ISSN 1936-5349 (print)
ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS
FELLOWS' DISCUSSION PAPER SERIES

HAVE INTER-JUDGE SENTENCING DISPARITIES INCREASED IN AN ADVISORY GUIDELINES REGIME? EVIDENCE FROM *BOOKER*

Crystal S. Yang

Discussion Paper No. 52

08/2013

Harvard Law School
Cambridge, MA 02138

Contributors to this series are John M. Olin Fellows or Terence M. Considine Fellows in Law and Economics at Harvard University.

This paper can be downloaded without charge from:

The Harvard John M. Olin Fellow's Discussion Paper Series:
http://www.law.harvard.edu/programs/olin_center/

Have Inter-Judge 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. *What has been the result of reintroducing greater judicial discretion on inter-judge disparities? This Article utilizes new data to undertake the first national empirical analysis of inter-judge disparities post* Booker. *Relying on the random assignment of cases to judges, I find that inter-judge 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. The application of mandatory minimums by prosecutors, potentially through the use of superseding indictments, is another prominent source of disparities.*

JEL Classifications: J15, K14, K40

*Olin Fellow and Instructor in Law, University of Chicago Law School. Support was provided by Harvard Law School's John M. Olin Center for Law, Economics, and Business and the Multidisciplinary Program on Inequality and Social Policy at Harvard. I also thank Gary Chamberlain, Raj Chetty, Will Dobbie, Judge Nancy Gertner, Claudia Goldin, Paul Hofer, Louis Kaplow, Lawrence Katz, Ilyana Kuziemko, J.J. Prescott, Shayak Sarkar, Steven Shavell, and Kate Stith for helpful comments and suggestions, and participants in the Harvard Labor Economics/Public Finance Seminars, Harvard Law School Law and Economics Seminar, and Chicago Junior Faculty Workshop. 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: csyang@uchicago.edu.

TABLE OF CONTENTS

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 judge would have been sentenced to twenty years had he been assigned to another judge.[2] Decrying these disparities and championing sentencing reform, federal district judge Marvin Frankel 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 (SRA) of 1984, 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 guidelines aimed at the

---

[1] *See* S. REP. NO. 97-307, at 5 (1981).

[2] Anthony Partridge & William B. Eldridge, Federal Judicial Center, *The Second Circuit Study: A Report to the Judges of the Second Circuit* 36 (1974).

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

[4] *See* S. Rep. No. 98-225, at 38 (1983) (Senate Report on precursor to federal Sentencing Reform Act of 1984) ("[E]very day Federal judges mete out an unjustifiably wide range of sentences to offenders with similar histories, convicted of similar crimes, committed under similar circumstances. . . . These disparities, whether they occur at the time of the initial sentencing or at the parole stage, can be traced directly to the unfettered discretion the law confers on those judges and parole authorities responsible for imposing and implementing the sentence"); *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 service in prison.").

[5] Pub. L. No. 98-473, 98 Stat 1837, 1987 (1984) (codified as amended at 18 USC §3551 et seq., 28 USC §991 et seq.). For a discussion of the efforts leading up the the promulgation 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 (1993).

[6] U.S.S.G. §1A.1, intro to comment., pt. A, ¶2 (Congress "sought uniformity in sentencing by narrowing the wide disparity in sentences imposed by different federal courts for similar criminal conduct"); S. REP. NO. 225, at 45 (1983) ("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). The Commission was placed in the Judicial Branch because Congress concluded that "sentencing should remain primarily a judicial function," and because sitting judges would serve on the Commission. The Commission is comprised of seven voting members. 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 States" and no more than four members of the Commission could be members of the same political party. *See id.* The Commission later withstood separation-of-powers challenges, as the Court rejected several constitutional challenges to the Commission and its delegated authority. *See Mistretta v. United States*, 488 U.S. 361 (1989).

primary goal of 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 Anderson, Stith, and Kling revealed that the Guidelines were somewhat successful in reducing inter-judge sentencing disparity.[10] The authors estimate that the expected difference in sentence length between judges fell from 17% (4.9 months) in 1986-1987 to 11% (3.9 months) in 1988-93 in 25 cities where case assignment was found sufficiently random.[11] Another study by Hofer, Blackwell, and Ruback also found evidence of reduced inter-judge sentencing disparities after the promulgation of the Guidelines.[12] The study concludes that the Guidelines achieved "modest success" in reducing inter-judge 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.[13] Convergence in findings by outside researchers and the Commission led the Commission to conclude that "the federal sentencing guidelines have made significant progress toward reducing disparity caused by judicial discretion."[14]

After almost two decades of mandatory Guidelines sentencing, the Guidelines were struck down in *United States v. Booker*,[15] dramatically altering the sentencing landscape. In *Booker*, the Supreme Court found that the Guidelines violated the Sixth Amendment right to a jury trial and rendered the Guidelines advisory. Subsequent Supreme Court decisions furthered weakened the effect of the Guidelines on criminal sentencing by reducing the degree of appellate scrutiny for both within and outside Guidelines sentences. In *Rita v. United States*, the Court directed courts of appeals to apply a presumption of reasonableness to within Guidelines sentences.[16]

---

[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) (Congress sought to reduce "unjustifiably wide" sentencing disparity.); 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 effectively mandatory, the Guidelines did provide a permissible range of sentence lengths. Thus, one should expect that the Guidelines would ameliorate, but not completely eliminate, inter-judge 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] Paul J. Hofer, Kevin R. Blackwell, & R. Barry Ruback, *The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity*, 90 J. CRIM. L. & CRIMINOLOGY 239 (1999). *See id.* at 284-86 for a discussion of the statistical techniques employed in the study.

[13] *Id.* at 241, 287 ("Together with the other research reviewed below, [our] findings suggest that the sentencing guidelines have had modest but meaningful success at reducing unwarranted disparity among judges in the sentences imposed on similar crimes and offenders.")

[14] United States Sentencing Commission, FIFTEEN YEARS OF GUIDELINES SENTENCING: AN ASSESSMENT OF HOW WELL THE FEDERAL CRIMINAL JUSTICE SYSTEM IS ACHIEVING THE GOALS OF SENTENCING REFORM 99 (Nov. 2004).

[15] 543 U.S. 220 (2005).

[16] 551 U.S. 338, 347-50 (2007).

Later in *Gall v. United States*, the Court held that appellate courts could not presume that a sentence outside the Guidelines range was unreasonable, reducing the degree of appellate review to a more deferential abuse of discretion standard.[17] Concurrent with the *Gall* decision, the Court in *Kimbrough* held that federal district court judges have the discretion to impose sentences outside the recommended Guidelines range due to policy disagreements with the Sentencing Commission.[18]

Following *Booker* and its progeny, *Rita*, *Gall*, *Kimbrough*, the legal community expressed concerns on the impact of such decisions on inter-judge sentencing disparities. Congressman Tom Feeney wrote that "the Supreme's Court 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."[19] U.S. Attorney for the Northern District of Illinois, Patrick Fitzgerald, stated that *Booker* has "re-introduced into federal sentencing both substantial district-to-district variations and substantial judge-to-judge variations."[20] Similarly, scholars commented on the huge increase in the degree of judicial discretion afforded to judges,[21] and predicted an increase in unwarranted sentencing disparities.[22]

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 Attorney General Alberto Gonzales claimed that in light of "increasing disparity in sentences" since *Booker*, the Guidelines needed to be fixed.[23] As a potential "fix," former Chair of the Sentencing Commission, Judge William K. Sessions III, has suggested "resurrecting" the mandatory Guidelines in order to give them greater weight during sentencing.[24]

---

[17] 552 U.S. 38, 52-53 (2007).

[18] *Kimbrough v. United States*, 552 U.S. 85, 101 (2007).

[19] Carl Hulse & Adam Liptak, *New Fight Over Controlling Punishments Is Widely Seen*, N.Y. TIMES, Jan. 13, 2005, at A29.

[20] Patrick J. Fitzgerald, Testimony before the U.S. Sentencing Commission Regional Hearing in Chicago, Illinois, at 3 (Sept. 10, 2009).

[21] *See* 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."); 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"); Luiza Ch. Savage, *Chaos Ahead After Sentencing Guidelines Decision*, N.Y. SUN, Jan. 13, 2005, at 1 (quoting Frank Bowman) (arguing that Booker resulted in "the most amount of judicial discretion ever afforded to sentencing judges").

[22] *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) ("*Kimbrough*, *Gall* ... 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").

[23] Federal Sentencing Guidelines Speech (June 21, 2005), 17 FED. SENTENCING REP. 324, 325-26 (2005).

[24] 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, 346-50 (2011).

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."[25] Indeed, the vast majority of federal district court judges also indicate that they prefer the current advisory Guidelines system to alternative regimes. In a 2010 Sentencing Commission survey of district court judges, 75% of judges indicated that the current advisory Guidelines system "best achieves the purposes of sentencing," while only 3% preferred the mandatory Guidelines regime in place before *Booker*.[26]

This Article asks the question: What has been the result of reintroducing greater judicial discretion after *Booker* on inter-judge disparities? The primary empirical work on the impact of *Booker* on sentencing disparities is suggestive of increases in inter-judge sentencing disparity. Using sentencing data from the District of Massachusetts, Scott observes an increase in judge differences.[27] While a first step, the study is limited to ten judges in the Boston courthouse. Therefore, a comprehensive analysis of disparities post *Booker* is essential to informing recent policy debates.[28]

---

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

[26] U.S. Sentencing Commission, RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES JANUARY 2010 THROUGH MARCH 2010 (June 2010) (Question 19, Table 19).

[27] *See* Ryan W. Scott, *Inter-judge Sentencing Disparity After Booker: A First Look*, 63 STAN. L. REV. 1, 4-5 (2010). Scott finds almost a doubling of the effect of judge on sentence length post *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 16%), while others increased their rate of sentencing below-range to as high as 53%. *Id.* The Transactional Records Access Clearinghouse (TRAC) recently compiled a dataset of the sentencing records of over 800 federal judges from fiscal year 2007 to 2011. *See* Susan B. Long & David Burnham, *TRAC Report: Examining Current Federal Sentencing Practices: A National Study of Differences Among Judges*, 25 FED. SENTENCING REP. 6, 7 (2012); *see also Big Sentencing Disparity Seen for Judges*, N.Y. TIMES, March 5, 2012, at A23. Relying on the random assignment of cases to judges within district courthouses, the TRAC study found statistically significant, unexplained differences in the typical sentences of judges in over half of the courthouses studied. *Id.* at 15. The most recent Commission report (2012) also finds suggestive evidence 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*. United States Sentencing Commission, REPORT ON THE CONTINUING IMPACT OF UNITED STATES V. BOOKER ON FEDERAL SENTENCING (2012). The Commission concludes that "sentencing outcomes increasingly depend upon the judge to whom the case is assigned." *Id.* at 7. However, the Commission does not account for caseload composition differences across judges within the same district, and analyzes all 94 districts, despite evidence by previous researchers that random assignment of cases is not universal. Thus, the Commission's findings are only suggestive.

[28] 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* United States Sentencing Commission, DEMOGRAPHIC DIFFERENCES IN FEDERAL SENTENCING PRACTICES: AN UPDATE OF THE BOOKER REPORT'S MULTIVARIATE REGRESSION ANALYSIS (2010) (providing evidence that demographic differences were significantly less when the Guidelines were binding, particularly during the PROTECT Act when appellate review of departures involved *de novo* review); United States Sentencing Commission, FINAL REPORT ON THE IMPACT OF UNITED STATES V. BOOKER ON FEDERAL SENTENCING, at 105-108 (2006). However, other scholars have found no significant change in racial disparities, at least in sentence length. *See* J.T. Ulmer et al., *Racial Disparity in the Wake of the Booker/Fanfan Decision: An Alternative Analysis to the USSC's 2010 Report*, 10 CRIM. &

This Article fills this gap by undertaking the first national, multi-district empirical analysis of inter-judge sentencing disparities in federal sentencing after *Booker* by utilizing a new and comprehensive dataset constructed for this study. The Article proceeds in five parts. Part I provides a brief background of the legal landscape. Part II describes the framework, dataset and empirical methods. Matching three data sources, I construct a dataset of over 600,000 criminal defendants linked to sentencing judge 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 inter-judge disparities. A defendant who is randomly assigned a one standard deviation "harsher" judge in the district court received a 2.6 month longer prison sentence compared to the average judge before *Booker*, but received a 5.3 month longer sentence following *Kimbrough/Gall*, a doubling of inter-judge disparities. Similarly, a defendant randomly assigned to a one standard deviation more "lenient" judge faced a 4.8% chance of receiving a below range departure before *Booker*, but over 6.6% chance after *Kimbrough/Gall.*

Part III.D undertakes an analysis of the sources of increases in inter-judge disparities. Many scholars have suggested that judges have different sentencing philosophies,[29] which may be affected by the standard of appellate review.[30] Sentencing practices are correlated with judge demographic characteristics such as race,[31] gen-

---

PUB. POL'Y 1077 (2011). Some scholars argue that judicial discretion may actually mitigate recent increases in racial disparities. *See* Joshua Fischman & Max Schanzenbach, *Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, 9 J. EMPIRICAL LEGAL STUD. 729 (2012) (arguing that recent increases in racial disparities are mainly due to the increased relevance of mandatory minimums).

[29] *See* John S. Carroll et al., *Sentencing Goals, Casual Attributions, Ideology, and Personality*, 52 J. PERSONALITY & SOC. PSYCHOL. 107 (1987) (arguing that judicial ideology is reflected in how a judge thinks about the causes of crime and the goals of sentencing); Shari S. Diamond & Hans Zeisel, *Sentencing Councils: A Study of Sentence Disparity and Its Reduction*, 43 U. CHI. L. REV. 109, 114 (1975) ("it is reasonable to infer that the judges' differing sentencing philosophies are a primary cause of the disparity"); *see also*, Paul Hofer, Kevin Blackwell, & R. Barry Ruback, *The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity*, 90 J. CRIM. L. & CRIMINOLOGY 239 (1999) (claiming that there are differences between how liberals and conservatives view the goals of sentencing which can drive different sentencing practices).

[30] Joshua Fischman & Max Schanzenbach, *Do Standards of Review Matter? The Case of Federal Criminal Sentencing*, 40 J. LEG. STUD. 405 (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.

[31] *See* Thomas Uhlman, RACIAL JUSTICE: BLACK JUDGES AND DEFENDANTS IN AN URBAN TRIAL COURT (1979) (claiming that 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 (1988) (finding that black judges do not differ much in their incarceration decisions from white judges based on one city's criminal cases).

der,[32] and political affiliation.[33] In particular, the inevitable shift in the composition of the federal district courts may have profound consequences on unwarranted disparities as judges who have no experience sentencing under a presumptive Guidelines regime take the federal bench.[34] 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."[35]

I find that female judges and Democratic appointed judges issue shorter sentences and are more likely to depart downwards from the Guidelines than their male and Republican appointed peers, respectively. Also striking are the differential sentencing practices of post *Booker* and pre-*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, potentially suggesting that newer judges are less anchored to the Guidelines.

In addition to analyzing the impact of philosophy on sentencing practices of individual judges, this paper also contributes to the literature on geographical variations in sentencing patterns, which has found that court cultures can affect sentencing through both exercise of judicial discretion and also regional differences in policies among prosecutors.[36] In Part III.E, I present evidence of substantial

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

[33] For instance, Max M. Schanzenbach & Emerson H. Tiller, *Strategic Judging Under the United States Sentencing Guidelines: Positive Theory and Evidence*, 23 J.L. ECON. & ORG. 24 (2007) explore the impact of ideology on federal criminal sentencing decisions from 1992 through 2001. They find that sentences for serious crimes in districts comprised of more Democrat appointed judges are lower than sentences in districts with more Republican appointed judges. The alignment of the ideology of the reviewing court also increased departures from the Guidelines. 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 (2008) reveals that Republican appointed judges give longer sentences for the same crime compared to their Democratic appointed counterparts. Moreover, Democratic-appointed judges are more likely to depart downwards from the Guidelines when the reviewing circuit court is majority Democratic appointed.

[34] Until now, prior studies have been unable to identify the impact of post *Booker* appointed judges on inter-judge sentencing disparities. The Scott study, which only looks at the Boston courthouse, is unable to take into account changes in judicial composition because the Boston courthouse did not experience turnover during the years in his study. Recent work suggest that racial disparities in sentencing are greater among judges appointed after *Booker*. *See* Crystal S. Yang, *Free At Last? Judicial Discretion and Racial Disparities in Federal Sentencing*, Harvard John M. Olin Center for Law, Economics, and Business Fellows' Discussion Paper Series No. 47 (2012).

[35] *See* James Felman, *How Should Congress Respond if the Supreme Court Strikes Down the Federal Sentencing Guidelines*?, 17 FED. SENTENCING REP. 97, 98-99 (2004).

[36] *See, e.g.,* Paul L. Sutton, Department of Justice, *Federal Sentencing Patterns: A Study of Geo-*

inter-district differences in sentencing outcomes. Significant differences exist in sentence length, rates of below range departures, rates of mandatory minimums, and rates of substantial assistance motions, with inter-district differences expanding after *Booker*.

In Part III.F, I present some evidence on the contribution of prosecutorial decisions on inter-judge 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 inter-judge 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.[37]

In particular, I analyze the impact of mandatory minimums on inter-judge disparities. If mandatory minimums were charged prior to judge assignment, one would in fact expect the application of mandatory minimums to reduce inter-judge disparities.[38] 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.[39] Indeed, I find evidence that the application of a mandatory minimum is a large contributor of inter-judge and inter-district 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 suggest 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 substantial differences in the rates of

---

graphical Variations (1978); William M. Rhodes & Catherine Conly, Department of Justice Federal Justice Research Program, *Analysis of Federal Sentencing* (1981); Charles D. Weisselberg & Terrence Dunworth, *Inter-District Variation under the Guidelines: The Trees May Be More Significant Than the Forest*, 6 FED. SENTENCING 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).

[37] *See, e.g.,* Franklin E. Zimring, *Making the Punishment Fit the Crime*, 6 HASTINGS CENTER REP. 13, 13-14 (1976) (arguing that there is "multiple discretion" in the criminal justice system from the legislature, prosecutor, judge and parole board); Ilene H. Nagel & Steven J. Schulhofer, *A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices Under the Sentencing Guidelines*, 66 S. CAL. L. REV. 501, 502 (1992) (noting that "both Congress and the U.S. Sentencing Commission were well aware that plea bargaining posed a potential threat to the success of guidelines sentencing"). As a result, the Guidelines system has been attacked by many for its rigidity and for shifting power to prosecutors in their charge and plea bargaining decisions. *See, e.g.*, Daniel J. Freed, *Federal Sentencing in the Wake of the Guidelines: Unacceptable Limits on the Discretion of Sentencers*, 101 YALE L.J. 1681, 1719-20, 1725-27 (1992); Kate Stith, *The Arc of the Pendulum: Judges, Prosecutors, and the Exercise of Discretion*, 117 YALE L.J. 1420, 1430 (2008); Albert W. Alschuler, *Sentencing Reform and Prosecutorial Power: A Critique of Recent Proposals for "Fixed" and "Presumptive" Sentencing*, 126 U. PA. L. REV. 550 (1978); Kate Stith & Jose A. Cabranes, FEAR OF JUDGING: SENTENCING GUIDELINES IN THE FEDERAL COURTS (1998).

[38] *See infra* Part III.F.

[39] For instance, prosecutors often bring superseding indictments under various statutes, such as 20 U.S.C. §851 and 18 U.S.C. §924(c). A literature documents the large degree of prosecutorial discretion in bringing superseding indictments and potential 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.").

substantial assistance motions filed by prosecutors across judges and district courts. Such results indicate that prosecutorial charging is a measurable contributor to disparities in sentencing.

In Part IV, I describe recent proposals to reform federal sentencing and apply the empirical findings in this paper to shed light on the soundness of the proposals. Given the finding that a substantial portion of inter-judge disparities and regional disparity is attributable to the application of mandatory minimums, any proposal that contemplates shifting power to prosecutors will likely exacerbate the presence of disparities. Indeed, many judges and scholars have suggested that mandatory minimums are "fundamentally inconsistent with the sentencing guidelines system."[40] Instead, I argue that strengthened appellate review and elimination of mandatory minimums are potential steps in the direction of reducing unwarranted disparities in sentencing. Part V concludes.

I BRIEF LEGAL BACKGROUND OF FEDERAL SENTENCING

A. Adoption of the Federal Sentencing Guidelines

In the early twentieth century, criminal justice was premised on the notion of rehabilitation.[41] 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.[42] In this regime of "free at last" sentencing,[43] federal judges had essentially unlimited authority in imposing sentences, restrained only by statutorily prescribed minimum and maxi-

[40] *See* Sessions, *supra* note 24, at 329 (citing Senator Edward M. Kennedy, *Sentencing Reform-An Evolutionary Process*, 3 FED. SENTENCING REP. 271, 272 (1991) ("Mandatory minimum sentencing statutes have . . . 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. SENTENCING 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.").

[41] *See* David Rothman, *Sentencing Reform in Historical Perspective*, 29 CRIME & DELINQUENCY 631, 637-41 (1983) (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).

[42] *See* Rothman, *supra* note 41, at 638 ("the 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)").

[43] A term coined by Judge Nancy Gertner to describe the state of indeterminate sentencing prior to 1984. *See United States v. Mueffelman*, 400 F. Supp. 2d 368, 372 (D. Mass 2005); *United States v. Jaber*, 362 F. Supp. 2d 365, 370 (D. Mass. 2005).

mum sentences.[44] Lack of appellate review of sentences meant that judges faced no meaningful check to ensure uniformity and consistency in sentencing.[45]

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.[46] The legal community and 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.[47] Others were distressed by inequitable and arbitrary treatment within the criminal justice system as studies showed that similar offenders were often punished very differently. A 1977 study of 47 Virginia state district court judges revealed that while judges generally agreed on the verdict in legal cases, they applied radically different sentences.[48] A Federal Judicial Center Second Circuit Study found large inter-judge differences in the sentences imposed based on identical presentence reports of defendants.[49] The same defendant was sentenced to three years by one judge, and twenty years by another.[50] Some concluded that this disparate treatment of defendants by judges produced racial inequities in sentencing. The American Friends Service Committee claimed that decreasing discretion among judges and parole boards was the only way to eliminate racial discrimination and sentencing disparities in the criminal justice system.[51]

Other studies identified large inter-court differences. A 1988 study of federal courts found that white collar offenders who committed similar offenses received very different sentences depending on the court in which they were sentenced,[52] with one study observing that " a defendant sentenced by a Southern judge was

---

[44] *See* United States Sentencing Commission, 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, Vol. I at 9 (December 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 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.")

[45] *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).

[46] *See, e.g.*, Francis A. Allen, THE DECLINE OF THE REHABILITATIVE IDEAL: PENAL POLICY AND SOCIAL PURPOSE, 25-30 (1981); *see also* Bowman, *Debacle*, *supra* note 22, at 374 (attributing demand for social controls to rising crime rates and social upheaval).

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

[48] 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 (1977).

[49] *See* Partridge & Eldridge, *supra* note 2, at 36.

[50] *Id.*

[51] *See* Am. Friends Serv. Comm., Struggle for Justice: A Report on Crime and Punishment in America 130 (1971) (claiming that discretion allowed judges and parole boards to control minority groups); *see also* Bowman, *Quality of Mercy*, *supra* note 41, at 686-88 (critics arguing discretion produced unjustifiable disparities open to conscious or unconscious racial and other biases, demanding "truth in sentencing").

[52] Wheeler et al., SITTING IN JUDGMENT: THE SENTENCING OF WHITE-COLLAR CRIMINALS (1988).

likely to serve six months more than average, while a defendant sentenced in Central California was likely to serve twelve months less."[53]

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.[54] Judge Frankel called for the creation of an independent sentencing commission that would replace judicial and parole board discretion.[55]

In response, Congress created the United States Sentencing Commission to adopt and administer the Sentencing Guidelines, aimed at eliminating unwarranted sentencing disparities "among defendants with similar records who have been found guilty of similar criminal conduct."[56] Part of the SRA of 1984, the Guidelines were applied to all federal offenses committed after November 1, 1987. The SRA has been viewed by some as creating a regime that preserves judicial discretion,[57] while others have viewed the SRA has substantially increasing the severity of punishment and dramatically reducing the discretion afforded to judges to consider the particular circumstances of each offender.[58]

Notwithstanding disagreement about the degree to which sentencing reform changed the legal landscape, the new SRA introduced a shift from a regime of virtually unfettered judicial discretion to more restricted discretion within a system of determinate sentencing.[59] By requiring judges to sentence within the recommended Guidelines range unless the court found aggravating or mitigating circumstances,[60] the Guidelines were treated as presumptively mandatory, although the particular standards for departure were ambiguous.[61] 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.[62]

---

[53] *See* Justice Stephen Breyer, *Federal Sentencing Guidelines Revisited*, 11 FED. SENTENCING REP. 180 (1988).

[54] 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 (1972).

[55] Frankel, *supra* note 3.

[56] 28 U.S.C. §991(b)(1)(B).

[57] Marc L. Miller & Ronald F. Wright, *Your Cheatin' Heart(land): The Long Search for Administrative Sentencing Justice*, 2 BUFF. CRIM. L. REV. 723 (1999).

[58] *See* Stith & Koh, *supra* note 5, at 284-85 ("It should come as a surprise to no one that in 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. It is no accident that judges have found it difficult to depart from the guidelines; this is precisely what Congress intended.").

[59] In addition to creating the Guidelines, the SRA also abolished federal parole and instituted supervised release in its place. 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.

[60] H.R. REP. NO. 1017, 98th Congress, 2nd Session, at 93-94 (1984).

[61] *See* Miller & Wright, *supra* note 57, at 730 (The Commission allowed judges to depart from the Guidelines when the case fell within the "heartland," but the concept was left highly general.)

[62] 518 U.S. 81, 99 (1996). The Court in *Koon* stated that Congress "did not intend, by establish-

Under the Guidelines, federal district court judges assign each federal crime to one of 43 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.[63] For example, trespass offenses are assigned a base offense level of four,[64] while kidnapping is assigned a base offense level of 32.[65] From a base offense level, the final offense level is calculated by adjusting for applicable offense characteristics and adjustments. Relevant adjustments include the amount of loss involved in the offense, use of a firearm, the age or condition of the victim, etc.[66] Chapter Three of the Guidelines allows for further adjustments based on aggravating or mitigating factors, such as a defendant's acceptance of responsibility.[67]

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, 50 state systems, all territories and foreign or military courts.[68] For example, three points are added for each prior sentence of imprisonment exceeding one year and one month, and two points are added for each prior sentence of imprisonment of at least 60 days and less than one year and one month.[69] Two points are also added if the defendant committed the instant offense under any criminal justice sentence.[70]

The intersection of the final offense level and criminal history category yields a fairly narrow Guidelines recommended sentencing range, where the top of the range exceeds the bottom by the greater of either six months or 25%. If a judge determines that there are aggravating or mitigating circumstances that warrant a departure from the Guidelines, she would have to justify her reasons for departure to the appellate court,[71] but in general the Guidelines were treated as sufficiently

---

ing limited appellate review, to vest in appellate courts wide-ranging authority over district court sentencing decisions." *Id.* at 97.

[63]United States Sentencing Commission, Guidelines Manual, Chapter Two.

[64]*Id.* §2B2.3.

[65]*Id.* §2A4.1.

[66]*Id.*

[67]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.

[68]United States Sentencing Commission, Guidelines Manual, §4.1 ("A defendant with a record of prior criminal behavior is more culpable than a first offender and thus deserving of greater punishment. Greater 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. To protect the public from further crimes of the particular defendant, the likelihood of recidivism and the future criminal behavior must be considered. Repeated criminal behavior is an indicator of a limited likelihood of successful rehabilitation.").

[69]*Id.* §4.1.

[70]*Id.* §4.1.

[71]18 U.S.C. §3553(b) ("the 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 in formulating the guidelines that should result in a sentence different from that described").

mandatory prior to *Booker*.[72] 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.[73]

There are numerous other ways in which Congress has attempted to limit unwarranted disparities in sentencing. Beginning in 1984, and subsequently 1986 and 1988, Congress enacted a series of mandatory minimum statutes directed at drug and firearms offenses.[74] Mandatory minimums were also applied to recidivist offenders, through the Armed Career Criminal Act,[75] enhancements for career offenders,[76] and enhancements for repeat and dangerous sex offenders.[77]

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.[78] Under the Feeney Amendment to the PROTECT Act, judicial departures were only allowed for certain reasons outlined in the Guidelines Manual.[79] Additionally, the Feeney Amendment to the PROTECT Act replaced the prior abuse of discretion standard of review for downward departures with *de novo* review by overturning the Supreme Court's holding in *Koon*.

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*.[80] 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.[81]

---

[72]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." *Booker*, 543 U.S. at 233.

[73]18 U.S.C. §3742 (a)-(b).

[74]Anti-Drug Abuse Act of 1986 (Pub. L. 99-570, 100 Stat. 3207); Pub. L. No. 98-473, §1005(a), 98 Stat. 1837, 2138-39 (1986).

[75]Guidelines Manual §4B1.2; 18 U.S.C. §924(e). The Armed Career Criminal Act (ACCA) imposes a minimum 15-year term of imprisonment for defendants convicted of unlawful possession of a firearm under 18 U.S.C. §922(g), with three prior state or federal convictions for violent felonies or serious drug offenses.

[76]§4B1.1; 28 U.S.C. §944(h) (mandating that the Commission impose imprisonment "at or near the maximum term authorized" for defined "career" offenders).

[77]§4B1.5; 18 U.S.C. §§2247, 2426.

[78]Pub. L. 108-21, 117 Stat. 650, S. 151, 2003.

[79]For certain offenses, such as child abduction and child sex offenses, the PROTECT Act amended 18 U.S.C. §3553(b) to only allow the sentencing court to depart downwards if there are mitigating circumstances of a kind or to a agree that has been "affirmatively and specifically identified" as permissible grounds for downward departure. The PROTECT Act also amended the Guidelines Manual §5K2.0 to state that the "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 departure."

[80]530 U.S. 466, 425 (2000) (O'Connor, J., dissenting) (calling the *Apprendi* decision a"watershed change in constitutional law").

[81]*Id.* at 468-69, 490.

These principles were subsequently applied to the constitutionality of mandatory sentencing guidelines, first questioned in reference to 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.[82] As a result, Washington's mandatory sentencing guidelines were struck down. While the majority opinion in *Blakely* emphasized that the decision was not passing judgment on the constitutionality of the Federal Sentencing Guidelines,[83] the parallels were apparent and shortly after, the reasoning of *Blakely* was applied to the Guidelines.

In *United States v. Booker*, the mandatory Federal Sentencing Guidelines were also found to be unconstitutional under the Sixth Amendment by mandating judicial fact-finding in determining sentencing ranges.[84] The *Booker* ruling, however, did not apply to Congressionally-enacted mandatory minimum sentences.[85] Rather than invalidate the Guidelines wholly, or prescribe an enhanced role for jury fact-finding, the Court held in a separate remedial decision led by Justice Breyer, that the remedy for the Sixth Amendment violation was to declare the Guidelines no longer mandatory but "effectively advisory."[86] The Court explained that "district courts, while not bound to apply the Guidelines, must consult those Guidelines and take them into account when sentencing."[87]

In the immediate aftermath of *Booker*, district courts took varied approaches in applying *Booker*.[88] Some courts sentenced with minimal consideration of the applicable Guidelines range, while others treated the Guidelines as a dominant factor.[89] Circuit courts later reached a consensus that sentencing must begin with the calculation of the applicable Guidelines range.[90] Today, after a sentencing judge has

---

[82] 542 U.S. 296 (2004).

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

[84] 543 U.S. 220, 226-27, 243-44 (2005) (Stevens, J., writing for the Court).

[85] *See* United States Sentencing Commission, REPORT TO CONGRESS: MANDATORY MINIMUM PENALTIES IN THE FEDERAL CRIMINAL JUSTICE SYSTEM, at 37 (Oct. 2011) (discussion of the compatibility of mandatory minimums and the Guidelines after *Booker*).

[86] *Booker*, 543 U.S. at 245 (Breyer, J., writing for the Court). Similarly, the provisions on supervised release also became advisory, although the USSC states that the majority of courts continue to impose at least the minimum terms in §5D1.2.

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

[88] *See* Gilles R. Bissonnette, *"Consulting" the Federal Sentencing Guidelines After Booker*, 53 UCLA L. REV. 1497, 1521-22 (2006) (arguing that Booker left open how much sentencing judges could deviate from the Guidelines).

[89] *See id.* at 1522-32 (claiming that district courts have taken two approaches in applying Booker, the "substantial-weight" approach and the "consultative" approach). For an example of the two type of approaches taken by district courts, *see, e.g. 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," but that courts should consider all the factors in §3553(a)); *cf. United States v. Wilson*, 350 F. Supp. 2d 910, 925 (D. Utah 2005) (suggesting courts give "heavy weight" to the Guidelines after *Booker*).

[90] *See, e.g., United States v. Howard*, 454 F.3d 700, 703 (7th Cir. 2006) (court noting that 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, 113-114 (2d Cir. 2005) ("consideration of the Guidelines will normally

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.[91] After consideration of all the factors, the sentencing judge is instructed to "impose a sentence sufficient, but not greater than necessary, to comply with" the basic goals of sentencing.[92]

Subsequent Supreme Court decisions furthered 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, it did not clarify the meaning of this standard until a sequence of decisions in 2007. In *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 consistent with the Commission's judgment in general."[93] 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.[94] 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."[95] 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, the Supreme Court confirmed the holding in *Booker* as applied to cases involving possession, distribution and manufacture of crack cocaine.[96] The Court in *Kimbrough* held that federal district court judges have the discretion to impose sentences outside the recommended Guidelines range due to policy disagreements with the Sentencing Commission.[97]

---

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 to select any sentence within the applicable statutory maximum and minimum").

[91] 18 U.S.C. §3553(a).

[92] *Id.*

[93] 551 U.S. 338, 347-50 (2007) (holding that "a court of appeals may apply a presumption of reasonableness to a district court sentence that reflects a proper application of the Sentencing Guidelines").

[94] 552 U.S. 38 (2007).

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

[96] *Kimbrough v. United States*, 552 U.S. 85 (2007).

[97] *Id.* (granting sentencing judges explicit permission to deviate from the Guidelines based on disagreement with the disparate treatment of crack and powder cocaine offenses - the so called "100-to-1 ratio").

II FRAMEWORK, DATA, AND METHODS

A. Judicial Behavior in Criminal Sentencing

While judges have an obligation to "follow the law,"[98] they have additional motivations that affect their decision-making.[99] Scholars have suggested that judges care about a variety of factors such as public recognition, leisure, and reputation.[100] In addition, judges have preferences for sentencing according to their personal tastes.[101] 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.[102]

How might the sentencing regime affect judicial behavior? Given individual judge-specific preferences for sentencing, one would likely observe large inter-judge 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 were likely due to differing judicial attitudes regarding rehabilitation and deterrence.[103] At the other end of the spectrum, judges who are restricted in the exercise of discretion would be unable to fully sentence according to their preferences. The type of sentencing

[98] *See, e.g.*, Lewis A. Kornhauser, *Judicial Organization and Administration*, 7 Encyclopedia of Law and Economics 27 (2000)

[99] The economic analysis of judicial behavior builds on work by Judge Richard A. Posner. *See, e.g.* Hon. Richard A. Posner, *Judicial Behavior and Performance: An Economic Approach*, 32 FLA. ST. U. L. REV. 1259 (2005); *see also* Hon. Richard A. Posner, HOW JUDGES THINK (2008); Lee Epstein, William M. Landes & Richard A. Posner, THE BEHAVIOR OF FEDERAL JUDGES: A THEORETICAL AND EMPIRICAL STUDY OF RATIONAL CHOICE (2013). *See* Nicola Gennaioli & Andrei Shleifer, *Judicial Fact Discretion*, 37 J. LEG. STUD. 1 (2008) for a theoretical economic model of judicial discretion in fact determination.

[100] Federal district judges occupy a unique position because most district judge appointments are terminal, 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, *Judicial Behavior and Performance: An Economic Approach*, *supra* note 99, at 1260, 1269 (noting that a judge likely cares more about leisure and public recognition relative to income, compared to average practicing attorneys).

[101] *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").

[102] Indeed, federal district court judges have expressed a great degree of dissatisfaction with the Guidelines. In a 2010 survey of federal district judges, 65% 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 Commission, RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES JANUARY 2010 THROUGH MARCH 2010 (June 2010), (Question 14, Table 14).

[103] *See* Posner, *Judicial Behavior and Performance: An Economic Approach*, *supra* note 99, at 1270 (inferring from the "extraordinary variance" in federal sentences prior to the promulgation of the Guidelines that differences in sentence 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, 801-804 (1981) (documenting disagreement between judges regarding five goals of sentencing: general deterrence, special deterrence, incapacitation, rehabilitation, and just deserts).

regime can place a constraint on judicial discretion. For instance, the adoption of determinate sentencing under the Guidelines introduced a mechanism by which to constrain judges, likely explaining studies finding reduced inter-judge sentencing disparity after the promulgation of the Guidelines.[104]

In addition to a restriction on free sentencing imposed by the sentencing regime, another constraint comes from the prospect of appellate review. Judges who depart from the Guidelines incur economic and social costs from deviating. A high reversal rate is not only administratively burdensome, but also potentially harms a trial judge's prospects for promotion to the appeals court.[105] Indeed, under the mandatory sentencing regime, departures from the Guidelines were relatively rare.[106] After the Feeney Amendment of the PROTECT Act, which subjected district court judges to *de novo* review for departures from the Guidelines, departures were reduced even further, suggesting that judges respond to changes in appellate scrutiny.[107]

Given the countervailing forces of (1) judge sentencing preferences versus (2) costs of exercising discretion, what is the theoretical prediction of the impact of *Booker*, *Rita*, *Gall*, and *Kimbrough* on inter-judge disparities? Following *Booker*, the total cost of exercising discretion fell substantially for judges who wanted to depart from the Guidelines sentence as the Guidelines were rendered advisory. This major shift in sentencing may have been accompanied by increases in inter-judge disparity.

However, to the extent the the relative cost associated with *de novo* appellate review was still binding, judges may have been hesitant to alter their practices. Indeed, not until *Rita*, *Gall*, and *Kimbrough* did the Court reduce the level of appellate review from *de novo* to substantial abuse of discretion, intuitively lowering the probability of appellate reversal.[108] Thus, one might expect larger increases in inter-judge 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.

There are other reasons why judicial behavior and inter-judge disparities may have not changed much after *Booker* and its progeny. First, judges may become acculturated to the Guidelines if they have had substantial experience sentencing

---

[104] *See supra* notes 9-14.

[105] *See* Posner, *Judicial Behavior and Performance: An Economic Approach*, *supra* note 99, at 1270-71; *see also* Stephen J. Choi, Mitu Gulati, & Eric A. Posner, *What do Federal District Judges Want?: An Analysis of Publications, Citations, and Reversals*, University of Chicago John M. Olin Law & Economics Working Paper Series Paper No. 508, at 3-4 (2009) (judges care about minimizing workload and maximizing reputation by avoiding appellate reversal); 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 due to professional reputation, advancement, judicial power); Richard S. Higgins & Paul H. Rubin, *Judicial Discretion*, 9 J. LEGAL STUD. 129, 130 (1980).

[106] The rate of departure from the Guidelines was less than 15% in the early 1990s.

[107] Recall that the Commission found that demographic differences under the mandatory Guidelines regime were lower during the PROTECT Act. *See supra* note 28.

[108] The probability of reversal on sentencing matters fell from 36% in 2006 (under *de novo* review), to 26% in 2008 (under abuse of discretion review). I calculate rate of appellate reversals using yearly data on the universe of criminal appeals from the United States Sentencing Commission. Reversal is defined as all reversals and remands on appeals arising out of *Booker* sentencing issues.

under the previous Guidelines regime.[109] 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*.

Another potential mechanism is anchoring, a type of cognitive bias in which decision-makers rely heavily on one piece of information and fail to make rational adjustments.[110] Judge Nancy Gertner of the District of Massachusetts predicted that the Guidelines would still play a predominant role for all judges post *Booker* because "appellate courts have insisted that district court judges begin with - effectively, 'anchor' their decisions - in the Guidelines before considering anything else."[111] 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.[112]

B. Sentencing Data

This Article provides the first comprehensive empirical evidence on the impact of *Booker* and its progeny on inter-judge sentencing disparities. As noted previously, the Scott study is the only empirical study thus far, but is limited to the 2,262 cases sentenced by judges who served continuously from 2001 to 2008 in the Boston courthouse of the District of Massachusetts.[113] While the study is a first step in characterizing the extent to which inter-judge sentencing practices have changed in the aftermath of *Booker*, the Boston courthouse is likely unrepresentative of sen-

---

[109] *See* Judge 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 and are likely to shape the way judges view 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*, *The Arc of the Pendulum*, *supra* note 37, at 1496-97 ("incumbent sentencing decision makers may be reluctant to regard as unreasonable the sentences they were obliged to seek and impose for two decades"); Frank O. Bowman, *Beyond Band-Aids: A Proposal for Reconfiguring Federal Sentencing After Booker*, 2005 CHI. LEGAL F. 149, 187 (2005) (arguing that advisory guidelines might still constrain judicial discretion "if for no other reason than that the federal bench has become acculturated to the Guidelines over the last seventeen years").

[110] *See, e.g.*, Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1130-31 (1974); *see also* 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, 188 (2006) (experimental results showing that criminal sentences were higher if participants were confronted with a randomly high rather than a low anchor).

[111] *See* Judge Nancy Gertner, *What Yogi Berra Teaches About Post-Booker Sentencing*, 115 YALE L.J. POCKET PART (2006) 137, 138-40 (describing the Guideline Manual as a "ready-made anchor").

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

[113] Scott, *supra* note 27, at 17.

tencing patterns in other courthouses across the United States.[114] As a result, the presence of growing inter-judge sentencing disparities after *Booker* in the Boston courthouse may be the result of the particular caseload and judicial composition of that court, making conclusions that *Booker* has increased inter-judge sentencing disparities likely not generalizable across other courts. In fact, a comparison of the Boston courthouse to other district courthouses reveals that the Boston courthouse experienced a greater increase in inter-judge disparities following *Booker* than the average court in the nation.

Prior empirical research on inter-judge disparity and the impact of judicial demographics on sentencing practices has been hampered by the lack of judge identifiers in available data.[115] Because cases are generally randomly assigned to judges within a district courthouse, judge identifiers allow one to compare judges within the same court and in the same time period, capturing judge differences in sentencing rather than different caseloads.[116] However, the Sentencing Commission data does not identify the sentencing judge.[117] In response, most researchers have resorted to using aggregate district-level variation in judicial demographics to control for judge sentencing preferences,[118] but this methodology is flawed if districts with different judicial compositions differ in ways in that affect all judges within the district.[119] A few researchers have resorted to hand matching sentencing data from the Sentencing Commission with Public Access to Court Electronic Records (PACER), but due to the intensive matching process, sample sizes have been severely limited.[120]

This paper is the first in over 25 years to match sentencing data from fiscal years 2000-2009 to judge identifiers in all 94 district courts, allowing for a comprehensive look at inter-judge sentencing disparities after *Booker.* In order to overcome the lack of judge identifiers in sentencing data, I utilize data from three sources: The United States Sentencing Commission, the Transactional Records Access Clearinghouse, and the Federal Judicial Center. I describe each dataset in turn.

United States Sentencing Commission - I use data from the United States Sentencing Commission (USSC) on records of all federal offenders sentenced pursuant

---

[114] *See supra* note 36 for a discussion of the large geographical differences in sentencing and Guidelines treatment.

[115] The Anderson et al. study is the only empirical work with comprehensive sentencing data with judge identifiers in the past 25 years. *See* Anderson et. al, *supra* note 10, at 287.

[116] According to the Administrative Office of the United States Courts, "The majority of courts use some variation of a random drawing."

[117] United States Sentencing Commission, GUIDE TO PUBLICATIONS & RESOURCES 2007-2008 45 (2007), available at http://www.ussc.gov/publicat/Cat2005.pdf ("Pursuant to the policy on public access to Sentencing Commission documents and data, all case and defendant identifiers have been removed from the data." (internal citation omitted)).

[118] Fischman & Schanzenbach, *Strategic Judging*, *supra* note 33 (relying only on district-level variation in observable characteristics of judges).

[119] For instance, a district with a greater percentage of Democratic judges could be different from other districts. It may be that both Democratic and Republican judges within the district sentence differently from judges in other districts, so any effect cannot be solely attributable to Democratic judges.

[120] *See, e.g.,* Schanzenbach & Tiller, *Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform*, 75 U. CHI. L. REV. 715 (2008); Scott, *supra* note 27, at 15-16 (describing the PACER method used to match records to sentencing data).

to the Sentencing Guidelines and Policy Statements of the SRA of 1984 in fiscal years 1994-2009 (October 1, 1993 - September 30, 2009). The USSC provides detailed sentencing data on federal defendants, but defendant and judge identifiers are redacted.[121] The dataset 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.[122] 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.[123] Data is also available on whether the offense carries a 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.[124]

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. From these variables, I construct indicator variables for above range and below range departures from the Guidelines.[125]

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. Defendant, offense characteristics and Guidelines application information is not included, but defendants are linked to sentencing judge.[126]

Federal Judicial Center - Finally, I obtain demographic information on federal district court judges. Federal district judges are Article III judges who serve

[121] Over 90% of felony defendants in the federal criminal justice system are sentenced pursuant to the SRA of 1984 and all cases are assessed to be constitutional.

[122] USSC data prior to 2004 includes information on the exact sentencing day, but this variable is not available in later years. 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.

[123] There are a total of 35 offense categories in the dataset. The most common offense is drug trafficking, followed by immigration, fraud, firearms, and larceny.

[124] Data is also collected on the total number of criminal history points applied and the final criminal history category.

[125] An above range departure is defined as 1 for a defendant who received a sentence above the maximum Guidelines recommended range. Similarly, a below range departure is defined as 1 for a defendant who received a sentence below the minimum Guidelines recommended range.

[126] TRAC has compiled records on the criminal cases and the civil matters handled by federal district court judges in each of the 94 federal judicial districts through over 20 years of FOIA requests.

life term tenures. New appointments are generally made when a judge retires or dies.[127] As of the current day, there are a total of 677 authorized Article III district judgeships.[128] The number of federal district judgeships in each district court varies. The largest district court is the Southern District of New York with 28 authorized judgeships. The majority of other district courts have between 2-7 district court judgeships. 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*.

C. Matching

In order to connect defendants to judges, I match observations across these three datasets. 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 each record: sentencing year, sentencing month, offense type,[129] 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.[130] Then, I match the USSC and TRAC combined data to judge biographical data from the Federal Judicial Center. I successfully match over 90% of all USSC cases from fiscal years 2000-2009, resulting in a matched dataset of 636,063 cases representing 91 district courts (hereinafter "full sample").[131]

D. Testing for Random Assignment

In an ideal experiment to test the impact of *Booker*, one would randomly allocate the treatment - sentencing under an 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. 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 an advisory Guidelines), and the judges in the control group (those who sentence under a mandatory Guidelines), would capture the causal effect of greater judicial discretion via *Booker* on inter-judge 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-

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

[128] 67 positions are currently vacant. Authorized judgeships only refer to full-time non-senior status judges.

[129] Data from USSC are coded to correspond with the offense categories in the TRAC data.

[130] These match variables are comparable to those used by previous scholars under the PACER method. *See* Schanzenbach & Tiller, *supra* note 120, at 729 (matching USSC data with PACER records on date and length of the sentence, and when necessary, the amount of any fine, the offense type, and the Hispanic ethnicity of the defendant).

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

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 inter-judge disparities before *Booker* to inter-judge disparities after *Booker*. If there are no other contemporaneous changes that affect inter-judge sentencing, an increase in inter-judge disparities is attributable to changes in judicial behavior, rather than underlying differences in case composition.[132] 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.[133] 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. While each district court is authorized to specify its own methods of case assignment,[134] "[m]ost district and 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."[135]

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,[136] and some courts assign certain types of cases to particular judges.[137] Moreover, even if a court has local rules and orders that specify the use

[132]Previous research indicates that *Booker*, and in particular *Rita*, *Gall*, and *Kimbrough* did have a causal impact on judicial behavior, after controlling for unobservable year and monthly changes that might affect sentencing. *See* Yang, *supra* note 34, at 17-18.

[133]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.

[134]Under 28 U.S.C. §137, "[t]he business of a court having more than one judge shall be divided among the judges as provided by the rules and orders of the court. The chief judge of the district court shall be responsible for the observance of such rules and orders, and shall divide the business and assign the cases so far as such rules and orders do not otherwise prescribe." For example, in the Arizona district court, "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." AZ. L. R. Crim. 5.1(a). In the Northern District of California, "[c]ases shall be assigned blindly and at random by the clerk by means of a manual, automated or combination system approved by the judges of the court." CA General Order No. 44. In Colorado,"[t]he clerk shall maintain a computerized program to assure random and public assignment of new cases on an equal basis among the judicial officers." D.C. Colo. L Civ R 40.1.

[135]For a description of judge assignment methods, see the Federal Judicial Center FAQS, available at http://www.fjc.gov/federal/courts.nsf/.

[136]28 U.S.C. §294 governs the assignment of cases to senior status judges. *See, e.g.,* MA General Order 10-04 §4.2 ("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.")

[137]For instance, the Southern District of New York assigns civil and criminal cases such that all judges, "except the chief judge, shall be assigned substantially an equal share of the categories of cases of the court over a period of time. There shall be assigned or transferred to the chief judge such matters as the chief judge is willing and able to undertake, consistent with the chief judge's administrative duties." Thus, assignment is based on equal caseload, rather than pure random assignment, and the chief judge is exempted from the rules. *See* Rules for the Division of Business

of random assignment, empirically testing for random assignment is important because random assignment can be violated if individuals seek to game the system.[138] For instance, some courts have decried situations in which high profile cases are given to judges viewed as favorably disposed to one side.[139]

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 beginning of the dataset in 2000 to remove the possibility of non-random 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 less than 25 cases in order to obtain a sufficient number of cases per judge for statistical power.[140] Finally, I drop district courthouses with only one active judge.

With this restricted sample, I test for randomly assignment using the matched USSC, TRAC, and Federal Judicial Center data from 2000-2009.[141] If cases are randomly assigned to judges, then judges should see on average, cases with the same distribution of predetermined defendant characteristics. I test random assignment based on five fixed defendant characteristics: gender, age, a black race indicator, number of dependents, and an indicator for less than a high school degree.[142] Intuitively, there should be no significant correlation between a particular judge and defendant characteristics if cases are randomly assigned. I drop all courthouses that violate random assignment, resulting in a subsample of 163 courthouses from 73 district courts representing about 50% of the cases from 2000-2009, for a total of

---

Among District Judges, available at http://www.nysd.uscourts.gov/courtrules.php.

[138] *See* J. Robert Brown Jr. & Allison Herren Lee, *Neutral Assignment of Judges at the Court of Appeals*, 78 TEXAS L. REV. 1037 (2000) (describing nonrandom assignment in federal courts of appeals).

[139] *See United States v. Pearson*, 203 F.3d 1243, 1264 (10th Cir. 2000) ("If a judge receives case assignments not through some neutral system, but rather because of prosecutors' opinion that he or she is more favorably disposed to the government's arguments than another judge in the same district, then a judge's caseload might be based in part on prosecutors' evaluations of judicial performance."). For recent allegations of "gaming" the random assignment system, *see* William Safire, *Essay; Norma The Plumber*, N.Y. TIMES, July 31, 2000, available at http://www.nytimes.com/2000/07/31/opinion/essay-norma-the-plumber.html (allegations that the Chief Federal Judge of the U.S. District Court (U.S.D.C.) went "off the wheel" to assign a politically sensitive case in a non-random fashion); *see also* Dan Fitzpatrick, *Bank of America Manages to Avoid Judge Rakoff*, WALL ST. J., May 17, 2010, available at http://online.wsj.com/article/SB10001424052748703699804575247132437874588.html (non-random assignment of Bank of America case in the Southern District of New York).

[140] Results are robust to choice of caseload minimums, but follow the convention in prior literature. The Scott study excludes judges who imposed less than 25 sentences in a two year period. *See* Scott, *supra* note 27, at 17. Similarly, Anderson et. al exclude judges who sentence less than 30 cases during a two year period. Anderson et al., *supra* note 10, at 287.

[141] *See* Appendix A for details.

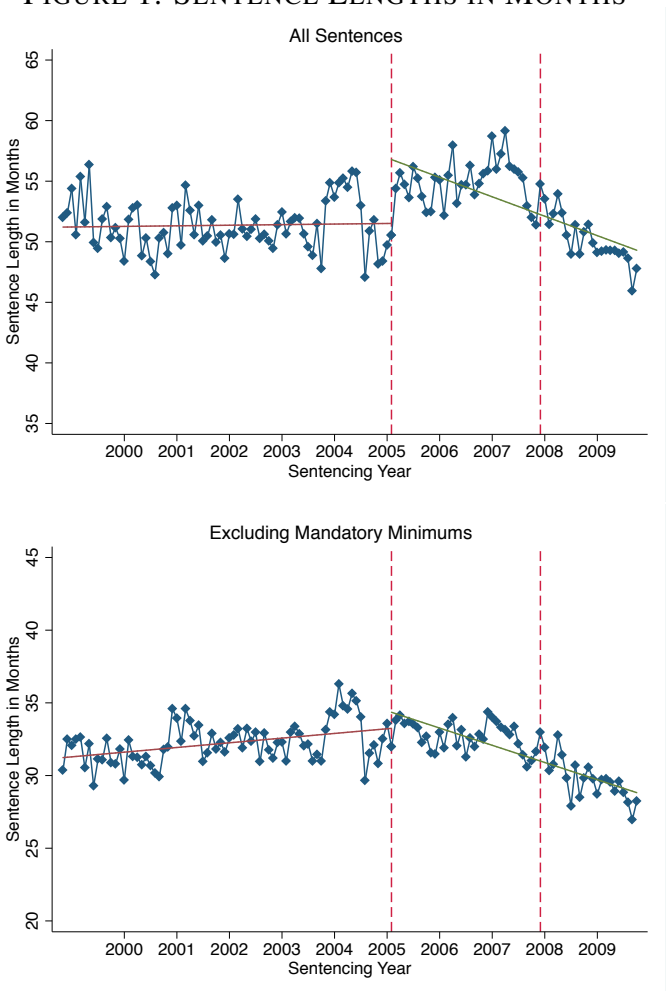
[142] For each of the five defendant characteristics, I regress the characteristic on district courthouse by sentencing year fixed effects, sentencing month fixed effects and judge fixed effects. I test the hypothesis of no judge effects (the null hypothesis) using an F-test for whether the judge fixed 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 AEJ: Applied Economics 96, 106-107 (2010). *See* Appendix A, Table A1.

270,334 cases (hereinafter "random sample").

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 presents a graph of average sentence lengths in months over time, with the timing of *Booker* delineated by the first vertical line and *Kimbrough/Gall* by the second vertical dash, along with predicted trend lines before and after *Booker*. Figure 1 indicates 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.[143]

FIGURE 1. SENTENCE LENGTHS IN MONTHS

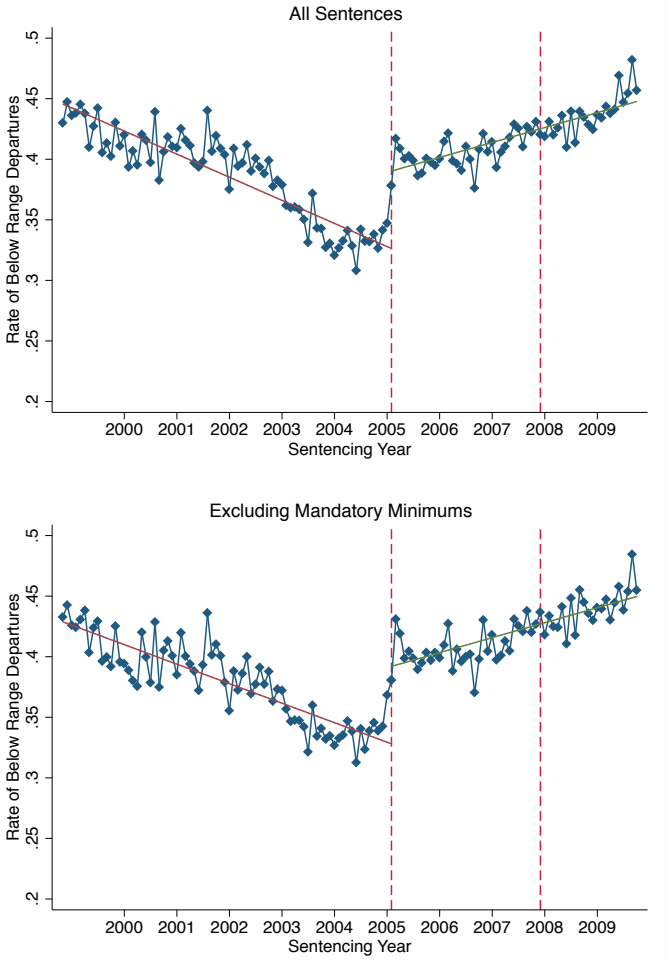


Notes: Data is from the full sample 2000-2009 on all incarcerated defendants. Data points are monthly averages.

[143]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.

Figure 2 presents a graph of the 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 recommended minimum sentence. Figure 2 reveals a trend of decreasing rates of below range departures prior to *Booker*, characterized by very low relative rates of departures in the PROTECT Act era. The decreasing trend in below range departures was significantly changed following *Booker*, which induced a sudden jump in the rate of departure, as well as an increasing trend throughout *Booker*, *Rita*, *Gall*, and *Kimbrough*, back to pre-PROTECT era levels.

FIGURE 2. BELOW RANGE DEPARTURE RATES



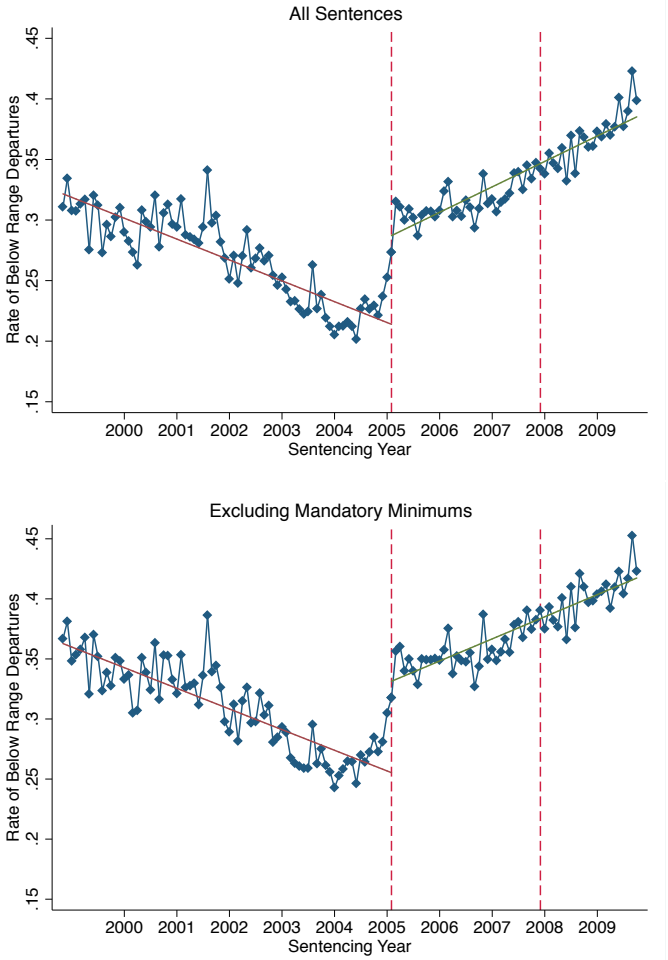
Notes: Data is from the full sample 2000-2009 on all incarcerated defendants. Data points are monthly averages.

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 a result of a government sponsored substantial assistance motion.[144] Figure 3 indicates a similar trend with respect to rates

[144] 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.

of non-government sponsored below range departures, suggesting that judicial behavior has 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 inter-judge variation has increased.

FIGURE 3. NON-GOVERNMENT SPONSORED BELOW RANGE DEPARTURES



Notes: Data is from the full sample 2000-2009 on all incarcerated defendants. Data points are monthly averages.

F. Measuring Inter-Judge Disparity: Analysis of Variance

To identify changes in inter-judge disparity, I employ an analysis of variance methodology. Variants of this methodology has been used in the federal sentencing litera-

ture,[145] as well as in the economics literature on teacher value added.[146] The analysis of variance technique measures inter-judge dispersion in sentencing outcomes based on the variance of a judge-specific random variable.[147]

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 = 0 and variance = σ^2.[148] 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 variance or standard deviation in judge effects, σ. The greater the inter-judge variation in outcomes, the larger the 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. A finding that σ = 5 implies a defendant who is assigned to a judge that is one standard deviation "harsher" than the average judge receives a five month longer sentence. In order to capture changes in inter-judge disparity, this paper measures σ in periods before *Booker* and after *Booker*. An increase in σ after *Booker* implies an increase in inter-judge sentencing disparity after the Guidelines were rendered

[145] Studies using a similar methodology include Anderson et al., *supra* note 10; Joel Waldfogel, *Aggregate Inter-Judge Disparity in Sentencing: Evidence from Three Districts*, 4 FED. SENTENCING REP. 151 (1991); Abigail Payne, *Does Inter-judge Disparity Really Matter? An Analysis of the Effects of Sentencing Reforms in Three Federal District Courts*, 17 INT'L J. LAW & ECON. 337 (1997).

[146] *See, e.g.* Raj Chetty et al., *How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star*, 126 Q. J. ECON. 1593 (2011).

[147] This paper does not employ the statistical technique used by Scott. *See* Scott, *supra* note 27. Scott regresses sentencing outcomes on dummy indicators for each sentencing judge, such that the corresponding R-squared measures the percentage of variance in the dependent variable that is explained by sentencing judge identity. Scott, *supra* note 27, at 58. The author interprets an increase in the R-squared in time periods following *Booker* as indicative of growing inter-judge disparities. For instance, with regards to sentence lengths for cases excluding mandatory minimums, the author finds an increase in R-squared from 0.014 pre-*Booker* to 0.080 post *Kimbrough/Gall*. *Id.* at Table 2, at 34. However, the R-squared measure is problematic for two main reasons. First, the measure of R-squared does not have a straightforward interpretation in terms of actual inter-judge variation, in contrast to a measure of the variance in a judge-specific random variable. Second, the magnitude of an R-squared cannot be taken literally without some discussion of its statistical significance, which is proxied by the linear regression model's significance. Scott's linear regression models are often statistically insignificant, suggesting that judge fixed effects are a poor predictor of sentencing outcomes, but he 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, and four out of five of the studied periods in Table 3, at 34-40. In contrast, measures of inter-judge 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." Scott, *supra* note 27, at 34-35 FN 177.

[148] *See* Appendix B for details on the empirical methodology.

advisory. In particular, I separate the sample timeframe of 2000-2009 into four main periods: (1) *Koon* (October 2000-April 2003), (2) PROTECT Act (May 2003 - January 2005),[149] (3) *Booker* (January 2005 - November 2007), and (4) *Kimbrough*/*Gall* (December 2007 - September 2009).[150]

III Results on Inter-judge and Regional 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.[151] A boxplot captures the distribution of judge sentencing practices, with a more narrow spacing of the boxplot evidencing less inter-judge disparity. In particular, the top and bottom of the box capture the spread between the 75 percentile and 25 percentile mean judge effect, also known as the interquartile range (IQR). More extreme judge effects are represented in the whiskers of the boxplot, as well as by outliers. The greater the IQR and presence of outliers, the larger the inter-judge disparity within district courts.

The first panel of Figure 4 indicates that over 50% 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 25th and 75th percentile (IQR) expands modestly but noticeably across the time periods. Following *Kimbrough/Gall*, the number of outliers also increases.

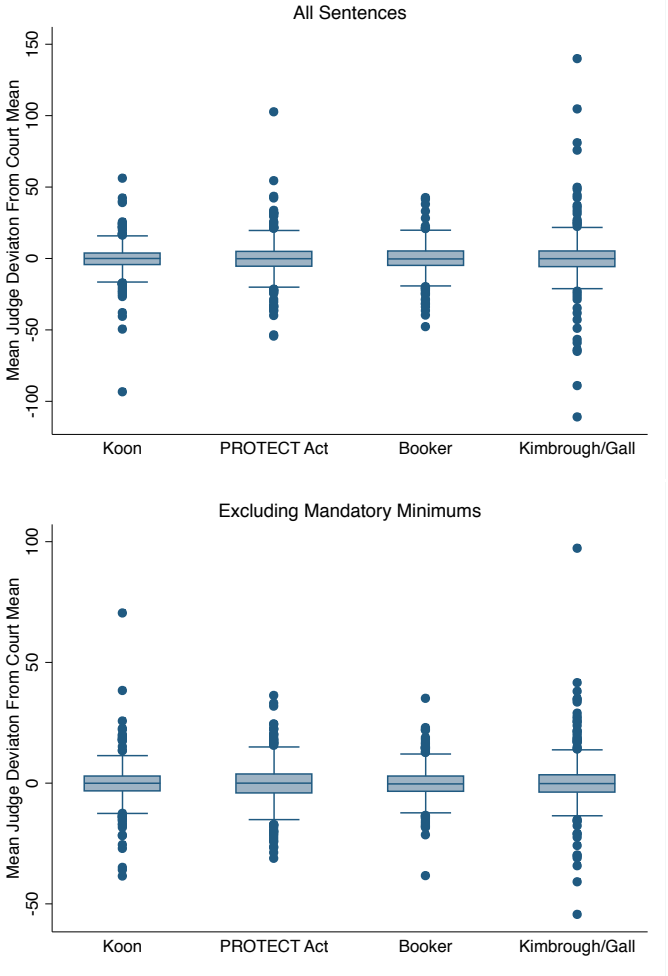
Note, however, that some of the inter-judge disparities may be attributable to uneven applications of mandatory minimums by prosecutors. The second panel of Figure 4 presents distributions of average judge sentences for those cases in which a mandatory minimum was not charged, approximately two-thirds of all cases. These cases better represent disparities more likely attributable to judicial behavior. As the figure shows, following *Kimbrough* and *Gall*, judge sentence lengths begin to depart more radically from court averages, with substantially more outlier sentence lengths on both sides of the distribution.

[149] The PROTECT Act became effective as of April 30, 2003.

[150] *Booker* was decided on January 12, 2005, and *Kimbrough/Gall* were decided on December 10, 2007. Although the USSC data only contains information on sentencing month and year, the data is coded to denote which January 2005 cases were pre and post *Booker* and which December 2007 cases were pre and post *Kimbrough/Gall*.

[151] Case composition likely varies across district courts. Thus, I use a measure of judge sentence length that is comparable across all districts, which can be accomplished by calculating average sentence length by judge relative to the mean district courthouse sentence.

FIGURE 4. AVERAGE JUDGE SENTENCE LENGTHS IN MONTHS



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

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 sentencing district courthouse.[152] Each measure of σ is also accompanied by a 95% 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 2.6 month longer sentence 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 almost a 4 month longer sentence. Inter-disparities increased further following *Booker*. *Booker* and *Kimbrough*/*Gall* induced almost a doubling of inter-judge disparity compared to the *Koon* period. A harsher judge sentenced a defendant 4.6 months longer than the court average in the immediate aftermath of *Booker* and over 5.2 months longer after *Kimbrough* and *Gall*.

[152]Table 1 analysis includes all defendants that received a prison sentence, excluding those who received probation.

The second panel of Table 1 excludes from the analysis those cases in which a mandatory minimum was charged. During *Koon*, a one standard deviation "harsher" judge sentenced a defendant to 1.6 months more than the court average, and 3.2 months longer during the PROTECT Act. Interestingly, inter-judge disparity for non-mandatory minimum cases falls to 2.4 months during *Booker*, rising back up to 3.3 months after *Kimbrough* and *Gall*. Changes in σ are not significant from the PROTECT Act to *Kimbrough* and *Gall*, suggesting that on average, inter-judge disparities in sentence lengths of non-mandatory minimum cases have not changed starkly during this period. However, inter-judge disparities are significantly larger following *Gall/Kimbrough* compared to the *Koon* period, more than doubling. This evidence suggests that even in cases in which mandatory minimums were not charged, inter-judge 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. This finding suggests that a large proportion of inter-judge disparities may be driven by the disparate application of mandatory minimums by prosecutors.

TABLE 1. INTER-JUDGE VARIATION IN SENTENCE LENGTHS

ALL SENTENCES				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	2.616	1.951	3.509	45407
PROTECT Act	3.975	3.118	5.068	38316
Booker	4.651	3.923	5.513	68129
Kimbrough/Gall	5.282	4.519	6.175	48605
EXCLUDING MANDATORY MINIMUMS				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	1.599	1.115	2.293	30193
PROTECT Act	3.217	2.633	3.931	27571
Booker	2.392	1.903	3.008	46160
Kimbrough/Gall	3.296	2.783	3.902	33732

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.

Table 2 presents evidence of inter-judge variation in the decision to incarcerate, disentangling the decision to incarcerate from the sentence length decision.[153] 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 non-zero. Indeed, Table 2 reveals that inter-judge disparity has increased throughout the four time periods, and most significantly following *Kimbrough* and *Gall*. During *Koon*, a one standard deviation "harsher" judge was 2.9% more likely

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

to incarcerate than the courthouse average. The effect increased to 3.3% during the PROTECT Act, 3.5% during *Booker* and almost 5% following *Kimbrough/Gall*.

TABLE 2. INTER-JUDGE VARIATION IN INCARCERATION RATE

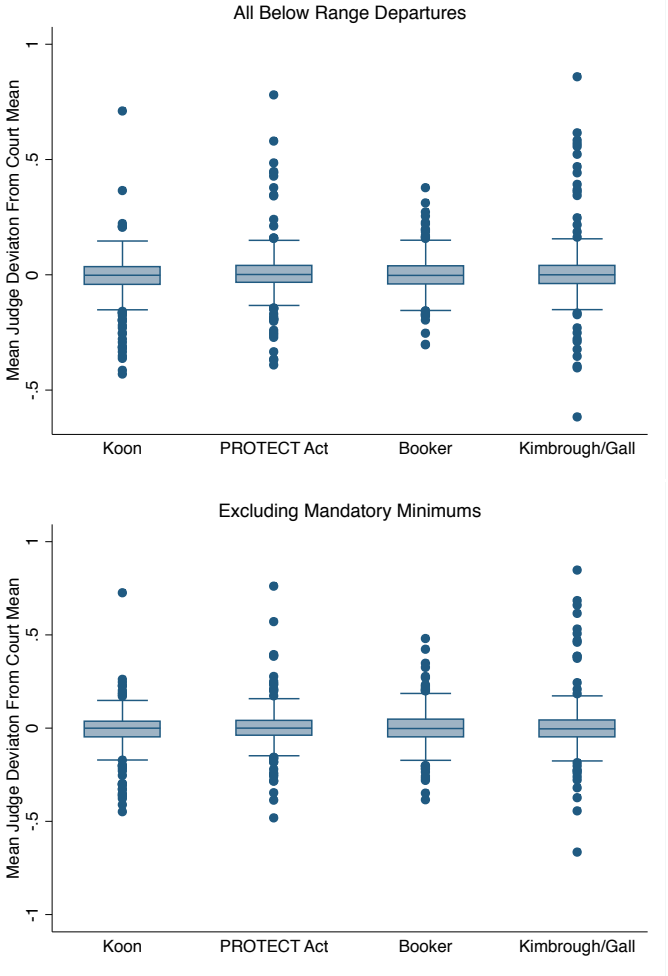
	ALL SENTENCES			
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0291	.0239	.0353	51122
PROTECT Act	.0335	.0271	.0414	41713
Booker	.0349	.0297	.0412	73782
Kimbrough/Gall	.0499	.0433	.0575	52586

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

Figure 5 shows the boxplots of average rates of below range departures by judge, relative to the district courthouse mean, for all incarcerated defendants. While the rate of below range departures in the aggregate was lowest during the PROTECT Act period (Figure 2), the distribution of judge below range departure rates does not appear to be significantly different between *Koon* and the PROTECT Act. In fact, there are far fewer outliers during the first two years following *Booker*. However, inter-judge deviations from the court mean expand visibly following *Kimbrough/Gall*. Figure 5 suggests that increasing inter-judge disparities in sentence length as described in Part IV.A are partly attributable to growing inter-judge disparities in the rate of below range departures.

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 4.8% more likely to be sentenced below the Guidelines recommended minimum.[154] During the PROTECT Act, a similar judge was 4.2% more likely to sentence below range. Following *Booker*, the "lenient" judge's practices deviated more greatly from the courthouse average, with a 5.4% rate immediately following *Booker* and 6.7% rate after *Kimbrough/Gall*. The increased likelihood of below range departures following *Kimbrough/Gall* is statistically significant from the *Koon*-era rate, revealing markedly higher inter-judge disparities.

Excluding cases with mandatory minimums reveals a very similar trend, with the one standard deviation more "lenient" judge being 5.1% more likely to sentence below range during *Koon*, rising to 7.1% following *Kimbrough/Gall*. 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

[154]Here, I define a judge who sentences defendants at greater rates below range as more "lenient." Leniency is used solely to connote a tendency to impose shorter sentence lengths.

that inter-judge disparities in below range departures are real and substantial, and not the mere product of mandatory minimums. If anything, measures of inter-judge 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 minimum, becoming the statutorily binding minimum, thus potentially reducing inter-judge disparity. The findings suggest that the application of mandatory minimums may yield the appearance of inter-judge *consistency.*[155]

TABLE 3. INTER-JUDGE VARIATION IN BELOW RANGE DEPARTURES

ALL SENTENCES				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0477	.0396	.0576	44338
PROTECT Act	.0427	.0337	.0542	36613
Booker	.0538	.0459	.0632	64781
Kimbrough/Gall	.0668	.0576	.0774	45267

EXCLUDING MANDATORY MINIMUMS				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0511	.0415	.0629	29369
PROTECT Act	.0522	.0411	.0662	26080
Booker	.0500	.0405	.0618	43390
Kimbrough/Gall	.0712	.0599	.0845	30793

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.

Not only do mandatory minimums confound the accurate determination of inter-judge disparities, so do below range departures that are government sponsored. Table 4 analyzes inter-judge variation in only those below range departures that are judicially initiated, rather than the result of a government substantial assistance motion. Table 4 indicates that the increasing inter-judge disparities in below range departures evidenced in Table 3 persist in this subset of departures. Inter-judge disparities increased from 4.3% during *Koon* to 5.9% after *Booker* and over 7.4% after *Kimbrough/Gall*, with the lowest inter-judge disparities during the PROTECT Act. Inter-judge disparities similarly increased throughout the period for the subset of cases not subject to a mandatory minimum, from 4.3% during *Koon* to over 7.4% after *Kimbrough/Gall*. These results indicate that in the subset of departures that are most likely attributable to judicial behavior, the PROTECT Act was not only associated with the lowest aggregate rates of downward departures, but also the lowest inter-judge disparities.

[155] *See* Scott, *supra* note 27, at 26 ("mandatory minimums may interfere with accurate assessment of inter-judge sentencing disparity by creating the illusion of inter-judge consistency.").

TABLE 4. INTER-JUDGE VARIATION IN BELOW RANGE DEPARTURES
NON-GOVERNMENT SPONSORED DEPARTURES

ALL SENTENCES				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0426	.0349	.0519	37449
PROTECT Act	.0321	.0247	.0416	33432
Booker	.0588	.0507	.0682	59386
Kimbrough/Gall	.0743	.0645	.0856	42701

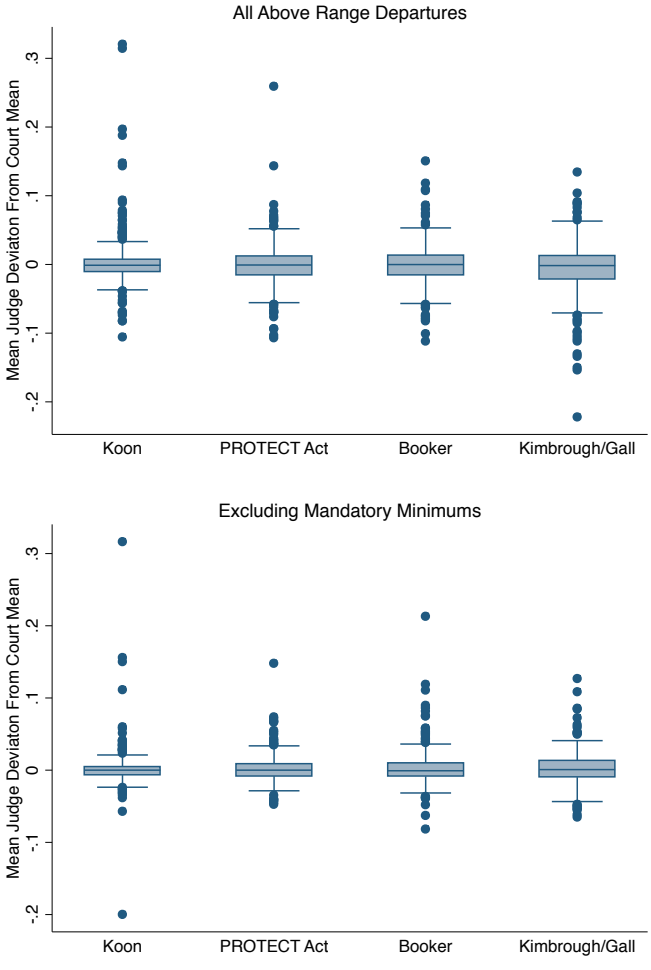
EXCLUDING MANDATORY MINIMUMS				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0434	.0345	.0547	27051
PROTECT Act	.0369	.0274	.0496	25505
Booker	.0553	.0457	.0669	43230
Kimbrough/Gall	.0742	.0628	.0878	31796

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.

C. Above Range Departures

Inter-judge disparities have increased not only in the rate of below range departures, but also the rate of above range departures, which comprise approximately 5% of cases. Figure 6 presents the distribution of average rates of above range departures by judge, relative to their district courthouse mean, for all incarcerated defendants. The graphs reveal an expansion in the distribution of above range departure rates within district courts, particularly between the 25th and 75th percentile of the distribution. Increased inter-judge deviations are also reflected in the rate of above range departures for cases with no mandatory minimums charged. Although there appear to be fewer extreme outliers following *Kimbrough/Gall*, the spread between the 25th and 75th percentile is visibly larger compared to pre-*Booker* spreads.

FIGURE 6. AVERAGE JUDGE RATES OF ABOVE RANGE DEPARTURES



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

Table 5 presents measures of inter-judge variation from the analysis of variance and reveals significant and substantial increases in inter-judge disparities in above range departures. In the top panel where all sentences are analyzed, a one standard deviation "harsher" judge was 1.6% more likely to sentence a defendant above range compared to the average judge in the courthouse during *Koon*. While inter-judge variation did not change substantially from *Koon* to the PROTECT Act period, to the first few years after *Booker*, this harsher judge was over 2.7% more likely to sentence a defendant above range after *Kimbrough/Gall*. Inter-judge disparities in above range departures increased by over 70% from the beginning to the end of the time period.

Of course, mandatory minimums may explain a sizable fraction of defendants that are sentenced above range if the mandatory minimum trumps the maximum Guidelines recommended sentence. When cases with mandatory minimums are excluded, inter-judge disparities are approximately halved. During *Koon*, inter-judge disparities in above range departures were minimal, with a one standard deviation "harsher" judge being only .07% more likely to sentence above range. However, inter-judge disparities doubled by the end of the time period, to 1.3% after *Kim-*

brough/Gall, suggesting that increases in above range inter-judge disparities are not the mere byproduct of mandatory minimums.

TABLE 5. INTER-JUDGE VARIATION IN ABOVE RANGE DEPARTURES

		ALL SENTENCES		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0161	.0129	.0202	45184
PROTECT Act	.0182	.0134	.0247	38322
Booker	.0173	.0135	.0222	68120
Kimbrough/Gall	.0276	.0227	.0335	48596
		EXCLUDING MANDATORY MINIMUMS		
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0067	.0038	.0117	30002
PROTECT Act	.0102	.0072	.0144	27571
Booker	.0139	.0106	.0181	46144
Kimbrough/Gall	.0130	.0093	.0183	33734

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.

D. Sentencing Practices by Judge Demographics

The previous section finds that inter-judge disparities in sentence length, below range departures, and above range departures have increased significantly following *Booker*, in particular after *Kimbrough/Gall.* In this section, I analyze whether increases in inter-judge 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.[156] Recall that due to the random assignment of cases to judges within a district courthouse, any difference in judge sentencing practices can be solely attributable to a judge effect.

Consistent with previous research, 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.[157] These differences magnified in the aftermath of *Booker* and *Kimbrough/Gall*, suggesting that they are some of the sources of the growing inter-judge 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-year.

[156] *See* Appendix C.

[157] *See supra* notes 30-32.

TABLE 6. SENTENCING PRACTICES BY JUDGE CHARACTERISTICS
ALL SENTENCES

	(1) Sentence	(2) Below Range	(3) Below Range Non-Govt	(4) Above Range
Post Booker	1.010	0.0211	0.0462**	0.0242***
	(1.439)	(0.0240)	(0.0207)	(0.0091)
Tenure	-0.0556	0.0005	0.0003	-0.0005*
	(0.0446)	(0.0009)	(0.0009)	(0.0003)
Tenure*Booker	0.0461	-0.0006	0.0002	0.0010**
	(0.0614)	(0.0011)	(0.0009)	(0.0004)
Tenure*Kimbrough	0.0126	0.0011	0.0023	0.0009*
	(0.0672)	(0.0019)	(0.0019)	(0.0005)
Democratic	-0.522	0.0109	0.0066	0.0016
	(0.438)	(0.0093)	(0.0082)	(0.0027)
Democratic*Booker	-0.609	0.0084	0.0086	0.0019
	(0.571)	(0.0082)	(0.0081)	(0.0037)
Democratic*Kimbrough	-0.756	0.0223*	0.0269**	-0.0027
	(0.731)	(0.0115)	(0.0120)	(0.0051)
Female	0.537	-0.0025	-0.0073	0.0051
	(0.468)	(0.0126)	(0.0132)	(0.0042)
Female*Booker	-1.692***	0.0155	0.0164	-0.0100*
	(0.503)	(0.0121)	(0.0138)	(0.0054)
Female*Kimbrough	-0.401	0.0110	0.0145	0.0012
	(0.576)	(0.0192)	(0.0203)	(0.0055)
Black	-0.873	0.0192	0.0266	-0.0082*
	(0.581)	(0.0161)	(0.0173)	(0.0044)
Black*Booker	-0.624	-0.0106	-0.0104	0.0066
	(0.951)	(0.0176)	(0.0206)	(0.0071)
Black*Kimbrough	0.459	-0.0334	-0.0367	0.0137
	(1.405)	(0.0262)	(0.0315)	(0.0095)
Pre Guide	0.561	-0.0266**	-0.0218	0.0044
	(0.831)	(0.0134)	(0.0135)	(0.0053)
Pre Guide*Booker	-0.765	0.0254	0.0091	-0.0137**
	(0.952)	(0.0179)	(0.0176)	(0.0069)
Pre Guide*Kimbrough	-0.418	0.0129	-0.0027	-0.0003
	(1.224)	(0.0308)	(0.0313)	(0.0099)
Booker	0.276	0.0039	-0.0063	-0.0057
	(0.735)	(0.0139)	(0.0132)	(0.0050)
Kimbrough	-0.443	0.0038	-0.0076	-0.0017
	(1.193)	(0.0191)	(0.0182)	(0.0124)
Observations	206,292	205,160	178,150	205,160
R-squared	0.785	0.217	0.281	0.083

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. Female judges significantly altered their practices from their male counterparts within the same courthouse. Immediately after *Booker*, female judges sentenced observably similar defendants to approximately 1.7 months less than their male colleagues.[158] Columns 2 presents results for the rate of below range departures. Inter-judge disparities in rates of below range departures appear to be somewhat attributable to differences by judge political affiliation. Following *Kimbrough/Gall*, Democratic judicial appointees are significantly more likely to depart downwards from the Guidelines recommended range, compared to their Republican appointed colleagues. For a similar defendant and crime, Democratic judges were 2.2% more likely to depart downwards. Interestingly, pre-Guidelines (1987) appointees are significantly *less likely* to depart downwards from the Guidelines throughout the entire 2000-2009 period. Inter-judge differences by demographics are also prominent for the subset of below range departures not sponsored by the government, as seen in column 3. Democratic judicial appointees are 2.7% more likely to depart downwards compared to their Republican colleagues. Judges appointed post *Booker* are almost 5% more likely to depart downwards compared to pre-*Booker* appointees.

Finally, column 4 presents results for above range departures. Following *Booker*, female judges are 1% less likely to sentence above range compared to their male colleagues. Inter-judge differences also appear by judge tenure, defined as number of years of experience sentencing under the mandatory Guidelines regime for those judges appointed under the mandatory regime. In general, a judge with greater years of experience under the mandatory regime is significantly less likely to sentence above range, but this pattern reverses in the aftermath of *Booker* and *Kimbrough/Gall*, where judges with greater experience are *more* likely to sentence above range. Black judges are in general .08% less likely to sentence above range compared to white judges, and pre-Guidelines appointees are 1.4% less likely to depart upwards after *Booker*. Also striking are the inter-judge differences generated between judges appointed pre-*Booker* and judges appointed post *Booker*. In general, post *Booker* judicial appointees are 2.4% more likely to sentence above range their their pre-*Booker* appointed peers.

Table 7 presents the results excluding cases with mandatory minimums. Judge differences by political affiliation of appointing president and gender persist. Following *Kimbrough/Gall*, Democratic appointees issue sentences 0.8 months shorter than their Republican colleagues for observably similar defendants, are 3.1% more likely to depart downwards, and 3.1% more likely to depart downwards in cases without government substantial assistance motions. Similarly, female judges issue 0.8 month shorter sentences than their male colleagues following *Booker*. Differences between pre-*Booker* and post *Booker* appointees also remain. Post *Booker* are approximately 5% more likely to depart downwards in all cases, and over 7% more likely when there is no substantial assistance motion.

[158]The *Booker* indicator here represents only the period 2005-2007 prior to *Kimbrough*.

TABLE 7. SENTENCING PRACTICES BY JUDGE CHARACTERISTICS
EXCLUDING MANDATORY MINIMUMS

	(1) Sentence	(2) Below Range	(3) Below Range Non-Govt	(4) Above Range
Post Booker	-1.010	0.0496**	0.0713***	0.0033
	(0.683)	(0.0242)	(0.0228)	(0.0061)
Tenure	-0.0278	2.90e-05	-4.93e-05	-0.0001
	(0.0272)	(0.0008)	(0.0008)	(0.0002)
Tenure*Booker	0.0260	-0.0003	-0.0002	0.0002
	(0.0343)	(0.0010)	(0.0009)	(0.0003)
Tenure*Kimbrough	-0.0130	0.0022	0.0029	0.0006
	(0.0454)	(0.0019)	(0.0021)	(0.0005)
Democratic	-0.382	0.0077	0.0040	0.0011
	(0.267)	(0.0088)	(0.0086)	(0.0022)
Democratic*Booker	-0.291	0.0047	0.0069	-0.0023
	(0.290)	(0.0101)	(0.0104)	(0.0028)
Democratic*Kimbrough	-0.830*	0.0310**	0.0307**	-0.0030
	(0.456)	(0.0141)	(0.0152)	(0.0044)
Female	0.242	-0.0019	-0.0077	0.0019
	(0.324)	(0.0145)	(0.0156)	(0.0023)
Female*Booker	-0.811**	0.0145	0.0168	-0.0033
	(0.407)	(0.0158)	(0.0168)	(0.0032)
Female*Kimbrough	-0.217	0.0049	0.0123	0.0015
	(0.501)	(0.0219)	(0.0221)	(0.0047)
Black	-0.501	0.0297*	0.0309*	-0.0031
	(0.425)	(0.0175)	(0.0180)	(0.0021)
Black*Booker	0.220	-0.0258	-0.0227	0.0079*
	(0.637)	(0.0200)	(0.0223)	(0.0045)
Black*Kimbrough	0.587	-0.0533*	-0.0473	0.0009
	(0.811)	(0.0318)	(0.0380)	(0.0045)
Pre Guide	0.256	-0.0233	-0.0229	0.0037
	(0.400)	(0.0152)	(0.0157)	(0.0033)
Pre Guide*Booker	-0.312	0.0166	0.0145	-0.0019
	(0.566)	(0.0195)	(0.0200)	(0.0049)
Pre Guide*Kimbrough	0.364	-0.0073	-0.0137	-0.0017
	(0.949)	(0.0331)	(0.0370)	(0.0091)
Booker	0.492	-0.0062	-0.0174	0.0041
	(0.372)	(0.0145)	(0.0138)	(0.0035)
Kimbrough	0.184	-0.0201	-0.0277	-0.0030
	(0.670)	(0.0189)	(0.0192)	(0.0069)
Observations	141,647	141,386	131,417	141,386
R-squared	0.819	0.244	0.283	0.037

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.

Overall, these results suggest that sentencing differences associated with judge gender and political affiliation are magnified after *Booker* and/or *Kimbrough/Gall.* Such differences are likely sources of growing inter-judge disparities. Given these large changes in inter-judge disparities following *Booker*, judges do not appear to be completely "anchored" to the Guidelines.[159]

However, the finding that post *Booker* judicial appointees are more likely to depart downwards 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.[160] The "anchor" of the Guidelines sentence may be more prominent to pre-*Booker* appointees because these judges are more acculturated and experienced under the Guidelines. In contrast, the "anchor" is less prominent for post *Booker* appointees. These potential anchoring differences between pre and post *Booker* appointees suggests that defense lawyer James Felman's predictions may be true - that disparities may "increase as the years go by and the bench is filled with individuals who have no history with binding guidelines."[161] Yet, inter-judge disparities that are due to the entrance of new judges to the federal bench might only reflect a short-term surge in disparity. As time goes on and all judges have no history with binding Guidelines, inter-judge disparities attributable to this source may fall.

E. Regional Disparity: Inter-District Variation

Commentators have suggested that different political climates across districts and circuits can affect sentencing practices,[162] yielding empirical findings that jurisdictional effects are prominent in federal sentencing.[163] The recent 2012 Commission

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

[160] In robustness checks, I find that the behavior of post *Booker* appointees in my data is not due to the fact that they are George W. Bush appointees 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.

[161] *See* Felman, *supra* note 35, at 98-99.

[162] *See, e.g*, Nora Demlietner, *The Nonuniform Developments of Guideline Law in the Courts*, 6 FED. SENT. REP. 239 (1994) (describing district and circuit specific "personas" in sentencing case law); Jeffrey T. Ulmer & John Kramer, *Court Communities Under Sentencing Guidelines: Dilemmas of Formal Rationality and Sentencing Disparity*. 34 CRIMINOLOGY 383, 402-403 (1996) (Based on an analysis of three county courts in Pennsylvania, the authors argue that local courts operate under formal sentencing standards articulated by a guidelines regime and substantive, extralegal factors relevant to local courts, such as "perceptions of the defendant's characteristics, local concerns, and court actors' organizational and individual interests."); Jeffrey T. Ulmer, SOCIAL WORLDS OF SENTENCING: COURT COMMUNITIES UNDER SENTENCING GUIDELINES (1977).

[163] *See* Celesta 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 Everett & Roger Wojtkiewicz, *Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing*, 18 J. QUANTITATIVE CRIM. 189 (2002) (finding harsher sentencing in the southern circuits compared to other circuits); Paula Kautt, *Location, Location, Location: Interdistrict and Inter circuit Variation in Sentencing Outcomes for*

report finds that rates of non-government sponsored below range sentences increasingly depend upon the district court in which the defendant is sentenced and the influence of the Guidelines on sentence length varies significant by circuit court.[164] However, some researchers have found that between-district variation in the effects of *extralegal* factors on sentencing have not increased following *Booker*.[165]

Recall that the identification of the impact of *Booker* on inter-judge disparity within a district courthouse relies on the random assignment of cases to judges. Such random assignment does not exist between districts, such that differences in district sentencing practices are most 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 the 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.[166]

While I cannot control for unobservable differences across districts, the empirical methodology in this Article does control comprehensively for observable offender and case characteristics.[167] For the inter-district results, I utilize the full sample described in Section 3 as random assignment is no longer a prerequisite. In the context of inter-district disparities, analysis of variation 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 if he were sentenced in an average district court.

Figure 7 presents raw distributions of sentence lengths by circuit court, excluding life sentences.[168] While uncontrolled differences cannot be treated as regional effects because districts have very different case compositions, the raw data reveals substantial differences in sentence length, both in the distribution between the 25th percentile and 75th percentile, and presence of outliers. For instance, prior to *Booker*, defendants sentenced in the Third Circuit received an average sentence of 56 months, compared to an average sentence of 78 months for defendants sentenced in the neighboring Fourth Circuit. After *Booker*, the average Third Circuit defendant received 62 months in prison, while the average Fourth Circuit defendant

Federal Drug-Trafficking Offenses, 19 JUSTICE QUARTERLY 633, 659 (2002) ("despite the federal system's congressionally mandated return to determinate sentencing, extra-legal factors (specifically jurisdictional effects) continue to influence the federal sentencing system and its outcomes directly and indirectly").

[164]United States Sentencing Commission, REPORT ON THE CONTINUING IMPACT OF UNITED STATES V. BOOKER ON FEDERAL SENTENCING (2012).

[165]*See* Jeffery T. 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 JUSTICE QUARTERLY 799 (2011).

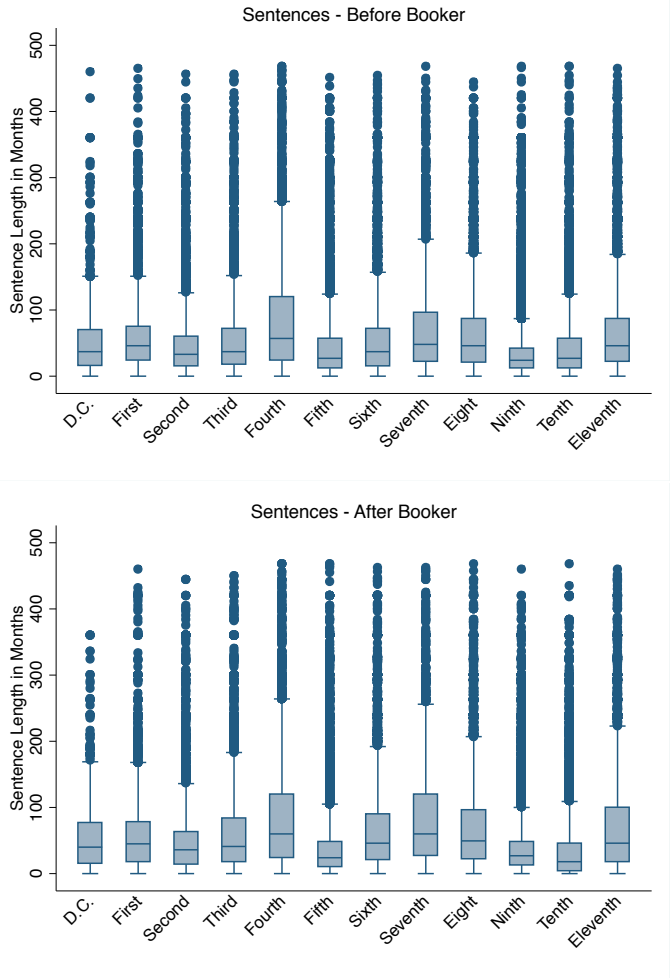
[166]United States Sentencing Commission, FIFTEEN YEARS OF GUIDELINES SENTENCING: AN ASSESSMENT OF HOW WELL THE FEDERAL CRIMINAL JUSTICE SYSTEM IS ACHIEVING THE GOALS OF SENTENCING REFORM at 99-100 (Nov. 2004) ("Similarly, variations in the rates of a particular type of departure among different districts 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.").

[167]Nevertheless, the results on inter-district variation should be interpreted with some caution to the extent that there are unobserved differences across district courts that cannot be captured.

[168]Life sentences are top coded as 470 months in the dataset.

received 84 months in prison.

FIGURE 7. DISTRIBUTION OF SENTENCE LENGTHS, BY CIRCUIT COURT



Notes: Data is from the full sample 2000-2009.

Table 8 shows that after controlling for case and defendant characteristics, there is substantial variation in the sentence that a defendant would receive depending on which district court he is sentenced in. During the *Koon* period, a defendant sentenced in a one standard deviation "harsher" district court received a 7.8 month longer prison sentence. This inter-district disparity increased to 8.4 months during the PROTECT Act, to 10.4 months immediately following *Booker*, reaching an 11.3 month difference after *Kimbrough/Gall*. By late 2007, inter-district disparities were significant 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 inter-district variation. The lower panel of Table 8 indicates that a one standard deviation "harsher" district court sentenced a defendant to 3.6 months longer than the average district court under *Koon*, 4.4 months longer after the PROTECT Act, 4.9 months longer after *Booker* and 5.2 months longer after *Kimbrough/Gall*. Once again, inter-district variation is statistically greater after *Kimbrough/Gall* compared to *Koon*. Nevertheless, the

finding the the magnitude of inter-district variation is reduced by over half when a statutory minimum is not charged indicates that the application of mandatory minimums is a large contributor to inter-district disparities, particularly in light of the fact that mandatory minimums represent only approximately one-third of the cases.

TABLE 8. INTER-DISTRICT VARIATION IN SENTENCE LENGTHS

ALL SENTENCES				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	7.799	6.701	9.077	159163
PROTECT Act	8.439	7.232	9.849	83829
Booker	10.397	8.961	12.063	148560
Kimbrough/Gall	11.262	9.692	13.087	106033
EXCLUDING MANDATORY MINIMUMS				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	3.578	3.057	4.188	104917
PROTECT Act	4.403	3.758	5.159	58104
Booker	4.920	4.229	5.724	97628
Kimbrough/Gall	5.157	4.419	6.018	72036

Notes: Data is from the full 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.

Table 9 reveals that district courts also significantly differ in their rates of below range departures. 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, during *Koon*. This measure of inter-district 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 inter-district disparity with regards to downward departures.

TABLE 9. INTER-DISTRICT 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 full 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.

F. Prosecutorial Contributions to Disparities

Prosecutors likely contribute to observed inter-judge disparities through their charging decisions. One area of great prosecutorial discretion is the decision to charge an offense that carries a mandatory minimum. Justice Breyer has stated that mandatory minimum statutes "transfer sentencing power to prosecutors, who can determine sentences through the charges they decide to bring."[169] Strategic charging of mandatory minimums are likely more prominent after *Booker* as some prosecutors charge mandatory minimums in order to narrow a judge's discretion.[170]

In a 2011 Congressional report on mandatory minimum penalties, the Sentencing Commission found significant variation in the extent which prosecutors applied enhancements for mandatory minimum penalties under drug trafficking offenses.[171] The report documented over 75% of eligible defendants receiving the statutory mandatory minimum penalty in some districts, but none of eligible defendants in other districts receiving the enhancement.[172] Furthermore, recent work by researchers shows evidence of significant racial disparities in prosecutorial charging.[173]

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

[170] *See* Testimony of Patrick J. Fitzgerald, U.S. Attorney, Northern District of Illinois, to the United States Sentencing Commission, at 252 (Sept. 2009) ("[A] prosecutor is far less willing to forego 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.").

[171] United States Sentencing Commission, REPORT TO CONGRESS: MANDATORY MINIMUM PENALTIES IN THE FEDERAL CRIMINAL JUSTICE SYSTEM, *supra* note 85, at 252-261.

[172] *Id.* at 111-113 (prosecutors reported wide variations in the district practices on seeking statutory minimum penalties).

[173] *See* M. Marit Rehavi & Sonja B. Starr, *Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences*, University of Michigan Law & Economics Working Paper 2012. Using data on 58,000 federal criminal cases from 2007-2009, the authors find significant racial disparities in severity of initial charges. In particular, they find that black offenders are on average more

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.[174] 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.[175] Scholars have commented that the substantial assistance departure provision affords prosecutors immense discretion over both plea bargaining and sentencing outcomes under the Guidelines.[176]

I find that the application of mandatory minimums appears to be a large contributor to inter-judge disparities. Given the random assignment of 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 10 reveal small, but significant differences in the percentage of cases with applicable mandatory minimums across judges. A judge one standard deviation out in the distribution was 2.3% more likely to see a case with a mandatory minimum during *Koon*, but 3.5% more likely after *Kimbrough/Gall*. The increase in the differential rates of mandatory minimums after *Kimbrough/Gall* coincide with substantial increases in inter-judge 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.

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 charge.

[174] 18 U.S.C. §3553(e). 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). Report to Congress: Mandatory Minimum Penalties in the Federal Criminal Justice System, *supra* note 85, states that in recent years, white defendants in drug cases are more frequently granted the safety valve exception compared to other defendants.

[175] According to the Sentencing Commission, substantial assistance motions reduce the average defendant's sentence length by 50%.

[176] *See* Nagel & Schulhofer, *supra* note 37, at 550 ("The use of the section 5K1.1 substantial-assistance motion varies from jurisdiction to jurisdiction....There is no limit on the amount of reduction once the motion is submitted. The section 5K1.1 motion is also used to avoid guideline ranges or mandatory minimum sentences for sympathetic defendants – even when there has been no genuine substantial assistance."); Michael H. Tonry, SENTENCING MATTERS (1996); Jeffrey Standen *Plea Bargaining in the Shadow of the Guidelines*, 81 CALIFORNIA L. REV. 1471 (1993).

TABLE 10. INTER-JUDGE VARIATION IN MANDATORY MINIMUMS

	ALL SENTENCES			
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0230	.0180	.0293	51077
PROTECT Act	.0188	.0130	.0272	41697
Booker	.0242	.0196	.0298	73706
Kimbrough/Gall	.0345	.0289	.0411	52551

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.

TABLE 11. INTER-JUDGE 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.

Appendix A Table A2 and Table A3 confirm that a large portion of inter-district 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 6% more likely to be charged with a mandatory minimum. Table A3 also presents evidence of large inter-district differences in the rates of substantial assistance motions, with a defendant being approximately 9% 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."[177] However, inter-district 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

This section describes three of the major proposals for reform of federal sentencing after *Booker*. I describe each in turn, and then apply the empirical findings in this paper to shed light on the desirability of the various proposals, assuming that a reduction in inter-judge disparities is a worthwhile goal of sentencing reform.

[177] Nagel & Schulhofer, *supra* note 37, at 550.

A. "Topless" Guidelines

Within a few months after *Booker*, the Department of Justice recommended a new "topless" Guidelines system, in which judges would be bound by the Guidelines minimum, but not the maximum.[178] Echoing the topless Guidelines regime first proposed by Professor Frank Bowman, this construction would still allow judicial fact-finding to facts that raised the minimum applicable sentence, remaining constitutional under the principles espoused in *Blakely*.[179] Recall that *Blakely* applied the Sixth Amendment to challenge judicial fact-finding which raised a defendant's maximum sentence.[180] As a result, the recommended "topless" Guidelines system appeared to comport with both *Blakely* and the Court's holding in *Harris v. United States* that facts triggering a mandatory minimum sentence could be found by a judge.[181] However, the constitutional viability of a "topless" Guidelines system has now been firmly rejected, with the Supreme Court's recent holding in *Alleyne v. United States*, in which the Court squarely overruled *Harris*.[182]

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."[183] By binding judges to the applicable minimum sentence, the proposal would likely re-introduce concerns associated with prosecutorial power in charging and plea bargaining.[184]

Indeed, the evidence from Part IV provides empirical support for the proposition that a "topless" Guidelines proposal would aggravate disparities that are attributable to prosecutorial charging decisions. Table 1 and Table 5, which present evidence of inter-judge disparities and inter-district disparities, 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 inter-judge differences, such that these decisions are not made equally across

---

[178] *See* Federal Guidelines Sentencing Speech, *supra* note 23, at 326 (favoring "the construction of a minimum guideline system").

[179] *See* Memorandum from Frank Bowman to U.S. Sentencing Comm'n (June 27, 2004), 16 FED. SENTENCING REP. 364, 367 (2004).

[180] *Blakely*, 542 U.S. 296.

[181] 536 U.S. 545, 567-69 (2002). However, some commentators have suggested that the Court's holding in *Harris* may not survive after *Booker*. *See, e.g.,* Erwin Chemerinsky, *Making Sense of Apprendi and its Progeny*, 37 MCGEORGE L. REV. 531, 541 (2006); Frank O. Bowman III, *Mr. Madison Meets a Time Machine: The Political Science of Federal Sentencing Reform*, 58 STAN. L. REV. 235, 261 (2005).

[182] *Alleyne v. United States*, 133 S. Ct. 420, slip op. at 15 (2012) ("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.").

[183] *See* Berman, *Tweaking Booker*, *supra* note 112, at 363. Berman also discusses potential constitutional challenges to a "topless" Guidelines system.

[184] *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.").

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 ascribe greater power to prosecutors, likely resulting in inequitable disparities. Furthermore, results in Table 6 suggest that there have been substantial increases in inter-judge 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 asymmetrical manner.

B. "*Blakely*-ized" Guidelines

Justice Breyer's ultimate remedy for the Sixth Amendment issues facing the Federal Sentencing Guidelines was to declare the Guidelines "effectively advisory." 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 proven by a jury beyond reasonable doubt, or admitted by the defendant.[185]

However, introducing jury fact-finding into a mandatory Guidelines system is likely particularly complex. Justice Breyer in his *Booker* remedial opinion mused over how jury fact-finding might work, asking "[w]ould 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...?"[186] Additionally, to the extent that the mandatory Guidelines regime contributed to prosecutorial discretion and disparity, jury fact-finding in the face of extensive plea bargaining "would move the system backwards in respect to both tried and 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."[187]

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.[188] Addressing some of Justice Breyer's concerns, a 2005 American Bar Association Report suggested a version of the *Blakely*-ized 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

[185] *Booker*, 543 U.S. at 284-84 (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 defendant's sentence under the Guidelines to a jury beyond a reasonable doubt.").

[186] *Id.* at 254.

[187] *Id.* at 256 ("plea bargaining would likely lead to sentences that gave greater weight, not to real conduct, but rather to the skill of counsel, the policies of the prosecutor, the caseload, and other factors that vary from place to place, defendant to defendant, and crime to crime...plea bargaining of this kind would necessary move 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, *Tweaking Booker*, *supra* note 112, at 365-71.

[188] *See* Bowman, *Beyond Band-Aids*, *supra* note 109, at 191 ("the consensus view is that the Guidelines as now written are simply too complex and confusing to operate through juries").

jury."[189]

This Article cannot 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 aggravating and mitigating factors, under the "*Blakely*-ized" Guidelines, 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 jury did not make a factual determination with respect to a potential mitigating factor, such as acceptance of responsibility by the defendant, a judge would not be allowed to consider this factor, even if it were applicable. This Article provides evidence suggesting that a large component of disparities stem from prosecutorial charging decisions. A requirement of jury fact-finding in a mandatory Guidelines regime may exacerbate such disparities.

C. Judge Sessions Proposal

Most recently, former Sentencing Commission Chair Judge William K. Sessions III has recommended adoption of a simplified presumptive Guidelines system.[190] 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.[191] Judge Sessions recommends a reduction in the number of possible sentencing ranges, but broader ranges to afford judges greater discretion.[192] 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.[193]

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.[194] 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.[195]

---

[189] ABA Criminal Justice Section, Report and Recommendation on Booker (Jan. 2005), reprinted in 17 FED. SENT. REP. 335, 339 (2005).

[190] Sessions, *supra* note 24, at 340.

[191] *Id.* at 339.

[192] *Id.* at 340-45 (describing recommended changes to the current Guidelines sentencing chart).

[193] *Id.* at 346.

[194] *Id.* at 347-48; *see also* Frank O. Bowman III, *The Failure of the Federal Sentencing Guidelines*, 105 COLUM. L. REV. 1315, 1341 (2005); Stephen Breyer, *Federal Sentencing Guidelines Revisited*, 11 FED. SENTENCING REP. 180 (1999) ("[T]he Guidelines are simply too long and too complicated.").

[195] Sessions, *supra* note 24, at 349; Stephen Breyer, *supra* note 53, at 184 ("... 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 ... .").

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," with within range sentences "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."[196] In contrast, Judge Sessions proposes "relatively strict scrutiny by the appellate court" for downward departures.[197]

Critics of the Sessions proposal argue that the proposal would eliminate "judicial feedback to the Commission and constructive evolution of the [G]uidelines would virtually cease" as judges would have limited authority in setting the applicable sentence range.[198] As a result, both scholars and district court judges have expressed the view that the current advisory Guidelines best achieves the goals of sentencing because it reflects the right balance between various actors in federal sentencing.[199] Of district judges surveyed in 2010, over 75% prefer the current advisory Guidelines system to other alternatives.[200] In contrast, 14% of judges favored a version of the *Blakely*-ized proposal - "[a] system of mandatory [G]uidelines that comply with the Sixth Amendment and have broader sentencing ranges than currently exist, coupled with fewer statutory mandatory minimums."[201] Only 3% of judges preferred a return to the pre-*Booker* Guidelines system, suggesting that the overwhelming majority of judges would be opposed to a return to presumptive Guidelines, as proposed by Judge Sessions.[202]

Undoubtedly, *Booker* has given judges the freedom to consider the particular circumstances of the offense and traits of the defendant. To the extent that growing inter-judge 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.[203]

The empirical findings in this Article reveal that inter-judge disparities have doubled from the period of mandatory Guidelines sentencing to post *Booker* sentencing, with a defendant potentially receiving a 5 month longer sentence due to the mere happenstance of the judge assigned. Undoubtedly, a return to "presumptive" Guidelines would mechanically reduce inter-judge disparities by greatly limiting

[196] Sessions, *supra* note 24, at 353-54.

[197] *Id.* at 353-54 ("District courts' choices of sentences within the applicable cells on the grid 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.").

[198] Baron-Evans & Stith, *supra* note 25, at 1716.

[199] *Id.* at 1681. *See also* Michael Tonry, *The U.S. Sentencing Commission's Best Response to Booker is to Do Nothing*, 24 FED. SENT. REP. 387 (2012); Sara Sun Beale, *Is Now the Time for Major Sentencing Reform?*, 24 FED. SENT. REP. 382 (2012).

[200] *See* U.S. Sentencing Commission, RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES JANUARY 2010 THROUGH MARCH 2010 (June 2010), at 23 (Question 19, Table 19).

[201] *Id.*

[202] *Id.*

[203] Sessions, *At the Crossroads*, *supra* note 24; Bowman, *Nothing is Not Enough*, 24 FED. SENT. REP. 356 (June 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.").

judicial discretion. However, the empirical evidence from Part IV seeks to ascertain the effect of the sentencing regime on inter-judge 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 curtail potentially desirable judicial discretion.

While this Article does not provide evidence supporting a return to "presumptive" Guidelines, it does suggest that strictness of appellate review is a potentially important constraint on judicial discretion in sentencing. Inter-judge disparities in below range departures were generally lowest during the PROTECT Act, which imposed *de novo* review. Furthermore, empirical evidence suggests that *Booker* alone did not contribute to recent increases in inter-judge disparities. In the first two years after *Booker*, inter-judge disparities were not statistically different from that during *Koon*. Rather, it appears to be the impact of *Booker* plus reduced appellate scrutiny following *Rita*, *Gall* and *Kimbrough* that are responsible for increases in inter-judge disparities.

Thus, reforms to strengthen the degree of appellate review could possibly reduce inter-judge sentencing disparities. In *Gall*, the Court did not require appellate courts to insist upon "extraordinary" circumstances to justify a sentence outside the Guidelines recommended range, specifically rejecting stronger justifications for sentences that departed more greatly from the Guidelines.[204] In order to constrain inter-judge 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 presumption of unreasonableness for sentences outside the Guidelines range.....[which] would not be consistent with *Booker*."[205]

V CONCLUSION

Exploiting the random assignment of cases to judges in district courthouses representing 73 U.S. district courts, this Article finds a significant increase in inter-judge disparities from the *Koon* period to after *Kimbrough/Gall*. A defendant sentenced by a "harsh" judge prior to *Koon* was sentenced to 2.6 months longer than the average, but over 5 months longer after *Kimbrough/Gall*, a doubling of inter-judge disparities. Increased inter-judge disparities persist even excluding cases in which mandatory minimums were charged, suggesting that judges are not completely anchored to the Guidelines.

Increases in between-judge differences following *Booker* and *Kimbrough/Gall* appear to be linked to observable judicial demographics such as gender, political

[204] *Gall*, 552 U.S. at 47 ("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. We also reject the use of a rigid mathematical formula that uses the percentage of a departure as the standard for determining the strength of the justifications required for a specific sentence.").

[205] *Id.* at 47.

affiliation of appointing president, and whether a judge has ever sentenced under the mandatory Guidelines regime. I also find modest evidence of increases in inter-district differences following *Kimbrough/Gall*, with large inter-district differences in sentence length, rates of below range departures, rates of mandatory minimums, and rates of substantial assistance motions. However, the magnitudes of both inter-judge and inter-district disparities are drastically smaller when mandatory minimums are excluded, suggesting that prosecutorial charging decisions are 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, have led to greater inter-judge disparities. Prosecutorial charging decisions, at least in the application of mandatory minimums, appear to play a substantial role in explaining disparities. While a first step in disentangling the sources of disparities ascribable to various actors, a primary limitation of this Article is its inability to thoroughly analyze all the disparities that can arise in earlier stages of the criminal justice systems, such as through charging and plea bargaining. Nevertheless, the results of this Article caution against recent proposals to move back towards a sentencing system in which judges are bound by the decisions of prosecutors. Instead, the Article suggests that it may be wise to modify standards of appellate review, as well as 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 include: gender, age, a black race indicator, number of dependents, and an indicator for 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 likely to have completed less than a high school degree. 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.[206]

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 tests 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 acceptance of the null hypothesis, and lead to the conclusion that random assignment was present.

See Table A1 for 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 **.

[206] Testing each characteristic individually would result in incorrect standard errors if the demographic characteristics are correlated. 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 AEJ: Applied Economics 96, 106-107 (2010).

TABLE A1. RANDOMIZATION TESTS 2000-2009

District Court	No. Obs.	P-value
ME (0)	1,668	0.1438
MA (1)	4,042	0.1054
NH (2)	1,617	0.9844
PR (4)	6520	0.2674
CT** (5)	664	0.0000
NY North - Syracuse (6)	1,148	0.1074
NY East** (7)	12,447	0.0004
NY South - White Plains (8)	1,338	0.4336
NY West - Rochester (9)	1,166	0.6226
VT (10)	1,400	0.2379
DE (11)	641	0.3831
NJ -Trenton (12)	476	0.2983
PA East** (13)	6,411	0.0000
PA Middle - Scranton (14)	969	0.6837
PA Middle - Williamsport (14)	234	0.2071
PA West - Erie (15)	609	0.0521
PA West - Pittsburgh (15)	2,917	0.0645
MD (16)	5,569	0.0631
NC East - Southern (17)	608	0.3847
NC Middle (18)	3,205	0.08086
NC West**(19)	5,563	0.0000
SC** (20)	8,848	0.0000
VA East -Alexandria (22)	4,500	0.3178
VA East -Norfolk (22)	1,105	0.1658
VA East -Newport News (22)	743	0.0662
VA West (23)	3,123	0.3250
WV North - Martinsburg (24)	639	0.4091
WV South (25)	1,778	0.0932
AL North** (26)	1,430	0.0189
AL Middle (27)	904	0.3242
AL South (28)	3,132	0.0702
FL North (29)	2,718	0.5783
FL Middle - Ft. Myers (30)	923	0.3824
FL Middle - Ocala (30)	465	0.3128
FL South - Ft. Pierce (31)	3,299	0.0541
FL South - Ft. Lauderdale (31)	649	0.2485
GA North** (32)	5,823	0.0000
GA Middle (33)	2,064	0.1396
LA East (35)	3,117	0.0606
LA West (36)	1,686	.6360
MS North (37)	925	0.4247
MS South (38)	3,057	0.0564

TABLE A1. RANDOMIZATION TESTS 2000-2009 (CONTINUED)

TX North - Forth Worth (39)	2,027	0.2386
TX East (40)	6,563	0.5598
TX South - Brownsville (41)	10,112	0.3364
TX South - Corpus Christi (41)	6,679	0.2767
TX South - Laredo (41)	19,079	0.6244
TX South - McAllen (41)	12,739	0.1093
TX West - Del Rio (42)	7,098	0.3500
TX West - Midland-Odessa (42)	3,567	0.4120
KY East - Covington (43)	717	0.5872
KY East - Pikeville (43)	139	0.0966
KY East - Lexington (43)	1,993	0.8694
KY West (44)	1,746	0.1114
MI East - Bay City (45)	458	0.4009
MI East - Flint (45)	673	0.3014
MI West (46)	3,313	0.0961
OH North - Toledo (47)	1,014	0.2105
OH South - Dayton (48)	1,300	0.9115
TN East (49)	5,200	0.0705
TN Middle** (50)	1,938	0.0126
TN West - Eastern (51)	831	0.3998
IL North - Rockford (52)	624	0.8929
IL Central (53)	2,618	0.1283
IL South (54)	2,736	0.1296
IN North - South Bend (55)	954	0.2764
IN North - Fort Wayne (55)	530	0.0741
IN South (56)	2,004	0.3266
WI East - Milwaukee (57)	2,206	0.4223
WI West (58)	1,571	0.1123
AR East (60)	2,739	0.1631
AR West**(61)	1,098	0.0001
IA North (62)	2,413	0.0561
IA South (63)	2,684	0.8265
MN** (64)	4,815	0.0001
MO East (65)	8,203	0.0785
MO West (66)	6,764	0.1191
NE - Omaha (67)	2,323	0.0532
ND (68)	1,888	0.2250
SD - Aberdeen (69)	309	0.1479
SD - Pierre (69)	1,010	0.8757
AZ - Tuscon (70)	23,677	0.0961
AZ - Yuma (70)	2,449	0.3392
CA North (71)	3,045	0.1970
CA East (72)	8,094	0.0646

TABLE A1. RANDOMIZATION TESTS 2000-2009 (CONTINUED)

CA Central - Riverside (73)	157	0.4520
CA South - El Centro (74)	8,664	0.3442
CA South - Yuma (74)	89	0.3502
HI** (75)	3,351	0.0012
ID (76)	1,526	0.0544
MT - Missoula (77)	516	0.1698
MT - Great Falls (77)	1,003	0.2206
NV (78)	4,867	0.6549
OR - Eugene (79)	954	0.2261
OR - Medford (79)	434	0.6618
WA East - Spokane (80)	1,401	0.3100
WA West** (81)	5,302	0.0001
CO** (82)	4,582	0.0000
KS (83)	5,509	0.2031
NM (84)	24,019	0.2924
OK North (85)	1,279	0.3240
OK East (86)	736	0.9312
OK West** (87)	1,809	0.0001
UT (88)	5,276	0.9421
WY** (89)	1,565	0.0002
DC (90)	346	0.5720
AK (95)	1,218	0.1105
LA Middle** (96)	1,112	0.0263

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. INTER-DISTRICT VARIATION

	APPLICATION OF MANDATORY MINIMUMS			
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0571	.0491	.0663	183732
PROTECT Act	.0652	.0559	.0760	94434
Booker	.0584	.0502	.0680	165139
Kimbrough/Gall	.0662	.0568	.0771	117536

Notes: Data is from the full sample 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, and sentencing month fixed effects.

TABLE A3. INTER-DISTRICT VARIATION

APPLICATION OF SUBSTANTIAL ASSISTANCE				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0868	.0747	.1009	176736
PROTECT Act	.0909	.0782	.1058	92736
Booker	.0904	.0779	.1049	163978
Kimbrough/Gall	.0789	.0679	.0918	117328

Notes: Data is from the full sample 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, 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 month *m*. The control variables include defendant and crime characteristics (X_i), 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 by ordinary least squares (OLS) regression. OLS 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, σ^2_μ). Intuitively, within judge variance in v_{ijdtm} is used to estimate the defendant variance:

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

The variance in the judge effect is the remainder:

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

The estimated standard deviation of judge effects on sentence length is σ_μ = 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 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 inter-disparities post *Booker* and post *Kimbrough/Gall* using a specification of the form:

$$\begin{aligned} Y_{icodtm} = \beta_0 + \alpha_1 * Judge_i * Booker + \beta_0 * Booker + \alpha_2 * Judge_i * Kimbrough \\ + \beta_0 * Kimbrough + \beta_1 * Race_i + \beta_3 * \mathbf{X}_i + Guide_{ico} + Offtype_i \\ + \gamma_d + \delta_t + \gamma_d * \delta_t + \lambda_m + \epsilon_{icodtm} \end{aligned}$$

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 vs. post Guidelines appointment, tenure under the Guidelines, and an indicator for pre vs. post *Booker* appointment. The main coefficients α_1 and α_2 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 but before *Kimbrough/Gall*. $Kimbrough$ is an indicator variable for defendants sentenced after *Kimbrough/Gall*.

$Race_i$ is a dummy variable for defendant *i*'s race: white, black, Hispanic, or other. $\mathbf{X}_i$ comprises a vector of demographic characteristics of the defendant including gender, age, age squared, educational attainment (less than high school, high school graduate, some college, college graduate), number of dependents, and citizenship status.

$Guide_{ico}$ 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. $Offtype_i$ is a dummy variable for offense type.

The specification also includes district court fixed effects (γ_d), sentencing year fixed effects (δ_t), and sentencing month fixed effects (λ_m). All standard errors are clustered at the district courthouse level to account for serial correlation.