines would be procedurally unworkable and overwhelm juries required to make findings of fact.[218]

Addressing some of Justice Breyer's concerns, a 2005 American Bar Association (ABA) report suggested a version of the *Blakely*-ized

---

[215] United States v. Booker, 543 U.S. 220, 245 (2005).

[216] *See id.* at 284–85 (Stevens, J., dissenting) ("Rather than engage in a wholesale rewriting of the SRA, I would simply allow the Government to continue doing what it has done since this Court handed down *Blakely*—prove any fact that is *required* to increase a . . . sentence under the Guidelines to a jury beyond a reasonable doubt.").

[217] *Id.* at 254 (majority opinion).

[218] *See* Bowman, *supra* note 119, at 191 ("[T]he consensus view is that the Guidelines as now written are simply too complex and confusing to operate through juries.").

system espoused by Justice Stevens in his *Booker* dissent, accompanied with "simplifying the Guidelines by reducing both the number of offense levels and the number of adjustments and presenting the remaining, more essential, culpability factors to the jury."[219] The ABA contemplates that these critical culpability factors would be charged in the indictment and presented to the jury, resulting in a sentencing range.[220] The Guidelines maximum associated with the jury-determined range would be binding on judges such that upward departures without jury fact-finding would be impermissible.[221]

This Article does not comment on the relative abilities of judges and juries to determine the applicability of aggravating and mitigating factors. Even supposing that juries are capable of fact determinations of complex aggravating and mitigating factors under the *Blakely*-ized Guidelines,[222] once a jury has made factual determinations as to conduct based on what a prosecutor chose to charge, a judge is bound by these determinations. For instance, if a prosecutor did not charge a potential aggravating factor, a judge would not be allowed to consider this fact, even if it were applicable. As Justice Breyer noted, incorporating the jury trial requirement into federal sentencing would "weaken the tie between a sentence and an offender's real conduct," because such a system would "effectively deprive the judge of the ability to use post-verdict-acquired real-conduct information."[223] Thus, to the extent that the mandatory Guidelines regime enhanced prosecutorial discretion and disparity, jury fact-finding in the face of extensive plea bargaining "would move the system backwards in respect both to tried and to plea-bargained cases" by effectively "prohibit[ing] the judge from basing a sentence upon any conduct other than the conduct the prosecutor chose to charge."[224]

---

[219] AM. BAR ASS'N, CRIMINAL JUSTICE SECTION, REPORT ON *Booker* AND RECOMMENDATION (2005), *reprinted in* 17 FED. SENT'G REP. 335, 339 (2005).

[220] *Id.*

[221] *Id.*

[222] Justice Breyer has raised concerns over how a judge would expect a jury to determine Guidelines factors such as "relevant conduct," "loss" in a securities fraud case, or a defendant's behavior at trial—something a prosecutor cannot even observe at the time of the indictment. *Booker*, 543 U.S. at 254–55. However, some district court judges have successfully implemented *Blakely*-ized Guidelines by requiring the jury to make a determination of whether each sentencing enhancement sought by the government has been proven beyond a reasonable doubt or admitted by the defendant. *See, e.g.*, United States v. Gurley, 860 F. Supp. 2d 95, 99–101 (D. Mass. 2012) ("On the most practical level, I have learned that guideline sentencing enhancements easily can be folded into the trial, jury instructions, and verdict slip in every criminal case.").

[223] *Id.* at 252, 256.

[224] *Id.* at 256; *see also id.* ("[P]lea bargaining would likely lead to sentences that gave greater weight not to real conduct, but rather to . . . other factors that vary from place to place, defendant to defendant, and crime to crime. . . . [This] would necessarily move

This Article provides evidence suggesting that a large component of sentencing disparities stems from prosecutorial charging decisions, and that prosecutors vary in their charging across similar defendants convicted of similar conduct.[225] A requirement of jury fact-finding in a mandatory Guidelines regime may exacerbate these disparities. The goal of sentencing uniformity may be even more compromised in a system controlled by prosecutorial discretion because, "[a]s long as different prosecutors react differently, a system with a patched-on jury factfinding requirement would mean different sentences for otherwise similar conduct, whether in the context of trials or that of plea bargaining."[226]

C. Judge Sessions's Proposal

Most recently, former Sentencing Commission Chair Judge William K. Sessions III has recommended adoption of a simplified presumptive-Guidelines system.[227] Judge Sessions argues in favor of a new sentencing regime that balances two goals: (1) the need to reduce unwarranted sentencing disparities, curbing the ability of judges to use subjective notions of justice to mete out punishment; and (2) giving judges discretion to tailor sentencing to the unique circumstances of offenders and offenses.[228] Judge Sessions recommends a reduction in the number of possible sentencing ranges, but broader ranges to afford judges greater discretion.[229] In order to comply with the constitutional requirements identified in *Blakely*, Judge Sessions suggests that any facts that would increase the base offense level would have to be proven by a jury beyond a reasonable doubt, unless admitted to by the defendant, potentially through a bifurcated jury trial.[230] Judge Sessions also proposes simplifying the Guidelines by reducing the number of aggravating or mitigating factors that increase or decrease the base offense level under Chapter Two and Chapter Three of the Guidelines Manual, which many have argued are overly complex.[231]

---

federal sentencing in the direction of diminished, not increased, uniformity in sentencing."). For a more thorough discussion of the potential problems with this particular recommendation, see Berman, *supra* note 122, at 365–71.

225 *See supra* Part III.F (discussing this evidence).

226 *Booker*, 543 U.S. at 257.

227 Sessions, *supra* note 27, at 340.

228 *Id.* at 339.

229 *See id.* at 340–45 (describing recommended changes to the current Guidelines sentencing chart).

230 *Id.* at 346.

231 *Id.* at 347–48; *see also* Frank O. Bowman III, *The Failure of the Federal Sentencing Guidelines: A Structural Analysis*, 105 COLUM. L. REV. 1315, 1341 (2005) (suggesting that its complexity allows Congress to intervene in the details of sentencing law); Breyer, *supra* note 44, at 184 ("[T]he Guidelines are simply too long and too complicated.").

In deciding which aggravating factors to keep within the Guidelines, Sessions argues in favor of the strategy suggested by Justice Breyer: empirically reviewing which enhancements in Chapter Two are commonly used.[232]

Finally, Judge Sessions suggests a new form of appellate scrutiny because "[t]he threat of reversal [on appeal] is a key component of [effective] guidelines."[233] Within-range sentences would be "essentially unreviewable on appeal . . . [unless] a district court refused to consider all relevant factors or instead considered a prohibited factor, such as a defendant's race or gender."[234] In contrast, Judge Sessions proposes "relatively strict scrutiny by the appellate court" for downward departures.[235]

Critics of the Sessions proposal argue that the proposal would eliminate "judicial feedback to the Commission and [that] constructive evolution of the [G]uidelines would virtually cease," as judges would have limited authority in setting the applicable sentence range.[236] Undoubtedly, *Booker* has given judges the freedom to consider the particular circumstances of the offense and traits of the defendant.[237] To the extent that growing interjudge disparities are reflective of these considerations, disparities are warranted and judicial discretion is desirable. On the other hand, some have suggested that the shift to advisory Guidelines has been accompanied by increases in unwarranted disparities.[238]

---

[232] *See* Breyer, *supra* note 44, at 186 ("I believe the Commission should review the present Guidelines, acting forcefully to diminish significantly the number of offense characteristics attached to individual crimes. The characteristics that remain should be justified for the most part by data that shows their use by practicing judges to change sentences . . . ."); Sessions, *supra* note 27, at 349 (recommending Justice Breyer's empirical approach for distinguishing offense characteristics).

[233] Sessions, *supra* note 27, at 353 (quoting Stephanos Bibas et al., *Policing Politics at Sentencing*, 103 Nw. U. L. REV. 1371, 1371 (2009)).

[234] *Id.* at 353–54.

[235] *Id.* at 354; *see also id.* at 353 ("District courts' choices of sentences . . . would be essentially unreviewable on appeal so long as the courts *considered* all of the relevant aggravating and mitigating factors identified in the application notes and all other relevant factors in the Guidelines Manual before imposing a particular sentence.").

[236] Baron-Evans & Stith, *supra* note 29, at 1716.

[237] *See id.* at 1742 (commenting on the ways in which the sentencing process has changed for the better).

[238] *See* Frank O. Bowman, III, *Nothing Is Not Enough: Fix the Absurd Post-*Booker *Federal Sentencing System*, 24 FED. SENT'G REP. 356, 356 (2012) ("[T]he post-*Booker* advisory system retains most of the flaws of the system it replaced, while adding new ones, and its sole relative advantage—that of conferring additional (and effectively unreviewable) discretion on sentencing judges—is insufficient to justify its retention as a permanent system."); Sessions, *supra* note 27, 329–31 (remarking on the signs of increased disparities).

The empirical findings in Part III of this Article reveal that interjudge disparities have doubled from the period of mandatory Guidelines sentencing to post-*Booker* sentencing, with a defendant potentially receiving a six-month longer sentence if assigned by happenstance to a harsh judge.[239] Certainly, a return to presumptive Guidelines would mechanically reduce interjudge disparities by greatly limiting judicial discretion. However, the empirical evidence seeks to ascertain the effect of the sentencing regime on interjudge disparities in outcomes that are most likely attributable to judge behavior. Differences in sentence lengths can be attributable to both judge disparities as well as differences in charging of mandatory minimums. The findings in this Article suggest that a return to a presumptive regime, without any changes in mandatory minimums, would only go partway in reducing disparities and would curtail potentially desirable judicial discretion.

While this Article does not provide evidence supporting a return to presumptive Guidelines, it does suggest that the strictness of appellate review is a potentially important constraint on judicial discretion in sentencing. Interjudge disparities in below-range departures were generally lowest during the PROTECT Act, which imposed de novo review.[240] Furthermore, empirical evidence suggests that *Booker* alone did not contribute to recent increases in interjudge disparities. Rather, it appears to be the impact of *Booker* plus reduced appellate scrutiny following *Rita*, *Gall*, and *Kimbrough* that are responsible for the largest increases in interjudge disparities.

Thus, reforms to strengthen the degree of appellate review could possibly reduce interjudge sentencing disparities. In *Gall*, the Court did not require appellate courts to insist upon "extraordinary" circumstances to justify a sentence outside the Guideline range, specifically rejecting stronger justifications for sentences that departed more greatly from the Guidelines.[241] In order to constrain interjudge disparities, the Commission could require district court judges to provide a heightened justification for more severe departures from the prescribed sentence, without coming too close to an "impermissible pre-

239 *Supra* Table 1.

240 *Supra* Tables 3–4.

241 Gall v. United States, 552 U.S. 38, 47 (2007) ("In reviewing the reasonableness of a sentence outside the Guidelines range, appellate courts may therefore take the degree of variance into account and consider the extent of a deviation from the Guidelines. We reject, however, an appellate rule that requires 'extraordinary' circumstances to justify a sentence outside the Guidelines range . . . .").

sumption of unreasonableness for sentences outside the Guidelines range . . . [that] would not be consistent with *Booker*."[242]

This Article's findings also suggest that mandatory minimums are likely a large contributor to disparities, stemming from judicial disagreement over the lengths of mandatory minimums, strategic charging of mandatory minimums by prosecutors, or both. Even absent a wholesale elimination of mandatory minimums, uniform charging policies could reduce disparities. Significantly, federal prosecutors and policymakers may be moving in this direction. In August 2013, Attorney General Eric Holder announced a new Department of Justice policy directing U.S. Attorneys to avoid charging "low-level, nonviolent drug offenders" with offenses carrying mandatory minimum sentences.[243] Following suit, the Sentencing Commission voted unanimously to continue its work on addressing concerns with mandatory minimum penalties.[244] While the impact of these policy directives remains to be seen, reforms to reduce sentencing disparities must ultimately include changes to mandatory minimum penalties.

Conclusion

Exploiting the random assignment of cases to judges in district courthouses representing seventy-four federal district courts, this Article finds a significant increase in interjudge disparities from the *Koon* period to after *Kimbrough* and *Gall*. A defendant sentenced by a harsh judge during the *Koon*-era was sentenced to a 2.8 month-longer term than if he had been sentenced by the average judge. The same defendant would have been sentenced to almost six months longer after *Kimbrough* and *Gall*, a doubling of interjudge disparities. Increased interjudge disparities persist even after excluding cases in which mandatory minimums were charged, suggesting that judges are not completely anchored to the Guidelines. These findings raise large equity concerns because the identity of the assigned sentencing judge contributes significantly to the disparate treatment of similar offenders convicted of similar crimes.

---

242 *Id.* at 47.

243 Memorandum from Eric H. Holder, Jr., U.S. Attorney Gen., to the U.S. Attorneys and Assistant Attorney Gen. for the Criminal Div. 2 & n.3 (Aug. 12, 2013), *available at* http://www.justice.gov/sites/default/files/oip/legacy/2014/07/23/ag-memo-department-policypon-charging-mandatory-minimum-sentences-recidivist-enhancements-in-certain-drugcases.pdf.

244 Press Release, U.S. Sentencing Comm'n, U.S. Sentencing Commission Selects Policy Priorities for 2013–2014 Guidelines Amendment Cycle (Aug. 15, 2013), *available at* http://www.ussc.gov/Legislative_and_Public_Affairs/Newsroom/Press_Releases/20130815_Press_Release.pdf.

Increases in between-judge differences following *Booker* and *Kimbrough*/*Gall* appear linked to observable judicial demographics such as gender, political affiliation of appointing president, and whether a judge has ever sentenced under the mandatory-Guidelines regime. I also find modest evidence of increases in interdistrict differences following *Kimbrough*/*Gall*, with large interdistrict differences in sentence length, rates of below-range departures, rates of mandatory minimums, and rates of substantial-assistance motions. However, the magnitudes of both interjudge and interdistrict disparities are drastically smaller when mandatory minimums are excluded, suggesting that prosecutorial charging decisions may be a major contributor to sentencing disparities.

Overall, these results suggest that the shift to an advisory Guidelines regime under *Booker*, coupled with lowered standards of appellate scrutiny after *Rita*, *Kimbrough*, and *Gall*, have led to greater interjudge disparities. Prosecutorial charging decisions, at least in the application of mandatory minimums, appear to play a substantial role in explaining disparities. While this Article is a first step in disentangling the sources of disparities ascribable to various actors, its primary limitation is its inability to thoroughly analyze all the disparities that can arise in earlier stages of the criminal justice system, such as charging decisions and plea bargaining. Nevertheless, the results of this Article caution against recent proposals to move back toward a sentencing system in which judges are bound by the decisions of prosecutors. Instead, this Article suggests that it may be wise to modify standards of appellate review, as well as to revisit the desirability of mandatory minimums.

APPENDIX

A. *Testing for Random Assignment*

To test for random assignment, I regress five defendant characteristics on district courthouse by sentencing-year fixed effects, sentencing-month fixed effects, and judge fixed effects. The five defendant characteristics are gender, age, a Black race indicator, number of dependents, and an indicator for education noting whether the defendant holds less than a high school degree. Intuitively, there should be no significant correlation between a particular judge and defendant characteristics if cases are randomly assigned.

However, in testing the random assignment of defendants across these five characteristics, I encounter the problem that defendant characteristics are not fully independent. For instance, black defendants are also more likely to have completed less than a high school degree.[245] To address the confounding nature of these characteristics, I use seemingly unrelated regression (SUR) to test for random assignment. SUR allows me to test random assignment simultaneously for all the five defendant characteristics, addressing correlations.[246]

SUR can be formally described as the regression model:

$$Y_{ijdtm} = \alpha_0 + \gamma_d + \delta_t + \gamma_d * \delta_t + \lambda_m + \kappa_j + \epsilon_{ijdtm}$$

where Y_{ijdtm} is a characteristic of defendant i, sentenced by judge j in district court d in year t and month m. The specification includes district-court fixed effects (γ_d), sentencing-year fixed effects (δ_t), sentencing-month fixed effects (λ_m), and sentencing-judge fixed effects (κ_j) to accurately compare cases assigned to judges in the same courthouse, and in the same year and month.

To formally test for random assignment, I test the null hypothesis of no judge effects—κ—using an F-test. The p-value for this F-test determines whether the defendant characteristics do not differ significantly among the cases that are assigned to district court judges in the same district courthouse, sentencing year, and sentencing month. A large p-value would signify the nonrejection of the null hypothesis, and lead to the conclusion that random assignment was present.

Table A1 presents the randomization checks done by district courthouse, along with associated p-values. I drop all courthouses with F-test p-values less than 0.05, but results are robust to other cutoffs. Dropped courthouses are indicated with "**."

---

[245] Summary statistics on federal defendants are available upon request to the author.

[246] Testing each characteristic individually would result in incorrect standard errors if the demographic characteristics are correlated. For a discussion of the SUR technique, see Autor & Houseman, *supra* note 156, at 106–07.

TABLE A1. RANDOMIZATION TESTS 2000–2009

District Courthouse	No. Obs.	P-value
ME (0)	1,722	0.3623
MA (1)	4,041	0.0666
NH (2)	1,457	0.5420
PR (4)	4,216	0.0820
CT** (5)	2,811	0.0002
NY North - Syracuse (6)	1,058	0.0941
NY North - Albany (6)	1,589	0.2107
NY East - Islip (7)	908	0.0920
NY South - White Plains (8)	832	0.1595
NY West - Rochester (9)	1,058	0.1368
VT (10)	1,304	0.0739
DE (11)	855	0.5426
NJ ** (12)	5,414	0.0000
PA East** (13)	7,208	0.0000
PA Middle - Scranton (14)	1,142	0.3819
PA Middle - Williamsport (14)	435	0.0691
PA West** (15)	3,088	0.0000
MD (16)	4,453	0.2884
NC East (17)	4,570	0.4106
NC Middle - Greensboro (18)	2,629	0.3887
NC West**(19)	4,301	0.0000
SC - Columbia (20)	1,295	0.7777
SC - Florence (20)	1,358	0.0786
VA East - Alexandria (22)	2,510	0.3583
VA East - Richmond (22)	2,446	0.0589
VA East - Newport News (22)	1,125	0.0709
VA West (23)	3,913	0.4506
WV North - Martinsburg (24)	594	0.4937
WV South (25)	2,297	0.5926
AL North (26)	3,867	0.0597
AL Middle (27)	1,522	0.2279
AL South (28)	2,552	0.2568
FL North - Tallahassee (29)	733	0.1338
FL North - Panama City (29)	432	0.1353
FL Middle (30)	8,206	0.0850
FL South (31)	11,747	0.3807
GA North** (32)	5,356	0.0000
GA Middle (33)	2,506	0.1643
LA East (35)	3,058	0.1635
LA West (36)	2,542	0.1538
MS North (37)	1,071	0.4247
MS South (38)	3,318	0.0900

TABLE A1. RANDOMIZATION TESTS 2000–2009 (CONTINUED)

TX North - Forth Worth (39)	1,709	0.0995
TX East - Beaumont (40)	1,732	0.9159
TX South - Brownsville (41)	2,636	0.7816
TX South - Corpus Christi (41)	2,824	0.4200
TX South - Laredo (41)	5,259	0.1849
TX West - Del Rio (42)	1,432	0.1721
TX West - Midland-Odessa (42)	1,097	0.6006
TX West - Austin (42)	1,560	0.7219
TX West - Pecos (42)	682	0.3520
TX West - San Antonio (42)	1,090	0.5264
KY East - Covington (43)	756	0.5020
KY East - Pikeville (43)	338	0.7396
KY East - Lexington (43)	1,415	0.2080
KY West - Bowling Green (44)	438	0.0554
KY West - Louisville (44)	1,179	0.0680
MI East - Bay City (45)	420	0.4750
MI East - Flint (45)	585	0.7528
MI West (46)	3,337	0.0571
OH North - Toledo (47)	829	0.1000
OH South (48)	3,884	0.1640
TN East - Knoxville (49)	1,293	0.3125
TN East - Greeneville (49)	871	0.0825
TN Middle** (50)	1,991	0.0457
TN West - Eastern (51)	703	0.5508
IL North** (52)	7,112	0.0000
IL Central (53)	2,640	0.1551
IL South (54)	3,148	0.1708
IN North (55)	2,944	0.0539
IN South (56)	2,290	0.0642
WI East - Milwaukee (57)	2,120	0.4223
WI West (58)	1,486	0.1221
AR East (60)	2,330	0.0838
AR West**(61)	1,451	0.0001
IA North** (62)	2,797	0.0003
IA South (63)	2,702	0.4151
MN (64)	3,825	0.2747
MO East (65)	5,659	0.0762
MO West (66)	5,034	0.0770
NE - Omaha (67)	3,605	0.1126
ND (68)	1,645	0.4508
SD - (69)	2,290	0.2520

TABLE A1. RANDOMIZATION TESTS 2000–2009 (CONTINUED)

AZ - Phoenix (70)	3,461	0.5011
AZ - Tuscon (70)	7,678	0.7057
AZ - Yuma (70)	662	0.3392
CA North (71)	2,652	0.3093
CA East (72)	5,180	0.7000
CA Central - Santa Ana (73)	1,057	0.3756
CA South - El Centro (74)	2,372	0.2427
HI** (75)	2,900	0.0015
ID (76)	1,856	0.2740
MT - Missoula (77)	452	0.1968
MT - Billings (77)	1,120	0.0624
NV (78)	4,171	0.0699
OR (79)	3,595	0.1110
WA East (80)	2,699	0.5287
WA West (81)	3,856	0.3358
CO** (82)	3,838	0.0000
KS (83)	3,987	0.1265
NM (84)	6,668	0.5484
OK North** (85)	1,529	0.0419
OK East (86)	849	0.5343
OK West** (87)	1,959	0.0001
UT (88)	3,763	0.8875
WY (89)	2,091	0.2775
DC (90)	2,746	0.5720
AK (95)	1,323	0.1546
LA Middle (96)	1,164	0.1805

Notes: Data is from the random sample from 2000–2009. I drop judges who retired or were terminated prior to 2000, and judges and district offices with an annual caseload of less than 25. For each district court, I control for district office by sentencing year, sentencing month, and judge fixed effects. P-values reported test whether judge fixed effects differ significantly from zero and are from a seemingly unrelated regression (SUR) on five defendant characteristics: defendant gender, age, black race indicator, number of dependents, and less than high school indicator. ** indicates dropped courthouses.

TABLE A2. INTERDISTRICT VARIATION

	Application of Mandatory Minimums			
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0557	.0480	.0645	196358
PROTECT Act	.0628	.0541	.0730	100492
Booker	.0573	.0493	.0665	171432
Kimbrough/Gall	.0652	.0562	.0758	120021

Notes: Data is from the USSC from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects and sentencing month fixed effects.

TABLE A3. INTERDISTRICT VARIATION

	APPLICATION OF SUBSTANTIAL ASSISTANCE			
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0837	.0722	.0971	186982
PROTECT Act	.0877	.0756	.1017	98596
Booker	.0850	.0734	.0983	170041
Kimbrough/Gall	.0765	.0660	.0886	119872

Notes: Data is from the USSC from 2000–2009. All regressions contain demographic controls and controls for offense type, offense level, and criminal history. Regressions also contain sentencing year fixed effects and sentencing month fixed effects.

B. Analysis of Variance

I implement an analysis of variance using a defendant-level random-effects specification of the form:

$$Y_{ijdtm} = X_i * \beta + \gamma_d + \delta_t + \gamma_d * \delta_t + \lambda_m + v_{ijdtm},$$

$$\text{where} \quad v_{ijdtm} = \mu_{jdtm} + \epsilon_{ijdtm}$$

The dependent variable Y_{ijdtm} is the sentence length in months for defendant i assigned to judge j in district court d, sentenced in year t and in month m. The control variables include defendant and crime characteristics (X_i),[247] sentencing-year fixed effects (δ_t), and sentencing-month fixed effects (λ_m). γ_d are indicator variables for the district courthouse in which sentencing occurred. The residual v_{ijdtm} is composed of a judge effect or value added that is constant for a judge over time, and an idiosyncratic defendant effect. I estimate the coefficients β and the judge effects μ via maximum likelihood (MLE). MLE estimation yields consistent estimates of β if the judge random effects are uncorrelated with the control variables X.

I estimate the magnitude of the judge effects under a mixed random effects specification, assuming that μ_{jdtm} is distributed $N(0, \sigma^2)$.[248] Intuitively, within judge variance in v_{ijdtm} is used to estimate the defendant variance:

$$\hat{\sigma}_\epsilon^2 = Var\left(v_{ijdtm} - \bar{v}_{jdtm}\right)$$

The variance in the judge effect is the remainder:

$$\hat{\sigma}_\epsilon^2 = Var\left(v_{ijdtm}\right) - \hat{\sigma}_\epsilon^2$$

[247] Previous researchers, such as Joshua Fischman and Max Schanzenbach have identified endogenous changes in Guidelines offense level calculations. *See* Fischman & Schanzenbach, *supra* note 32, at 730 (examining how doctrinal changes in sentencing law affected judges' tendency to depart from the Guidelines). Results are robust to exclusion of any measure of offense level and are on file with the *New York University Law Review*.

[248] Boxplots of judge effects relative to the district mean presented in Figures 4–6 support the assumption that judge effects are normally distributed.

The estimated standard deviation of judge effects on sentence length is $\sigma_\mu = X$, implying that a one standard deviation increase in judicial harshness raises a defendant's sentence by X months. Because the regression specification includes district-courthouse fixed effects, this measure represents the impact of being assigned a judge one standard deviation higher in harshness in the within–district-court distribution.

C. Judge Demographic Regression

To analyze the differential sentencing practices of certain types of judges, I use ordinary least squares (OLS) regression. The methodology captures how judges differ in their treatment of similar defendants in response to increased judicial discretion, compared to other judges within the same district courthouse. Because cases are randomly assigned to judges within a district court, judge identifiers allow one to compare judges within the same court, capturing judge differences in sentencing rather than different caseloads.

I identify the sources of increasing interjudge disparities post-*Booker* using a specification of the form:

$$Y_{icodtm} = \beta_0 + \alpha * Judge_i * Booker + \beta_1 * Booker + \beta_2 * Race_i + \beta_3 * X_i + Guide_{ico} + Offtype_i + \gamma_d + \delta_t + \gamma_d * \delta_t + \lambda_m + \epsilon_{icodtm}$$

Y_{icodtm} is a sentencing outcome for defendant i, with criminal history category c and offense level o, sentenced in district court d in year t and month m. Main outcomes include sentence length measured in months and binary indicators for below-range sentencing and above-range sentencing.

$Judge_i$ includes judicial demographics such as race, gender, political affiliation, an indicator for pre- or post-Guidelines appointment, tenure under the Guidelines, and an indicator for pre- or post-*Booker* appointment. The main coefficients α capture the impact of particular judicial characteristics on sentencing outcomes in the wake of *Booker* and its progeny. *Booker* is an indicator variable for defendants sentenced after the *Booker* decision.

$Race_i$ is a dummy variable for defendant i's race: White, Black, Hispanic, or other. 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.

$Guide_{ico}$ includes dummy variables for criminal history category c and offense level o, and each unique combination of criminal history category and offense level. The interaction captures differential sentencing tendencies at each unique 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. I also control for final criminal history category. $Off\ type_i$ 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). All standard errors are clustered at the district courthouse level to account for serial correlation.