

Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform

Max M. Schanzenbach[†] & Emerson H. Tiller^{††}

INTRODUCTION

Amid widespread belief that unwarranted sentencing disparity existed in federal criminal sentencing, Congress passed the Sentencing Reform Act of 1984,¹ which established the Sentencing Commission and authorized the creation of the United States Sentencing Guidelines.² The idea of creating guidelines to limit judicial discretion in sentencing was politically popular at the time,³ and was prompted by the belief among policymakers that significant disparities existed in sentences.⁴ The reaction of the federal judiciary was quite the opposite, with many federal judges objecting to the limitation of their historic sentencing discretion. Over one hundred district courts struck down the Sentencing Reform Act on a variety of constitutional grounds.⁵ The Supreme Court, however, upheld the Act in 1989.⁶

The calm lasted barely a decade. By 2000, the Court had begun to whittle away at the constitutionality of the Guidelines, beginning with the seminal case of *Apprendi v New Jersey*,⁷ and more recently with

[†] Professor of Law, Northwestern University School of Law.

^{††} Stanford Clinton Senior Research Professor of Law, Northwestern University School of Law. The authors would like to thank Jason Friedman, Ben Schaye, and Grace Tabib for excellent research assistance. The authors also thank participants in workshops at Harvard Law School, Yale Law School, and the University of Minnesota School of Law for helpful comments.

¹ Pub L No 98-473, ch II, 98 Stat 1837, 1987, codified as amended at 18 USC § 3551 et seq (2000) and 28 USC § 991 et seq (2000).

² See Kate Stith and Steve Y. Koh, *The Politics of Sentencing Reform: The Legislative History of the Federal Sentencing Guidelines*, 28 Wake Forest L Rev 223, 257–81 (1993); William W. Wilkins, Jr., Phyllis J. Newton, and John R. Steer, *The Sentencing Reform Act of 1984: A Bold Approach to the Unwarranted Sentencing Disparity Problem*, 2 Crim Law F 355, 362–64 (1991).

³ See Stith and Koh, 28 Wake Forest L Rev at 223, 258–59 (cited in note 2) (describing the “growing public concern with crime” and the shift in framing the creation of guidelines as an anticrime measure).

⁴ Senator Edward Kennedy was a sponsor of the Sentencing Reform Act, and President Reagan enthusiastically signed the legislation. See id at 223, 266.

⁵ See Gregory C. Sisk, Michael Heise, and Andrew P. Morriss, *Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning*, 73 NYU L Rev 1377, 1403, 1435 table 5 (noting that 179 district court judges invalidated the Guidelines on separation of powers, nondelegation, and due process grounds in less than one year after passage).

⁶ See *Mistretta v United States*, 488 US 361, 412 (1989).

⁷ 530 US 466 (2000).

United States v Booker,⁸ which made the Guidelines advisory rather than mandatory (thereby preserving their constitutionality).⁹ While the effect of this status change appears to be fairly small at present, new constitutional challenges continue to mount against Guidelines sentencing as the Supreme Court's sentencing jurisprudence remains fluid. In addition, the federal judiciary and, to a similar degree, the legal academy continue to show hostility to the Sentencing Guidelines.¹⁰

This study combines positive political theory with a unique data set on judges to appraise the effectiveness of the Sentencing Guidelines system. To our knowledge, ours is the first large-scale study on federal sentencing to match identified judges with offenders. In addition, we propose two reforms that could better accomplish the goals of

⁸ 543 US 220 (2005).

⁹ See *id.* at 245 (holding that the provision of the federal sentencing statute that makes the Guidelines mandatory should be excised to preserve the constitutionality of the Guidelines as a whole); *Apprendi*, 530 US at 476 (holding that any fact that increases the statutory maximum sentence for a crime must be proven to a jury beyond a reasonable doubt).

¹⁰ The judicial hostility toward the Guidelines is evidenced not only by public statements of many judges, but also by the fact that the Guidelines appear to have made judges more likely to retire. See Richard T. Boylan, *Do the Sentencing Guidelines Influence the Retirement Decisions of Federal Judges?*, 33 J Legal Stud 231, 234 (2004). In particular, Boylan found that post-Guidelines, judges retired 4.8 months after they became eligible for senior status (essentially right away), whereas the previous average had been three years after eligibility. See *id.* at 251 (concluding that the findings are consistent with the belief that "judges care about power, not being overturned, and imposing sentences proportional to the offense"). The differences between pre- and post-Guidelines retirement decisions are quite stark. See *id.* at 245 figure 1 (showing a drastic increase in judges' propensity to take senior status beginning in 1991).

Surveys of judges generally show hostility toward the Guidelines. A survey conducted by the Federal Judicial Center in 1996 revealed that almost three-quarters of trial judges and over two-thirds of appellate judges believed that mandatory federal guidelines are unnecessary. Molly Treadway Johnson and Scott A. Gilbert, *The U.S. Sentencing Guidelines: Results of the Federal Judicial Center's 1996 Survey* 3 (Fed Jud Center 1997). It is possible that the new, post-Guidelines generation of district court judges is more amenable to them. A later survey did not repeat the question by the Federal Judicial Center but indicated that opposition might have slackened. See Michael E. O'Neill, *Surveying Article III Judges' Perspectives on the Federal Sentencing Guidelines*, 15 Fed Sent Rptr (Vera) 215, 218 (2003) (showing that approximately 78 percent of judges reported "higher" or "middle" ratings for the Guidelines' achievements in furthering the purposes of sentencing). For example, roughly 50 percent of responding judges thought that "more" of their sentences were achieving the goals of the Guidelines. *Id.* at 215.

For a judge's defense of the Guidelines, see Honorable Patti B. Saris, *Below the Radar Screens: Have the Sentencing Guidelines Eliminated Disparity? One Judge's Perspective*, 30 Suffolk U L Rev 1027, 1029-30 (1997) (arguing that the Guidelines promote a more "deliberate, fair, and rational" sentencing process than the old regime of discretion, and discussing possible disparities that remain because of substantial-assistance downward departures).

Few academics defend the Guidelines as presently constituted. But see 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, 748 (arguing that the limitation of "judicial discretion is . . . a beneficial result" of the Guidelines and that an increase in prosecutorial discretion is "no cause for alarm").

the Guidelines while maintaining much of their present structure: (1) open data and (2) ideologically mixed circuit court panels.

The importance of this study is fourfold. First, it unifies theory with the type of real judge data and empirical testing that most scholars, policymakers, and judges would accept as theoretically and empirically valid. The theory opens up the black box of sentencing procedure and produces reasonable propositions for empirical testing. The empirical test brings together judge-specific data with actual sentencing data in a manner that directly tests the propositions from the theory. The study thereby enriches our understanding of criminal sentencing by focusing on the microanalytic details of criminal sentencing procedure.

Second, the study is important in how it relates to the Supreme Court's and circuit courts' evolving sentencing jurisprudence. The effects of recent Supreme Court decisions, including *Apprendi* and *Booker*, are uncertain, and more recent decisions continue to call Guidelines sentencing into question.¹¹ Our theory and evidence yield some interesting predictions—namely, that under an “advisory standard,” the circuit courts' role in reviewing sentences will become more important.

Third, this Article illustrates why transparency in sentencing data—in particular, data on the identity of sentencing judges—could be useful in identifying sources of judicial sentencing disparity that would inform the types of reform necessary to improve consistency. The Administrative Office of the United States Courts and the Sentencing Commission have collected a detailed database on all federal criminal sentences. But while the criminal sentencing database is generally available to scholars and the public (and we use available sentencing data in this study), the Administrative Office and Sentencing Commission have adopted a policy prohibiting scholars and the public access to one of the most important pieces of the data set—the judge identifiers for each sentence. Without such data, it is costly and difficult for most scholars to identify and measure interjudge disparities on the scale that we present here. Without such openness, empirical evaluation and the reform that could follow will likely be slow, incremental, and even misguided.

Finally, this study leads us to suggest a powerful and controversial reform—the requirement of political-ideological diversity on judicial panels reviewing criminal sentences, which would ensure hierarchical political-ideological diversity between the lower sentencing court and the higher court reviewing such decisions. Specifically, all sentencing

¹¹ See, for example, *Cunningham v California*, 127 S Ct 856, 870–71 (2007) (striking down California's guidelines as incompatible with *Booker* and *Apprendi* for authorizing a judge, not a jury, to find the facts that permitted a higher sentence).

review should involve both Democratic and Republican judicial appointees. This amounts to an engineering of the political structure of the judiciary that most scholars, judges, and policymakers would quickly reject at first mention for practical and jurisprudential reasons. We argue, nonetheless, that the benefits would be substantial and that such a proposal (or one that adopts its primary tenet—recognition of the role that political-ideological diversity within the judiciary can play) deserves serious consideration.

The rest of this Article is organized as follows. Part I explores the Sentencing Guidelines and surveys the theoretical and empirical literature on them; Part II presents our empirical results and discusses their implications for both the broader literature on judicial discretion and the Sentencing Guidelines; and Part III outlines proposals for reform.

I. THE SENTENCING GUIDELINES: POSITIVE POLITICAL THEORY AND EMPIRICAL SCHOLARSHIP

The Sentencing Act charged the United States Sentencing Commission to develop sentencing guidelines that would “provide certainty and fairness in meeting the purposes of sentencing, avoiding unwarranted sentencing disparities among defendants with similar records who have been found guilty of similar criminal conduct.”¹² Pursuant to this legislative directive, the Sentencing Commission developed a sentencing table containing recommended sentencing ranges for various offense level/criminal history combinations (see Appendix Figure 1).¹³ With the aid of the probation officer, a district judge uses the Sentencing Commission’s regulations, which rival the tax code in length, to calculate the defendant’s numeric “offense level.”¹⁴ The crime of conviction sets the “base” offense level; levels are added or subtracted based on a variety of factors, such as the use of a gun, the use of sophisticated means in a fraud, whether the crime affected a financial institution, whether the offender played a major or minor role in the crime, acceptance of responsibility, and the like.¹⁵ The offender’s criminal history category is calculated based on the prior of-

¹² 28 USC § 991(b)(1)(B).

¹³ Post-*Booker*, this basic structure has survived intact but is no longer “mandatory.” See *Booker*, 543 US at 259. The post-*Booker* world is discussed in greater detail below, but because our data are pre-*Booker*, we describe the pre-*Booker* framework.

¹⁴ See USSG § 1B1.1–1.3. See also Ilya Beylin, Comment, *Booker’s Unnoticed Victim: The Importance of Providing Notice prior to Sua Sponte Non-Guidelines Sentences*, 74 U Chi L Rev 961, 967–68 (2007) (describing the probation officer’s development of the presentence report and the role it plays in the calculation of defendant’s “offense level”).

¹⁵ See USSG §§ 2A1.1–3E1.1.

fenses committed by the offender.¹⁶ These two factors—offense level and criminal history—yield a sentencing range expressed in months.¹⁷ As can be seen from Appendix Figure 1, the sentencing range counts for roughly 25 percent of the total sentence imposed.¹⁸ All of these determinations are subject to appellate review.¹⁹ Pre-*Booker*, if the offense level and criminal history were properly calculated, a sentence within the specified range created a safe harbor for the sentencing judge and could not be reversed by the higher court.²⁰ Such protection from reversal is no longer certain in the aftermath of *Booker*.²¹

The Sentencing Reform Act authorized judges to depart from the calculated sentencing range if there was 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.”²² Under the Act, a judge had to justify a departure by making a statement in open court or in a written opinion.²³ The United States could appeal a downward departure, and the defendant could appeal an upward departure.²⁴ In 1994, the Commission made clear that factors “not ordinarily” relevant could still be considered in departure decisions if they removed the case from the “heartland” of the Guidelines.²⁵ The Supreme Court endorsed the “heartland” concept explicitly in *Koon v United States*.²⁶ *Koon* also required appellate courts to review a district court judge’s departure from the Guidelines for “abuse

¹⁶ See, for example, USSG §§ 4A1.1–1.2, 4B1.1–1.5 (instructing the sentencing judge as to which prior offenses are appropriate to consider and the proper weight to be accorded to various offenses and patterns of criminal behavior).

¹⁷ See USSG § 5A sentencing table.

¹⁸ While the Sentencing Commission ultimately determined the offense levels and criminal history categories, the Sentencing Reform Act mandated that unless the minimum exceeds thirty years, sentencing ranges be no more than the greater of 25 percent of the minimum guideline range or six months. See 28 USC § 994(b)(2).

¹⁹ See 18 USC § 3742 (2000) (instituting a clearly erroneous standard of review for factual findings and an abuse of discretion standard for reviewing the district court’s application of the Guidelines to the facts). But see *Booker*, 543 US at 259 (excising § 3742(e) because this section was predicated on the mandatory nature of the Sentencing Guidelines).

²⁰ See 18 USC § 3742(a)–(b) (allowing appeal only if a sentence is imposed in violation of the law, is imposed as a result of an incorrect application of the Guidelines, is greater than the maximum or below the minimum, or is imposed for an offense for which there is no guideline).

²¹ See text accompanying note 68–71.

²² 18 USC § 3553(b). See also USSG § 5K2.0 policy statement (specifically authorizing departure when certain Guidelines-identified circumstances are present that were not adequately taken into consideration by the Commission and limiting departure based on other circumstances to exceptional cases).

²³ 18 USC § 3553(c).

²⁴ See 18 USC § 3742(a)(3), (b)(3).

²⁵ USSG § 5K2.0 policy statement.

²⁶ 518 US 81, 98–110 (1996).

of discretion,” a substantially less deferential standard than the “clearly erroneous” standard used for reviewing offense-level adjustments.²⁷ In effect, a lower court was at greater risk of reversal if it used the more extreme departure mechanism to lengthen or shorten the sentence imposed on a defendant. After *Booker*, the standard of review for all sentences is “whether the sentence ‘is unreasonable’ with regard to [the Guidelines],”²⁸ though, as discussed below, it appears that departures continue to receive greater scrutiny.

Policymakers expected that the Guidelines, with their specific and mandatory formulas for calculating prison terms, would harmonize judicial practices and eliminate unwarranted interjudge disparities in sentencing. Empirical research since the promulgation of the Guidelines suggests, however, that considerable unwarranted disparity in federal sentencing remains. Some studies have focused on sentencing disparity related to characteristics of the criminal defendants.²⁹ Others have found that characteristics of judges, such as race, gender, and political affiliation, also affect sentences.³⁰ Those studies, however,

²⁷ Id at 97–100.

²⁸ See *Booker*, 543 US at 261.

²⁹ See, for example, David B. Mustard, *Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts*, 44 J L & Econ 285, 311–12 (2001) (finding that black and male offenders with lower levels of education and income receive longer sentences, primarily as a result of departures); Celesta A. Albonetti, *Sentencing under the Federal Sentencing Guidelines: Effects of Defendant Characteristics, Guilty Pleas, and Departures on Sentence Outcomes for Drug Offenses, 1991–1992*, 31 L & Socy Rev 789, 817 (1997) (finding disparity in sentencing related both to offense characteristics and to factors such as the defendant’s ethnicity, gender, education level, and noncitizenship); Douglas C. McDonald and Kenneth E. Carlson, *Sentencing in the Federal Courts: Does Race Matter? The Transition to Sentencing Guidelines, 1986–90* 178, 194 (DOJ 1993) (finding that racial disparity in sentences continues to exist largely as a result of legitimate factors, such as differing rates of gun use in robberies and the crack/powdered cocaine sentencing disparity). Mustard in particular finds that no matter how one cuts the data (looking at only within-range sentences, departures, drug crimes, and so forth), substantial unexplained differentials exist between the races and sexes. Substantial disparities exist even within the context of nonviolent crimes. See generally Max M. Schanzenbach and Michael L. Yaeger, *Prison Time, Fines and Federal White-collar Criminals: The Anatomy of a Racial Disparity*, 96 J Crim L & Criminol 757 (2006) (finding significant racial disparities in Guidelines sentences for white-collar crimes).

³⁰ A considerable amount of research suggests that judges have different sentencing philosophies. See, for example, John S. Carroll, et al., *Sentencing Goals, Casual Attributions, Ideology, and Personality*, 52 J Personality & Soc Psych 107, 110–17 (1987) (demonstrating how an individual’s ideology is reflected in how he or she thinks about the causes of crime and the goals of sentencing); Brian Forst and Charles Wellford, *Punishment and Sentencing: Developing Sentencing Guidelines Empirically from Principles of Punishment*, 33 Rutgers L Rev 799, 808–11 (1981) (finding that judges oriented towards utilitarian goals of incapacitation and deterrence gave sentences that were on average at least ten months longer than judges motivated by other goals); Shari S. Diamond and Hans Zeisel, *Sentencing Councils: A Study of Sentence Disparity and Its Reduction*, 43 U Chi L Rev 109, 114 (1975) (“[I]t is reasonable to infer that the judges’ differing sentencing philosophies are a primary cause of the disparity.”); Anthony Partridge and William B. Eldridge, *The Second Circuit Sentencing Study: A Report to the Judges of the Second Circuit* 36–40

were greatly hampered by the difficulty in identifying judges for the vast number of criminal sentencing cases. Studies of the effect of judicial characteristics on sentencing have relied either on state samples³¹ (raising issues of generalizability to federal sentencing) or nonspecific (that is, not judge-identified) variation in judicial characteristics aggregated at the federal district level³² (raising issues of accuracy). A few studies have addressed judicial disparity using coded judge identifiers provided by the Commission,³³ and one study employed a small sample of judges in two districts.³⁴ This permitted the study of inter-

(Jud Center 1974) (finding differences among judges in the hypothetical sentences they would impose on identical offenders). Paul J. Hofer, Kevin R. Blackwell, and R. Barry Ruback claim that:

“[L]iberals” tend to believe that factors external to the offender are responsible for criminal behavior. Rehabilitation is more of a sentencing goal for these judges, leading to greater reliance on probation and less concern with retribution. “Conservatives” believe that offenders choose to commit crimes. They are more punishment-oriented and tend to impose longer prison terms.

Paul J. Hofer, Kevin R. Blackwell, and R. Barry Ruback, *The Effect of the Federal Sentencing Guidelines on Inter-judge Sentencing Disparity*, 90 J Crim L & Criminol 239, 250 (1999).

³¹ See, for example, Darrell Steffensmeier and Chris Hebert, *Women and Men Policymakers: Does the Judge's Gender Affect the Sentencing of Criminal Defendants?*, 77 Soc Forces 1163 (1998). In this study of Pennsylvania criminal cases, the authors found that female judges tended to sentence offenders more severely than male judges, were more likely to incarcerate minorities, and were less likely to incarcerate women. Three studies on the effect of a judge's race on sentencing have found mixed results. However, they involved small sample sizes and examined state court judges in one city. *Id.* at 1181–86. Susan Welch, Michael Combs, and John Gruhl concluded, based on a northeastern city's criminal court outcomes, that black judges' incarceration decisions were moderately different from their white colleagues' decisions. Susan Welch, Michael Combs, and John Gruhl, *Do Black Judges Make a Difference?*, 32 Am J Polit Sci 126, 134 (1988). Two earlier studies concluded that both white and black judges treated black defendants more harshly. See Thomas M. Uhlman, *Racial Justice: Black Judges and Defendants in an Urban Trial Court* 63–72 (1967); Thomas M. Uhlman, *Black Elite Decision Making: The Case of Trial Judges*, 22 Am J Polit Sci 884, 889–91 (1978). These district- and city-level studies involved only a few minority judges. It would be hard to conclude from these studies that there are no judge race effects whatsoever.

³² See, for example, Max M. Schanzenbach, *Racial and Sex Disparities in Prison Sentences: The Effect of District-level Judicial Demographics*, 34 J Legal Stud 57, 85–90 (2005). Relying on variation at the district level, Schanzenbach found some evidence that minority and female judges sentence differently.

³³ See, for example, Hofer, Blackwell, and Ruback, 90 J Crim L & Criminol at 287–89 (cited in note 30) (concluding that the Guidelines slightly decreased interjudge sentence disparities); James M. Anderson, Jeffrey R. Kling, and Kate Stith, *Measuring Interjudge Sentencing Disparity: Before and after the Federal Sentencing Guidelines*, 42 J L & Econ 271, 298–99, 303–04 (1999) (finding a decrease in interjudge disparities in sentence length after the Guidelines, but cautioning that the advent of mandatory minimum sentences might have contributed to the decline). Anderson, Kling, and Stith had coded judge identifiers, so they could look at interjudge disparity, but could not study judicial characteristics. See *id.* at 287.

³⁴ Chantale LaCasse and A. Abigail Payne measured changes in interjudge disparity by examining whether plea bargain strategies changed after the Guidelines. Chantale LaCasse and A. Abigail Payne, *Federal Sentencing Guidelines and Mandatory Minimum Sentences: Do Defendants Bargain in the Shadow of the Judge?*, 42 J L & Econ 245, 247–61 (1999) (using a data set from the Southern and Eastern Districts of New York to analyze interjudge disparity). The au-

judge variation in sentencing, but not variation based on judge characteristics such as political party or background. Judges were not identified by name, so background characteristics of judges (such as political orientation) could not be used to study sentencing behavior.

While the existing studies tell us that “judges matter,” few measure how judicial biases are translated through the highly structured sentencing framework. Part of that neglect can be attributed to the general absence of theory regarding judicial behavior under the Sentencing Guidelines. There are fairly straightforward conjectures from political science and other fields that could be tested, such as whether a judge’s political orientation, race, or economic standing influence her sentencing outcomes. However, these propositions fail to incorporate the complexity of the Sentencing Guidelines or account for the strategic behavior of judges in accomplishing their sentencing goals.

Theory is a necessary guide here. We need to think critically about how the institutional structures built into the Guidelines—such as offense-level adjustments and departures—could be used strategically by judges to attain sentencing outcomes closer to their personal preferences instead of those outcomes intended by statute and the Guidelines. Indeed, given the complex formulas for determining criminal offense levels and the associated presumptive sentencing ranges from which judges determine prison sentences, empirical tests of judicial behavior require more nuanced construction to capture the limitations on and opportunities for judges created by such institutional complexity.

Recent progress on theory—in particular, positive political theory—has made more sophisticated and nuanced empirical scholarship regarding judicial behavior possible. In a recent article, we considered how judicial characteristics of judges interacted with the Guidelines’ structure and the political diversity within the judicial hierarchy (that is, lower court political alignment with higher courts).³⁵ To our knowledge, this was the first study to examine how the complexity of the Sentencing Guidelines affects the strategies judges employ to maximize their sentencing goals. Lacking judge-specific data, that empirical investigation instead relied on generalized statistical probabilities that a judge of a particular political persuasion would have heard a given case based on the proportion of Democratic versus Republican ap-

thors concluded that judge-specific effects on plea bargains actually increased post-Guidelines, implying that prosecutors and defense attorneys believed that judges mattered as much after the Guidelines as before. See *id.* at 267–68.

³⁵ See Max M. Schanzenbach and Emerson H. Tiller, *Strategic Judging under the U.S. Sentencing Guidelines: Positive Political Theory and Evidence*, 23 *J. L., Econ., & Org.* 24, 26 (2006) (examining the use of departures versus the use of offense-level adjustments in light of the political alignment between the sentencing judge and the reviewing court).

pointees sitting in a given district. We thus estimated the likelihood that a particular sentencing outcome was decided by a Democratic or Republican judicial appointee.³⁶ Under the assumption that cases were randomly assigned and that each judge sentenced approximately the same number of offenders, this identification strategy permitted us to estimate the effect on sentencing of increasing the probability that an offender was sentenced by a Democratic or Republican appointee.

We found that sentences for serious crimes in districts with more Democratic appointees were lower on average than in districts with more Republican appointees.³⁷ More importantly, we found that the politics of the circuit court was important in sentencing departures by lower court judges, but did not influence offense-level adjustments by those judges. For example, Democratic- and Republican-appointed district court judges used offense levels to adjust prison sentences to roughly the same degree (although in opposite directions) whether they were in majority Democratic- or Republican-appointed circuits. On the other hand, Democratic appointees in majority Democratic circuits departed to a degree greater than Democratic appointees in majority Republican circuits.³⁸ This result was in line with our theoretical prediction: because departures are reviewed more stringently than fact-oriented adjustments, the political alignment of the reviewing court was more important for departures.

That study, however, was limited by the lack of judge-level data to match sentencing data from the United States Sentencing Commission. The identification strategy, which relied on aggregated information on the general variation of political ideology at the district level, is open to criticism on several fronts. The first problem with aggregating information without knowing specific judge identifiers for particular sentences is that the district-level case assignment rules (perhaps imposed by the chief judge) may change as the district's ideology changes, or judges may use their colleague's sentences as a reference

³⁶ Id at 35–36. We used the political variation within the district to measure the impact of political ideology on sentencing. We let %*DEMOCRAT* be the percentage of active judges appointed by a Democratic president on the relevant district bench for the year of the observed sentence. Id at 36. The higher this percentage, the greater the chance that an individual offender is sentenced by a Democratic-appointed judge. Id. Provided that assignment is random, the coefficient on this measure is the same as comparing an all-Democratic-appointed bench to an all-Republican-appointed one.

³⁷ See id at 39 table 1, 40.

³⁸ Id at 52–53. Interestingly, Republican-appointed district judges did not depart upward more in majority-Republican circuits. Id at 49–50. We suspect several reasons for this, among them (1) the general high sentences under the Guidelines; (2) the ability to use adjustments to exponentially increase the sentences; and (3) the near-certainty of an appeal of an upward departure.

point. As such, our previous identification strategy captured some district-level effects apart from changes in individual judge ideologies.

A second criticism is that the aggregated identification strategy is bound to be “noisy” and this would tend to bias estimated effects toward zero (that is, make the results appear weaker than they actually are). For example, a newly appointed judge may not impose sentences for several months after confirmation. If judges become more efficient over time, they may also sentence more offenders. How to lag the change without knowing the individual judges who issued the sentences is therefore unclear. In addition, the strategy is only accurate if Democratic and Republican appointees sentence at similar rates. If Democratic appointees consistently sentence fewer individuals (perhaps because they have more time taken up by trials, or have higher acquittal rates), then the true district variation will be less substantial than the relative numbers of Democratic and Republican appointees indicate. Furthermore, local rules may mean that offenders are not randomly assigned to judges within a district but instead within a district division.

Finally, most of the variation was in one direction—that is, in the time frame of our previous study, most of the variation occurred as districts became more Democratic and less Republican as President Clinton replaced Bush and Reagan appointees. Contemporaneous time trends in sentencing could bias estimates of partisan effects in either direction. Trends could occur, for example, because we do not adequately control for changes in Guidelines policy toward certain crimes. We found some evidence, for example, that Democratic appointees were harsher on white-collar crime. However, this result did not survive when we included time trends for white-collar crimes. If we could actually match judges to the data, we could make direct comparisons between Republican and Democratic appointees over time.

The next Part summarizes a positive theory of sentencing under the Guidelines and presents estimates of differences in sentencing using a data set with judge-specific identities matched to sentenced offenders. This approach offers some fresh insights into the political dynamics of criminal sentencing.

II. EMPIRICAL STUDY WITH JUDGE-LEVEL DATA

A. Sentencing Framework and Theoretical Predictions

Positive political theories of judging suggest that much of the policy discretion exercised by judges is guided by the judges’ policy preferences, constrained by the prospect of higher court review, and accomplished through a variety of legal decision instruments available to judges when deciding cases. Judges are modeled as strategic policymakers who routinely manipulate doctrines, procedures, and other

decision instruments to advance their preferred policies when faced with review by higher courts, which may have competing policy preferences. Positive political theory suggests that lower court judges choose between fact- and law-based explanations for their decisions depending on the ideological alignment between the lower court and the supervising higher court. More specifically, because higher courts are more likely to review broad-ranging legal conclusions from lower courts than highly specific fact findings of lower courts, a lower court may emphasize factual findings in its decision when not politically aligned with a higher court (thereby ensuring minimal review by the higher court) and legal conclusions when the higher court is politically aligned (thereby enjoying the broad policy discretion associated with having likeminded higher court judges as reviewers).³⁹

The theory described above maps onto the federal criminal sentencing framework remarkably well. First, most scholars and observers agree that political-ideological preferences are at play when judges sentence criminals.⁴⁰ The conventional wisdom suggests that liberals (Democratic appointees) prefer more lenient sentences than do conservatives (Republican appointees) for “serious crimes” (violent, theft, and drug crimes).⁴¹

Second, the Sentencing Guidelines provide fact- and law-based instruments to judges for determining a defendant’s sentence—instruments that can be manipulated in setting sentence lengths. As mentioned above, these fact-oriented determinations relate to aggravating and mitigating factors set out in the Sentencing Guidelines. If these factors are found to exist, the sentencing judge can make upward or downward *adjustments* to the base offense level. While the base offense

³⁹ See Joseph L. Smith and Emerson H. Tiller, *The Strategy of Judging: Evidence from Administrative Law*, 31 J Legal Stud 61, 67, 81–82 (2002) (finding that federal appellate judges strategically employ fact- or law-based reasoning consistent with positive political theory when deciding administrative law cases); Emerson H. Tiller and Pablo T. Spiller, *Strategic Instruments: Legal Structure and Political Games in Administrative Law*, 15 J L, Econ, & Org 349, 362 (1999) (“[L]ower court judges are [] given toward protecting their decisions from higher court review through the strategic selection of . . . the legal grounds upon which they make their decisions.”).

⁴⁰ See, for example, Hofer, Blackwell, and Ruback, 90 J Crim L & Criminol at 250 (cited in note 30).

⁴¹ See, for example, *id.* (“[P]hilosophical differences among judges accounted for a significant portion of the differences in sentences imposed on offenders in the federal courts prior to enactment of the sentencing guidelines.”). We use the phrase “serious crimes” somewhat loosely. The crimes at issue here are federal crimes, so our criminals are not typical. Most of the crimes have interstate characteristics. In the time frame of the sample, 43 percent of those sentenced under the federal Guidelines were sentenced for drug trafficking, over 14 percent were sentenced for fraud, and 8.5 percent for immigration offenses. The crimes here are also federal in nature. For example, over 90 percent of the violent crimes in the sample are armed bank robbery and 96 percent of the “drug crimes” in the sample are for trafficking (less than 3 percent are for possession).

level is set by the crime of conviction⁴² (a determination made prior to and separate from the sentencing hearing), the Guidelines direct the sentencing judge in the postconviction proceedings to make modifications to the base offense-level calculation if the judge finds that certain adjustments or “specific offense characteristics” listed in the Guidelines—essentially, facts constituting aggravating and mitigating circumstances—are present in the case.⁴³ These adjustments produce a final offense level that results in a new sentencing range from which the judge can choose the precise prison term for the defendant. These fact-oriented adjustments are generally reviewed by the circuit court for clear error,⁴⁴ a legal standard that gives substantial discretion to the sentencing judge’s conclusion.⁴⁵ In other words, any sentence within the adjusted (or “final”) offense level would be relatively free from reversal.

Because adjustments are tied to factual circumstances listed in the Guidelines, there is a natural limit to the number of adjustments that can be made and, therefore, limits to the change in the length of sentence from the base level. To the extent that adjustments are insufficient to maximize the sentencing judge’s preferences, judges may choose the more dramatic law-based alternative to lengthen or shorten the presumptive sentence—a *departure* from the Guidelines’ sentencing range altogether.⁴⁶ In order to depart, a judge must find that as a matter of law the circumstances are so unusual that the case

⁴² As discussed in greater detail below, base offense levels in drug crimes are not set entirely by the crime of conviction, but also by the quantity of drugs involved.

⁴³ See USSG §§ 2A1.1–3E1.1. The sentencing judge uses the preponderance of the evidence standard to make these determinations, a standard considerably below the guilt phase standard of beyond a reasonable doubt. For a discussion of the scope of the beyond a reasonable doubt standard, see Jonathan F. Mitchell, Apprendi’s *Domain*, 2006 S Ct Rev 297, 305 (noting that the Supreme Court has limited the constitutional proof-beyond-a-reasonable-doubt requirement to facts characterized as “elements,” those facts that the prosecution must establish to secure a conviction, as opposed to facts that “affect only a defendant’s sentence”).

⁴⁴ See, for example, *United States v Rodriguez*, 278 F3d 486, 493 (5th Cir 2002) (holding that a district court’s valuation of funds for the purposes of making an adjustment to the offense level in a money laundering prosecution based on specific offense characteristics is a determination of fact reviewed for clear error).

⁴⁵ One Fourth Circuit judge characterized it this way: “The clear error standard is not concerned with the certainty of an appellate court regarding its own view of the facts. Where there are two permissible views of the evidence, the [sentencing judge’s] choice between them *cannot be clearly erroneous.*” *United States v Riggs*, 370 F3d 382, 391 (4th Cir 2004) (Duncan dissenting) (quotation marks and citations omitted), vacated and remanded as *Riggs v United States*, 543 US 1110 (2005).

⁴⁶ See 18 USC § 3553(b) (outlining grounds for imposing a sentence outside the Guidelines range). See also USSG § 5K2.0 policy statement (detailing further grounds for departures). In 1994, the Sentencing Commission adopted the position that factors “not ordinarily” relevant can still be considered if they remove the case from the “heartland” of the Guidelines. See *id.* The Supreme Court subsequently endorsed the “heartland” departures concept in *Koon*, 518 US at 98–110. The Court also held that departures from the Guidelines should be reviewed by circuit courts for abuse of discretion. See *id.* at 97–100.

lies outside the “heartland” of the Guidelines.⁴⁷ This determination requires significant legal conclusions about the reach of the Sentencing Guidelines in addition to factual findings. These law-oriented departures allow the district court judge to make significant enhancements or reductions to the calculated sentence because the judge is not limited to the more rigid boundaries associated with sentencing ranges determined by adjusted offense levels. The cost of such discretionary power is that a departure invites greater scrutiny from circuit courts because (1) the circuit courts are especially interested in determinations of law as they may set a legal precedent for future cases, and (2) the prevailing review doctrine (pre-*Booker*)—abuse of discretion—calls for a close review of the lower court’s law-oriented conclusions.⁴⁸ In addition, the Guidelines themselves strongly discourage departures on most grounds, and the legal standard for a departure is high—the judge must find that the case is sufficiently unusual to be removed from the “heartland” of the Guidelines. In short, the reversal risks for the sentencing judge increase if she uses a departure.

An application of positive political theory to federal criminal sentencing suggests that when the lower and higher courts have similar sentencing preferences, the sentencing judge has the ability to use both adjustments and departures in a cumulative manner to set the defendant’s sentence to the term most preferred by the sentencing judge. When the lower and higher courts are not so aligned, however, the risk of reversal increases, especially for departure (law-based) determinations by the lower court judge. Consequently, one would predict the district court judge to rely less on departures to maximize sentencing preferences under these conditions.

The empirical implications, thus, are as follows: (1) policy preferences matter in sentencing—liberal (Democratic-appointed) judges give different (generally lower) sentences than conservative (Republican-appointed) judges for certain categories of crime; (2) the length of the sentence given by sentencing judges depends on the amount of political-ideological alignment between the sentencing judge and the circuit court; and (3) sentencing judges selectively use adjustments and departures to enhance or reduce sentences, and the use of departures is influenced by the degree of political alignment between the sentencing judge and the overseeing circuit court, while the use of adjustments is not so influenced.

⁴⁷ USSG § 5K2.0 policy statement.

⁴⁸ See *Koon*, 518 US at 97–100.

B. Data Sources

Until recent developments in electronic recordkeeping by courts, access to sentencing decisions posed some challenges. Most decisions were not published in reporters; the only way to get the information was to make a trip to the courthouse and collect data by hand. During the 1990s, the Administrative Office created Public Access to Court Electronic Records (PACER) in order to make court documents easily accessible to parties and the general public. District participation varied somewhat over the first few years, but by the late 1990s almost all districts participated to some degree. PACER is not standardized across districts, and the types of documents that are available on PACER vary widely. Some districts merely report the docket sheets, while others include links to motions, plea agreements, and sentencing opinions. This information is not generally available on other electronic legal databases such as Westlaw or Lexis.

The Sentencing Reform Act requires the Commission to keep data on federal criminal sentencing and the Commission has generally provided substantial information on sentences to the public. From its inception, however, the Commission has refused to identify the sentencing judge in the data, and Congress recently balked at forcing the Commission to do so. Congress did require the Commission to release judge data to Congress when requested,⁴⁹ and even this raised a significant uproar among the judiciary.⁵⁰

The unavailability of judge data is one of the most frustrating aspects of the study of federal sentencing and has significantly impeded scholarly evaluation of the Guidelines' efficacy. The question of whether the Guidelines are serving their intended function of reducing dispar-

⁴⁹ See The Prosecutorial Remedies and Other Tools to End the Exploitation of Children Today Act of 2003 (PROTECT Act) requires the chief judge of every district to collect information on sentencing, including the stated reasons for downward departures. See Pub L No 108-21 § 401(h), 117 Stat 650, 672 (2003). This information must be made available to Congress or the Justice Department when requested. The Attorney General is required to make a report of every departure it opposes to the Judiciary Committees of both houses, including the name of the sentencing judge and whether an appeal will be made. PROTECT Act § 401(l)(2), 117 Stat at 674-75.

⁵⁰ Chief Justice William H. Rehnquist publicly expressed his dismay at the reporting requirements, worrying that the requirements "could appear to be an unwarranted and ill-considered effort to intimidate individual judges in the performance of their judicial duties." William H. Rehnquist, *2003 Year-end Report on the Federal Judiciary* (2003), online at <http://www.supremecourtus.gov/publicinfo/year-end/2003year-endreport.html> (visited Apr 16, 2008). See also Letter from William H. Rehnquist, S 151, 108th Cong, 1st Sess (Apr 3, 2003), in 149 Cong Rec S 5120 (Apr 10, 2003) (opposing "any amendment to 28 USC § 994(w) that would impose specific record keeping and reporting requirements on federal courts in all criminal cases or that would require the Sentencing Commission to disclose confidential court records to the Judiciary Committees on request"); Letter from Leonidas R. Mechem, S 151, 108th Cong, 1st Sess (Apr 2, 2003), in 149 Cong Rec S 5120-21 (Apr 10, 2003) (cautioning Congress to not subject judges to the "risk of unfair criticism" by requiring disclosure of sentencing records).

ity in sentencing requires the study of how judges behave under the Guidelines. For this purpose, the identification of judge characteristics is important for several reasons. First, sentences provide a unique and easily quantifiable measure of judicial decisionmaking. Instead of merely coding a “win” or a “loss” for a “conservative” or “liberal” position, studies of sentencing can measure the number of months that a judge imposes on an offender. Second, the extent to which judicial ideology or background influences sentencing may provide evidence of the sources of sentencing disparities. Third, the alignment of preferences between the circuit and district courts is important under the Sentencing Guidelines. For example, if disparities are largely driven by political differences, then attempts to cabin judges by providing rigid guidelines will only be successful if circuit courts can discipline the district judges to abide by them. Circuit courts will have an incentive to do so only to the degree that the sentencing preferences of the district court differ from those of the circuit court.

C. Matching and Reliability of the Data

The docket sheets from individual cases (which we accessed through PACER) provide most of the data we need to match cases (with the judge’s identity known) to the master data from the Sentencing Commission. Docket sheets almost always report three items of interest: (1) the date of the sentence; (2) the length of the sentence; and (3) the sentencing judge. In addition, the docket sheets often record the amount of fine imposed and the general category of crime (for example, fraud or drug trafficking). The Guidelines data include the date, district, and length of sentence as well as the amount of fine and the category of crime. In the vast majority of districts, only a few sentences are imposed every day. Even in very large districts, there are only a few sentences within a given crime category per day, so the docket sheet information enables us to match data—including the identity of the sentencing judge—in a fairly straightforward manner. The matches were performed by matching date and length of sentencing, and, when multiple observations or differing sentences were encountered, we relied on type of offense and amount of fine. The offender’s Hispanic ethnicity, when identifiable from the docket sheet, was also used for identification and verification purposes. Cases were matched from eighty-two district courts out of the ninety-four districts in the country. For the most part, districts left out of the sample were those that adopted the PACER system late or did not regularly include docket sheets in their electronic files.

Searching the PACER records was not a straightforward process. PACER does not permit a search by anything other than party name and filing date under two broad categories: civil and criminal. We searched

twenty randomly chosen dates from the 1999–2000, 2000–2001, and 2001–2002 judicial terms under the “criminal” category and put in “U.S.,” “United States,” “U.S.A.,” and “United States of America” as the party name (the party name is a required field). Records returned were often dismissals, transfers, or motions that would not have resulted in a sentence. A number were also proceedings before magistrate judges who, as a general matter, do not impose sentences. We eliminated these cases, for a total of 2,815 “matchable” cases, of which 2,265 were successfully matched to sentences reported in the Guidelines data. Only 3 percent of the total matchable cases were not matched due to multiple similar sentences that could not be narrowed down by using fines or offender characteristics, and these were generally immigration cases. Another 14 percent were not matched because no sentence and no offender data reported in the master Sentencing Commission data looked similar to information from the docket. This is likely the result either of cases that were not reported to the Sentencing Commission or of errors in the docket sheet or Guidelines data that prevented a successful match. In our view, this represents a fairly high rate of missing data. As far as we know, this is the only attempt to verify the Commission’s data.

We coded whether the judge was appointed by a Republican president or a Democratic president. We failed to find any significant partisan effects for white-collar crimes.⁵¹ We therefore limited our sample to serious crimes, the majority of which were drug trafficking offenses.⁵² The table below presents summary statistics on relevant variables in both the matched and the master data.

⁵¹ This may be because there are no partisan differences and trends toward harsher white-collar sentencing simply biased our results. The weak white-collar results in the prison sentence regressions may also be due to smaller sample sizes and lower jail sentences for white-collar crimes (making it harder to detect any differences). For example, the average sentence for a white-collar crime was just over 9 months, with an average base offense level of 6.4. In contrast, the average sentence for drug crimes was 70.2 months, with an average base offense level of 28.46. It is particularly hard to detect changes in prison terms in the case of crimes with very low base offense levels because changes to these levels cause little change to the actual number of months in the sentencing range. For example, a decrease of two levels for a drug crime with a base offense level of twenty-eight reduces the minimum sentence by fifteen months. By comparison, in the case of the average white-collar crime with a base level of six, a change of two levels (up or down) does not change the actual sentencing range at all. We therefore draw no conclusions concerning the presence or absence of a partisan effect on white-collar crime.

⁵² The “serious offenses” (number of offenses in parentheses) were: murder (1); manslaughter (2); kidnapping/hostage taking (1); sexual abuse (7); assault (11); bank robbery (83); arson (3); drug trafficking (899); drugs: communication facilities (18); drugs: simple possession (9); firearms (215); and auto theft (16).

TABLE 1
VARIABLES OF INTEREST (SERIOUS CRIMES ONLY)

	Matched data	Master data
Offender variables		
Total prison (months)*	70.8 (70.86)	66.5 (70.3)
Final offense level	23.0 (7.61)	22.8 (8.04)
Drug trafficking	73%	75%
Age	32.3 (16.4)	32.5 (9.92)
Female	10.1%	11.6%
Black	41.7%	33.5%
Hispanic	27.8%	36.2%
Circuit breakdown		
First Circuit	1.7%	3.1%
Second Circuit	7.9%	7.3%
Third Circuit	6.1%	4.4%
Fourth Circuit	9.6%	10.3%
Fifth Circuit	17.1%	19.6%
Sixth Circuit	8.5%	7.6%
Seventh Circuit	3.9%	4.5%
Eighth Circuit	11.5%	7.6%
Ninth Circuit	10.3%	17.7%
Tenth Circuit	3.6%	5.7%
Eleventh Circuit	18.5%	11.4%
DC Circuit	1.0%	0.6%
Sample size	1,265	~63,000**

* only prison sentences between 0 and 989 months included.

** sample sizes vary slightly based on missing values.

Our sampling procedure requires us to rely on randomly chosen filing dates. There may be some concern that sampling based on dates will not yield a random sample. However, we sampled *filing* dates, not sentencing dates. Therefore, there is less concern about certain types of sentences or difficult cases being decided late or early in the week. As can be seen, the comparison of sample means and percentages in Table 1 strongly indicates that the sample was drawn randomly. One anomaly worth mentioning is the disparity between proportions of black and Hispanic convicts between the Sentencing Commission data set and our sample. This disparity appears to be largely driven by the fact that the Eleventh Circuit was oversampled relative to the Ninth; if

not for the oversampling, the Ninth Circuit would have included relatively more Hispanics. We suspect this difference was due to different rates of PACER adoption and completeness of records between circuits.

The most important measure of randomness is whether sentencing judges were randomly sampled. During the three judicial terms of the sample, 50 percent of active district court judges were appointed by Democratic presidents, and 49 percent of offenders in the matched data were sentenced by Democratic appointees. There were 389 judges in the sample (the federal district bench had around 1,000 active judges during each year of the sample).⁵³ Therefore, we have roughly one-third of the judiciary represented in the sample. In sum, we have a high degree of confidence that the data are representative and randomly drawn.

D. The Results

Tables 2 and 3 below present a basic summary of the empirical results. The coefficients in Table 2 are the average sentence or offense level of Democratic-appointed judges *relative* to Republican-appointed judges. Table 3 divides district court Democratic-appointed judges into “aligned” (circuit majority Democratic) and “unaligned” (circuit majority Republican) categories. The Appendix contains a detailed discussion of the results and also presents a more flexible specification for circuit level politics. In general, we remove from the analysis substantial-assistance departures, which are given for cooperation and generally controlled by the prosecution.

The outcome variables of interest are *change in offense level* (final offense level minus base offense level), *prison sentence*, *change in prison sentence from base offense level*, and *change in prison sentence from final offense level*. We run a very basic regression in the first instance, with no control variables except a dummy variable that equals one if the district judge was appointed by a Democrat. This is identical to testing the difference in means. We next include a set of controls including basic demographic variables, grid controls for criminal history and base offense level, and circuit dummies. (See the Appendix for a fuller explanation.) Thus, we can make comparisons between a raw difference between Democratic and Republican appointees (the first column) and the difference after we control for a variety of characteristics. The Appendix reports further specifications, which in general suggest that the results are surprisingly robust.

⁵³ Some judges were observed multiple times, which is inevitable given that many cases have codefendants. As a result, we cluster the data by judge last name.

TABLE 2
DEMOCRATIC JUDGE EFFECTS

	Change in offense level (1)	Change in offense level (2)	Prison sentence (3)	Prison sentence (4)	Base change (5)	Base change (6)	Final change (7)	Final change (8)
Democrat	-0.72* (0.34)	-0.31 (0.21)	-6.54 (6.40)	-6.65* (3.12)	-8.07+ (4.49)	-7.22* (3.40)	-6.88** (2.44)	-6.02** (2.27)
Offender controls	No	Yes	No	Yes	No	Yes	No	Yes
Circuit dummies	No	Yes	No	Yes	No	Yes	No	Yes
Grid controls	No	Yes	No	Yes	No	Yes	No	Yes
Sample	All	All	No Sub Asst	No Sub Asst	Non- departures	Non- departures	No Sub Asst	No Sub Asst
R-squared	0.01	0.33	0.01	0.63	0.01	0.33	0.01	0.21
Observations	1,241	1,225	936	889	762	732	936	887

Regressions are ordinary least squares. Standard errors are robust and reflect clustering by judge.
+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

TABLE 3
DEMOCRATIC JUDGE AND CIRCUIT EFFECTS

	Change in offense level (1)	Prison sentence (2)	Base change (3)	Final change (4)	Downward departure (5)
Aligned Democrat	-0.44 (0.31)	-7.76 (5.29)	-5.49 (6.19)	-8.50** (3.02)	0.089* (0.044)
Unaligned Democrat	-0.20 (0.28)	-5.75 (3.16)	-8.32** (3.88)	-4.03+ (2.65)	-0.022 (0.033)
Offender controls	Yes	Yes	Yes	Yes	Yes
Circuit dummies	Yes	Yes	Yes	Yes	Yes
Grid controls	Yes	Yes	Yes	Yes	Yes
Cases	All	No Sub Asst	Non- departures	No Sub Asst	No Sub Asst
R-squared	0.33	0.63	0.27	0.21	N/A
Observations	1,225	889	732	887	899

Standard errors are robust and reflect clustering by judge.

+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

The “downward departure” column is a probit model in which the dependent variable equals one in the event of a judge-induced downward departure. Marginal effects reported (so the probability of receiving a downward departure is almost 9 percentage points greater when a Democrat is in a Democratic circuit versus a Republican or unaligned Democrat).

Table 2 reveals consistent differences between Democratic and Republican appointees. The first two columns examine the change in offense level from the base offense level. The raw difference between Democratic and Republican appointees is on average -0.72 offense levels and is significant at the 5 percent level. When controls are added in Column (2), this difference drops to -0.31 and is not statistically significant. While these differences may not appear to be large at first, they can translate into large differences in prison sentences. For example, at an offense level of twenty-three (our sample average), reducing the offense level by one unit reduces the minimum prison sentence by five months, and similarly the maximum sentence can be increased by five months by increasing the offense level by one unit. At higher offense levels, the difference made by a single adjustment can be as much as two years. The effect of partisan affiliation on average prison sentences reported in Columns (3) and (4) suggests roughly 6.5-month-lower sentences on average if the offender is sentenced by a Democratic instead of a Republican appointee. In light of the differences in offense-level calculations, this appears to be a reasonable estimate. The coefficient suggests a nearly 10 percent difference given the average sentence of seventy months.

Columns (5) and (6) test partisan differences in the “base change,” or the change in prison sentences from the lowest possible sentence given the base offense level. Again, the average Democratic appointee sentences seven or eight months lower than the average Republican appointee. These changes come solely from changes to the base offense level and differences in sentencing within the Guidelines’ sentencing range. Columns (7) and (8) test partisan differences in the “final change,” or the change in prison sentences from the lowest possible final calculated range. The results again demonstrate consistent partisan differences of a magnitude similar to that estimated in Columns (3) and (4).

In sum, the results of Table 2 strongly suggest that there are consistent partisan differences in sentencing, and these differences are expressed both in terms of offense-level adjustments and departures. The strongest evidence of differences comes when we consider changes in prison sentences from “base” or “final” calculated sentencing ranges. Of course, there are situations in which the judge has little role to play in calculating the offense level and no opportunity to depart; for example, when there are no facts in dispute about the use of a gun or the quantity of drugs and nothing unique about the offender that would warrant a departure. Most observers agree, however, that

the sentencing hearing can play an important role, and does so when facts are in dispute.⁵⁴ Indeed, some justices on the Supreme Court have recently expressed uneasiness with regard to the Guidelines over the influence that judicial factfinding may have on sentencing.⁵⁵

Table 3 tests for circuit alignment effects by dividing Democratic appointees into “aligned” or “unaligned” circuits. For the most part, the difference between the aligned and unaligned Democratic appointees is not statistically significant.⁵⁶ This is not surprising given the sample size. However, the coefficients are suggestive and generally consistent with our hypotheses. First, taking the coefficients as given, prison sentences (Column (2)) are two months shorter in aligned Democratic circuits. However, when we measure the difference in changes from the base offense level in Column (3), sentences are actually shorter in *unaligned* circuits. This is counterbalanced by the magnitude of departures (Column (4)) and the probability of departures (Column (5)). These are much higher in aligned circuits. Indeed, the *probability of a Democratic appointee granting a departure is statistically significantly higher in aligned circuits* than in unaligned circuits. By contrast, in both probability and magnitude of departure, Democratic-appointed district court judges in Republican circuits are statistically indistinguishable from Republican-appointed district court judges.

Contrasted with Column (3), the departure results in Columns (4) and (5) are consistent with the idea that departures by Democratic appointees are easier to make under Democratic circuits. Therefore, Democratic appointees will rely more on adjustments in Republican circuits, and will make more use of departures in Democratic circuits. Thus, there is some evidence of substitution between adjustments and departures based on circuit court alignment.

This analysis is the first to use the actual party of the sentencing judge at the federal level to estimate sentencing effects. The results compare very closely to prior studies using cruder measures but greater sample sizes. In addition, we are able to identify the plausible impact of circuit review. Moreover, the results are entirely in line with

⁵⁴ See Stephanos Bibas, *Judicial Fact Finding and Sentence Enhancements in a World of Guilty Pleas*, 110 Yale L J 1097, 1152–67 (2001) (criticizing *Apprendi* on the grounds that it limits the ability of defendants who plead guilty to contest disputed enhancements at a sentencing hearing).

⁵⁵ See, for example, *Rita v United States*, 127 S Ct 2456, 2485 (2007) (Souter dissenting) (expressing concern over empowering judges “to find the facts necessary to sentence in the higher range” because doing so “would make the jury a mere gatekeeper to the more important trial before a judge alone”).

⁵⁶ We admit that simply dividing circuits into “aligned” and “unaligned” categories is a bit crude. In the Appendix, we extend the analysis and estimate these differences based on the percentage of the active circuit judges that are Democratic appointees, and we find results consistent with the cruder estimates presented here.

the theory. Republican and Democratic appointees sentence differently, and circuit court review clearly constrains their decisionmaking. Democratic appointees rely more on adjustments in Republican circuits, but make greater use of departures when they are aligned with the reviewing court. The Appendix details that the results are surprisingly robust to alternate empirical specifications.

III. IMPLICATIONS

A. Supreme Court Jurisprudence: *Apprendi*, *Booker*, and the Slow Unraveling of the Sentencing Guidelines

In *Apprendi*, decided in 2000, a deeply divided Court invalidated a New Jersey statute that allowed judges, at the sentencing phase, to make factual findings that would increase the sentence in excess of the statutory maximum.⁵⁷ Specifically, the Court held that due process and the Sixth Amendment jury right require that any fact (except that of prior conviction) that increases the sentence beyond the statutory maximum must be decided by a jury and proven “beyond a reasonable doubt.”⁵⁸ At the time, Justice O’Connor noted in dissent that the opinion represented a “watershed” event in constitutional law and hinted that the Guidelines regime created by Congress could be threatened.⁵⁹ After all, *Apprendi*’s reasoning could apply to the Guidelines’ sentencing ranges, requiring that the facts necessary for upward adjustments be proven to a jury.⁶⁰

Justice O’Connor’s prediction was almost borne out a few years later. In *Blakely v Washington*,⁶¹ the Supreme Court applied the reasoning of *Apprendi* to invalidate the state of Washington’s sentencing guidelines, holding that they violated due process and the Sixth Amendment’s guarantee of a right to a trial by jury.⁶² Again, dissenters, including O’Connor, predicted dire consequences for the federal Sentencing Guidelines, and the Court was widely expected to strike them down at the next opportunity.⁶³ However, in *Booker*, Justice Breyer, a

⁵⁷ See 530 US at 497 (holding that the New Jersey statute represents “is an unacceptable departure from the jury tradition that is an indispensable part of our criminal justice system”).

⁵⁸ See *id.* at 490. The Court thereby invalidated a New Jersey hate crime statute that authorized an increase beyond the maximum sentence based solely on a judge’s finding by a preponderance of the evidence that the defendant acted with the purpose to intimidate on the basis of particular characteristics of the victim. See *id.* at 491, 497.

⁵⁹ See *id.* at 524, 544 (O’Connor dissenting).

⁶⁰ See *id.* at 543–44.

⁶¹ 542 US 296 (2004).

⁶² See *id.* at 301–08.

⁶³ See *id.* at 323–26 (O’Connor dissenting) (“If the Washington scheme does not comport with the Constitution, it is hard to imagine a guidelines scheme that would.”).

leading proponent of Guidelines sentencing, crafted a majority opinion that simply excised the portion of the Sentencing Reform Act that made the Guidelines mandatory, thus preserving the Guidelines' general framework.⁶⁴ He did so by winning over Justice Ginsburg from the *Blakely* majority. As long as the Guidelines are simply advisory, the remedial opinion in *Booker* holds, they do not violate the right to a trial by jury or the right to have elements of a crime proven beyond a reasonable doubt.⁶⁵ In our view, the key element of the *Booker* opinion is not that it made the Guidelines advisory, but rather that circuit courts would now review district court sentences for "reasonableness" in light of the Guidelines.⁶⁶

What this new standard of review actually means, and what it means to have nonbinding, advisory Guidelines, was left to the courts of appeals. The opinion itself provided almost no guidance and did not choose to define what "reasonableness" review would entail, likely out of necessity given the 5-4 majority.⁶⁷

One key question—now only partially answered—is the extent to which Guidelines sentences still represent safe harbors. In *Rita v United States*,⁶⁸ the Supreme Court declined to clarify the *Booker* review standards further.⁶⁹ It held that circuit courts *may* provide a presumption of validity to a within-Guidelines sentence, but that this presumption still entails a "reasonableness" review.⁷⁰ The full effect of this reasoning is obviously far from clear because the majority opinion declined to provide clear guidelines on how "reasonableness" would be evaluated in light of the presumption. In addition, it is not clear from the opinion when courts can disregard the presumption.⁷¹

⁶⁴ See 543 US at 245–46.

⁶⁵ See *id.* at 245–46 (stating that the Court's approach would make the Guidelines constitutional "while maintaining a strong connection between the sentence imposed and the offender's real conduct," and thus preserving Congress's goal of promoting uniformity in sentencing).

⁶⁶ See *id.* at 261.

⁶⁷ However, the opinion did note that "reasonableness" had been the standard of review for departures from the Guidelines prior to a congressional amendment in 2003. See *id.*

⁶⁸ 127 S Ct 2456 (2007).

⁶⁹ See *id.* at 2465.

⁷⁰ See *id.* at 2464.

⁷¹ It is also important to note that *Rita* was a 6-3 opinion, and two of the justices in the majority (Justices Stevens and Ginsburg) had previously voted to declare the Guidelines unconstitutional on Sixth Amendment grounds. See *Booker*, 543 US at 232–33. Justice Stevens's concurring opinion tried to clarify reasonableness review and stated that it should be meaningful, see *Rita*, 127 S Ct at 2471–74, while the majority seemed to indicate that due deference would still apply to the district court's application of the Guidelines, compare *id.* at 2463 ("[T]he presumption reflects the fact that, by the time an appeals court is considering a within-Guidelines sentence on review, both the sentencing judge and the Sentencing Commission will have reached the *same* conclusion as to the proper sentence in the particular case.").

However, it was clear that, post-*Booker*, judges must still conduct a sentencing hearing, calculate an offense level, and justify a departure.⁷² But how should circuits review within-Guidelines sentences? Over half of the circuits have held that sentences within a properly calculated Guidelines range are presumptively reasonable.⁷³ Other circuits generally accord substantial weight to the Guidelines, but have made it clear that they do not consider them presumptively reasonable.⁷⁴ In practice, it is not clear that this split has made any difference. So far, only one sentence within the Guidelines range has been reversed by an appeals court, and it was reversed not on the grounds that it was unreasonable, but on the grounds that the district court did not adequately consider a number of factors.⁷⁵ Of course, *Rita* is a recent decision and we lack even anecdotal evidence of its impact at this point. However, based on the circuit courts' approach to *Booker*, and given that *Rita* largely validated this approach, it appears that the Guidelines ranges remain safe harbors as a matter of fact if not law.

Given that the Guidelines still represent safe harbors (or something close to it) as a matter of fact if not law, our prior analysis predicts that little would change post-*Booker*. District courts can protect themselves by properly calculating the offense level, and the reviewing circuit court can reverse departures as unreasonable instead of being an abuse of discretion. This prediction is largely borne out by the preliminary sentencing data. The United States Sentencing Commission has been closely following post-*Booker* developments and has released a thorough analysis of the data on sentencing under the new

⁷² For example, Justice Breyer was careful to note that judges must consider the Sentencing Guidelines range and continue to make fact findings in light of the Guidelines, conduct a sentencing hearing, and justify departures. See *Booker*, 543 US at 259–60, 264.

⁷³ See *United States v Green*, 436 F3d 449, 457 (4th Cir 2006); *United States v Alonzo*, 435 F3d 551, 554 (5th Cir 2006); *United States v Williams*, 436 F3d 706, 708 (6th Cir 2006); *United States v Mykytiuk*, 415 F3d 606, 608 (7th Cir 2005); *United States v Tobacco*, 428 F3d 1148, 1151 (8th Cir 2005); *United States v Kristl*, 437 F3d 1050, 1054 (10th Cir 2006). Compare *United States v Cooper*, 437 F3d 324, 332 (3d Cir 2006) (rejecting the rebuttable presumption approach but holding that a sentence within the Guidelines range is “more likely to be reasonable than one outside the guidelines range”); *United States v Talley*, 431 F3d 784, 788 (11th Cir 2005) (“[W]hen the district court imposes a sentence within the advisory Guidelines range, we ordinarily will expect that choice to be a reasonable one.”).

⁷⁴ See *United States v Cantrell*, 433 F3d 1269, 1279 (9th Cir 2006); *United States v Crosby*, 397 F3d 103, 114–16 (2d Cir 2005) (declining to declare the Guidelines presumptively reasonable, and identifying a number of errors that could render a sentence within the Guidelines range unreasonable).

⁷⁵ *United States v Lazenby*, 439 F3d 928, 933–34 (8th Cir 2006) (holding that the district court should have considered a downward departure in light of the defendant's cooperation with the investigation). The court also reversed the below-Guidelines sentence of a co-conspirator as unreasonable, suggesting that the conduct was similar and so the sentences should be as well. The court did not suggest that a lower sentence was the only reasonable course. See *id* at 932–33.

system. Although the data following *Booker* is preliminary, it appears, not surprisingly, that judge-induced downward departures have increased by about 4 percentage points (from 8.6 percent to 12.5 percent of all sentences).⁷⁶ While not earth-shattering, it does represent a nearly 50 percent increase in the probability of a judge-induced departure. The rate of upward departures doubled, from 0.8 to 1.6 percent. However, the sentence length⁷⁷ and probability of imprisonment increased only slightly.⁷⁸ Thus, the most important aspect of sentencing, the length of the sentence, changed very little. This is largely because, even though they were more frequent, the average size of departures remained the about same.⁷⁹ This conforms to the general predictions of our theory and the conclusion of most observers that the fundamentals of sentencing changed little post-*Booker*.⁸⁰

Some have expressed surprise that, given the widely noted judicial hostility to the Guidelines, the creation of an advisory system would change so little.⁸¹ Part of this is surely due to Justice Breyer's deft handling of the remedial opinion, which largely maintained the sentencing procedures of the Guidelines and preserved appellate court review.⁸² We acknowledge that it is entirely possible that sentences did not change post-*Booker* because *Booker* did not actually expand the discretion of the district courts, which was probably Justice Breyer's goal. But why the increase in departures? A likely answer, consistent with our theory, is that district courts now find it easier, as a matter of doctrine, to depart from the Guidelines, so they use departures relatively more than they did pre-*Booker*. They may nonetheless arrive at substantially the same sentence because the circuit court acts as the relevant constraint.

⁷⁶ The effect of *Booker* depends very much on the relevant comparison period. The PROTECT Act was effective in the year and a half prior to *Booker*, and it restricted the use of downward departures and encouraged prosecutors to appeal them more often. The Sentencing Commission's statistics indicate that in the seven months prior to the PROTECT Act, the average judge-induced downward departure rate was 8.6 percent. This fell to 5.5 percent during the PROTECT Act's period of applicability, but increased to 12.5 percent in the year following the *Booker* decision. United States Sentencing Commission, *Final Report on the Impact of United States v. Booker on Federal Sentencing* appendix E-1 (2006).

⁷⁷ *Id.* at 46.

⁷⁸ *Id.*

⁷⁹ *Id.* at 63.

⁸⁰ For a discussion of how sentencing has played out after *Booker*, see Douglas A. Berman, *Tweaking Booker: Advisory Guidelines in the Federal System*, 43 *Houston L Rev* 341, 347–55 (2006) (“Based on a year of experience with the *Booker* remedy . . . it now appears that Justice Breyer largely succeeded in preserving the fundamental pre-*Booker* features of federal sentencing.”). See generally Nancy J. King, *Reasonableness Review after Booker*, 43 *Houston L Rev* 325 (2006).

⁸¹ See Berman, 43 *Houston L Rev* at 347–58 (cited in note 80).

⁸² See note 72.

We are more hesitant than most other observers, however, to conclude that nothing meaningfully changed after *Booker*. The theory and empirical evidence presented here leave open the possibility that sentences in fact did change post-*Booker*, but such changes are masked by the interplay between district and circuit courts. It is not clear whether “reasonableness” review is stricter or less strict than the “abuse of discretion” review that prevailed before. Indeed, *reasonableness* may suggest a broader role for the circuit courts. If this is the case, then after *Booker* sentences will more closely match the preferences of the circuit than before. While sentences may have stayed the same or even increased in some circuits, they may have decreased in others. In addition, the changes in sentences are most likely to be apparent only for certain crimes for which long sentences are controversial, such as drug trafficking.

A second possibility is that changes only occurred in circuits that wished to transfer more discretion to the district courts. It is not surprising that the more conservative circuits (for example, the Fourth, Fifth, and Seventh) decided to declare the generally high Guidelines sentences presumptively reasonable,⁸³ while the more liberal Second and Ninth declined to do so.⁸⁴ Thus, the presumptively reasonable circuits encourage Guidelines sentences while maintaining the ability of an appeals panel to reverse a departure as unreasonable. District judges in the Second and Ninth Circuits, however, are relatively more free from the Guidelines’ strictures, knowing that the reviewing court favors lower sentences and has signaled its intent to be lenient. If most judges wish to reduce sentences, then the preferences of the circuit are met. In addition, these circuits retain the ability to reduce long sentences as unreasonable.

In sum, the evidence on judicial preferences pre-*Booker* suggests that we cannot say much about how sentences have changed after *Booker* without (1) identifying categories of crime in which judges have strong sentencing preferences; (2) considering the preferences of the district judges *and* the reviewing circuits; and (3) accounting for how each circuit has interpreted reasonableness review and the weight each accords to the Guidelines in determining reasonableness.

B. Transparency Reform

The unavailability of judge-identifying data in criminal sentencing is one of the most frustrating aspects of the study of federal sentencing and has significantly impeded scholarly evaluation of the

⁸³ See *Green*, 436 F3d at 457; *Alonzo*, 435 F3d at 554; *Mykytiuk*, 415 F3d at 608.

⁸⁴ See *Cantrell*, 433 F3d at 1279; *Crosby*, 397 F3d at 114–16.

Guidelines' efficacy. The Sentencing Commission's policy of concealing in its sentencing database the identity of judges and the sentences they impose in individual cases raises the suspicion that sentencing bias and sentencing disparity are related and that judges are being protected from public and academic scrutiny in imposing sentences. Determining whether the Guidelines are serving their intended function of reducing disparity in sentencing requires the study of how judges behave under the Guidelines.

The empirical test above illustrates the importance of transparency for identifying bias in criminal sentencing. We suspect that if such transparency was the norm rather than the exception, it would ultimately translate into less bias as judges become more aware of their own tendencies and realize the power of empirical scholarship in exposing inconsistencies in sentencing.

In 1986, the Sentencing Commission requested baseline sentencing data from the Administrative Office, including judge identifiers. The Administrative Office initially supplied the data absent information that would identify the judge that issued a sentence in any specific case. Only later, under an agreement that prevented the public release of any individual's name associated with the sentencing decision, did the Administrative Office provide the Sentencing Commission with judge identifiers.⁸⁵ But scholars and the public were not given access because of stated concerns for defendant privacy. Professor Marc Miller has argued that the agreement between the Administrative Office and the Sentencing Commission is "inconsistent with the spirit and purpose of the Congressional mandates to the Commission to collect and release information . . . [and] obfuscates the decision to withhold judge identifiers behind an apparent (but logically distinct) concern for defendant privacy."⁸⁶ We agree.

The public availability of judge identifiers would allow scholars to check judicial discretion in criminal sentencing. In particular, judge identifiers would allow scholars to evaluate whether, and how much,

⁸⁵ Judge identifiers remained secret until Congress enacted the PROTECT Act, which, among other things, made judge identifiers available to the Department of Justice and Congress. See PROTECT Act § 401(h), 117 Stat at 672. See also Marc L. Miller, *Domination & Dissatisfaction: Prosecutors as Sentencers*, 56 Stan L Rev 1211, 1245–47 (2004) (describing the information dissemination requirements of the PROTECT Act).

⁸⁶ Marc L. Miller, *A Map of Sentencing and a Compass for Judges: Sentencing Information Systems, Transparency, and the Next Generation of Reform*, 105 Colum L Rev 1351, 1356–57 n 19 (2005) (arguing in favor of sentencing information systems that would increase the access of the public to sentencing data). The agreement states that "[n]o information that will identify an individual defendant or other person identified in the sentencing information will be disclosed to persons or entities outside of the Commission without the express permission of the court for which the information was prepared." United States Sentencing Commission, *Public Access to Sentencing Commission Documents and Data*, 54 Fed Reg 51279, 51282 (1989).

judges engage in sentencing disparity under the Sentencing Guidelines, and whether such disparities relate to political and ideological orientation, other demographics of the judiciary, or the characteristics of the defendant.⁸⁷ There is considerable literature linking political ideology to decision outcomes, and there is little reason to think that criminal sentencing would escape such bias. Given the establishment of the Sentencing Commission and the imposition of the Sentencing Guidelines, making judge identifiers available would help to answer questions about the success or failure of such institutional reforms.⁸⁸

At least one judge has called for transparency, arguing that making data available to academic researchers would provide for a better understanding of the operation of the Sentencing Guidelines. According to Judge Avern Cohn of the Eastern District of Michigan,

[T]here are variations in the “sentencing persona” of district judges. Experience under the Civil Justice Reform Act using judge identifiers . . . resulted in substantial improvement in shortening the time that judges take to dispose of motions and cases. Thus far, similar data on individual judges have not been made available as to guideline sentences out of deference to the perceived sensitivity of judges. The Commission collects these data under its statutory powers and there is no good reason to continue to maintain the data in secret. A policy of full disclosure is in the public interest.⁸⁹

Judge Cohn points out that the Commission likely overrates the extent to which district judges are opposed to publishing judge identifiers. He notes that an informal poll of judges in the Eastern and Western Districts of Michigan showed a more than two-to-one vote in favor of publication. The judges opposing the identifiers were con-

⁸⁷ See Miller, 105 Colum L Rev at 1385 (cited in note 86) (“The identification of the individual judge is essential to evaluate the most basic justification for modern sentencing reform: that individual judges may introduce their own biases across cases or in individual cases.”).

⁸⁸ See Marc L. Miller, *Sentencing Reform “Reform” through Sentencing Information Systems*, in Michael Tonry, ed, *The Future of Imprisonment* 121, 146–48 (Oxford 2004) (proposing that “wide availability of sentencing data including judge identifiers and detailed offense and offender information” will improve transparency in sentencing). See also Mark H. Bergstrom and Joseph S. Mistick, *The Pennsylvania Experience: Public Release of Judge-specific Sentencing Data*, 16 Fed Sent Rptr (Vera) 57, 63 (2003) (describing with general approval Pennsylvania’s experience with public release of judge-specific sentencing data and concluding that “[m]any of the negative outcomes predicted during the development of the policy have not materialized”); Paul J. Hofer and William P. Adams, *Using Data for Policymaking, Litigation, and Judging*, 16 Fed Sent Rptr (Vera) 8, 11–12 (2003) (describing the current methods used to identify individual judges and weighing whether it would be good policy to have “full public release of judge-identifying sentencing data”).

⁸⁹ Avern Cohn, *Advice to the Commission—A Sentencer’s View*, 8 Fed Sent Rptr (Vera) 14, 14 (1995).

cerned about “the harm to collegiality that might be occasioned by such information; its lack of usefulness and possible misuse; and because decisions in sentencing are too fact specific to make such information of value.”⁹⁰ While we cannot address the harm to collegiality, the empirical study provided here suggests that judge identifiers are very useful in identifying systematic bias attached to ideology and that the aggregation of information makes it especially valuable compared to an individual case approach.

C. Political-institutional Reform

An important conclusion from this Article is that doctrine and appellate review matter in sentencing. Sentencing judges are constrained from the excessive use of departures by circuit court review, and the excessive use of adjustments is less constrained. In our view, political alignment essentially relaxes the doctrinal and jurisprudential constraints judges face when they want to depart from the Guidelines. Democratic appointees can depart more when in Democratic-dominated circuits. Republican appointees have a similar freedom in Republican-dominated circuits, although they are less willing to depart than Democratic appointees based on the makeup of the overseeing circuit. This reluctance of Republican appointees is entirely consistent with the structure of the Guidelines, as (1) Guidelines sentences are generally high already and (2) most adjustments, which are difficult to review regardless of the circuit court’s political makeup, are (compared to departures) sentence-increasing. Of course, upward departures being rarer events, it is possible that we simply do not have a large enough sample size to detect differences for Republican appointees along these lines.

Nonetheless, an extension of this basic insight is that the lack of judicial political diversity in any given sentencing event increases the likelihood of greater disparity in sentencing. The current practice of randomly assigning circuit judges to panels often produces ideologically unbalanced panels with either three Democratic or three Republican appointees controlling the outcome (horizontal political alignment). When such unified panels review a lower court judge who holds the same political orientation as the panel (vertical political alignment), there appears to be little check on the severity or leniency of the sentence. Consequently, district court judges sitting under circuits that have a strong tendency towards ideological alignment (such as a Republican-appointed district court judge sitting within the Fifth Circuit Court of

⁹⁰ Avern Cohn, *The Sentencing Commission’s 1993 Annual Report*, 7 Fed Sent Rptr (Vera) 137, 138 (1994).

Appeals, which is made up of eleven Republican appointees and four Democratic appointees) likely feel little vertical constraint in choosing the severity or leniency of the sentence imposed on defendants.

Other scholarship has demonstrated that the political structure of the judicial hierarchy, as well as horizontal relationships within a court, can have profound effects on the consistency of application of legal doctrines to similarly situated cases. Frank Cross and Emerson Tiller have shown that a politically diverse panel of judges is more likely to ensure obedience to established legal doctrine than a panel made up of judges from the same political orientation.⁹¹ They found that the presence of a political minority member on a three-judge federal appellate court panel (that is, a Democratic appointee with two Republican appointees, or a Republican appointee with two Democratic appointees) led to more consistent application of the *Chevron U.S.A. Inc v NRDC*⁹² doctrine in administrative law cases than when the panel was politically unified (all Democratic or all Republican appointees).⁹³ Cass Sunstein found that judges on like-minded panels appeared to vote far less moderately than judges on divided panels.⁹⁴ Richard Revesz found similarly in the environmental context.⁹⁵ These results support the idea that judges within panels may learn from one another's ideas and world-views or otherwise act as a check on unwarranted discretion.

To the extent that reducing sentencing disparity remains a goal of reform, we suggest that structural changes in judicial review hierarchy—changes that acknowledge the political-ideological forces at work in sentencing—be considered. In particular, we suggest institutionalizing

⁹¹ See Frank B. Cross and Emerson H. Tiller, *Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals*, 107 Yale L J 2155, 2159 (1998) (arguing that judges are less likely to make decisions based on political preferences when the panel is ideologically divided). See also Cass R. Sunstein, David Schkade, and Lisa M. Ellman, *Ideological Voting on Federal Courts of Appeals*, 90 Va L Rev 301, 319–25 (2004) (finding evidence of ideological dampening in decisionmaking when a judge sits with two judges of the opposite party).

⁹² 467 US 837, 842–43 (1984) (holding that “[w]hen a court reviews an agency’s construction of the statute which it administers,” and the intent of Congress as to the meaning of the statute is not clear, the court must defer to the agency’s construction if that construction “is based on a permissible construction of the statute”).

⁹³ See Cross and Tiller, 107 Yale L J at 2168–73 (cited in note 91) (conducting an empirical examination of all DC Circuit Court of Appeals cases between 1991 and 1995 that cited *Chevron* and finding that mixed panels were almost twice as likely to defer to nonaligned agency interpretations as compared with unified panels).

⁹⁴ Cass R. Sunstein, *Why Societies Need Dissent* 168–82 (Harvard 2003) (discussing panel effects in cases involving affirmative action, sex discrimination, environmental regulation, liability of directors for corporate wrongdoing, and the Americans with Disabilities Act).

⁹⁵ See Richard L. Revesz, *Environmental Regulation, Ideology, and the D.C. Circuit*, 83 Va L Rev 1717, 1719 (1997) (finding that, on the DC Circuit, judges vote ideologically in general, tend to vote more ideologically when their decision is unlikely to be reviewed by the Supreme Court, and vote in a manner that is heavily influenced by the composition of the panel in a given case).

political diversity for every sentencing event by ensuring that for any criminal sentencing the lower court and higher court not share a uniform political (partisan) orientation.

In practice, political diversity in sentencing could be achieved in a variety of ways. First, the system could ensure that a three-judge circuit panel reviewing a district court's sentencing decision would have no more than two judges sharing the same political orientation (that is, no more than two Democratic appointees or two Republican appointees on a panel reviewing a sentencing decision). Consequently, at least one circuit judge would have a political orientation different from that of the sentencing judge, thereby eliminating politically uniform vertical alignment.⁹⁶ In terms of implementation, this could be achieved by a congressional statute mandating the diversity, or the appellate courts could impose such a rule on themselves. In effect, this would impose an ideological cap on the normal practice of random assignment of circuit judges to panels—no more than two circuit judges on a panel can share the same political orientation. If a Democratic (Republican) appointee was the sentencing judge, then at least one, if not two, of the circuit judges assigned to the case would come from a pool of Republican (Democratic) appointees. We would expect the consequence to be less disparity in sentencing, to the extent that such is due to ideological bias.⁹⁷

A second possible mechanism would be to set up a separate appellate review court for criminal sentencing (possibly made up of current federal circuit judges brought together in an alternative forum) and ensure that such court is politically diverse. Consideration of an appellate court with specialized jurisdiction is not without precedent. The Federal Circuit Court of Appeals has specialized jurisdiction over patent appeals from lower district courts. And Supreme Court justices, legislators, and special commissions have, at times, called for a new national court of appeals to handle intercircuit conflicts or other issues that would reduce the Supreme Court's caseload.⁹⁸

⁹⁶ In fact, one circuit panel judge with a different political orientation than the district court judge may be enough to prevent sentencing disparity, because that judge can act as a whistleblower on the other two judges if they do not properly scrutinize the lower court's sentencing decision. See Cross and Tiller, 107 *Yale L J* at 2173–74 (cited in note 91).

⁹⁷ See Emerson H. Tiller and Frank B. Cross, *A Modest Proposal for Improving American Justice*, 99 *Colum L Rev* 215, 215–16 (1999) (proposing politically diverse circuit court panels more generally).

⁹⁸ See, for example, Warren E. Burger, *Annual Report on the State of the Judiciary*, 69 *ABA J* 442, 447 (1983) (proposing the creation of a “special, but temporary panel of the new United States Court of Appeals for the Federal Circuit” with the sole duty of reviewing intercircuit conflicts); Court Improvements Act of 1983 (CI Act), S 645, 98th Cong, 1st Sess, in 129 *Cong Rec S* 3402–03, 3410–11 (Mar 1, 1983) (proposing the creation of an “Intercircuit Tribunal of the United States Courts of Appeals” situated between the Supreme Court and the courts of appeals

As far as the use of political ideology or partisan identifiers for constructing the specialized reviewing court, the use of such devices to ensure policy and decisional balance is not unique in our system of government. Numerous independent regulatory commissions—such as the Surface Transportation Board (formerly the Interstate Commerce Commission), the National Transportation and Safety Board, the SEC, and the Nuclear Regulatory Commission—are designed to limit partisan excess through the use of split partisan arrangements. Congress designed these agencies as quasi-judicial institutions to be filled with experts, but with partisan checks to protect against ideological excess by one party. Most telling in this regard is the design of the United States Sentencing Commission itself, which the law requires to have seven voting members, *no more than four of whom may be members of the same political party*.⁹⁹

Restructuring courts to minimize political-ideological bias is not without its critics. Former DC Circuit Court of Appeals Judge Patricia Wald claims that using political orientation as a control on panel makeup would have the effect of making judges act more partisan in their decisionmaking.¹⁰⁰ In a sense, judges would become role players for their partisan identification.¹⁰¹ While that is certainly possible, the political diversity of the mixed panel should nonetheless offset the extremes of either group. In other words, a highly partisan yet mixed panel is more likely to sentence consistently than a less charged group of like-minded partisans sitting together (unified panel) that arose randomly.

An additional criticism by Wald is that mandating political diversity on appellate panels is contrary to the notion of a judiciary free of political constraints and might undermine the authority and independence of the judiciary.¹⁰² However, the judiciary's *lack* of freedom from politics and ideology is exactly what is exposed by this empirical

and presenting the remarks of Senator Dole, who argued that the tribunal would “relieve the drastically increased workload of the Supreme Court”); CI Act, 129 Cong Rec at D 484 (June 29, 1983) (noting the subcommittee’s approval of the bill); Federal Judicial Center, *Report of the Study Group on the Caseload of the Supreme Court*, 57 FRD 573, 590–95 (1972) (recommending a National Court of Appeals that would decide many conflicts between circuits); Commission on Revision of the Federal Court Appellate System, *Structure and Internal Procedures: Recommendations for Change*, 67 FRD 195, 208 (1975) (recommending the creation of a National Court of Appeals to “increase the capacity of the federal judicial system for definitive adjudication of issues of national law”).

⁹⁹ See 28 USC § 991(a).

¹⁰⁰ See Patricia M. Wald, *A Response to Tiller and Cross*, 99 Colum L Rev 235, 255 (1999) (criticizing Tiller and Cross’s proposed mandatory mixed panels as both unnecessary and likely to be counterproductive).

¹⁰¹ *Id.* (arguing that Tiller and Cross’s “proposal will turn conferences into a type of political jousting match, where each side is required to wear the colors of its party sponsor and to defend its name to the death”).

¹⁰² See *id.* at 256–57.

study, and exactly what our proposal is intended to alleviate. If the judiciary truly was an independent, nonpolitical body in the way envisioned by the Constitution, then of course such a proposal would not be necessary. Unfortunately, this is not the case.

We have framed this discussion as one of Republican appointees versus Democratic appointees. This is a natural distinction and, given the attention that appellate courts have received recently, a reasonable one. Of course, we have focused on what is merely a proxy for judicial philosophy and attitudes toward punishment. Other characteristics, perhaps more palatable to policymakers, could also be considered. For example, previous prosecutorial or judicial experience may also be indicators of judicial attitudes toward sentencing.

CONCLUSION

This Article is the first to unify theory and judge-specific data to describe sentencing in a political and hierarchical framework. We combine judge-level data with actual sentencing data in a manner that directly tests propositions derived from positive political theory. In addition, we believe that the methodology employed is one that most scholars, policymakers, and judges can accept as theoretically and empirically valid. The study enriches our understanding of criminal sentencing by focusing on the procedural and doctrinal details of criminal sentencing through which judges can exercise sentencing discretion in a biased, if not strategic, manner. In addition, this Article explains how the role of circuit courts will become more important in the face of *Booker*, in which the Supreme Court declared the Guidelines to be advisory.

Our results directly support two reforms—one informational and the other structural. The first relates to transparency in sentencing data, in particular data on the identity of the sentencing judge, and how such data could be useful in helping to identify the sources of judicial sentencing disparity that would lead to the types of reform necessary to eliminate such bias. The second reform is more powerful and controversial—the requirement of political-ideological diversity on judicial appellate panels reviewing criminal sentences (which would ensure hierarchical political-ideological diversity between the lower sentencing court and the higher court reviewing such decisions). Specifically, all sentencing review should involve both Democratic and Republican judicial appointees. This amounts to an engineering of the political structure of the judiciary that most scholars, judges, and policymakers would quickly reject at first mention for practical and jurisprudential reasons. We argue, nonetheless, that the benefits would be substantial and that such a proposal (or one that adopts its primary tenet—recognition of the role that political-ideological diversity within the judiciary can play) deserves serious attention.

APPENDIX

The purpose of this Appendix is to test the robustness of the results presented in the main text to a number of different specifications. In general, we find that the results are quite robust and compare favorably to our previous work. As in the Article, we explore five dependent variables: (1) *change in offense level*; (2) *total prison sentence*; (3) *change in prison sentence from minimum base offense level*; (4) *change in prison sentence from minimum final offense level*; and (5) *judge-induced departures*.

We present a number of different specifications to aid in interpretation of the data. The first specification is:

$$(1) \text{ Total Prison} = \alpha + \beta \text{Democrat} + \varepsilon$$

The regression is run at the level of an individual offender. *Democrat* is simply an indicator variable that takes on the value one if the sentencing judge was appointed by a Democratic president and zero if appointed by a Republican. The coefficient of interest in Equation (1) is β , which merely measures the average difference in sentences between Republican and Democratic appointees. It does not take account of offender or offense characteristics, which may vary systematically across districts and, consequently, across judges. Next, we run the following regression to determine whether offender and offense characteristics can explain differences between Republican and Democratic appointees:

$$(2) \text{ Total Prison} = \alpha + \beta \text{Democrat} + \delta \text{OffenderVars} + \gamma \text{OffenseVars} + \varepsilon$$

Offender variables are reported in the tables and are *age of offender*, *age of offender squared*, and indicator variables for *female*, *black*, and *Hispanic origin*.¹⁰³ Offense variables should attempt to control for position on the sentencing grid. We do so by using *base offense level*, *base offense level squared*, *base offense level's* interaction with *criminal history category*, and five dummies for *criminal history category*. This specification is quite flexible and should take good account of the Guidelines' structure. Because drug trafficking offenses constitute nearly three-quarters of the sample, we also include an indicator variable for whether the offense of conviction was drug trafficking. As such, β now measures the difference between Republican and Democratic appointees conditional on offender and offense characteristics.

¹⁰³ We employed a richer set of offender variables in our previous paper. We use fewer here in order to preserve sample size because some are missing.

Another important concern is whether systematic differences across circuits could drive the results. For example, if the Ninth Circuit were largely comprised of Democratic appointees and had a liberal departure policy, these circuit effects may be reflected in the *Democrat* coefficient, β , even though Republican appointees may behave similarly to Democratic appointees in the Ninth Circuit. We next add circuit dummies to the regressions:

$$(3) \text{ Total Prison} = \alpha + \beta \text{Democrat} + \delta \text{OffenderVars} + \gamma \text{OffenseVars} + \gamma \text{Circuit} + \varepsilon$$

Circuit is a matrix of eleven dummy variables.¹⁰⁴ Therefore, Equation (3) compares differences between Democratic appointees and Republican appointees *within* a circuit and conditional on offender and offense characteristics.

An important innovation of the Sentencing Guidelines was the creation of appellate review of criminal sentences. To examine the influence of district and circuit court alignment, we separate our *Democrat* indicator into two new variables: *Aligned Democrat* and *Unaligned Democrat*. This mimics the specification of our previous work, in which we simply had a dummy for whether the circuit was majority Democratic-appointed, and then interacted it with the percentage of the district court that was Democratic-appointed. In addition, we attempt a new specification that was not tractable previously: we test the effect of changes in the overall political composition of the circuit courts. We will discuss this approach in greater detail below.

¹⁰⁴ There are twelve circuit courts, including the DC Circuit. In our previous work, we conditioned on district dummies. However, given the small sample of offenders available here and the fact that only 389 judges are in the sample, it was generally not tractable to include 82 district dummies. Doing so generally decreased the estimated effect of *Democrat*.

A. Appendix Table 1: Change in Offense Level

Appendix Table 1 presents the results for *change in offense level*, so the dependent variable is thus *final calculated offense level* minus *base offense level*.¹⁰⁵ The results for *Democrat* are fairly weak across the board. Model 1 demonstrates that Democratic appointees reduce offense levels by -0.72 more than Republican appointees. This difference is significant at the 5 percent level. Model 2 controls for offender and offense variables, but the coefficient on *Democrat* is reduced to -0.39, and is significant at only the 10 percent level. Adding circuit dummies in Model 3 further reduces the magnitude of the coefficient to -0.31. Given the weakness of these initial results, it is not surprising that we get even less precise results when we separate *Democrat* into *Aligned Democrat* and *Unaligned Democrat* groups in Model 4.

Despite their weakness, these results compare favorably to our previous work, which found that an all-Democratic bench would reduce offense-level calculations for serious offenses by about 0.5 offense levels. We found no evidence of circuit alignment effects. The coefficients reported here suggest that there may be an alignment effect, but the standard errors are too large to suggest a difference.

¹⁰⁵ Regressions taking *final offense level* as the dependent variable did not yield any significant results.

2008]

Reviewing the Sentencing Guidelines

751

APPENDIX TABLE 1
CHANGE IN OFFENSE LEVEL

	Model 1	Model 2	Model 3	Model 4
Judge variables				
Democrat	-0.72* (0.34)	-0.39+ (0.22)	-0.31 (0.21)	
Aligned Democrat				-0.44 (0.31)
Unaligned Democrat				-0.20 (0.28)
Offender variables				
Drug offense		-3.11** (0.22)	-3.00** (0.38)	-3.00** (0.38)
Age		-0.03 (0.07)	-0.05 (0.06)	-0.05 (0.06)
Age sq		0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Female		-0.84** (0.29)	-0.80** (0.28)	-0.80** (0.27)
Black		-0.11 (0.24)	-0.12 (0.24)	-0.11 (0.24)
Hispanic		-0.82** (0.22)	-0.65** (0.22)	-0.65** (0.22)
Constant	-1.31** (0.25)	-0.38 (0.22)	6.45** (1.68)	6.44 (1.69)
Circuit dummies	No	No	Yes	Yes
Grid controls	No	Yes	Yes	Yes
R-squared	0.01	0.21	0.33	0.33
Observations	1,241	1,225	1,225	1,225

Standard errors are robust and reflect clustering by judge.

+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

B. Appendix Table 2: Total Prison Sentence in Months

Appendix Table 2 presents the results taking *total prison sentence* as the dependent variable. We include all prison sentences less than 990 months in the initial regressions¹⁰⁶ and then remove substantial assistance cases. The first three columns provide very weak evidence of partisan differences, although the estimate of the coefficient on *Democrat* of roughly -6 in Models 1 and 2 compares quite favorably to our previous work. Removing substantial assistance cases in Model 4 creates a clearer picture: Democratic appointees impose 6.7-month-lower sentences relative to Republican appointees, and the difference is significant at the 5 percent level. This is nearly identical to our previous results of six- to seven-month-lower prison sentences when the district is entirely Democratic-appointed. Separating Democratic appointees into aligned and unaligned circuits in Column (5) does not present a clear picture of circuit court influence.

¹⁰⁶ Prison sentences of 990 months represent significant outliers. There were two 990-month sentences, and the next highest sentences were 572 months and 346 months. There was one life sentence, which we also excluded.

2008]

Reviewing the Sentencing Guidelines

753

APPENDIX TABLE 2
TOTAL PRISON SENTENCE IN MONTHS

	Model 1	Model 2	Model 3	Model 4	Model 5
Judge variables					
Democrat	-6.54 (6.40)	-6.14+ (3.52)	-3.12 (2.69)	-6.65* (3.12)	
Aligned Democrat					-7.76 (5.29)
Unaligned Democrat					-5.75 (3.16)
Offender variables					
Drug offense		-26.5** (4.30)	-27.7** (12.08)	-24.7** (4.30)	-24.6** (4.28)
Age		0.39 (1.26)	-0.28 (0.99)	-0.20 (1.21)	-0.23 (1.26)
Age sq		-0.003 (0.02)	0.01 (0.01)	0.006 (0.02)	0.006 (0.02)
Female		-13.7** (3.57)	-12.9** (2.96)	-11.2** (3.41)	-11.2** (3.40)
Black		6.24 (4.32)	5.78 (3.59)	5.65 (4.24)	5.73 (4.30)
Hispanic		-5.52 (3.43)	0.81 (2.59)	-0.13 (3.14)	-0.21 (3.39)
Constant	79.7** (5.31)	34.4 (21.6)	41.4 (16.7)	39.0 (20.4)	38.5 (20.4)
Circuit dummies	No	No	Yes	Yes	Yes
Grid controls	No	Yes	Yes	Yes	Yes
Substantial assistance cases	Yes	Yes	Yes	No	No
R-squared	0.01	0.41	0.43	0.63	0.63
Observations	1,254	1,212	1,212	889	889

Standard errors are robust and reflect clustering by judge.

+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

C. Appendix Table 3: Change in Prison Sentence from Minimum of Base Range

The best way to study the full effect of adjustments is to examine changes in prison sentences that result from that type of manipulation. Appendix Table 3 presents this specification, although we limit cases to those that did not involve a departure, so that we may focus on changes that were exclusively the result of offense-level manipulations. The dependent variable is the difference between the minimum prison sentence in the base sentencing range and the final prison sentence. Including departures tends to make the results stronger but confounds the alignment analysis, because we predict that departures will be more likely in aligned circuits. We want to focus the alignment analysis on cases that relied only on departures.

The results are similar across the first three specifications, and are actually the most significant when we include the full set of controls in Model 3. Democratic appointees reduce base sentences by 6.5 to 8 months relative to Republican appointees. There were a fair number of outliers in the data, and a concern with OLS is that outliers (particularly in small samples) could be responsible for the results. In Model 4, we eliminated observations in which the final sentence was more than one hundred months away from the base sentence, removing forty-three observations. The coefficient on *Democrat* shrank to -4 months but remained significant at the 10 percent level.

Models 5 and 6 separate *Democrat* into aligned and unaligned circuits. An “aligned circuit” is 50 percent or more Democratic-appointed, and an “unaligned circuit” is less than 50 percent Democratic-appointed. Model 5 includes offender and offense controls but not circuit dummies, while Model 6 adds circuit dummies. Both columns suggest that there is little difference between aligned and unaligned Democratic appointees. In fact, in Model 6 the coefficient on *Unaligned Democrat* is actually larger than the coefficient on *Aligned Democrat* and is independently significant. These results again conform nicely to our previous work, which failed to find any alignment effects for changes in base offense levels.

2008]

Reviewing the Sentencing Guidelines

755

APPENDIX TABLE 3
CHANGE IN PRISON SENTENCE FROM MINIMUM OF BASE RANGE

	Model 1	Model 2	Model 3	Model 4 (no outliers)	Model 5	Model 6
Judge variables						
Democrat	-8.07+ (4.49)	-6.56+ (3.71)	-7.22* (3.40)	-4.08+ (2.25)		
Aligned Democrat					-7.32 (4.46)	-5.49 (6.19)
Unaligned Democrat					-6.06 (4.27)	-8.32** (3.88)
Offender variables						
Drug offense		-24.26** (4.89)	-23.08** (4.72)	-24.26** (4.89)	-24.23** (4.86)	-23.08** (4.72)
Age		-0.39 (1.64)	-0.76 (1.63)	-0.25 (0.65)	-0.41 (1.65)	-0.73 (0.99)
Age sq		0.01 (0.02)	0.02 (0.02)	0.003 (0.01)	0.011 (0.02)	0.017 (0.02)
Female		-10.11* (3.84)	-10.23* (4.00)	-7.76** (2.19)	-10.17** (3.82)	-10.12** (4.00)
Black		1.50 (4.88)	1.72 (4.72)	1.49 (2.74)	1.47 (4.84)	1.67 (4.74)
Hispanic		-9.93* (3.71)	-5.12 (3.73)	-2.28 (2.86)	-9.99** (3.70)	-5.02 (3.70)
Constant	0.21 (3.58)	-17.61 (29.17)	-22.16 (26.80)	7.21 (13.39)	-17.09* (29.37)	-22.26* (26.77)
Circuit dummies	No	No	Yes	Yes	No	Yes
Grid controls	No	Yes	Yes	Yes	Yes	Yes
Cases	Non-departures	Non-departures	Non-departures	Non-departures	Non-departures	Non-departures
R-squared	0.01	0.24	0.33	0.33	0.27	0.27
Observations	762	732	732	689	732	732

Standard errors are robust and reflect clustering by judge.

+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

D. Appendix Table 4: Change in Prison Sentence from Minimum of Final Range

Theory predicts that district court judges inclined to give lower sentences should make greater use of departures when they face a sympathetic appellate body. The first three columns of Appendix Table 4 simply test differences between Republican and Democratic appointees. We find that Democratic appointees depart to a greater extent than Republican appointees (by roughly five to seven months). This is in contrast to our previous finding that there were no partisan differences in departure magnitudes unless we considered circuit alignment. There is an easy explanation for this difference, however, because the present study covers a much more recent timeframe, during which circuits are relatively more Democratic. Strong evidence for this position is found in the next few columns.

Models 4 and 5 both indicate that the Democratic appointees in Democratic-majority circuits grant departures on average twice as large as Democratic appointees in Republican circuits. In other words, the coefficient on *Democrat* in Models 1 through 3 comes almost entirely from Democratic appointees in Democratic circuits. Given the fairly large standard deviations, we cannot say as a statistical matter that the coefficient on *Aligned Democrat* is larger than the coefficient on *Unaligned Democrat*. Given the sample size, however, the results are as suggestive as can be hoped. Most importantly, they again conform well to our previous findings that Democratic appointees in aligned circuits gave 3- to 5.5-months-lower sentences relative to Democratic appointees in Republican circuits.

Models 6 and 7 employ a more flexible specification for circuit court politics: the percentage of Democratic appointees on the active circuit court bench. The variable *%Circuit Dem* is added to the regression and interacted with *Democrat*. The signs of the ideology variables are, as expected, negative. The coefficient on *Democrat* is a small -1 month, implying that, without alignment, there is little difference between Democratic appointees and Republican appointees on departure magnitude. The *%Circuit Dem* and *Democrat* coefficients are both negative and similar in magnitude, but are not independently significant. Their joint significance was 0.001, which is consistent with the strong alignment effects observed previously. Model 7 adds circuit dummies, and we remove the *%Circuit Dem* because there was so little variation in our timeframe.

2008]

Reviewing the Sentencing Guidelines

757

APPENDIX TABLE 4
CHANGE IN PRISON SENTENCE FROM MINIMUM OF FINAL RANGE

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Judge variables							
Democrat	-6.88** (2.44)	-5.47** (2.40)	-6.02** (2.27)			-1.03 (5.73)	-3.10 (5.73)
Aligned Democrat				-7.81* (3.13)	-8.50** (3.02)		
Unaligned Democrat				-3.40 (2.42)	-4.03+ (2.65)		
%Circuit Dem						-9.49 (10.38)	
Dem * %Circuit Dem						-8.00 (12.29)	-5.68 (11.29)
Offender variables							
Drug offense		-8.21** (2.06)	-7.64** (2.79)	-8.64** (3.02)	-7.40** (2.82)	-8.90** (3.03)	-8.90** (3.03)
Age		0.10 (1.01)	-0.22 (1.03)	0.31 (1.03)	-0.27 (1.02)	-0.01 (1.01)	0.17 (0.73)
Age sq		-0.001 (0.02)	0.004 (0.02)	-0.004 (0.01)	0.005 (0.02)	0.001 (0.02)	-0.002 (0.01)
Female		-8.21** (2.06)	-6.92** (1.92)	-9.37** (1.96)	-6.96** (1.90)	-8.13** (1.96)	-6.28** (2.03)
Black		3.55 (3.07)	3.03 (2.75)	3.07 (3.09)	3.21 (2.81)	3.20 (3.08)	7.93* (3.03)
Hispanic		-1.88 (2.42)	0.82 (2.32)	-2.71 (2.46)	0.66 (2.19)	-1.82 (2.43)	0.48 (2.55)
Constant	6.33** (1.94)	-16.42 (16.50)	-21.62 (16.37)	-23.75+ (12.72)	-29.60* (13.47)	-9.97 (17.50)	-29.65* (13.67)
Circuit dummies	No	No	Yes	No	Yes	No	Yes
Grid controls	No	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.01	0.06	0.21	0.06	0.21	0.06	0.06
Observations	936	887	887	887	887	887	887

Standard errors are robust and reflect clustering by judge.

+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

E. Appendix Table 5: Probability of Judge-induced Downward Departures

Appendix Table 5 presents probit regression results, in which the dependent variable equals one if the judge chooses to depart downward (substantial assistance departures are excluded from the sample). Thus, Appendix Table 4 measures the *degree* of departure, while Appendix Table 5 measures the *probability* of a departure. The two tables mesh very nicely, with likelihood of departures being greater when the judge is an aligned Democratic appointee and no greater otherwise. Again, in Columns (4) and (5), the proportion of the reviewing court that is Democratic-appointed, while not statistically significant, is positively correlated with the probability of downward departures. These results are actually stronger than those of our earlier work, which failed to find statistically significant ideological effects on sentencing (although the signs in general were, as expected, positive).

2008]

Reviewing the Sentencing Guidelines

759

APPENDIX TABLE 5
PROBABILITY OF JUDGE-INDUCED DOWNWARD DEPARTURES

	Model 1	Model 2	Model 3	Model 4	Model 5
Judge variables					
Democrat				-0.077 (0.096)	-0.039 (0.086)
Aligned Democrat	0.17** (0.052)	0.15** (0.047)	0.088* (0.044)		
Unaligned Democrat	-0.044 (0.029)	-0.067* (0.031)	-0.021 (0.034)		
%Circuit Dem				0.30 (0.17)	
Dem * %Circuit Dem				0.22 (0.18)	0.14 (0.16)
Offender variables					
Drug offense		0.022 (0.035)	0.004 (0.033)	0.004 (0.033)	0.006 (0.032)
Age		-0.013* (0.006)	-0.009 (0.006)	-0.009 (0.006)	-0.010 (0.006)
Age sq * 100		0.02* (0.009)	0.012 (0.009)	0.012 (0.009)	0.014 (0.008)
Female		0.15** (0.055)	0.16** (0.057)	0.16** (0.057)	0.16** (0.056)
Black		-0.087** (0.029)	-0.043 (0.027)	-0.043 (0.027)	-0.040 (0.034)
Hispanic		0.031 (0.034)	-0.0019 (0.034)	-0.0019 (0.034)	-0.0039 (0.034)
Circuit dummies	No	No	Yes	No	Yes
Grid controls	Yes	Yes	Yes	Yes	Yes
Observations	921	899	899	899	899

The estimates are probit regressions in which the dependent variable equals one in the event of a judge-induced departure. Marginal effects are reported (so the probability of receiving a downward departure is 8.8 percentage points greater for an aligned Democrat than for a Republican or unaligned Democrat). Standard errors are robust and reflect clustering by judge.

+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

APPENDIX FIGURE 1
SENTENCING TABLE (IN MONTHS OF IMPRISONMENT)

Offense Level	Criminal History Category (Criminal History Points)					
	I (0 or 1)	II (2 or 3)	III (4, 5, 6)	IV (7, 8, 9)	V (10, 11, 12)	VI (13 or more)
1	0-6	0-6	0-6	0-6	0-6	0-6
2	0-6	0-6	0-6	0-6	0-6	1-7
3	0-6	0-6	0-6	0-6	2-8	3-9
4	0-6	0-6	0-6	2-8	4-10	6-12
Zone A 5	0-6	0-6	1-7	4-10	6-12	9-15
6	0-6	1-7	2-8	6-12	9-15	12-18
7	0-6	2-8	4-10	8-14	12-18	15-21
8	0-6	4-10	6-12	10-16	15-21	18-24
9	4-10	6-12	8-14	12-18	18-24	21-27
Zone B 10	6-12	8-14	10-16	15-21	21-27	24-30
11	8-14	10-16	12-18	18-24	24-30	27-33
Zone C 12	10-16	12-18	15-21	21-27	27-33	30-37
13	12-18	15-21	18-24	24-30	30-37	33-41
14	15-21	18-24	21-27	27-33	33-41	37-46
15	18-24	21-27	24-30	30-37	37-46	41-51
16	21-27	24-30	27-33	33-41	41-51	46-57
17	24-30	27-33	30-37	37-46	46-57	51-63
18	27-33	30-37	33-41	41-51	51-63	57-71
19	30-37	33-41	37-46	46-57	57-71	63-78
20	33-41	37-46	41-51	51-63	63-78	70-87
21	37-46	41-51	46-57	57-71	70-87	77-96
22	41-51	46-57	51-63	63-78	77-96	84-105
23	46-57	51-63	57-71	70-87	84-105	92-115
24	51-63	57-71	63-78	77-96	92-115	100-25
25	57-71	63-78	70-87	84-105	100-25	110-37
26	63-78	70-87	78-97	92-115	110-37	120-50
27	70-87	78-97	87-108	100-25	120-50	130-62
Zone D 28	78-97	87-108	97-121	110-37	130-62	140-75
29	87-108	97-121	108-35	121-51	140-75	151-88
30	97-121	108-35	121-51	135-68	151-88	168-210
31	108-35	121-51	135-68	151-88	168-210	188-235
32	121-51	135-68	151-88	168-210	188-235	210-62
33	135-68	151-88	168-210	188-235	210-62	235-93
34	151-88	168-210	188-235	210-62	235-93	262-327
35	168-210	188-235	210-62	235-93	262-327	292-365
36	188-235	210-62	235-93	262-327	292-365	324-405
37	210-62	235-93	262-327	292-365	324-405	360-life
38	235-93	262-327	292-365	324-405	360-life	360-life
39	262-327	292-365	324-405	360-life	360-life	360-life
40	292-365	324-405	360-life	360-life	360-life	360-life
41	324-405	360-life	360-life	360-life	360-life	360-life
42	360-life	360-life	360-life	360-life	360-life	360-life
43	life	life	life	life	life	life