

Do Standards of Review Matter? The Case of Federal Criminal Sentencing

Joshua B. Fischman and Max M. Schanzenbach

ABSTRACT

We study whether changes to standards of review affect district court sentencing decisions under the U.S. sentencing guidelines. Departures from the guidelines by district judges have at times been reviewed strictly or deferentially. If review standards are constraining, then differences among judges should be larger when review is deferential. We find that Democratic appointees are more lenient than Republican appointees under deferential review, but this difference significantly narrows when review is strict. We conclude that district judges are meaningfully constrained by the prospect of appellate reversal. By contrast, judges appointed before the adoption of the guidelines are more likely to depart and issue shorter sentences, but their decisions are not significantly affected by the standard of review. We suggest that the constraining effect of appellate review varies with a judge's respect for the underlying legal regime.

1. INTRODUCTION

Prior to the enactment of the U.S. sentencing guidelines, federal district judges exercised virtually unreviewable sentencing discretion within broad statutory ranges. When Congress enacted the guidelines over 2 decades ago, it imposed a highly structured sentencing regime on district courts and, for the first time, opened federal sentencing decisions to

JOSHUA B. FISCHMAN is Associate Professor of Law, University of Virginia. MAX M. SCHANZENBACH is Benjamin Mazur Professor of Law, Northwestern University. The authors thank Rachel Harmon, Peter Low, Thomas Miles, Lydia Brashear Tiede, an anonymous referee, and conference and workshop participants at the Massachusetts Institute of Technology/Harvard University Positive Political Economy Workshop, the University of Virginia School of Law, the University of Southern California, the American Law and Economics Association annual meeting, the Political Economy and Public Law Conference, and the Conference on Empirical Legal Studies.

[*Journal of Legal Studies*, vol. 40 (June 2011)]

© 2011 by The University of Chicago. All rights reserved. 0047-2530/2011/4002-0012\$10.00

meaningful appellate review. In particular, the guidelines permit district judges to depart from specified sentencing ranges under certain circumstances, but this decision is subject to appellate scrutiny. Congress and the Supreme Court have modified the standard of review for departures on three occasions, alternating between strict and deferential review.

The changes to the standard of review applied to departures provide a unique opportunity to study the impact of appellate review on district judges' decision making. Although standards of review are often the subject of intense political and judicial debate, there is little empirical evidence that these standards have an effect on either district or circuit judges. Indeed, some guidelines scholars have doubted that changing from a *de novo* to an abuse-of-discretion standard would have much impact (see, for example, Stith and Cabranes 1998; Lee 1997).

We estimate the impact of changes in the standard of review on district judges by examining interjudge disparities in departure rates and prison sentences. Because we lack data on individual judges, we rely on district-level variation in judicial characteristics to identify the effect of changes to the standard of review. Consistent with prior studies, we find a significant difference between Democratic and Republican appointees, with Democratic appointees preferring lower sentences. However, this difference becomes significantly larger when review is deferential. These results are the first direct evidence that standards of review have an effect on substantive outcomes at the district court level. We also find that judges appointed prior to the adoption of the guidelines are more likely to depart from the guidelines and issue shorter sentences than judges appointed after the adoption of the guidelines. However, this difference is not significantly affected by the standard of review. We suggest that the constraining effect of appellate review may vary with a judge's respect for the underlying legal regime.

Two conditions are necessary for the standard of appellate review to substantively affect sentencing decisions at the district court level: district judges must be averse to reversal, and circuit court panels must themselves be constrained by review standards. Thus, our results have two important implications. First, we interpret district court sensitivity to standards of review as evidence that district judges are averse to reversal and respond prospectively to changes in standards of review. Second, our results provide indirect evidence that circuit court panels are also constrained by standards of review.

No single model of judicial behavior can fully explain our findings. The difference in sentencing behavior between Democratic and Repub-

lican district judges reflects differences in judicial attitudes toward sentencing policy. The sensitivity of district judges' sentencing decisions to the standard of review suggests that these judges are strategically modifying their behavior in response to the likelihood of reversal. Even this response is nuanced: we find that district judges may be less averse to reversal if they lack respect for the legal regime. Finally, we conclude that doctrine matters. The constraining effect of standards of review on circuit court panels is most consistent with preferences for adherence to doctrine, given the negligible likelihood of en banc or Supreme Court reversal of appellate sentencing decisions.

The paper is organized as follows. Section 2 analyzes the implications of changes to the standard of review in a positive political theory framework, Section 3 lays out the guidelines framework and recent changes to the standards of review, Section 4 explains the data and estimation strategy, Section 5 discusses the results, and Section 6 concludes.

2. POSITIVE POLITICAL THEORY OF STANDARDS OF REVIEW IN SENTENCING

The positive political theory literature on judicial hierarchy typically models judges as strategic actors who maximize policy preferences subject to constraints imposed by other actors (Shavell 2006; Cameron, Segal, and Songer 2000; McCubbins, Noll, and Weingast 1995). Lower court judges, in particular, are assumed to decide cases in a manner that maximizes conformity with their policy preferences while avoiding reversal by a higher court. In some models, the lower court judges suffer an exogenous disutility from reversal (for example, Shavell 2006); in others, a reversal penalty may arise endogenously if the higher court moves doctrine further from the lower court's ideal point (for example, McCubbins, Noll, and Weingast 1995; Spitzer and Talley 2000).

While these theories are intuitively appealing, they have little empirical support. Some studies have demonstrated that measures of higher court ideology are correlated with lower court decision making (for example, Songer 1987; Songer and Sheehan 1990). However, these results are consistent both with strategic behavior and with lower courts faithfully applying precedents issued by a superior court.

Other studies have attempted to disentangle strategic and legal motivations for lower court compliance, but none of these have found compelling evidence of the former. Cross (2005) examines the impact of current and lagged Supreme Court preferences on the behavior of circuit court judges but does not find evidence that the current prefer-

ences of the Supreme Court justices have an impact. Songer, Segal, and Cameron (1994) find that circuit court panels adhere more closely to Supreme Court preferences in cases that are more likely to be appealed, but Klein and Hume (2003) do not. However, the approach taken in these papers has been criticized for failing to account for selection effects (Kim 2007; Lax and Kastellec 2008).

Much of the empirical literature on judicial hierarchy focuses on the relationship between the circuit courts and the Supreme Court. In this context, however, the constraining effects of appellate review will be weakest, given the low rate of Supreme Court review. For this reason, many commentators doubt that the Supreme Court can maintain control over the circuit courts through the threat of reversal (see, for example, Posner 2008; Kim 2007).

Schanzenbach and Tiller (2007) study the interplay between circuit and district courts under the sentencing guidelines. They find that Democratic district judges calculated lower offense levels than Republicans and that Democratic district judges also granted larger departures when supervised by circuit courts dominated by Democrats. This is evidence that circuit courts influence district judges' sentencing decisions, but the results are not directly attributable to more searching review standards for departures. It is possible that district judges were simply applying circuit court precedent concerning departures and that such precedent became more favorable to defendants in more liberal circuits. Moreover, although departures are reviewed more strictly than adjustments, Schanzenbach and Tiller note that the results are also potentially attributable to other institutional factors that shield adjustment calculations from review, such as their fact-intensive nature and the lower stakes involved. Finally, their results say nothing about whether circuit courts are themselves constrained by verbal formulations of review standards; conceivably, circuit courts could reverse sentences as freely under an abuse-of-discretion standard as under a *de novo* standard.

In contrast to these previous studies, we directly examine the effect of a change in review standards on district judges' departure decisions—a context in which review standards are most likely to be effective and in which the underlying substantive law was largely unchanged.¹ The

1. A subtle but important distinction is that we examine the threshold decision whether to depart, whereas Schanzenbach and Tiller (2007) examine the magnitude of departures. The magnitude of a departure has always been reviewed for reasonableness, whereas whether a departure was warranted in the first place has been subject to changing standards of review. In Schanzenbach and Tiller's data time frame, there was only one change to the

changes in standards of review for sentencing departures provide a reasonable basis for disentangling motivations for district judges' adherence to higher court preferences. If district judges are unconcerned with the prospect of reversal, then changes in standards of review unaccompanied by changes in substantive law should have no effect on district judges' decision making. On the other hand, if district judges are averse to reversal, then their decisions should be influenced by standards of review, and interparty disparities should be larger under more lenient standards of appellate review.

The above discussion relies on a presumption that deferential review standards have a constraining effect on circuit courts in sentencing decisions.² Some studies have sought to measure the effect of review standards on circuit courts directly by examining affirmance rates under varying levels of appellate deference. These results have been mixed.³ However, the study of affirmance rates is problematic because affirmance rates are endogenous, being determined jointly by the degree of appellate deference and the behavioral response of litigants and trial courts (Spitzer and Cohen 1994). We do not directly examine circuit court decisions in this study but note that our findings indirectly support an inference that standards of review have an impact on circuit courts' sentencing decisions.

3. THE GUIDELINES

The Sentencing Reform Act of 1984 charged the United States Sentencing Commission with developing guidelines that would “provide certainty and fairness in meeting the purposes of sentencing, avoiding unwarranted disparities among defendants with similar records who have been found guilty of similar criminal conduct” (28 U.S.C. sec. 991[b]). Pursuant to this legislative directive, the Sentencing Commission developed a sentencing table (see Table A1). The recommended sentencing ranges

standard of review—the *Koon* decision (*Koon v. United States*, 518 U.S. 81 [1996]). With 4 fewer years of data, they were not able to identify differences along party lines with respect to the departure decision.

2. Although recent empirical literature on appellate review suggests that circuit judges are subject to ideological influences (for example, Revesz 1997; Sunstein et al. 2006; Miles and Sunstein 2006, 2008), these findings are not inconsistent with a constraining effect of legal doctrine (Kastellec 2007; Tiller and Cross 1998).

3. Eskridge and Baer (2008) and Schuck and Elliott (1990), for example, find that more deferential standards of review correspond to higher affirmance rates for administrative agencies, but Verkuil (2002) and Zaring (2010) find only a weak association between review standards and agency win rates across a broader set of agency decisions.

in the table are determined by the offense level and criminal history category. A district judge, with the aid of the probation officer, uses the Sentencing Commission's regulations to calculate the defendant's numeric offense level. The crime of conviction sets the base offense level, with adjustments being made to this offense level on the basis of a variety of factors, such as the use of a gun, the use of sophisticated means in a fraud, whether the offender played a major or minor role in the crime, and acceptance of responsibility, resulting in a final offense level. The offender's criminal history category is calculated on the basis of the prior offenses committed by the offender. These two factors—final offense level and criminal history—yield a sentencing range expressed in months. The lower bound of the sentencing range is roughly 75 percent of the maximum sentence.

If a judge determines that the guidelines are not appropriate to the unique factual situation before the court, he or she may depart from the guidelines. The judge must defend the departure in open court or in a written opinion and specify the reasons for the departure. Both the government and the defendant may appeal adverse decisions. Prior to *United States v. Booker* (543 U.S. 220 [2005]), a sentence within a properly calculated guidelines range was not reviewable by a circuit court, and the guidelines therefore provided a safe harbor. Calculation of the sentencing range itself was reviewable, but mostly under a clear-error standard. Since *Booker*, a judge must still calculate a sentencing range, but within-guidelines sentences are now reviewable. However, because post-*Booker* circuit courts have almost uniformly declined to reverse within-guidelines sentences, they remain de facto safe harbors, and the Supreme Court has allowed circuit courts to attach a presumption of "reasonableness" to all within-guidelines sentences (see *Rita v. United States*, 551 U.S. 338 [2007]).

Although over 90 percent of convictions are the product of plea bargains, judges still have a significant influence on offenders' final sentences. Following a plea agreement, there is a sentencing hearing at which the sentencing judge may make adjustments to the plea-bargained base offense level on the basis of additional findings of fact. The judge may also grant a departure if he or she deems the circumstances to be outside the "heartland" of the guidelines. For example, after pleading guilty to fraud, a defendant could dispute the amount stolen at the sentencing hearing, which could result in an adjustment to the offense level. Or the defendant could plead guilty to drug trafficking, and at the sentencing hearing the judge could adjust the offense level downward on the

grounds that the defendant accepted responsibility.⁴ Even if the plea bargain stipulates specific facts that bear on sentence enhancements or reductions, the judge need not accept them. Indeed, the judge may throw the plea out altogether and proceed with sentencing on the basis of his or her own evaluation of the case.

When the plea agreement is made, the identity of the sentencing judge is usually known to the parties. Therefore, the plea bargain occurs in the shadow of the sentencing judge. LaCasse and Payne (1999) find that judges exert a strong influence over the substance of a plea agreement, and this influence changed little postguidelines. In addition, any stipulations of fact by the prosecution and defense are likely made with an eye toward the judge who will be conducting the sentencing hearing and, hence, are still reflective of the judge's preferences.

In sum, the U.S. sentencing guidelines provide judges with two instruments with which to alter criminal sentences: adjustments and departures. Adjustments to the sentencing range are largely fact-bound decisions explicitly authorized by the guidelines and have a predetermined effect on the sentencing range. For the most part, these adjustments are reviewed for clear error. On the other hand, when a judge departs from the presumptive range specified by the guidelines, a stricter standard of review applies, though this standard has changed over time. The changes in the standard of review for departures are the core of our identification strategy.

Since the guidelines became effective in 1987, there have been several changes to how circuit courts review departures. The Sentencing Reform Act directed courts of appeals to "give due deference to the district court's application of the guidelines to the facts" but did not specify a standard of review (18 U.S.C. sec. 3742[e][4]). The decision to depart is typically a mixed question of law and fact. In order for a judge to depart from the guidelines, he or she must make findings of fact concerning the uniqueness of the case and then make a legal argument that the findings place the case outside the heartland of the guidelines. Thus, as a matter of law, there must be factors present in the case that were not taken into account by the Sentencing Commission or are present to such an unusual degree that the guidelines could not have been intended to apply.

Courts often struggle with the correct standard of review to apply

4. For these reasons, some have asserted that the most important part of the modern criminal process is the sentencing hearing. See, for example, Bibas (2001).

to mixed questions of law and fact. In an early influential case, *United States v. Diaz-Villafane* (874 F.2d 43 [1st Cir. 1989]), the First Circuit resolved this problem for departures by adopting a different standard of review for each portion of the departure decision.⁵ First, the factual findings justifying a departure were reviewed under the “clearly erroneous” standard. Second, the district court’s determination of whether the circumstances justified a departure was reviewed de novo. Third, the magnitude of the departure was reviewed for abuse of discretion. By the early 1990s, all of the other circuits had adopted this framework.

The First Circuit, however, abandoned its own standard in 1993. In an opinion written by then-judge Stephen Breyer, it applied a reasonableness standard to the entire departure decision (*United States v. Rivera*, 994 F.2d 942, 951–52 [1993]).⁶ The reasoning behind this change was that district courts have greater institutional competence to make decisions about whether or not a case is unusual, and consequently a deferential standard of review should apply to the decision to depart (994 F.2d 950). The First Circuit’s reasonableness standard was adopted by the Supreme Court in *Koon v. United States* in 1996, although the Court used the phrase “abuse of discretion,” rather than “reasonableness,” to describe the standard of review (518 U.S. 98–99). In effect, *Koon* relaxed the standard of review for all circuits but the First Circuit.

The *Koon* abuse-of-discretion standard was repealed by Congress in the PROTECT Act (Prosecutorial Remedies and Other Tools to End the Exploitation of Children Today Act of 2003, Pub. L. 108–21, sec. 401 [codified and amended at 18 U.S.C. secs. 2423 (a)–(g), 2246, 2516(1)(c), 1591, 3142(e), 3283, 3559(e)]), which reinstated a de novo standard of review for the determination that a departure was justified. In addition, the PROTECT Act placed new limitations on grounds for departures and imposed reporting requirements on prosecutors that encouraged appeals of departures.

Less than 2 years after its adoption, the PROTECT Act was partially

5. For a collection of cases and analysis, see Johnson (1998).

6. *Rivera* was also an influential opinion, and application of the tripartite standard of review in other circuits appears to have become less consistent after *Rivera*. See, for example, *United States v. Canoy*, 38 F.3d 893, 908 (7th Cir. 1994): “When a district court clearly explains the basis for its finding . . . [it] is entitled to considerable respect on appeal” (citing *Rivera*); *United States v. Simpson*, 7 F.3d 813, 820 (8th Cir. 1993) (favorably citing *Rivera*).

struck down on constitutional grounds in *United States v. Booker*.⁷ *Booker* held that in order to avoid running afoul of the Sixth Amendment guarantee of trial by jury, sentencing judges must take the guidelines as advisory only, and appellate courts must review the sentence under a “reasonableness” standard. One could view *Booker* as mandating a return to the abuse-of-discretion standard. However, the *Booker* standard is even more deferential because the Court’s reasoning encouraged more departures, undermined the legitimacy of the guidelines, and made them advisory, at least to the extent that within-guidelines sentences became reviewable.

The changes to standards of review over time combined with natural variation in the composition of the district and circuit courts have created multiple natural experiments. We exploit these natural experiments to test whether district judges react strategically to the changes in standards of review.

4. THE DATA AND IDENTIFICATION STRATEGY

A great advantage to studying federal sentencing is the detailed data on all sentences issued pursuant to the sentencing guidelines. The important outcome variables—sentence length, adjustments, and departures—are all coded objectively. In addition, all convicted federal offenders must be sentenced under the guidelines, and the federal conviction rate (including plea bargains) is around 95 percent. Thus, there is no concern that selection effects in litigated cases could confound our estimates (Priest and Klein 1984; Lax and Kestelc 2008). Table A2 gives the descriptive statistics for relevant outcome and control variables in our sample.

We use data from 1991 to 2007.⁸ In our preferred regressions, we exclude several categories of cases. However, we perform robustness checks to verify that our results are not sensitive to the selection of the subsample. First, we exclude departures on substantial assistance

7. Six months before *Booker*, the Supreme Court struck down a portion of a state guidelines system in *Blakely v. Washington*, 542 U.S. 296 (2004), which cast doubt on the validity of the federal guidelines. We considered whether there was an anticipatory effect in light of *Blakely*. A review of the data suggests some possibility for an anticipatory effect, but given our strong findings that there was an impact on departures and prison sentences, we conservatively code *Booker* as a pure shock.

8. The guidelines years prior to 1991 are excluded because of uncertainty about the standard of review, coding issues, and the presence of substantial numbers of nonguidelines cases.

grounds, since they require both that the defendant provide useful evidence and that the prosecution move for such a departure. The role of judicial discretion in these cases is strongly curtailed. To account for the possibility that substantial assistance departures could be partial substitutes for judge-induced departures or adjustments, we report some specifications including substantial assistance departures. We find that their inclusion does not greatly affect the coefficients of interest. In addition, we examine whether changes to the standard of review for judge-induced departures affect substantial assistance departure rates. We find no evidence that substantial assistance departures are correlated with changes to the standard of review.

Second, we exclude all cases from five border districts: Arizona, southern California, New Mexico, and south and west Texas. These districts maintained an informal system of so-called fast-track departures that granted a departure pursuant to deportation. These departures were codified by the PROTECT Act and thus might confound the estimation. In addition, these districts experienced a relatively rapid growth in case-loads over the sample time frame, which contributes to concerns about compositional changes in the sample. Nonetheless, when border districts are included in the estimation, the magnitude of the coefficients of interest increases.

Finally, we focus on serious crime categories, which have substantially higher sentences and are likely more politically salient. These crime categories—murder, manslaughter, sex crimes, robbery (primarily bank robbery), drug trafficking, firearms offenses, racketeering, arson, and auto theft—constitute the majority of prison sentences and almost 90 percent of all prison time imposed under federal criminal law. Less serious crimes, such as fraud and larceny, often start with low base offense levels resulting in guidelines sentencing ranges that do not require a prison sentence. There is no reason for a judge to depart in such a case, since the range prescribed by the guidelines already includes the minimum sentence.

4.1. Departure Rates and Guidelines Sentencing over Time

Initial studies of the effect of the PROTECT Act and *Booker* suggest that both affected departure rates. For example, the United States Sentencing Commission (2006) concluded that the PROTECT Act reduced the judge-induced departure rate. Freeborn and Hartmann (2010) confirm this finding and also conclude that the PROTECT Act did not affect offense-level calculations. The Sentencing Commission also concluded

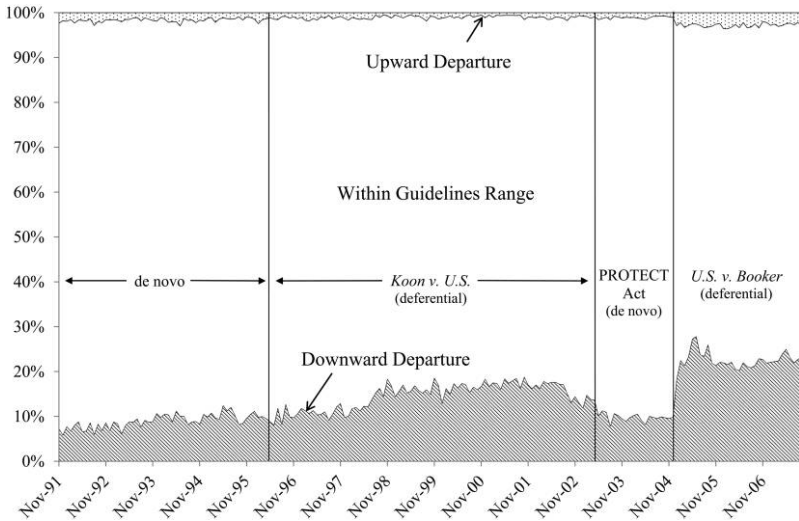


Figure 1. Upward and downward departures over time

that *Booker* substantially increased both downward and upward departures. However, *Booker* appears to have had little effect on overall prison sentences (United States Sentencing Commission 2006), although its impact may have varied among the circuits (Tiede 2009). We are aware of only one examination of the *Koon* decision, which concluded that it had little effect on aggregate departure rates (Hofer, Martin, and Montgomery 1997).

Figure 1 divides sentences into three categories—sentences within the guidelines range, downward departures, and upward departures—and graphs how the share of sentences for serious crimes represented by each category changes over time (substantial assistance departures are excluded). In the period before *Koon*, when most circuits reviewed departures *de novo*, 5–10 percent of sentences were downward departures from the guidelines. This level gradually increased to 15–18 percent after *Koon*, which relaxed the standard of review, and then declined below 10 percent after the PROTECT Act, which reinstated *de novo* review. There was an immediate jump in downward departures after *Booker*, reaching a rate of about 25 percent. Upward departures are quite rare

but increased from 1 percent to about 3 percent immediately after *Booker*.⁹

Figure 1 provides some perspective on changes in sentencing practices. The changes in departure rates after the PROTECT Act and *Booker* are quite intuitive, although the post-*Koon* change is less dramatic and more gradual. Before making a causal inference, we must disentangle changes in review standards from other trends in sentencing, such as the changing political composition of the bench and changes in the composition of the caseload. For example, the increase in departure rates post-*Koon* coincides with an increasingly Democratic bench, while a decrease in departure rates post-PROTECT Act coincides with an increasingly Republican bench. Indeed, it appears that departures may have begun declining in the years prior to the PROTECT Act, coinciding with the increasing proportion of Republican judges during the administration of George W. Bush.

4.2. Judicial Characteristics

Ideally, we would match the sentencing judge to each sentencing outcome, which would permit a straightforward examination of interjudge disparity. However, the sentencing data do not identify the sentencing judge, and the Sentencing Commission will not release judge identifiers. Instead of judge identifiers, we rely on district-level variation in the observable characteristics of sentencing judges to proxy for judges' likely sentencing preferences. Figure 2 offers a graph of changes in district and circuit characteristics over time.

A large prior literature has established that judicial ideology, as proxied by the party of the president who appointed the judge, is an important predictor of judicial decisions.¹⁰ Some studies have found that other characteristics such as race, sex, and professional background are also

9. The departure rates we report are nearly twice as high as the changes reported by the United States Sentencing Commission (2006, app. E, p. E-1). This is because we remove substantial assistance departures from the denominator, while the Sentencing Commission includes them, and we consider only serious crimes.

10. Although party of appointment has been criticized as a crude measure of judicial ideology (Epstein and King 2002; Edwards 1998), it has also been shown to have predictive power across a wide range of contexts (Pinello 1999). Some scholars have advocated the use of "judicial common space scores" (Giles, Hettinger, and Peppers 2001) instead of the party variable, but recent research has shown that these two measures have nearly equal predictive power in many applications (Fischman and Law 2009; Sisk and Heise 2005). We employ the party variable here because it is easy to interpret and because the district-level variation requires a dichotomous specification, which is more straightforward with the presidential party formulation.

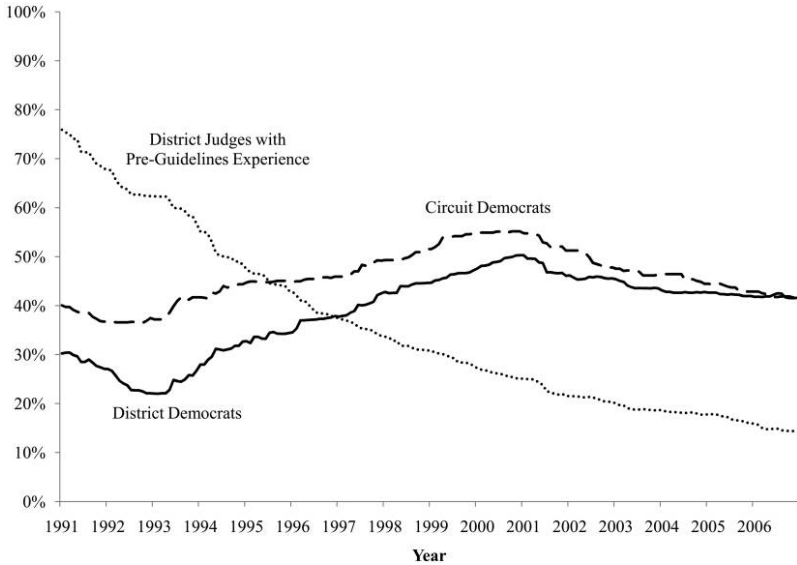


Figure 2. Percentages of Democrats and pre-guidelines judges

significant predictors of judicial decisions in circumstances in which the characteristic is relevant, such as racial or sex discrimination.¹¹ Using data from the Federal Judicial Center (2010), we obtained the party of the president who appointed a judge and the judge's race, gender, age, prior prosecutorial experience, and prior experience as an academic. We calculated the monthly average of each of these variables by district using data derived from the Federal Judicial Center biographical data on federal judges. As expected, only the party of the appointing president was significantly correlated with sentencing outcomes.

We also identify an additional factor that may be important in the unique context of the guidelines: whether the judge's appointment pre-

11. For example, a judge's race may matter in voting rights cases (Cox and Miles 2008), and a judge's sex has been found to influence outcomes in sex discrimination cases (Boyd, Epstein, and Martin 2010; Peresie 2005). However, Schanzenbach (2005) fails to find that district-level judge characteristics such as race and gender affect average sentences, though such characteristics may matter in some cases. For surveys, see Fischman and Law (2009) and George (2001).

dated the guidelines' adoption in 1987.¹² The guidelines substantially reduced district judicial discretion in a sphere in which, subject to statutory maximums and minimums, it had previously been unfettered. Judges appointed after the guidelines would not have experienced the preguidelines regime, except possibly as a prosecutor or defense counsel.

Two district judges, Nancy Gertner (2005) and Jack Weinstein (1992), have suggested that judges appointed after the introduction of the guidelines would be more accepting of guidelines sentencing because they were not accustomed to exercising judicial discretion in sentencing. In addition to having prior experience with discretion, many of the preguidelines judges evinced a strong hostility to the guidelines when they were first enacted. Within 1 year of passage, an astonishing 179 district judges had invalidated the guidelines on constitutional grounds.¹³ In the parlance of models of judicial behavior, preguidelines judges who opposed the guidelines may have faced lower legitimacy costs if their sentencing decisions were overturned. The guidelines also appear to have made judges more likely to retire. Boylan (2004) found that after the sentencing guidelines were introduced, judges retired 4.8 months after they became eligible for senior status, whereas the previous average had been 3 years after eligibility.¹⁴

4.3. Regression Model and Identification

The random assignment of offenders to sentencing judges within each district ensures that the proportion of Democratic and preguidelines

12. We thank Thomas Miles for suggesting some of these possibilities to us, in particular the possibility that preguidelines judges may have had a different view of the guidelines. Preguidelines experience is highly correlated with duration of service, and these two variables are difficult to disentangle. When included separately in regressions, both were significant predictors of downward departures. When included together, neither was independently significant, but they were jointly significant. However, we conducted another regression using only districts with no preguidelines judges, and duration of service was no longer significant. We interpret our results to be driven by preguidelines experience but cannot be certain that they are not driven in part by judicial experience.

13. For details, see Sisk, Heise, and Morriss (1998), who did not find a significant partisan effect in the probability of a court striking down the guidelines. The Supreme Court upheld the guidelines as constitutional in *United States v. Mistretta*, 488 U.S. 361 (1989).

14. In addition, surveys of judges suggest that hostility toward the guidelines may have declined over time. In 1996, a survey conducted through the Federal Judicial Center found that almost three-quarters of trial judges opposed the sentencing guidelines (Johnson and Gilbert 1997). A survey in 2003 did not repeat the exact question asked by the Federal Judicial Center but indicated that 78 percent of judges surveyed thought the guidelines were doing a reasonable job of furthering the purposes of sentencing (O'Neill 2003). These

appointees in a district accurately reflects the probability that any given offender will be assigned to a Democratic or preguidelines judge. Random assignment also ensures that an offender's characteristics will be uncorrelated with the characteristics of the assigned sentencing judge.

In addition, it is reasonable to assume that the observable characteristics of the judges in each district are not correlated with the aggregate characteristics of offenders in that district; that is, Republican judges do not self-select into particular districts. Changes in the political composition of each district are determined by whether vacancies occur during Republican or Democratic administrations, and the literature on judicial retirement has found (at least postguidelines) that district judges' retirement is driven primarily by apolitical factors, such as health and the vesting of pensions (Yoon 2006). The proportion of judges appointed prior to the guidelines would be driven by similar factors. Some districts may systematically have more liberal judges than other districts, but that is not a difficulty for our analysis. Presumably, a Democratic appointee would be, in expectation, more liberal than the district average and a Republican appointee would be more conservative than the district average. Since our model includes fixed effects for each district, we estimate the effect of each district having a composition that is more liberal or conservative at any particular point in time than the average for that district.

We define our dependent variable *DownDepart* to equal one if the judge issued a downward departure and zero otherwise, and let *DemDistrict* denote the proportion of Democratic appointees in each district. It follows that

$$\begin{aligned} \Pr(\text{DownDepart}) &= \Pr(\text{DownDepart}|\text{DemJudge}) \times \text{DemDistrict} \\ &+ \Pr(\text{DownDepart}|\text{RepJudge}) \times (1 - \text{DemDistrict}). \end{aligned} \quad (1)$$

Therefore, the regression is linear in *DemDistrict*, the primary variable of interest. This equation shows that employing variation at the district level will identify the effect of Democratic appointees on downward departures, provided that the underlying assignment is random.¹⁵ The same identity would hold with regard to the proportion of preguidelines judges.

survey results are also suggestive evidence that preguidelines judges were more hostile to the guidelines.

15. The use of district-level variation is the same approach as in Schanzenbach (2005) and Schanzenbach and Tiller (2007). More recently, Price and Wolfers (2010) employ a similar approach, using average characteristics among National Basketball Association referees in a game to identify the effects of referee and player race on calls, where specific calls are not attributable to particular referees in the data.

Since equation (1) is linear in DemDistrict, we employ a linear probability model and report heteroskedasticity-robust standard errors clustered at the district level.¹⁶ Our regression equation takes the following form:

$$\begin{aligned}
 \text{DownDepart}_{ijt} = & \alpha + \beta_1 \text{DemDistrict}_{jt} + \beta_2 \text{Preguidelines}_{jt} \\
 & + \gamma \text{Deferential}_t + \delta_1 \text{Deferential}_t \times \text{DemDistrict}_{jt} \\
 & + \delta_2 \text{Deferential}_t \times \text{Preguidelines}_{jt} \\
 & + \zeta \text{PostBooker}_t + \eta_1 \text{PostBooker}_t \times \text{DemDistrict}_{jt} \quad (2) \\
 & + \eta_2 \text{PostBooker}_t \times \text{Preguidelines}_{jt} + \theta \text{DemCircuit}_{jt} \\
 & + \kappa \text{Year}_t + \lambda \text{OffenderDemog}_{ijt} + \mu \text{OffenseType}_{ijt} \\
 & + \nu \text{Grid}_{ijt} + \rho \text{District}_j + \varepsilon_{ijt},
 \end{aligned}$$

where i indexes offender, j indexes district, and t indexes the year and month of sentencing. The variables DemDistrict and Preguidelines represent the proportion of Democratic appointees and judges with preguidelines experience in a district for a given month. The variable Year contains dummies for the guideline year (November through October);¹⁷ OffenderDemog is a vector of offender demographic variables including age, race, gender, and number of dependents; OffenseType is a set of dummy variables for broad offense categories as defined by the Sentencing Commission; and Grid is a set of controls for position on the sentencing grid. To approximate the sentencing grid in the Appendix, the Grid controls are entered as dummies for base offense level and criminal history category and an interaction term between criminal history category and base offense level. The variable District is a matrix of district fixed effects, which means that DemDistrict measures the effect of changes in the percentage of Democratic appointees in the district. We will also report some specifications that include district-specific linear trends to verify that preexisting trends correlated with DemDistrict do not account for the results. The variable

16. Marginal effects from probit models yielded very similar results. However, a probit model imposes a nonlinear relationship between DemDistrict and the probability of departure, which is inappropriate. The linear probability model also provides easier interpretation of the marginal effect of the interaction between the standard of review and the district's composition.

17. The guidelines are frequently amended by the Sentencing Commission, and the rules that apply are published in the Federal Sentencing Guidelines Manual, which applies from November to October.

DemCircuit measures the proportion of Democratic appointees in the circuit in which the sentence would be reviewed.¹⁸

The changes in the standard of review are captured by Deferential, which takes on a value of one under the abuse-of-discretion regimes (*Koon* and *Booker*).¹⁹ PostBooker indicates whether a sentence was issued after *Booker*. Because Deferential also covers the post-*Booker* period, the Booker variable measures the difference between deferential review under mandatory guidelines and deferential review under advisory guidelines. Since changes to the standard of review are mostly uniform across circuits and we include year effects, our focus is not on the Deferential variable but rather its interaction with the judge characteristics.

The coefficients β_1 and β_2 can be interpreted as the impact of a Democratic judge and a judge with preguidelines experience, respectively, on the likelihood of a departure. The primary coefficients of interest are δ_1 and δ_2 , which measure the impact of these judge characteristics under deferential review relative to de novo review. For example, if δ_1 is statistically significant and of the same sign as β_1 , then deferential review amplifies differences between Republican and Democratic appointees. Similarly, η_1 and η_2 measure the change in the effect of judicial characteristics as a result of the guidelines being rendered advisory.

There are two caveats to the approach outlined here. The first is that sentencing judges may be influenced by their peers in the same district, which might yield nonlinearities in the effect of DemDistrict. We directly tested for this by adding the square of DemDistrict in our base models and found little evidence of nonlinearities (the coefficient on the squared term was very small and statistically insignificant, while the linear term remained basically unchanged and strongly significant).

The second caveat concerns the presence of senior judges. The results we report include only active district judges, but judges who take senior status remain on the bench usually at a substantially reduced caseload. A senior district judge who has a distaste for sentencing or the guidelines may remove himself from the criminal draw. According to the Federal

18. We weighted senior circuit court judges half as much as active circuit court judges. We collected a random sample of judges across circuits who took senior status during the guidelines period and estimated that they heard approximately half as many sentencing cases subsequent to taking senior status.

19. We set Deferential to equal one under post-*Koon* abuse-of-discretion review, post-*Booker* reasonableness review, and the First Circuit's *Rivera* standard. Deferential equals zero when review is de novo.

Judicial Center (2010), roughly 20 percent of combined civil and criminal cases are resolved by senior district court judges, who maintain on average a 50 percent caseload. Results weighting senior district judges at 50 percent were qualitatively very similar, though the magnitude of the coefficients tended to increase slightly.

5. RESULTS

Our first estimates examine the effect of standards of review on the two instruments available to the sentencing judge: departures and adjustments. Review of departures was directly affected by *Koon*, the PROTECT Act, and *Booker*. Although the standard of review for adjustments did not change, adjustments are potential substitutes for departures because they also have the effect of shortening or lengthening sentences. Second, we estimate the effect of standards of review on overall prison sentences. We estimate the prison sentence regressions in reduced-form versions using the basic specification given above. This specification conditions on the base offense level, which is plausibly exogenous in that base offense levels are largely set prior to sentencing by the crime of conviction. We also present results that condition on final offense level, which will help us to isolate the effect of departures on prison sentences.

Table 1 presents results for departures, with the dependent variable equal to one if a judge granted a departure and zero otherwise. In the first specification, we include characteristics of the district and circuit judges but not legal changes. The coefficients on the proportion of Democrats and preguidelines judges are both positive and significant.²⁰ These results suggest that Democratic appointees are almost 6 percentage points more likely to depart downward than are Republican appointees ($p < .01$). Given that the overall departure rate for serious crimes in nonborder districts is 13 percent, this is a sizable difference. Similarly, preguidelines judges are over 5 percentage points more likely to depart downward than judges who joined the bench after the guidelines ($p < .05$).

The proportion of Democratic appointees in a circuit court also has a significant impact on its departure rates. Moving from a circuit with

20. Schanzenbach (2005) found no difference in departure probabilities by district court characteristics. This result seems to have been due to a lack of power, as the present paper includes 6 more years and finds a robust effect. In addition, the present paper explicitly considers the effect of review standards, and stricter review tends to mute the ideological impact.

Table 1. Downward Departures (Judge-Induced Downward Departure Equals One)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DemDistrict	.057** (.021)	.024 (.024)	.030 (.024)	-.015 (.025)	.024 (.016)	.057 (.041)	.020 (.030)
Preguidelines	.052* (.020)	.086** (.024)	.069** (.024)	.029 (.017)	.061** (.018)	.066 (.059)	.038 (.045)
CircuitDem	.13* (.058)	.12* (.058)	.13* (.058)	.003 (.054)	.070 (.04)	.16 (.10)	.14* (.078)
Deferential	.013* (.006)	-.000 (.014)	.001 (.012)	-.001 (.008)	.001 (.008)	-.021 (.019)	-.018 (.012)
Deferential × DemDistrict		-.047* (.022)	.042* (.018)	.031** (.012)	.028* (.013)	.074** (.024)	.057* (.025)
Deferential × Preguidelines		-.021 (.019)	-.016 (.017)	-.003 (.011)	-.008 (.012)	.018 (.031)	.019 (.029)
Booker	.14** (.008)	.14** (.008)	.14** (.008)	.099** (.012)	.097** (.012)	.14** (.008)	.089** (.020)
Booker × DemDistrict			.014 (.030)	.018 (.028)	.031 (.026)	-.008 (.032)	.013 (.032)
PostBooker × Preguidelines			-.020 (.033)	-.001 (.024)	-.012 (.030)	-.020 (.033)	-.031 (.040)
(Deferential × DemDistrict) – (De novo × DemDistrict)		.046** (.013)	.043** (.011)	.030** (.010)	.030** (.008)	.053** (.014)	.041** (.013)
(Deferential × RepDistrict) – (De novo × RepDistrict)		-.000 (.014)	.000 (.011)	-.001 (.008)	.001 (.008)	-.021 (.019)	-.018 (.016)
Crime categories	Serious	Serious	Serious	Serious	Serious	Serious	All
District trends	No	No	No	Yes	No	No	No
Substantial assistance	No	No	No	No	Yes	No	Yes
Border	No	No	No	No	No	Yes	Yes
N	255,182	255,182	255,182	255,182	352,720	341,343	815,648

Note. Robust standard errors clustered by district are in parentheses. All regressions condition on dummies for district, year, base offense level, criminal history, race (black or Hispanic), sex, citizenship status, numbers of dependents (zero, one, two, or more), and continuous variables age, age squared, and base offense level × criminal history.

[†] $p < .10$.
^{*} $p < .05$.
^{**} $p < .01$.

25 percent Democrats to a circuit with 75 percent Democrats (corresponding roughly to the observed range in the data) increases departures by about 6 percentage points. The impact of the circuit composition on district judges could reflect either reversal aversion or the faithful application of circuit court precedent. This result, however, does not survive the inclusion of district-specific trends in column 4.

Column 2, which accounts for the changes in the standard of review, presents a more nuanced picture. The political differences among district judges are evident only when appellate review is deferential. The coefficient on DemDistrict under de novo review is not significant. However, the coefficient on Deferential \times DemDistrict is statistically significant and, at 4.7 percentage points, roughly twice the size of the coefficient on DemDistrict. Thus, the difference between Democrats and Republicans is significant only under deferential review. In addition, Democrats are 4.6 percentage points more likely to depart under an abuse-of-discretion review than under a de novo standard ($p < .01$).

By contrast, deferential review appears to have little effect on the behavior of preguidelines judges or Republican appointees, and these two results are consistent across all specifications and outcome variables. Preguidelines judges tend to grant more departures but do so equally under a de novo or abuse-of-discretion regime. The invariance of Republican judges does not appear to be due to hostility to the guidelines but more likely reflects a weaker inclination to depart in the first place—that is, they are largely inframarginal. This may be because Republican appointees favor higher sentences or, given that all judges depart more after *Booker*, are simply more willing to assume a limited judicial role in sentencing by adhering to the guidelines sentencing ranges.

Column 3 of Table 1 includes the changes in the standard of review and an additional term for cases decided after the *Booker* decision. The Deferential dummy is as before and equals one under *Booker* and in the period between *Koon* and the PROTECT Act. Thus, the PostBooker dummy measures the marginal effect of *Booker* beyond the effect of deferential review. We estimate a marginal effect for *Booker* because it not only reverted to a deferential standard but also changed substantive law by declaring the guidelines to be advisory. The basic results from the previous regressions survive: Democratic district judges depart more under deferential review, and departures still increase with the proportion of Democratic judges in the circuit. The PostBooker coefficient is large and positive, suggesting that the departure rate nearly doubles.

This is in line with the raw numbers detailed in Figure 2. Thus, both Republicans and Democrats were much more likely to depart after *Booker*. However, the coefficient on $\text{DemDistrict} \times \text{Booker}$ is small and not statistically significant, which means that the relative differences between Republicans and Democrats did not change much after *Booker*.

Column 4 adds district-specific linear time trends. The coefficient on $\text{Deferential} \times \text{DemDistrict}$ decreases to 3.1 percentage points but is more precisely estimated than before. In unreported regressions, the results are not meaningfully changed by the inclusion of quadratic or cubic terms in the district trends. In short, we are confident that preexisting trends do not explain the changes observed.

Columns 5, 6, and 7 expand the sample to confirm the robustness of our results. Column 5 includes substantial assistance cases,²¹ column 6 includes border districts, and column 7 includes all federal sentences. The interaction between Deferential and DemDistrict remains positive and statistically significant in all regressions.²²

Table 2 reports results using substantial assistance departures as the dependent variable, taking the value of one if the defendant was granted a substantial assistance departure and zero otherwise. We use these regressions as placebo tests; in principle, substantial assistance departures should not be affected by a change in the standard of review applied to judge-induced departures. The results confirm that substantial assistance departures are unaffected by changes to the standard of review. The interactions between DemDistrict and Deferential or Booker have small coefficients that are never statistically significant. The term DemDistrict is correlated with fewer substantial assistance departures in some regressions, but this result disappears when district-specific trends are included.

Table 3 reports the same specifications as Table 1 but uses final offense level as the dependent variable. Thus, Table 3 measures how the standards of review for departures influenced judges' decisions regarding adjustments. In models 1–4, the impact of the proportion of Democrats

21. When substantial assistance departures are included, the dependent variable is coded as zero if the defendant is granted a substantial assistance departure or no departure at all; it is coded as one if the defendant receives a judge-induced downward departure.

22. Additional results from specific subsamples not reported here, such as drug crimes or serious nondrug crimes, yielded similar coefficients but tended to be less precise. For example, when the sample was restricted to drug crimes for the regression in column 3, Table 1, the coefficient on $\text{Deferential} \times \text{DemDistrict}$ was 4.4, with a p -value of .026; for serious nondrug crimes, the coefficient was 3.5, with a p -value of .13.

Table 2. Substantial Assistance Departures (Substantial Assistance Departure Equals One)

	(1)	(2)	(3)	(4)	(5)
DemDistrict	-.096** (.032)	-.030 (.026)	-.10** (.037)	-.10** (.036)	-.029 (.036)
Preguidelines	-.053 (.035)	-.027 (.024)	-.038 (.038)	-.036 (.059)	-.009 (.045)
CircuitDem	.087 (.083)	-.039 (.087)	.083 (.082)	.16 (.10)	-.056 (.087)
Deferential	-.001 (.006)	-.003 (.006)	.001 (.016)	.005 (.015)	.012 (.013)
Deferential × DemDistrict			.013 (.022)	.010 (.023)	-.008 (.021)
Deferential × Preguidelines			-.022 (.019)	.018 (.031)	-.036+ (.019)
Booker	-.017+ (.009)	-.014+ (.009)	-.020* (.010)	-.031 (.021)	-.037+ (.019)
Booker × DemDistrict				.005 (.032)	.028 (.030)
PostBooker × Preguidelines				.041 (.046)	.039 (.036)
District trends	No	Yes	No	No	Yes

Note. The sample is limited to serious crimes and includes substantial assistance departures. Robust standard errors clustered by district are in parentheses. All regressions condition on dummies for district, year, base offense level, criminal history, race (black or Hispanic), sex, citizenship status, numbers of dependents (zero, one, two, or more), and continuous variables age, age squared, and base offense level × criminal history. $N = 352,720$.

+ $p < .10$.

* $p < .05$.

** $p < .01$.

in the district is highly significant. On average, a Democratic judge calculates a .45 to .67 lower final offense level than a Republican judge. However, judges appointed before enactment of the guidelines are not significantly different from those appointed afterward.

When interaction terms are added for the standard of review in columns 3–5, the standard of review for departures has no effect on party differences in offense level calculations. We find some evidence of substitution between adjustments and departures after *Booker*, although the party effect is not statistically significant in columns 3–5. There is a large positive *Booker* effect on final offense levels, which was not specific to any judicial group. We will return to this result in our discussion of prison sentences.

Table 4 considers the impact of these changes on the log prison sen-

Table 3. Final Offense Level

	(1)	(2)	(3)	(4)	(5)
DemDistrict	-.67** (.20)	-.45* (.18)	-.59** (.22)	-.57** (.22)	-.26 (.22)
Preguidelines	-.069 (.15)	.11 (.18)	-.086 (.18)	-.048 (.18)	.10 (.24)
CircuitDem	.61 (.49)	.90 (.81)	.57 (.48)	.58 (.48)	.88 (.78)
Deferential	-.10 (.095)	-.058 (.041)	.046 (.095)	.13 (.11)	.11 (.11)
Deferential × DemDistrict			-.14 (.16)	-.25 (.17)	-.27 ⁺ (.16)
Deferential × Preguidelines			-.052 (.17)	-.16 (.19)	-.16 (.20)
Booker	.31** (.074)	.60** (.10)	.31** (.074)	.16 (.14)	.24 (.13)
Booker × DemDistrict				.21 (.23)	.31 ⁺ (.18)
PostBooker × Preguidelines				.26 (.23)	-.18 (.28)
(Deferential × DemDistrict) - (De novo × DemDistrict)			-.095 (.11)	-.12 (.12)	-.17 (.11)
(Deferential × RepDistrict) - (De novo × RepDistrict)			.046 (.093)	.13 (.11)	.11 (.11)
District trends	No	Yes	No	No	Yes
R ²	.828	.828	.828	.828	.892

Note. Robust standard errors clustered by district are in parentheses. All regressions condition on dummies for district, year, base offense level, criminal history, race (black or Hispanic), sex, citizenship status, numbers of dependents (zero, one, two, or more), and continuous variables age, age squared, and base offense level × criminal history. $N = 255,182$.

⁺ $p < .10$.

* $p < .05$.

** $p < .01$.

tence.²³ When we condition on base offense levels, Democratic judges give sentences that are on average 17 percent shorter than those of Republican judges, translating into a difference of about 10 months.²⁴ The impact of preguidelines judges is significant at the 10 percent level ($p = .068$) in column 1, and the coefficient suggests that preguidelines judges give sentences that are 8 percent shorter than those of judges appointed after the guidelines.

23. In order not to drop sentences coded as 0 months (about 8 percent of the total) in the log regressions, we take the log of months sentenced plus 1.

24. Comparable effects were observed in regressions that used total prison sentence as the dependent variable.

Table 4. Prison Sentences

	log(Sentence) Base Offense Controls				log(Sentence) Final Offense Controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DemDistrict	-.17** (.052)	-.14* (.058)	-.14* (.057)	-.038 (.049)	-.100** (.034)	-.059 (.055)	-.058 (.041)	.002
Preguidelines	-.081+ (.045)	-.096+ (.055)	-.097+ (.057)	-.036 (.047)	-.055 (.043)	-.055 (.041)	-.054 (.056)	-.042 (.038)
CircuitDem	-.097 (.125)	-.098 (.13)	-.10 (.12)	.12 (.10)	-.097 (.13)	-.16 (.11)	-.16 (.113)	-.002 (.073)
Deferential		.008 (.024)	.023 (.025)	.029 (.024)	-.010 (.010)	.014 (.019)	.020 (.025)	.020 (.016)
Deferential × DemDistrict		-.063+ (.039)	-.083* (.039)	-.10** (.037)		-.049* (.024)	-.052* (.024)	-.064** (.023)
Deferential × Preguidelines		.006 (.04)	-.012 (.042)	-.012 (.042)		.006 (.04)	-.010 (.042)	-.012 (.026)
Booker			-.041 (.029)	-.036 (.030)	-.064** (.014)	-.065** (.016)	-.018 (.029)	-.087** (.023)
Booker × DemDistrict			.051 (.049)	.081+ (.047)			-.032 (.034)	.051 (.037)
PostBooker × Preguidelines			.052 (.054)	-.018 (.050)			.042 (.041)	.030 (.036)
(Deferential × DemDistrict) – (De novo × DemDistrict)		-.054* (.024)	-.060* (.025)	-.071** (.025)		-.035* (.016)	-.034* (.017)	-.044** (.016)
(Deferential × RepDistrict) – (De novo × RepDistrict)		.008 (.024)	.023 (.025)	.029 (.025)		.015 (.019)	.023 (.025)	.020 (.016)
District trends	No	No	No	Yes	No	No	No	Yes
R ²	.676	.676	.676	.677	.786	.786	.786	.790

Note. Robust standard errors clustered by district are in parentheses. All regressions condition on dummies for district, year, criminal history, race (black or Hispanic), sex, citizenship status, numbers of dependents (zero, one, two, or more), and continuous variables age and age squared. Columns 1–3 include base offense level dummies and base offense level × criminal history dummies. Columns 6–8 include final offense level dummies and final offense level × criminal history. $N = 255,182$.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

As in the case of departures, a deferential standard of review widens the sentencing disparity between Democratic and Republican judges. In columns 2–4, the interparty sentencing disparity increases by roughly 50 percent under deferential review. In addition, Democrats under deferential review give lower sentences than Democrats constrained under *de novo* review, while Republicans are relatively insensitive to review standards.

Columns 3 and 4 suggest that *Booker*'s deferential standard did not have an impact on party differences in prison sentences beyond that of *Koon*'s deferential standard. (The same was true for departures as well.) However, *Booker* had little or no independent effect on sentence length, whereas it was associated with increases in departures and offense levels. In other words, post-*Booker* judges calculated higher offense levels but compensated with more downward departures. Because of these offsetting effects, the net effect of *Booker* on average prison sentences is not significant. Thus, we find evidence for substitution between departures and adjustments post-*Booker*, though the substitution does not appear to be confined to a particular group of judges.²⁵

These offsetting changes lead to an intriguing conclusion: deferential review under mandatory guidelines results in prison sentences that are similar to those of deferential review under advisory guidelines.²⁶ In other words, under deferential review, mandatory guidelines served only as a weak constraint on judges' ability to use adjustments and departures to reach target sentences.

Because final offense levels are endogenous, the results in columns 5–8 are merely suggestive. However, by conditioning on final offense level, we remove the effect of manipulations to the offense level and can isolate the effect of changes in the departure rate on prison sentences. Note first that the coefficient on DemDistrict in column 5 disappears when we include the interactions with Deferential in columns 6–8. This is similar to the departure results in Table 1—party differences concerning probability and magnitude of departures are statistically significant and roughly twice the magnitude under a deferential standard of

25. Freeborn and Hartmann (2010) examine the PROTECT Act period and conclude that departures declined under the act but do not find evidence that adjustments were used to offset the act's restrictions on departures.

26. This result is similar to that of Pfaff (2006), who finds that advisory state-level guidelines regimes may still act as substantial constraints on judicial discretion. Our results suggest that mandatory guidelines, with weak review standards, are similar to advisory guidelines.

Table 5. District- and Circuit-Level Interactions

	Departures (1)	Offense Level (2)	log Prison (3)
DemDistrict	.030 (.020)	-.57** (.21)	-.14* (.057)
Preguidelines	.059** (.020)	-.092 (.160)	-.10* (.047)
CircuitDem	.14 (.100)	.44 (.53)	-.10 (.17)
Deferential	.008 (.037)	-.31* (.13)	-.035 (.044)
Deferential × DemDistrict	.044* (.019)	-.25 (.16)	-.086* (.039)
Deferential × CircuitDem	-.033 (.093)	.98** (.33)	-.012 (.042)
Booker	.12** (.038)	.62** (.18)	.048 (.063)
Booker × DemDistrict	.012 (.029)	.19 (.22)	.047 (.047)
PostBooker × CircuitDem	.043 (.090)	-1.00* (.41)	-.19 (.14)
R ²	N.A.	.83	.67

Note. Robust standard errors clustered by district are in parentheses. All regressions condition on dummies for district, year, base offense level, criminal history, race (black or Hispanic), sex, citizenship status, number of dependents (zero, one, two, or more), and continuous variables age, age squared, and base offense level × criminal history. N.A. = not applicable.

* $p < .10$.

* $p < .05$.

** $p < .01$.

review. Note also that conditioning on final offense level suggests that *Booker* lowered sentences by roughly 6 percent. However, this is misleading because higher offense levels were calculated post-*Booker*. Columns 1–4 give the effect of *Booker* conditional on base offense levels, which is indistinguishable from zero.

Table 5 includes interactions between circuit court composition and standards of review. Since there is much less variation at the circuit court level, it is harder to identify the effect of circuit court composition, and consequently these results should be interpreted with caution. In these regressions, our results for DemDistrict and its interactions with standards of review are little changed from our prior regressions. We do not detect significant interaction effects with circuit court composition for departures or log prison sentences. However, the final offense-level regression (column 2) suggests that under a deferential standard of review there is some substitution toward higher offense-level calculations in

more Democratic circuits, though this is not true under *Booker* (the coefficient on CircuitDem \times PostBooker is negative and offsets the coefficient on CircuitDem \times Deferential).²⁷

6. CONCLUSION

Changes to standards of review clearly have an impact on district judges' sentencing behavior. Further study is needed to determine how our results generalize to other areas of law and other systems of adjudication. However, the response observed here is robust across a variety of specifications and measures of sentencing behavior. Democrats are more likely than Republicans both to depart downward and to give lower sentences, and these differences are amplified when review is more deferential. The behavior of Republicans is invariant to the standard of review, either because the guidelines do not constrain their sentencing preferences or because they are simply more inclined to honor the guidelines' restrictions on discretion. Review standards do not affect offense-level calculations with one exception: *Booker* is associated with higher calculated offense levels. This finding suggests that making the guidelines advisory resulted in a substitution between adjustments and departures.

Another intriguing result is that preguidelines judges are insensitive to the standard of review—even though they exhibit a desire to give lower sentences and depart more frequently. We posit that preguidelines judges were more hostile to guidelines sentencing and hence may have incurred a lower disutility from reversal. Although our results show that aversion to reversal acts as a constraint on the behavior of district judges, its impact may be heterogeneous and depend upon the perceived legitimacy of the guidelines.

Our findings are best explained by a combination of models of judicial behavior. Differences between Democratic and Republican judges show that policy preferences play a role in sentencing decisions. The significant impact of standards of review also suggests that district judges modify their decisions to avoid reversal, consistent with strategic models of judging used in positive political theory. However, the constraining power of review standards depends on context; reversal aversion may be contingent on the judge's respect for the underlying legal regime.

The response on the part of district judges implies that they anticipate

27. Adding three-way interactions between DistrictDem, CircuitDem, and standards of review resulted in large standard errors for the interaction terms and thus prevents us from making inferences about alignment effects and standards of review.

behavioral changes at the circuit level. Thus, the results also provide indirect evidence that review standards constrain circuit courts. Given the negligible chance of reversal en banc or by the Supreme Court, this response is most likely driven by judicial preferences for adherence to legal doctrine. It may be possible to more precisely measure the constraining effect of review standards on circuit courts using data on sentencing appeals. However, our finding that district judges respond strategically to these changes cautions against examining affirmance rates in isolation.

Table A2. Descriptive Statistics

	All Crimes, Including Border Districts	Serious Crimes, No Border Districts
Judicial variables:		
Preguidelines	.318	.324
DemDistrict	.415	.415
CircuitDem	.387	.376
Deferential	.664	.646
De novo	.336	.354
Booker	.237	.232
Guidelines variables:		
Total prison (months)	49.6 (67.5)	88.4 (84.0)
Final offense level	17.6 (8.8)	23.2 (8.2)
Base offense level	16.4 (10.0)	24.2 (24.2)
Any imprisonment	.830	.940
Downward departure	.192	.131
Upward departure	.013	.014
Criminal history I	.510	.442
Criminal history II	.112	.120
Criminal history III	.144	.156
Criminal history IV	.086	.093
Criminal history V	.052	.055
Criminal history VI	.094	.132
Offender demographics:		
Age	34.5 (10.8)	33.3 (10.0)
Male	.856	.910
Black	.258	.416
Hispanic	.389	.232
White (non-Hispanic)	.320	.330
Citizen	.658	.81
N	658,029	255,182

Note. Standard deviations are in parentheses. Substantial assistance departures are excluded.

REFERENCES

- Anderson, John M., Jeffrey R. Kling, and Kate Stith. 1999. Measuring Interjudge Sentencing Disparity: Before and after the Federal Sentencing Guidelines. *Journal of Law and Economics* 42:271–307.
- Bibas, Stephanos. 2001. Judicial Fact-Finding and Sentence Enhancements in a World of Guilty Pleas. *Yale Law Journal* 110:1097–1186.
- Boyd, Christina, Lee Epstein, and Andrew Martin. 2010. Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science* 54:389–411.
- Boylan, Richard T. 2004. Do the Sentencing Guidelines Influence the Retirement Decisions of Federal Judges? *Journal of Legal Studies* 33:231–53.
- Cameron, Charles M., Jeffrey A. Segal, and Donald Songer. 2000. Strategic

- Auditing in a Political Hierarchy: An Informational Model of the Supreme Court's Certiorari Decisions. *American Political Science Review* 94:101–16.
- Cohen, Linda R., and Matthew L. Spitzer. 1994. Solving the *Chevron* Puzzle. *Law and Contemporary Problems* 57:65–110.
- Cross, Frank B. 2005. Appellate Court Adherence to Precedent. *Journal of Empirical Legal Studies* 2:369–405.
- Cross, Frank B., and Emerson Tiller. 1998. Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeal. *Yale Law Journal* 107:2155–76.
- Edwards, Harry T. 1998. Collegiality and Decision Making on the D.C. Circuit. *Virginia Law Review* 84:1335–70.
- Epstein, Lee, and Gary King. 2002. The Rules of Inference. *University of Chicago Law Review* 69:1–133.
- Eskridge, William N., Jr., and Lauren E. Baer. 2008. The Continuum of Deference: Supreme Court Treatment of Agency Statutory Interpretations from *Chevron* to *Hamdan*. *Georgetown Law Journal* 96:1083–1226.
- Federal Judicial Center. 2010. *History of the Federal Judiciary: Biographical Directory of Federal Judges*. <http://www.fjc.gov/history/home.nsf/page/judges.html>.
- Fischman, Joshua B., and David S. Law. 2009. What Is Judicial Ideology, and How Should We Measure It? *Washington University Journal of Law and Policy* 29:133–214.
- Freeborn, Beth A., and Monica E. Hartmann. 2010. Judicial Discretion and Sentencing Behavior: Did the Feeney Amendment Rein In District Judges? *Journal of Empirical Legal Studies* 7:355–78.
- George, Tracey. 2001. Court Fixing. *Arizona Law Review* 43:9–62.
- Gertner, Nancy. 2005. Sentencing Reform: When Everyone Behaves Badly. *Maine Law Review* 57:569–84.
- Giles, Micheal W., Virginia A. Hettinger, and Todd Peppers. 2001. Picking Federal Judges: A Note on Policy and Partisan Selection Agendas. *Political Research Quarterly* 54:623–41.
- Hofer, Paul J., Ken Blackwell, and R. Barry Ruback. 1999. The Effect of the Federal Sentencing Guidelines on Inter-judge Sentencing Disparity. *Journal of Criminal Law and Criminology* 90:239–322.
- Hofer, Paul J., Willie Martin, and Pamela Montgomery. 1997. Departure Rates and Reasons after *Koon v. U.S.* *Federal Sentencing Reporter* 9:284–91.
- Johnson, Barry L. 1998. Discretion and the Rule of Law in Federal Guidelines Sentencing: Developing Departure Jurisprudence in the Wake of *Koon v. United States*. *Ohio State Law Journal* 58:1697–1751.
- Johnson, Molly Treadway, and Scott A. Gilbert. 1997. *The U.S. Sentencing Guidelines: Results of the Federal Judicial Center's 1996 Survey*. Washington, D.C.: Federal Judicial Center.
- Kastellec, Jonathan P. 2007. Panel Composition and Judicial Compliance on the

- U.S. Courts of Appeal. *Journal of Law, Economics, and Organization* 23: 421–41.
- Kastellec, Jonathan P., and Jeffrey R. Lax. 2008. Case Selection and the Study of Judicial Politics. *Journal of Empirical Legal Studies* 5:407–46.
- Kim, Pauline T. 2007. Lower Court Discretion. *New York University Law Review* 82:383–442.
- Klein, David E., and Robert J. Hume. 2003. Fear of Reversal as an Explanation of Lower Court Compliance. *Law and Society Review* 37:579–606.
- LaCasse, Chantale, and A. Abigail Payne. 1999. Federal Sentencing Guidelines and Mandatory Minimum Sentences: Do Defendants Bargain in the Shadow of the Judge? *Journal of Law and Economics* 42:245–70.
- Lee, Cynthia K. Y. 1997. A New “Sliding Scale of Difference” Approach to Abuse of Discretion: Appellate Review of District Court Departures under the Federal Sentencing Guidelines. *American Criminal Law Review* 35:1–55.
- McCubbins, Mathew, Roger Noll, and Barry Weingast. 1995. Politics and the Courts: A Positive Theory of Judicial Doctrine and the Rule of Law. *Southern California Law Review* 68:1631–83.
- Miles, Thomas J., and Cass R. Sunstein. 2006. Do Judges Make Regulatory Policy? An Empirical Investigation of *Chevron*. *University of Chicago Law Review* 73:823–82.
- . 2008. The Real World of Arbitrariness Review. *University of Chicago Law Review* 75:761–814.
- O’Neill, Michael Edmund. 2003. Surveying Article III Judges’ Perspectives on the Federal Sentencing Guidelines. *Federal Sentencing Reporter* 15:215–19.
- Peresie, Jennifer L. 2005. Female Judges Matter: Gender and Collegial Decision-making in the Federal Appellate Courts. *Yale Law Journal* 114:1759–90.
- Pfaff, John F. 2006. The Continued Vitality of Structured Sentencing Following *Blakely*: The Effectiveness of Voluntary Guidelines. *UCLA Law Review* 54: 235–308.
- Pinello, Daniel R. 1999. Linking Party to Judicial Ideology in American Courts: A Meta-Analysis. *Justice System Journal* 20:219–54.
- Posner, Richard A. 2008. *How Judges Think*. Cambridge, Mass.: Harvard University Press.
- Price, Joseph, and Justin Wolfers. 2010. Racial Discrimination among NBA Referees. *Quarterly Journal of Economics* 125:1859–87.
- Priest, George L., and Benjamin Klein. 1984. The Selection of Disputes for Litigation. *Journal of Legal Studies* 13:1–55.
- Revesz, Richard L. 1997. Environmental Regulation, Ideology, and the D.C. Circuit. *Virginia Law Review* 83:1717–72.
- Schanzenbach, Max. 2005. Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics. *Journal of Legal Studies* 34: 57–92.
- Schanzenbach, Max M., and Emerson H. Tiller. 2007. Strategic Judging under

- the U.S. Sentencing Guidelines: Positive Political Theory and Evidence. *Journal of Law, Economics, and Organization* 23:24–56.
- Schuck, Peter H., and E. Donald Elliott. 1990. To the *Chevron* Station: An Empirical Study of Federal Administrative Law. *Duke Law Journal* 1990: 984–1077.
- Shavell, Steven M. 2006. The Appeals Process and Adjudicator Incentives. *Journal of Legal Studies* 35:1–29.
- Sisk, Gregory C., and Michael Heise. 2005. Judges and Ideology: Public and Academic Debates about Statistical Measures. *Northwestern University Law Review* 99:743–803.
- Sisk, Gregory C., Michael Heise, and Andrew P. Morriss. 1998. Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning. *New York University Law Review* 73:1377–1500.
- Songer, Donald R. 1987. The Impact of the Supreme Court on Trends in Economic Policy Making in the United States Courts of Appeals. *Journal of Politics* 49:830–41.
- Songer, Donald R., Jeffrey A. Segal, and Charles M. Cameron. 1994. The Hierarchy of Justice: Testing a Principal-Agent Model of Supreme Court–Circuit Court Interactions. *American Journal of Political Science* 38:673–96.
- Songer, Donald R., and Reginald S. Sheehan. 1990. Supreme Court Impact on Compliance and Outcomes: Miranda and New York Times in the United States Courts of Appeals. *Western Political Quarterly* 43:297–316.
- Spitzer, Matt, and Eric Talley. 2000. Judicial Auditing. *Journal of Legal Studies* 29:649–83.
- Stith, Kate, and José A. Cabranes. 1998. *Fear of Judging: Sentencing Guidelines in the Federal Courts*. Chicago: University of Chicago Press.
- Sunstein, Cass R., David Schkade, Lisa M. Ellman, and Andres Sawicki. 2006. *Are Judges Political? An Empirical Analysis of the Federal Judiciary*. Washington, D.C.: Brookings Institution Press.
- Tiede, Lydia Brashear. 2009. The Impact of the Federal Sentencing Guidelines and Reform: A Comparative Analysis. *Justice System Journal* 30:34–49.
- United States Sentencing Commission. 2006. *Final Report of the Impact of United States v. Booker on Federal Sentencing*. Washington, D.C.: United States Sentencing Commission.
- Verkuil, Paul R. 2002. An Outcomes Analysis of Scope of Review Standards. *William and Mary Law Review* 68:679–735.
- Weinstein, Jack B. 1992. A Trial Judge’s Reflections on Departures from the Federal Sentencing Guidelines. *Federal Sentencing Reporter* 5:6–9.
- Yoon, Albert. 2006. Pensions, Politics, and Judicial Tenure: An Empirical Study of Federal Judges, 1869–2002. *American Law and Economics Review* 8:143–80.
- Zaring, David. 2010. Reasonable Agencies. *Virginia Law Review* 96:135–97.