Unintended Consequences of Constraining Discretion: Evidence from Criminal Sentencing

Spencer Cooper* & Andy Ye Yuan†

Abstract

We study how institutional reforms may produce unintended consequences through behavioral responses by public officials in the context of criminal sentencing. We find that a reform constraining judicial and prosecutorial discretion led to an unexpected 2-month sentence increase despite a 30% reduction in sentence enhancements. This discrepancy between mechanical and realized effects suggests behavioral adjustment by prosecutors and/or judges. We find no evidence of prosecutorial adjustment across thirteen charging and bargaining outcomes. Instead, judges drive the effect with an asymmetric response: lenient judges increase sentences within the remaining sentencing guideline while strict judges reduce them, compressing the distribution of punishment.

* Assistant Professor of Economics, University of Connecticut, spencer.cooper@uconn.edu

† Assistant Professor of Law, University of Florida Levin College of Law, yuan@law.ufl.edu

We thank Desmond Ang, Jason Baron, Diego Battiston, Katie Bollman, Remy Levin, David Rosenkranz, Steve Ross, David Simon, Michal Šoltés, and numerous conference participants for their helpful comments and advice.

1. Introduction

How do institutional decisionmakers respond to reforms that constrain their discretion? Policies that limit discretion often leave other decision margins unchanged, allowing actors to adapt in ways that offset or reshape the policy's intended effects. Understanding these behavioral responses is essential for evaluating legal reforms since their consequences depend not only on the mechanical effects of new rules but also on how institutional actors adjust their behavior. In this paper, we study these responses in the context of criminal sentencing, where prosecutors and judges exercise substantial discretion across multiple stages of the legal process.

Measuring these behavioral responses is challenging. Many policies and court decisions are multi-dimensional reforms, affecting multiple aspects of the justice system simultaneously. When multiple policy levers are pulled or multiple actors are affected, isolating the causal source of observed changes becomes difficult. Furthermore, highly salient or impactful reforms may induce changes in case composition, such that prosecutors and judges face systematically different defendants before and after the reform. These two challenges make it difficult to distinguish behavioral responses from selection effects. We address these concerns by exploiting the Supreme Court decision *Blakely v. Washington*, which generated a sharp increase in the evidentiary standard required for sentencing enhancements. This allows us to cleanly identify how prosecutors and judges respond to a procedural change that constrains discretion along the punishment severity margin.

Blakely provides an attractive setting for studying behavioral responses to institutional constraints for two reasons. First, the decision only constrained a single punishment margin: sentencing enhancements. Enhancements are dependent on facts that allow for increases in the statutory maximum sentence. *Blakely* held that such facts must be found by a jury beyond a reasonable doubt rather than by a judge under the preponderance-of-the-evidence standard. Therefore, the decision both raised the evidentiary burden for enhancements and removed judges' authority to determine enhancement-related facts. However, *Blakely* left the underlying charges, convictions, and other features of sentencing unchanged. This allows us to isolate the consequences of constraining this particular margin of punishment without contamination of other policy impacts. Said another way, because *Blakely* only applied to cases in which enhanced sentences were at issue, its direct mechanical effect should be concentrated among those cases.

Second the impact of *Blakely* is easy to predict. By making sentencing enhancements more difficult to obtain, *Blakely* should mechanically reduce the frequency with which enhanced sentences are imposed. Because an enhanced sentence is longer than the corresponding non-enhanced sentence, the decline in enhancement rate should reduce average sentence lengths if prosecutors and judges otherwise leave their behavior unchanged. Any increase in sentence lengths following the reform must therefore reflect behavioral responses along other margins rather than the direct mechanical effect of the decision.

Our strategy to detect a behavioral response to *Blakely* comes in four steps. First, we measure the direct effect of *Blakely* on enhancement rate using a regression discontinuity design with time acting as the running variable. Second, we consider the average impact of enhancement on sentence length. Third, we combine these two results to create the predicted, no-response impact of *Blakely* on sentence length. Finally, we estimate the actual impact of *Blakely* on sentence length using regression discontinuity and compare these effects against the no-response prediction. If the difference between these two effects is significant, we consider this evidence that prosecutors and/or judges are changing their behavior in some way to cause an increase in sentence lengths.

We find that *Blakely* caused a decrease in enhancement rate of approximately 30% from the pre-decision mean. By estimating the effects of enhancement on sentence, we predict this reduction in enhancement should decrease overall average sentence length by 0.362 months. Instead, *Blakely* causes an increase of nearly 2 months. The difference between these actual and predicted effects is statistically significant and moves in the opposite predicted direction. We check for any discontinuities in defendant demographics, case density, and enhancement propensity score and find smoothness through the timing of *Blakely*. This indicates that the surprising and significant sentence length effects are not driven by changes in case composition but are instead evidence of a behavioral change in charging, bargaining, or sentencing.

Following this baseline analysis, we turn to identifying potential mechanisms for the surprising and counter-intuitive increase in sentence lengths. To do so, we first estimate the impact of *Blakely* across a wide set of prosecutor-related outcomes. Specifically, we consider changes in 13 dependent variables across three main categories: charges, plea bargaining, and sentence outcomes plausibly impacted by prosecutor discretion. These are again estimated using a regression discontinuity design. Next, we turn our attention to judges. To consider judge effects, we construct a measure of judge leniency based on pre-*Blakely* average sentence lengths. Using

this measure, we estimate the heterogeneous effects of *Blakely* across judge types using a difference-in-differences design.

We find no evidence of systematic behavioral responses by prosecutors. Across all 13 prosecutorial outcomes, estimated effects are either small and statistically insignificant or move in the opposite direction of the observed increase in sentence lengths. These null results are also highly informative. The estimated confidence intervals are sufficiently tight to rule out economically meaningful adjustments along the primary prosecutorial margins through which adjustment might occur, including outcomes such as the number and severity of charges, bargaining concessions, and downstream outcomes like incarceration. The findings suggest that prosecutors did not respond to the loss of enhancements by systematically pursuing harsher charges or plea agreements. However, the absence of charging and bargaining responses does not imply that prosecutors were unaffected by *Blakely*. We estimate a 6.5 percent decline in case duration, consistent with the reform reducing litigation over sentencing enhancements. Therefore, prosecutors appear to have adapted to the procedural change by resolving cases more quickly rather than by increasing punishment.

We then turn our attention to judges. We first show that cases proceeding to trial, where judges have greater scope to directly influence sentencing outcomes, exhibit an even larger discontinuous increase in sentence lengths following *Blakely*. We then consider heterogeneity in judicial response. Comparing direct effects by judicial leniency types, we find that cases assigned to the most lenient judges exhibit little change in enhancement use, yet account for nearly all of the increase in sentence length after *Blakely*. These patterns stand in sharp contrast to those for strict judges, who experience substantial declines in both enhancements and sentence lengths. As a result, *Blakely* compresses the distribution of sentences across judges, reducing inter-judge sentencing variance.

The heterogeneous response reflects differences in the remaining discretion available to judges. After *Blakely*, judges respond by increasing sentences within the applicable guideline range. But before *Blakely*, strict judges were substantially more likely to impose sentences at the guideline maximum. Thus, strict judges have less scope to increase sentences along the remaining punishment margin than lenient judges. Consistent with this interpretation, we find that the increase in sentence lengths is concentrated among cases with a high pre-*Blakely* probability of receiving an enhancement, with little evidence of spillover to cases where enhancements were

unlikely. Together, these findings suggest that judges respond to the loss of enhancement authority by adjusting sentences along remaining discretionary margins, with the magnitude of the response determined by both the discretion available to the judge and the relevance of the lost enhancement margin to the case.

1.1. Contributions to Literature

Our first contribution is to show that the consequences of institutional constraints depend on the alternative margins available to the constrained actor. A common theoretical framework for these responses is Holmström and Milgrom (1991), in which restricting one dimension of performance induces substitution toward others. Empirical work documents such behavioral adjustment in settings ranging from education to health care (Jacob and Levitt 2003; Shearer, Somé, and Fortin 2018; Alexander 2020). Among these, our paper relates most closely to Alexander (2020), which shows that changing bonus qualifications for doctors has the unintended effect of inducing doctors to admit heathier patients. Similarly, we find that constraining one margin of punishment produces an unintended behavioral response along another margin, as judges increase sentence lengths following *Blakely*. Related work shows that changes in institutional incentives can generate significant behavioral responses within public institutions (Khan et al. 2019; Bertrand et al. 2020). Most notably, Petkun (2026) shows that federal judges respond to a non-monetary institutional reform designed to alter bureaucratic behavior. Our setting similarly examines how judicial behavior responds to institutional change, but focuses on a procedural reform that constrains judicial authority rather than an administrative policy designed to influence judicial performance.

Our second contribution is to provide evidence on how prosecutors and judges respond differently to the same institutional constraint. Existing work largely shows that prosecutors respond to changes or constraints in incentives by adjusting charging, plea bargaining, and sentencing decisions (Rehavi and Starr 2014; Didwania 2022; Agan, Doleac, and Harvey 2023; Shaffer 2023; Shaffer and Harrington 2025; Yuan and Cooper 2025; Sloan 2026; Tuttle 2026). Both empirical and theoretical work also suggests that prosecutors may substitute across margins when incentives or constraints change (Boylan 2005; Silveira 2017). Importantly, this literature does not provide a clear prediction about how prosecutors should respond to a constraint on one punishment margin: prior studies find both increases and decreases in charging severity depending

on the institutional environment (Bjerk 2005; Tuttle 2026; Cooper 2026). In contrast to many papers in the literature, we find remarkably little evidence of prosecutorial response to an institutional constraint. Instead, our results show that judges respond strongly to the constraint in punishment discretion and this response is focused on affected cases, with no evidence of spill over to cases for which the constrained punishment margin was largely irrelevant.

Our third contribution is to show that institutional design can change the extent of heterogeneity in judicial decision making. A large literature establishes substantial differences across judges in sentencing decisions and uses quasi-random judge assignment to identify their effects (Kling 2006; Dobbie et al. 2018; Arnold, Dobbie, and Yang 2018; Frandsen, Lefgren, and Leslie 2023). Less is known about how institutional rules affect these differences. We show that sentencing outcomes compress across judges following *Blakely*, such that judges who were more likely to use enhancements reduce their reliance on that margin, while judges with greater discretion to increase sentences within the guideline range offset more of the resulting decline. Our findings suggest that judge fixed effects capture not only persistent differences across judges but also differences in how judges respond to the institutional environment.

Finally, our fourth contribution is to the empirical law and economics literature examining institutional changes in criminal sentencing resulting from the Supreme Court's jury-trial jurisprudence. Our paper is among the first in this literature to examine these doctrinal developments in a state sentencing system, where the vast majority of criminal prosecutions in the United States occur. Prior work has focused primarily on the federal effects of *Booker*, which expanded judicial discretion by making the Federal Sentencing Guidelines advisory and appears to have increased prosecutors' use of mandatory minimums (Starr and Rehavi 2013; Yang 2015). The distinction between *Blakely* and *Booker* is important for our analysis. *Blakely* did not alter the mandatory nature of North Carolina's sentencing guidelines, allowing us to measure the decline in sentence enhancements and calculate its mechanical effect on sentence length. We can then compare this mechanical effect with the actual change in sentences to identify behavioral responses by judges and prosecutors. In contrast, because *Booker* made the federal guidelines non-binding, the mechanical sentencing effects of changes in guideline enhancements cannot be calibrated in the same way.

2. Background

2.1. North Carolina Sentencing

The institutional setting for this study is the North Carolina structured sentencing system. After conviction, either through a plea bargain or trial, judges determine a minimum sentence within the applicable guideline range, which mechanically determines the corresponding maximum sentence. Judges then impose a sentence within the resulting minimum-maximum interval. We focus on minimum sentences because they provide a cleaner measure of judicial decision-making than final sentence lengths, which may also reflect post-sentencing behavior and administrative decisions by other actors.

Appendix E.1 illustrates North Carolina's structured sentencing grid. The sentencing guidelines are mandatory throughout the sample period. Each defendant is assigned to a specific grid cell based on offense severity and criminal history. Every cell contains three sentencing ranges: a presumptive range, an aggravated range, and a mitigated range. Judges are generally expected to select minimum sentences from the presumptive range. A sentence enhancement occurs when the presence of aggravating factors permits a shift from the presumptive range to the aggravated range. Conversely, mitigating factors may justify a sentence within the mitigated range.

The decision to pursue an aggravated sentence originates with the prosecutor, who must allege the relevant aggravating factors in connection with the charged offense. To obtain an aggravated sentence, the prosecutor must establish the existence of one or more statutory aggravating factors in addition to proving the underlying offense. Some aggravating factors are relatively objective and straightforward to verify, such as whether the defendant possessed a deadly weapon during the commission of the crime. Others are more subjective and potentially more difficult to establish, such as whether the offense was committed in an especially cruel manner. North Carolina General Statute identifies twenty aggravating factors that may support sentence enhancement[1]. A complete list of these factors is provided in the Online Appendix E.2.

Sentence enhancements can substantially increase defendants' expected punishment. The magnitude of the increase varies with offense severity. For example, among defendants with no prior criminal history convicted of a low-level offense, the presumptive minimum sentence range is 4 to 6 months, whereas the aggravated range is 6 to 8 months. In contrast, for a more serious

[1] § 15A-1340.16.(d)

Class C felony, the presumptive range is 58 to 73 months, and the aggravated range is 73 to 92 months. Both the absolute and proportional differences between the presumptive and aggravated ranges tend to increase with offense severity. Consequently, sentence enhancements are particularly consequential in serious cases, where they can increase expected incarceration by several years. This feature of the sentencing guidelines makes enhancements an important margin of prosecutorial and judicial discretion.

2.2. *Blakely v. Washington* and its implication to North Carolina

Our identification strategy exploits the Supreme Court's decision in *Blakely v. Washington*. In *Blakely*, the defendant was convicted of second-degree kidnapping involving domestic violence. Under Washington's sentencing guidelines, the presumptive range was 49 to 53 months. However, the judge imposed a 90-month sentence after finding that the offense involved "deliberate cruelty," an aggravating factor permitting an exceptional sentence under Washington law.[2] The central issue in *Blakely* was whether judges could constitutionally increase sentences based on their own finding of facts. In a 5-4 decision, the Supreme Court held that Washington's sentencing scheme violated the defendant's Sixth Amendment right to a jury trial. The Court ruled that any fact increasing the "maximum sentence" must be proven to a jury beyond a reasonable doubt unless admitted by the defendant. The decision was surprising and generated profound impacts on the sentencing guidelines and practices at both the federal and state level (Prescott & Starr 2006).

Blakely substantially altered sentencing procedures in North Carolina. Before the decision, prosecutors could obtain sentence enhancements by proving aggravating factors to a judge under a preponderance-of-evidence standard unless the defendant admitted the relevant facts. After *Blakely*, prosecutors seeking an enhanced sentence were required either to obtain an explicit admission from the defendant or to prove aggravating factors to a jury beyond a reasonable doubt. The decision therefore increased the evidentiary burden associated with sentence enhancements and reduced prosecutors' ability to obtain enhanced sentences based on judicial fact-finding.

North Carolina provides a particularly useful setting for studying the effects of this change because its sentencing guidelines remained mandatory throughout both the pre-*Blakely* and post-*Blakely* periods. As a result, sentence enhancements applied to a clearly defined set of cases and operated through a single institutional channel: prosecutors' ability to establish legally prescribed

[2] § 9.94A.120(2)

aggravating factors. This institutional structure allows us to measure changes in enhancement rates cleanly and, more importantly, to quantify both the mechanical and realized effects of enhancements on sentence lengths. Because sentences imposed from the aggravated range are always weakly greater than those imposed from the presumptive range, a decline in enhancement rates mechanically reduces sentence lengths. Any divergence between this mechanical effect and the realized effect therefore reflects behavioral responses by prosecutors, judges, or both along other margins of decision-making. The remainder of the paper develops this framework and empirically evaluates the extent of such responses.

3. Empirical Framework & Data

To show how *Blakely* creates predictable variation in sentence length, we outline a simple conceptual model. This model illustrates how we obtain a predicted benchmark that we use as a null hypothesis for the effect of *Blakely* on sentence length. We then provide an empirical strategy to show how we estimate these predicted values in practice.

3.1. Empirical Outline

Suppose $\bar{X}_1$ is the average sentence length for defendants with sentence enhancements. $\bar{X}_2$ is the average sentence length for defendants without sentence enhancements. α is the share of defendants with sentence enhancements. The overall average sentence length $\bar{X}$ can be written as

$$\bar{X} = \alpha * \bar{X}_1 + (1 - \alpha) * \bar{X}_2$$

where $\bar{X}$ is an increasing function in α. Note that $\bar{X}_1 - \bar{X}_2 > 0$ by assumption. By this monotonicity assumption, $\bar{X}$ will decrease if α decreases.

Let α_{pre} and α_{post} represent the share of defendants with sentence enhancement pre- and post-*Blakely*. Then the effect of *Blakely* on sentence length, which we denote as $\Delta_{Blakely}$, can be written as follows:

$$\Delta_{Blakely} = \bar{X}_{post} - \bar{X}_{pre}$$
$$= \alpha_{post} * \bar{X}_1 + (1 - \alpha_{post}) * \bar{X}_2 - \alpha_{pre} * \bar{X}_1 + (1 - \alpha_{pre}) * \bar{X}_2$$
$$= (\alpha_{post} - \alpha_{pre})\bar{X}_1 - (\alpha_{post} - \alpha_{pre})\bar{X}_2$$
$$= (\alpha_{post} - \alpha_{pre})(\bar{X}_1 - \bar{X}_2)$$

In theory, assuming that *Blakely* does not change the average sentence lengths for defendants with and without sentence enhancements, the effects of *Blakely* on sentence length can be written as above, which has two components: (1) $\alpha_{post} - \alpha_{pre}$, which are the effects of *Blakely* on sentence enhancements, (2) the difference in sentence length between defendants with sentence enhancements and those without sentence enhancements. Each of these parts can be estimated empirically to derive our predicted benchmark, $\widehat{\Delta}_{blakely}$.

3.2. Empirical Strategy

For a case to receive enhancement, prosecutors must explicitly file intent to prove specific aggravating factors at least 30 days before a plea deal is reached or a trial is held.[3] Thus, the prosecutor must choose whether to pursue enhancement or not for each case. For some cases, pursuing enhancement might be obvious if evidence of a specific aggravating factor is strong and clear.[4] However, for some aggravating factors, such as "The offense was especially heinous, atrocious, or cruel" or the offense created "…damage causing great monetary loss…," the proof of such factors will be more discretionary. For cases of this type, a prosecutor will weigh the cost of proving such factors against the benefit of a potentially higher sentence length. Because *Blakely* increased the burden of proof required to prove aggravating factors, the cost of pursuing enhancement is higher post-*Blakely*. Therefore, *Blakely* is expected to decrease enhancements as the cost for pursuing enhancement for some marginal cases will be too high post-*Blakely.* Because enhancements strictly increase sentence lengths, *Blakely* is expected to decrease average sentence length.

However, enhancements make up only a small proportion of all cases (only 2.88% of cases received enhancement pre-*Blakely* between the years of 2001 and 2004). Thus, any effects on sentence length are likely to be lost in noise created by a large number of cases unaffected by *Blakely.* Our goal is to narrow our sample to those cases that are most likely to receive enhancement. Doing so is difficult; enhancement is primarily based upon the existence of aggravating factors unobserved in our data. We attempt to predict enhancement likelihood using a logistic regression with crime type and criminal history points as predictors:

[3] See *N.C. G.S.* §15A-1340.16(a6).
[4] For example, carrying a firearm during the crime or involving an individual under the age of 16 in the crime are relatively clear cut to prove compared to other, more subjective aggravating factors.

Equation 1: $\Pr(enhancement_i = 1 | \, crime\ type_i, history\ points_i) = F(crime\ type_i + history\ points_i)$

where F is the cumulative standard logistic distribution function. We then build a propensity score using the logistic regression results. While more regressors could be included, we find that the best prediction comes from this simpler model. For the majority of our analysis, we reduce our sample to cases with a propensity score of 5% or higher. This reduces our sample of cases from 117,222 cases to 27,263 and raises our pre-*Blakely* mean in aggravation from 3.79% to 7.21%. The purpose of this restriction is not to estimate causal treatment effects of enhancement, but to increase statistical power in detecting changes driven by *Blakely*. Because the decision affects only cases plausibly eligible for enhancement, focusing on this subset sharpens the first stage while preserving the interpretation of our estimates. We show in robustness checks that results are qualitatively similar when considering the full sample of cases.

Because enhancements always increase sentence lengths, we can predict the mechanical expected decrease in sentence length following *Blakely.* We do this by estimating the average sentence length increase caused by enhancement, holding other court factors constant, and multiplying this effect by the estimated effect of *Blakely* on enhancement rate. The product of these two estimates provides a benchmark that represents the expected effect of *Blakely* if criminals, prosecutors, and judges do not change their behavior in a way that impacts sentence length (except through this decrease in enhancement). This benchmark serves as a sharp null hypothesis: any observed change in sentence length that differs from this predicted value must reflect behavioral responses by prosecutors or judges along margins other than enhancement use. This process is outlined in Equation 2, which gives our three main models for our econometric analysis:

Equation 2:

(1) $$enhancement_{itg} = \beta_0 + \beta_1 Blakely_t + \beta_2 month_t + \beta_3 (Blakely \cdot month_t) + \eta_g + u_{it}$$

(2) $$sentence_length_{itg} = \delta_0 + \delta_1 enhancement_{it} + \eta_g + u_{it}$$

(3) $$sentence_length_{itg} = \gamma_0 + \gamma_1 Blakely_t + \gamma_2 month_t + \gamma_3 (Blakely \cdot month_t) + \eta_g + u_{it}$$

$$H_0: \quad \gamma_1 = (\beta_1 \times \delta_1)$$

$$H_1: \quad \gamma_1 \neq (\beta_1 \times \delta_1)$$

The dependent variable in the first equation, $enhancement_{ictg}$, is a binary measure that equals one if case i in county c at half-year t sentenced in grid cell g receives enhancement at sentencing. Note that $\beta_1 \times \delta_1$ is equivalent to $\Delta_{blakely}$ in the conceptual model. The first model in Equation 2 estimates the change enhancement rates, such that $\beta_1 = (\alpha_{post} - \alpha_{pre})$ from the conceptual model. Similarly, the second model estimates the average increase in sentence length for enhanced cases over non-enhanced cases, such that $\delta_1 = (\bar{X}_1 - \bar{X}_2)$. This second model is estimated using only pre-*Blakely* cases. The third model gives the observed, actual effect of *Blakely* on sentence length. We then consider how this observed effect differs from the predicted benchmark. If the null hypothesis is rejected, this is evidence of changed behavior from either prosecutors or judges.

This strategy is effectively a regression discontinuity in time analysis. The validity of this method depends upon *Blakely* occurring without other shocks or policies happening at the same time. In our research of North Carolina sentencing procedure, we find no evidence of any relevant changes or policies within the time bandwidths we use. Empirically, we check for major concurrent changes or potential case-timing manipulation through a series of McCrary tests. First, we consider whether case average propensity score is smooth across the *Blakely* timing using both the propensity score restricted sample and the entire sample of cases. We then check for smoothness of case density, enhancement propensity, and other defendant characteristics for the propensity score reduced sample. These smoothness tests are shown in Figures A.1 and A.2 in the online appendix. All tests show smoothness through the *Blakely* timing with no evidence of any discontinuous changes in the sample composition. To ensure results are not driven by specific bandwidth or kernel choices, we vary bandwidths between 9, 12, and 15 months and consider results under rectangular, triangular, and Epanechnikov kernels. We also show results hold using quadratic fits rather than standard local linear regressions.

Because our running variable is time, one might worry that our results are driven by an anomaly in time trends that happen to occur near the timing of *Blakely.* To ensure we are capturing the causal effects of *Blakely* and not momentary deviations in enhancement or sentence lengths, we perform additional analyses under a standard event study model using a 6-year time frame (3 years before and after the timing of *Blakely*). We likewise perform a placebo test to check for discontinuities at other years. Graphs and tables for these results are presented in the online appendix but are mentioned in Section 5 to show consistency of our estimated effects.

3.3. Data

We use records provided by the Administrative Office of the Courts of North Carolina that give all post-arrest charges from 2001 to 2007. These records cover the universe of state-level criminal cases in North Carolina during this period, allowing us to observe charging, bargaining, and sentencing decisions for a large and representative population of defendants. The data contain detailed defendant and case characteristics throughout the legal process, including race, sex, age, criminal history points, type of counsel, charged and convicted statutes, plea status, and minimum sentence length. We also observe the sentencing judge. These data allow us to examine prosecutorial responses through charging and bargaining decisions and to distinguish them from heterogeneous judicial responses.

It is important to note that we do not explicitly observe whether a case receives enhancement or what factors may be driving enhancement petitions from the prosecutor. We instead identify enhancements by comparing the minimum sentence at disposition with the prescribed guideline ranges. As described in Section 2.1, each grid-cell has a prescribed guideline sentence length for cases with and without enhancement. We consider a case to be enhanced if the minimum sentence length at disposition is within the enhanced sentence length range. While we do not directly observe the aggravating factor findings themselves, any misclassification would tend to attenuate estimated effects, biasing against our main results rather than generating spurious findings.

As described in Section 3.2, our main analysis restricts the full sample to cases with an enhancement propensity score above 5 percent. This restriction substantially changes the offense composition of the sample but has little impact on defendant selection. In the full sample of cases, the three most common crime types are drugs (33.3% of sample), burglary (12.2%), and larceny (9.5%). The primary crime types in the reduced sample are robbery (24.7% of sample), assault (22.6%), and sexual assault (21%). These crime types, along with arson, homicide, and kidnapping, retain almost all the cases from the full sample. These differences reflect the fact that enhancements are rarely applied to lower-severity drug and property offenses. The propensity-score restriction therefore focuses the analysis on cases for which *Blakely* plausibly altered the incentives surrounding enhancement decisions.

Appendix Table B.1 reports summary statistics for the full and restricted samples. Defendant characteristics are similar across samples, while the restricted sample has higher

sentence lengths, charge severity, criminal history, incarceration, and enhancement rates. More than 95 percent of cases in both samples are resolved through plea bargains and the mean sentence length in the restricted sample is 38.85 months.

Our main sample, which includes cases with propensity scores over 5%, includes cases heard by 143 judges across 73 different districts, allowing us to precisely estimate heterogeneity in judicial responses. Case timing is measured using disposition month, which we use as the running variable in our regression discontinuity design. In all of our sentence length analysis, we focus on minimum sentence length, which determines time to parole eligibility and is the margin directly affected by enhancements under North Carolina law.

4. Mechanical Effects and the Sentencing Puzzle

The direct effects of *Blakely* are presented in Table 1 across three separate panels, one for each component of our main models outlined in Equation 2. Across each panel, the first three columns report discontinuity results for the sample of cases with propensity scores greater than or equal to 0.05, while varying the bandwidth across 9, 12, and 15 months. Columns 4 and 5 utilize a 12-month bandwidth but include the full sample and a sample with propensity scores over 0.03, respectively. These specifications show robustness across bandwidth selections and propensity restrictions.

Panel A reports the impact of *Blakely* on enhancement rate. In each specification, enhancement decreases dramatically following *Blakely.* Column 2 gives the result for the preferred specification, which includes a 12-month bandwidth on the 0.05 propensity score reduced sample. The estimate for this model is a decrease of 2.08 percentage points, which is a 29.5% drop from the pre-*Blakely* mean. This discontinuity is illustrated in Figure 1. While the graph shows a negative trend in enhancement pre-*Blakely*, the decrease in enhancement rate is still discontinuous and large. These effects are corroborated in the event study version of the analysis, displayed in the online appendix Table B.2.

This large decrease in enhancement suggests that sentence lengths should decrease following *Blakely.* As discussed in the methodology section, we can create a rough prediction of the magnitude of this decrease by estimating the impact of enhancement on sentence lengths and multiplying these estimates with the above discontinuity results. Panel B reports the association between enhancement and sentence length, using the same specification schedule as in Panel A.

The preferred specification returns an estimate of 18.26 months, which is a 47.3% increase over the pre-*Blakely* mean.

Multiplying the enhancement discontinuity from Panel A by the estimated enhancement–sentence length relationship from Panel B yields the benchmark prediction, $\Delta_{blakely}$. For the preferred specification, this is a 0.38 month decrease following *Blakely.* However, the observed sentence length does not follow this predicted decrease. Instead, sentence length increases after the *Blakely* decision by up to 1.996 months. Figure 2 shows the discontinuity for sentence length with the estimated benchmark of -0.38 marked using a dashed orange line. This effect is fairly consistent across our five specifications and within the event study analysis. This increase in sentence length is surprising and motivates our subsequent analysis of prosecutorial and judicial responses.

The results are robust to alternative bandwidths, kernels, polynomial specifications, standard-error procedures, and placebo cutoff dates. These robustness tests are shown in online appendix tables B.5-B.7. These results suggest that our findings are not driven by specific specification choices, seasonality, unrelated policy changes, or other spurious timing effects.

We also examine cases with low enhancement propensity scores. Because these cases have little ex ante likelihood of receiving an enhancement, they provide a useful test of whether the observed sentence length effects are concentrated among cases most directly affected by Blakely. Results in Table B.8 show main effects for cases with propensities below 0.05 and 0.03. We find enhancement rates still decrease, but with much smaller magnitudes, and noisy null impacts for sentence lengths. While estimates are imprecise, they provide little evidence of sentence length increases among low-propensity cases, suggesting that the main effects are concentrated among cases for which enhancements were plausibly relevant.

5. Prosecutorial Response: Evidence of No Compensation

In this section, we consider whether prosecutorial decisions can explain the increase in sentence lengths following *Blakely*. By making sentence enhancements more costly to pursue, *Blakely* could have induced prosecutors to compensate through other charging, bargaining, or sentencing decisions. We therefore examine three sets of outcomes: charging decisions, plea bargaining, and other sentencing practices. We estimate the same discontinuity specifications used above across 9-, 12-, and 15-month bandwidths in both the propensity-score-restricted and full

samples. The results are reported in Tables 2–4, with corresponding event studies in online appendix Figures A.5 and A.6.

5.1. Charging Decisions

If prosecutors respond to the higher burden of proof by substituting toward harsher charges, we would expect charge severity to increase following *Blakely*. We find no evidence of such a response. As shown in Table 2, across five measures of charging behavior, including charge severity, index-crime charges, number of charges, and dismissals, estimated effects are small and statistically insignificant across specifications. The largest 95 percent confidence interval permits an increase of only about 0.2 in charge severity, relative to a pre-*Blakely* mean of 5.9. Moreover, several estimates have signs that would predict lower rather than higher sentence lengths. The event studies similarly show no discernible discontinuity. Thus, prosecutors did not respond to reduced access to enhancements by pursuing more severe or numerous charges.

5.2. Plea Bargaining Outcomes

Prosecutors could also respond through plea bargaining by offering less favorable charge reductions or otherwise negotiating more stringent outcomes. We therefore examine guilty pleas, incarceration conditional on a guilty plea, reductions from initial felony charges to lower felony charges or misdemeanors, and case duration.

Table 3 shows little evidence of changes in the four bargaining outcomes. Estimated effects are small and tightly centered around zero across specifications. The only exception is a decline in plea bargains in the 12-month specification, but this effect disappears in the other bandwidths and is economically small: its 95 percent confidence interval rules out effects larger than 3.2 percentage points relative to a pre-*Blakely* mean of 95 percent.

We do, however, find a notable effect on case duration. In the propensity-score-restricted sample, cases resolve 23.57 to 26.86 days faster following *Blakely*, a 6.5 to 7.5 percent decline relative to the pre-*Blakely* mean. This reduction is consistent with *Blakely* lowering the need to gather and negotiate over enhancement-related evidence, although our data does not allow us to identify the precise stage of the process at which time is saved. Importantly, shorter case duration would predict lower rather than higher sentence lengths, making this response incapable of

explaining the increase in punishment we observe. Still, the significant impact on case duration highlights the fact that prosecutors were not unaffected by *Blakely.*

5.3. Other Sentencing and Case Outcomes

Finally, we examine whether prosecutors respond by reallocating toward other sentencing outcomes. Table 4 considers incarceration, probation length, and mitigation. With the exception of two specifications for probation, which do not persist in the event-study analysis, the estimates are small and statistically insignificant. In our preferred propensity-score-restricted specification, the 95 percent confidence interval rules out an increase in incarceration larger than 3.83 percentage points. The preferred specification also shows a noisy decline in mitigation that disappears across alternative bandwidths.

Taken together, the evidence provides little indication that prosecutors adjusted charging, bargaining, or other sentencing practices in ways that could explain the post-*Blakely* increase in sentence lengths. The one systematic prosecutorial response is a reduction in case duration, which would, if anything, predict shorter sentences. We therefore turn to judicial behavior as the source of the unexpected increase in punishment.

6. Judicial Response and Mechanism

The main results present a surprising finding: *Blakely* drastically reduced enhancement rates, yet sentence lengths increased rather than declined. The preceding section finds no evidence that observable prosecutorial responses can explain the increase, pointing instead to judicial behavior.

Judges may impact sentence lengths both directly and indirectly. If a case goes to trial, upon conviction the judge has full discretion in choosing the sentence length from within the guideline window. For cases settled by a plea deal, the judge must sign off on the agreed upon deal, a key part of which is the agreed upon sentence length. The judge may also assign a sentence length outside of the recommended, agreed upon sentence length in the deal. If this occurs, the prosecution or defense has the right to withdraw the plea and enter negotiations or move towards trial (North Carolina Defender Manual 2018). But even outside of these direct impacts, the assigned trial judge is likely to impact plea negotiations by reputation. This is evidenced in LaCasse and Payne (1999), which finds plea negotiations occur under the shadow of the judge, particularly for cases with minimum sentencing.

We use two approaches to assess whether judicial behavior drives the post-*Blakely* increase. First, because judges have greater direct control over sentencing in jury trials, we repeat the main sentence-length discontinuity analysis for cases resolved by jury trial. Second, we examine whether the effects vary systematically with judge leniency. If the increase is driven by judicial behavior, we would expect it to be concentrated among judges with greater scope to adjust sentences within the remaining discretionary range. We describe the construction of our judge-leniency measure below before presenting these heterogeneous effects.

6.1. Measuring Judicial Leniency

There are two separate ways we consider judge leniency – by average enhancement rate and by average sentence length. Judges do not directly impact enhancement but may do so indirectly through reputation effects. That is, a prosecutor may be more likely to pursue enhancement in a given case depending on the assigned judge. For each of these leniency measures, we residualize enhancement/sentence length by baseline defendant characteristics, grid fixed effects, and year fixed effects using only pre-*Blakely* data. Residualizing the measures is important; judges who see especially severe crime types or offenders with a large prior history more often than others will look more severe without controls. By controlling for these factors, we effectively compare judges' sentencing and enhancement propensities for cases of a similar type.

This residual measure is then averaged for each judge and standardized to have a mean of 0 and a standard deviation of 1. Given the resulting continuous measure of averages, we then consider leniency by quartiles, with the lowest quartile representing the most lenient judges and the highest quartile representing the most strict ones. That is, the most lenient judges adjudicate cases with the fewest enhancements/lowest sentence lengths, after controlling for offense class, criminal history, defendant characteristics, and unit-invariant variation over time.

The two measures capture related but distinct dimensions of judicial leniency. The enhancement measure captures heterogeneity in judges' propensity to adjudicate cases with enhancements, while the sentence-length measure captures heterogeneity in sentence severity at conviction. These measures need not coincide; judges who impose longer sentences may not be more likely to allow enhancements. Figure A.7 in the online appendix shows that the two measures are nevertheless strongly related, with most judges who are strict on enhancements also strict on sentence length. We therefore use the sentence length measure in our main analysis and report

results using the enhancement measure in the online appendix. We prefer the sentence length measure because sentence lengths are observed in most cases, providing a more precise measure of judge behavior than the relatively rare enhancement outcome, and because sentence length is more likely to capture indirect effects operating through plea bargaining.

It's worth noting that because we separate by quartile of judge and not of case, it's possible that judges in one quartile see far more cases than in others. Indeed, we find that with both measures of leniency, the number of cases is nearly equal for the bottom three quartiles, but higher for the strictest judges. This is especially true when using the sentence length measure, where the top quartile has around 27 percent more cases than the other quartiles. One might wonder whether judges in each quartile have observably different characteristics. While we do not observe many judge characteristics, we can say that judge tenure is nearly identical across the quartiles.

Because the heterogeneous analysis effectively splits effects across four groups, we consider a larger time horizon of 6 years (3 years pre and post) for the heterogeneous regressions in order to maintain statistical power. Thus, we now employ a difference-in-differences design as outlined in the below equation:

Equation 3:
$$Y_{itg} = \alpha + \beta_1 Blakely_t + \beta_2 leniency_50_{it} + \beta_3 leniency_75_{it} + \beta_4 leniency_100_{it} + \beta_5(Blakely \cdot leniency_50_{it}) + \beta_6(Blakely \cdot leniency_75_{it}) + \beta_7(Blakely \cdot leniency_100_{it}) + \eta_g + \rho_t + u_{it}$$

where Y_{itg} measures enhancement or sentence length for case i in year t, sentenced in guideline grid g. Coefficient β_1 gives the effect of *Blakely* for the most lenient judges. Coefficients $\beta_5, \beta_6,$ and β_7 give the differential impact of *Blakely* for each leniency group compared to the most lenient group.

6.2. Heterogeneous Judicial Responses to *Blakely*

We first briefly present the jury-trial discontinuity results before discussing the heterogenous leniency findings. Figure 3 presents the discontinuity in sentence length at the timing of *Blakely* for cases that were resolved by trial. The graph shows a large increase in sentence length immediately following *Blakely*. Table B.9 presents the estimates for the full set of sentence length discontinuity specifications for the jury trial cases, which shows significantly larger magnitudes compared to the main analysis results. While statistical significance is dependent on specification

due to small samples, each regression shows a large increase in sentence length. Figure A.8 in the online appendix illustrates the discontinuity for the full sample of cases and the residual sentence length graph. In each case, sentence length increases greatly following *Blakely.* These findings suggest that the post-*Blakely* increase in sentence lengths is driven by judicial behavior, motivating our analysis of how the response varies with judicial leniency.

Table 5 gives the heterogeneous effects of *Blakely* on both enhancement and sentence length across the four judge leniency quartiles. Columns 2 and 3 give the estimates for the enhancement effect of *Blakely*. Judges who are most lenient (based on the residualized, sentence length-leniency score) exhibit a 2.38 to 2.52 percentage point decrease in enhancement rate following *Blakely.* The next two quartiles of judges show no significant difference from the most lenient quartile. However, the strictest quartile of judges shows a decrease that is nearly double the most lenient group. The difference between the most strict and lenient judge enhancement effects is illustrated in Figure 4. Note that *Blakely* completely eliminates the pre-period discrepancy in enhancement rate.

Columns 3 and 4 of Table 5 report the estimates for the sentence length effects. The most lenient judges see significant increases in sentence length, with an estimated 2.23 to 2.68 month increase in average sentence length following *Blakely,* despite enhancement rates falling. The other quartiles show increasingly negative effects, with the second quartile exhibiting small increases following *Blakely,* the third quartile showing null effects, and the most strict group of judges showing significant decreases in sentence length. Figure 5 visualizes these effects for the most lenient and most strict judicial groups. We also show these results are robust to a discontinuity design with similar specifications to the mechanical analysis in Section 4. These RD results are shown in Tables B.11 and B.12 in the online appendix.

6.3. Interpretation and Mechanisms

These results uncover three important findings: (1) the sentence length increase shown in Section 4 of this paper is driven by judges, (2) judicial responses are highly heterogeneous: most judges increase sentence lengths following *Blakely*, while the strictest judges reduce them, and (3) these asymmetric responses generate a pronounced compression of sentencing outcomes across judges, narrowing disparities in both sentence lengths and enhancement use. In regard to the first finding, we interpret the scaling of the effect size with judicial leniency as evidence that the

increase in sentence lengths is driven by judicial behavior rather than prosecutorial behavior. While we cannot fully rule out unobserved prosecutorial responses, such as more intensive efforts to seek enhancement-related evidence in each case, it is not clear why such behavior would vary systematically with judicial leniency. Furthermore, we repeat the prosecutor analysis in Section 5 separately by judicial leniency quartile and again find consistent null effects.

Our second main result is somewhat surprising. One might expect cases assigned to the strictest judges to exhibit the largest increase in sentence length, since these judges should be most directly affected by the loss of enhancement authority. Instead, we find significant decreases in sentence length, which align with the observed decline in enhancement rates. This pattern suggests that judges differ in their ability to respond along the remaining sentencing margin. Following *Blakely*, judges retain discretion to select a sentence within the applicable guideline range. Because we find no evidence of changes in prosecutorial charging behavior, the increase in sentence lengths does not appear to reflect changes in the severity of cases reaching judges. Instead, judges adjust the sentence imposed within the remaining guideline range. Lenient judges have greater scope to increase sentences along this margin, while strict judges are more likely to be constrained by the upper bound of the guideline range.

To assess whether this explanation is consistent with the data, we examine the share of non-enhanced cases in which the imposed sentence equals the maximum guideline sentence. We estimate the following regression:

Equation 4: $$Y_{itg} = \alpha + \beta_1 leniency_50_{it} + \beta_2 leniency_75_{it} + \beta_3 leniency_100_{it} + \eta_g + \rho_t + u_{it}$$

where $Y_{itg}$is an indicator equal to one if the minimum sentence imposed equals the maximum amount permitted under the guidelines. The results indicate that judges in higher strictness quartiles are substantially more likely to sentence at the guideline maximum. Judges in the strictest quartile are 8.69 percentage points more likely to impose the maximum sentence than judges in the most lenient quartile, a 28 percent increase relative to the latter's pre-*Blakely* mean. We repeat this analysis with varied controls and with the full sample and find similar results (Online Appendix Table B.13). These findings suggest that guideline ceilings bind more frequently for strict judges, limiting their ability to offset the decline in enhancement use, while leaving greater scope for lenient judges to increase sentences within the remaining guideline range.

Our third key finding speaks to how changes in procedural constraints reshape the distribution of sentencing outcomes. Consistent with prior work on judicial discretion, restricting discretion through higher standards of proof is expected to compress sentencing outcomes by limiting the ability of some judges to impose especially severe punishments. In our setting, however, this compression arises through two distinct behavioral responses: harsh judges reduce sentence lengths as their ability to rely on enhancements becomes constrained, while more lenient judges increase sentences within the remaining discretionary range. As a result, sentencing outcomes converge toward the middle of the distribution. As illustrated in Figure 5, both lenient and strict judges move toward a common median sentence length following *Blakely*, generating substantial compression across judges. This compression result illustrates how constraining one margin of judicial discretion can reshape the distribution of outcomes along another, rather than simply reducing punishment.

6.3.1. Does *Blakely* Spill Over to Other Cases?

The evidence above establishes why the response is concentrated among lenient judges: strict judges have less room to increase sentences within the guideline range. We next examine whether the response is concentrated among the relevant cases – those with a high pre-policy propensity of enhancement. If judges are responding to *Blakely* in an effort to maintain pre-reform sentencing patterns, we would expect larger responses among defendants who would have been likely to receive an enhancement before the reform, but little response among defendants for whom enhancements were unlikely.

We repeat the judicial heterogeneity difference-in-differences analysis for cases with a low predicted probability of enhancement. These results are presented in Table B.14. This analysis provides a sharper test of the results in Table B.7, which examines the overall effects of *Blakely* among low-probability cases. Here, we ask whether the heterogeneous sentence-length responses across judges extend broadly across cases or are concentrated among cases for which enhancement was more likely before *Blakely*. The estimates show that heterogeneity across judges largely dissipates among low-probability cases. Even among the judges with the largest sentence-length increases in the full sample, the estimated effects are statistically insignificant and small in magnitude. We therefore find little evidence of spillover to cases that were unlikely to receive an enhancement before *Blakely*. Instead, the results suggest that the judges who

increase sentence lengths following *Blakely* do so primarily among cases for which enhancement was a meaningful pre-policy possibility.

6.3.2. Consistency and Information

One potential explanation for the judicial response is that judges have a desire for consistency in sentencing. This interpretation is consistent with a growing literature showing that judges use other cases in their caseloads and statutory reference points when determining sentence lengths (Leibovitch 2016; Drápal and Šoltés 2024; Burýšková 2026). In our setting, judges may use their pre-*Blakely* sentencing patterns as a reference point when determining how to sentence cases affected by the loss of enhancement authority. For cases that they believe would have received an enhancement absent *Blakely*, judges may seek to maintain their pre-reform sentencing levels by increasing sentences within the remaining guideline range. This interpretation also predicts that judicial responses should be concentrated among cases with a high pre-*Blakely* probability of enhancement, rather than extending broadly to cases that were unlikely to receive an enhancement. Consistent with this prediction, the preceding spillover analysis finds little evidence of heterogeneous judicial responses among low-probability cases.

A complementary mechanism operates through the information conveyed by the enhancement process. An enhancement filing may provide judges with information about crime severity because prosecutors seeking an enhancement must present evidence of statutory aggravating factors. Because *Blakely* increased the cost of obtaining an enhancement, prosecutors may have less incentive to gather and present such evidence through the enhancement process. If the resulting signal becomes noisier, judges may face greater uncertainty about which cases warrant lower or higher sentences within the guideline range and rely more heavily on a pooling equilibrium, moving sentences toward the center of the distribution. This mechanism is consistent with the compression we observe, in which lenient judges increase sentences while harsher judges reduce them. Importantly, the mechanism does not require prosecutors to possess or present less underlying information: evidence that would previously have been part of an enhancement filing may still be available elsewhere in the case. Rather, the formal enhancement signal may change how judges receive or weight that information. A simple model of this mechanism is presented in online appendix C.

The two mechanisms are not mutually exclusive and both imply that judges perceive at least some cases affected by *Blakely* as warranting higher sentences than the mechanical loss of

enhancement authority would imply. Under the consistency interpretation, judges may overestimate the share of cases that would have received an enhancement before *Blakely*; under the information interpretation, judges may place greater weight on aggravating circumstances when selecting a sentence within the guideline range. Either mechanism can generate upward pressure on sentences among lenient judges, while guideline ceilings limit the ability of harsher judges to respond along the same margin. Thus, the institutional environment can transform similar underlying behavioral responses into the asymmetric effects we observe.

We also consider alternative explanations for the heterogeneous judicial response, including differences in judicial characteristics and other judicial decisions. We discuss these possibilities in online appendix D.

7. Conclusion

This paper studies how institutional actors respond to reforms that constrain discretion in the U.S. judicial system. We examine the Supreme Court's decision in *Blakely v. Washington*, which limited one important margin of punishment by increasing the evidentiary burden required for sentence enhancements. Although *Blakely* sharply reduced the use of sentence enhancements, we find that average sentence lengths increased rather than declined. This increase runs counter to the mechanical implications of the sentencing regime and highlights the importance of behavioral responses to procedural reforms.

We show that this result is driven by judicial, rather than prosecutorial, behavior. Across a wide range of charging and bargaining outcomes, we find no evidence of prosecutorial response. Instead, judges respond heterogeneously: stricter judges reduce sentences as enhancements become harder to obtain, while more lenient judges increase sentences within the remaining discretionary range. These opposing responses generate a compression of sentences across judges and raise average sentence lengths overall. This compression arises because guideline ceilings constrain the strictest judges, leaving more room for lenient judges to increase sentences along the remaining discretionary margin. We also find little evidence that these responses spill over to cases with little ex ante likelihood of receiving an enhancement, suggesting that judicial responses are concentrated among cases most directly affected by the reform. Together, these findings indicate that judges adapt to procedural reforms in ways that can offset their intended effects. The

heterogeneous responses across judges show that the effects of institutional constraints depend on the discretion that remains available to decision-makers.

These findings illustrate that constraining one margin of discretion need not reduce punishment as intended. Instead, institutional actors may adapt their behavior within the remaining scope of discretion, altering both the level and distribution of punishment. Our findings suggest that measured judicial heterogeneity depends in part on the institutional environment in which judges operate. More broadly, our results illustrate that evaluating institutional reforms requires accounting for both their direct mechanical effects and the behavioral responses they induce. Ignoring these responses can substantially misrepresent the ultimate effects of legal and policy change.

References

Agan, Amanda., Jennifer L. Doleac, and Anna Harvey. (2023). Misdemeanor Prosecution. *Quarterly Journal of Economics*, *138*(3).

Alexander, Diane. 2020. "How Do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs." *Journal of Political Economy* 128(11): 4046-4096.

Angelova, Victoria, Will Dobbie, and Crystal Yang. 2026. "Algorithmic Discretion and Human Discretion." *Review of Economic Studies* (forthcoming).

Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Quarterly Journal of Economics* 133 (4).

Bertrand, M., Burgess, R., Chawla, A., and Xu, G. (2020). The glittering prizes: Career incentives and bureaucrat performance. *The Review of Economic Studies,* 87(2):626-655.

Bjerk, David. 2005. "Making the Crime Fit the Penalty: The Role of Prosecutorial Discretion under Mandatory Minimum Sentencing." *Journal of Law and Economics* 48 (2).

Boylan, Richard T. 2005. "What Do Prosecutors Maximize? Evidence from the Careers of U.S. Attorneys." *American Law and Economics Review* 7 (2).

Burýšková, Tereza. 2026. "Categorization and Reference-Dependent Decision-Making under Statutory Thresholds: Evidence from a Criminal Law Reform." Working paper.

Cooper, Spencer. 2026. "How Much Do Mandatory Minimums Matter?" *Journal of Law, Economics, and Organization* (forthcoming).

Didwania, Stephanie. 2022. "Gender Favoritism Among Criminal Prosecutors." *The Journal of Law and Economics* 65(1): 77-104.

Dobbie, Will, Jacob Goldin, and Crystal S Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108 (2).

Drápal, Jakub, and Michal Šoltés. 2024 "Sentencing decisions around quantity thresholds: theory and experiment." *Journal of Experimental Criminology.* 20: 1323-1367.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie. 2023. “Judging Judge Fixed Effects: Testing the Identifying Assumptions in Judge Fixed-Effects Designs.” *American Economic Review* 113(1): 253-77.

Holmström, Bengt and Paul Milgrom. (1991). “Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design.” *Journal of Law, Economics, and Organization,* 7 (Special Issue):24-52.

Jacob, Brian A., and Steven Levitt. 2003. “Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating.” *Quarterly Journal of Economics.* 118 (3): 843-877.

Khan, A. Q., Khwaja, A.I., and Olken, B. A. 2019. Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings. *American Economic Review,* 109(1):237-270.

Kling, Jeffrey R. 2006. “Incarceration Length, Employment, and Earnings.” *American Economic Review*, 96(3): 863-876.

Kuziemko, Ilyana, 2013. “How Should Inmates Be Released From Prison? An Assessment of Parole Versus Fixed Sentence Regimes.” *Quarterly Journal of Economics,* 128(1): 371-424.

Leslie, Emily and Nolan G. Pope. 2017. “The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments.” *The Journal of Law and Economics*, 60(3):529–557.

Leibovitch Adi. 2016. “Relative Judgements” *The Journal of Legal Studies*, 45(2).

Rehavi, M. Marit, and Sonja B. Starr. 2014. “Racial Disparity in Federal Criminal Sentences.” *Journal of Political Economy* 122 (6).

North Carolina Defender Manual. 2018. Vol. 2, Trial. "Ch. 23: Guilty Pleas (The Plea Procedure)." UNC School of Government. Accessed September 3, 2025. https://defendermanuals.sog.unc.edu/sites/default/files/pdf/23.4%20The%20Plea%20Procedure_0.pdf.

North Carolina Administrative Office of the Courts. 2001–2007. *Administrative Records on Criminal Charges and Sentencing*. Raleigh, NC.

Petkun, Jonathan. 2026. “Non-monetary Incentives and Bureaucratic Performance: Evidence from U.S. Courts.” *Journal of Public Economics* 262.

Prescott, JJ. “Essays in Empirical Law and Economics.” (Ph.D. Thesis, Massachusetts Institute of Technology, 2006), http://hdl.handle.net/1721.1/37408.

Prescott, JJ, and Sonja Starr. 2006. “Improving Criminal Jury Decision Making After the Blakely Revolution.” *University of Illinois Law Review* 301.

Shaffer, Hannah. 2023. “Prosecutors, Race, and the Criminal Pipeline.” *The University of Chicago Law Review* 90 (7): 1889–1965.

Shaffer, Hannah, and Emma Harrington. 2025. “Brokers of Bias in the Criminal Justice System: Do Prosecutors Compound or Attenuate Racial Disparities Introduced by Police?” Working paper.

Shearer, Bruce, Nibene H. Somé, and Bernard Fortin. 2018. “Measuring Physicians’ Response to Incentives: Evidence on Hours Worked and Multitasking.” *IZA Discussion Paper* 11565.

Silveira, Bernardo S. 2017. “Bargaining With Asymmetric Information: An Empirical Study of Plea Negotiations.” *Econometrica* 85 (2).

Sloan, CarlyWill. 2026. “Do Prosecutor and Defendant Race Pairings Matter? Evidence from Random Assignment.” *Journal of Public Economics* (forthcoming).

Starr, Sonja, M. Marit Rehavi. 2013. “Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of Booker.” *Yale Law Journal* 123: 2.

Tuttle, Cody. 2026. “Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums.” *Review of Economic Studies* (forthcoming).

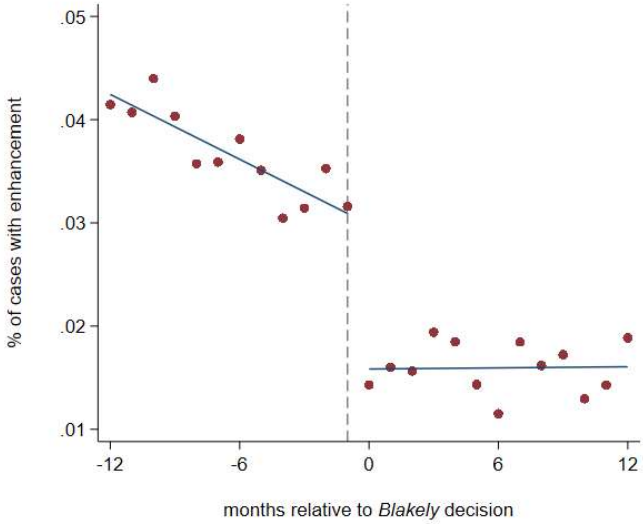
Yang, Crystal S. 2015. “Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing.” *The Journal of Legal Studies* 44 (1): 75-11.

Yang, Crystal S. 2016. “Resource Constraints and the Criminal Justice System: Evidence from Judicial Vacancies.” *American Economic Journal: Economic Policy* 8 (4).

Yuan, Andy and Spencer Cooper. 2025. “Prosecutorial Incentives and Outcome Disparities.” Working paper.

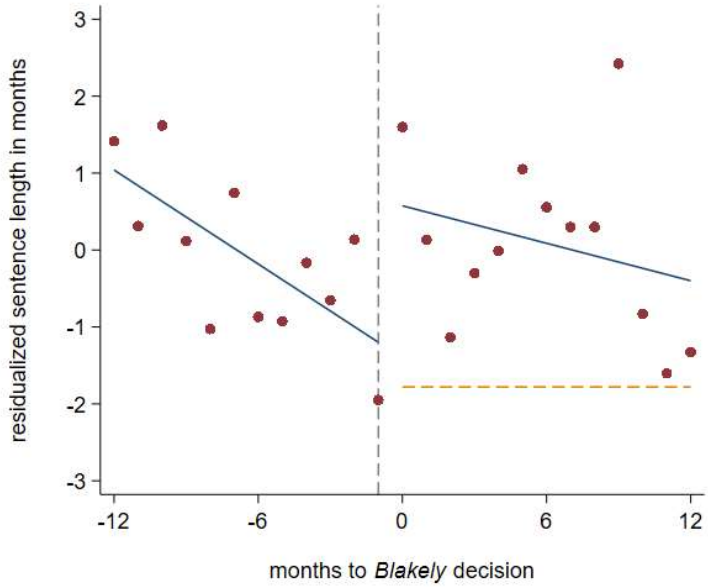
Figures

Figure 1. Effect of *Blakely* on enhancement



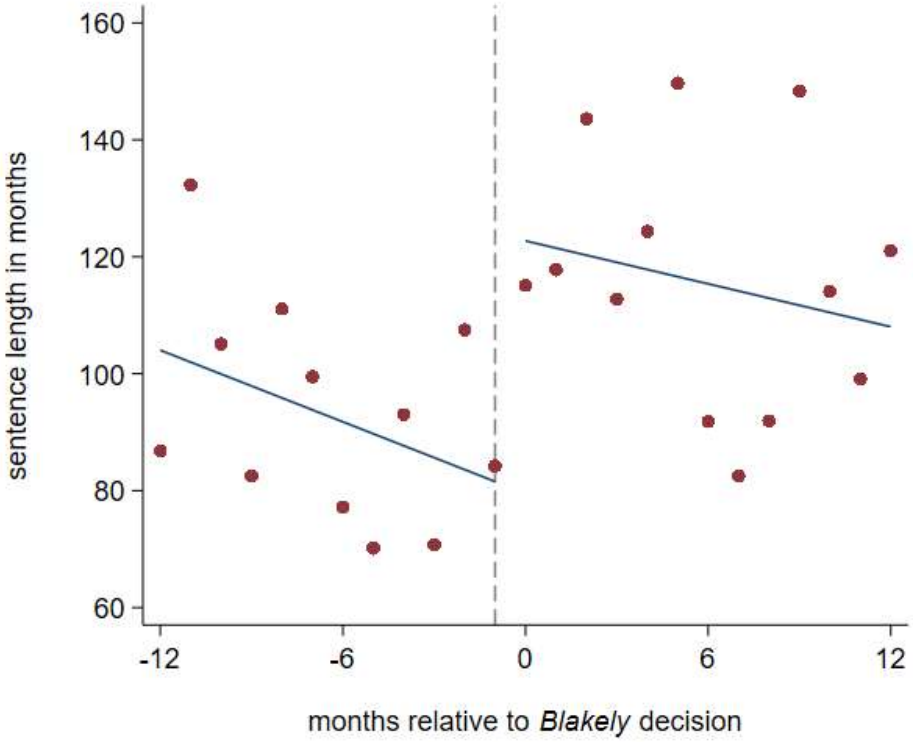
Notes: This graph shows the discontinuity for the percentage of cases with enhancement under a 12-month bandwidth with linear fits for the full sample of cases. The dashed vertical line indicates the timing of the *Blakely* decision.

Figure 2. Effect of *Blakely* on sentence length



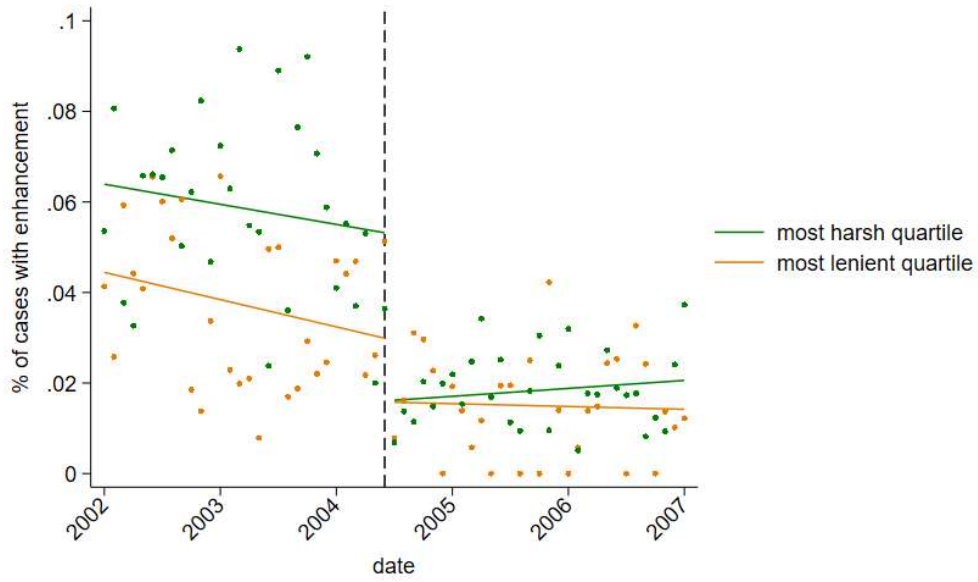
Notes: This graph gives the discontinuity for residualized sentence length using a 12-month bandwidth and linear fit. Sentence length is residualized on grid fixed effects. The vertical dashed line indicates the timing of the *Blakely* decision. The orange, horizontal dashed line gives the predicted decrease in sentence length following *Blakely,* termed the benchmark value. The sample includes cases with propensity scores greater than 0.05.

Figure 3. Effect of *Blakely* on sentence length – jury trial cases



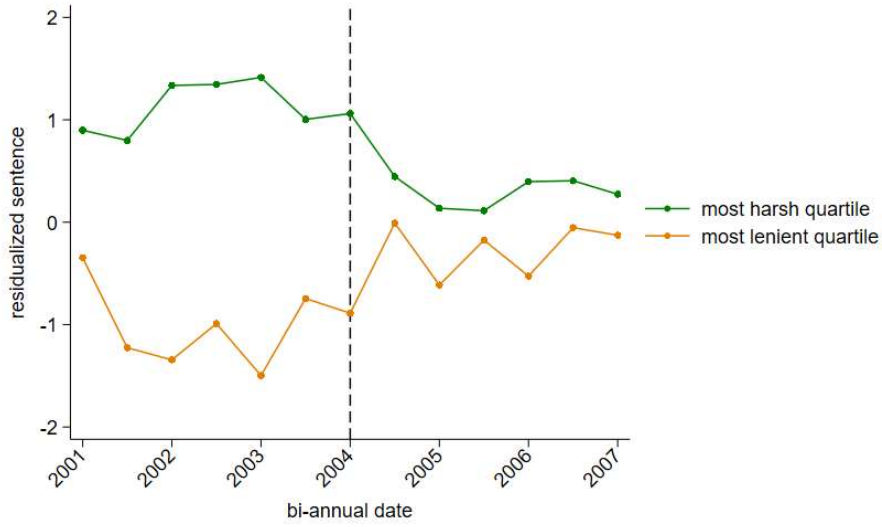
Notes: This graph gives the discontinuity for sentence lengths for cases resolved by jury trial and with a propensity scores greater than 0.05. The fit is linear and bandwidth is 12 months.

Figure 4. Effects of *Blakely* on Enhancement – split by judge leniency



Notes: This graph illustrates heterogeneity in the effects of *Blakely* on enhancement rates by judge type. Specifically, the green points give averages for the harshest quartile of judges while the orange points are for the most lenient quartile. The sample is restricted to cases with a propensity score higher than 0.05.

Figure 5. Effects of *Blakely* on Residual Sentence – split by judge leniency



Notes: This graph gives average residual sentence lengths for the most lenient and harshest judge quartiles. Sentence lengths are residualized on grid and year fixed effects. The sample only includes cases with propensity scores greater than 0.05.

Tables

Table 1. The main effects of *Blakely*

Panel A. The effect of *Blakely* on enhancement rate

	(1)	(2)	(3)	(4)	(5)
Blakely	-0.0200*	-0.0208**	-0.0311***	-0.0140***	-0.0201***
	(0.0119)	(0.0091)	(0.0086)	(0.0017)	(0.0056)
Pre-*Blakely* mean	0.0653	0.0704	0.0670	0.0368	0.0564
p-hat restriction	≥ 0.05	≥ 0.05	≥ 0.05	none	≥ 0.03
bandwidth	9	12	15	12	12
N	6952	9314	11556	39955	16332

*Panel B***.** The effect of enhancement on sentence length

	(1)	(2)	(3)	(4)	(5)
Aggravated	17.33***	17.39***	17.92***	8.82***	13.26***
	(1.846)	(2.016)	(1.751)	(0.871)	(1.547)
Pre-*Blakely* mean	37.55	38.57	38.59	18.55	29.27
p-hat restriction	≥ 0.05	≥ 0.05	≥ 0.05	none	≥ 0.03
Bandwidth	9	12	15	12	12
N	3260	4502	5601	19141	7905

Panel C. The effect of *Blakely* on sentence length

	(1)	(2)	(3)	(4)	(5)
Blakely	1.024	1.996***	1.082	0.493	0.929
	(0.927)	(0.719)	(0.735)	(0.349)	(0.695)
pre-*Blakely* sentence length	37.55	38.57	38.59	18.55	29.27
$\widehat{\Delta}_{blakely}$ (benchmark)	-0.347	-0.362	-0.557	-0.123	-0.267
t-score (benchmark)	1.48	3.28	2.24	1.77	1.72
p-hat restriction	≥ 0.05	≥ 0.05	≥ 0.05	none	≥ 0.03
bandwidth	9	12	15	12	12
N	6952	9314	11556	39955	16332

Notes: This table presents the main three effects as described in Equation 2, with Panel A giving estimates of β_1 in equation 2.1, Panel B giving estimates of δ_1 in Equation 2.2, and Panel C giving estimates of γ_1 in Equation 2.3. $\widehat{\Delta}_{blakely}$ is the predicted, mechanical decrease calculated using the coefficients in Panels A and B. The t-score (benchmark) row gives the t-score value calculated using the benchmark value as the null hypothesis. Across all three panels, standard errors are clustered at the month level, a uniform kernel function is used, and each specification includes grid fixed effects. The p-hat restriction is based on the propensity score described in Section 4. * p < 0.1, ** p < 0.05, *** p < 0.01

Table 2. Effect of *Blakely* on Charging Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables:						
charge severity	0.0014	-0.0303	-0.0043	0.0161	0.0182	0.0445
	(0.0544)	(0.0959)	(0.0502)	(0.0868)	(0.0444)	(0.0797)
	{3.597}	{5.910}	{3.600}	{5.917}	{3.597}	{5.900}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
index crime	-0.0206	-0.0278	-0.0175	-0.0165	-0.0148	-0.0147
	(0.0138)	(0.0208)	(0.0126)	(0.0154)	(0.0117)	(0.0140)
	{0.415}	{0.828}	{0.412}	{0.830}	{0.411}	{0.831}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
# of charges (pre-consol.)	-0.150	-0.273	-0.116	-0.211	-0.0284	-0.115
	(0.106)	(0.253)	(0.0926)	(0.186)	(0.111)	(0.157)
	{3.240}	{2.837}	{3.212}	{2.842}	{3.195}	{2.800}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
# of charges (post-consol.)	0.0103	-0.0845	0.0028	-0.0925*	0.0093	-0.0534
	(0.0286)	(0.0568)	(0.0263)	(0.0444)	(0.0259)	(0.0458)
	{1.754}	{1.674}	{1.750}	{1.685}	{1.742}	{1.675}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
dismiss	-0.0045		-0.0033		0.0082	
	(0.0090)		(0.0075)		(0.0062)	
	{0.104}		{0.105}		{0.104}	
	[34519]		[46059]		[57302]	
p-hat restriction	none	≥ 0.05	none	≥ 0.05	none	≥ 0.05
bandwidth	9	9	12	12	15	15

Notes: This table presents results of RD analysis for five different charging outcome dependent variables across six different specifications. The reported effect is the coefficient representing the effect of *Blakely* on each dependent variable. All specifications include criminal history category fixed effects, besides for dismissed cases which do not have that data. Standard errors, clustered at the month level, are presented in parentheses. Pre-*Blakely* means are presented in curly braces. Sample sizes are presented in brackets. The p-hat restriction is based on the propensity score described above. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3. Effect of *Blakely* on Plea Bargain Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables:						
plea bargain	-0.0016	-0.0104	-0.0053**	-0.0169**	-0.0023	-0.0109
	(0.0019)	(0.0081)	(0.0021)	(0.0079)	(0.0019)	(0.0071)
	{0.979}	{0.952}	{0.979}	{0.953}	{0.979}	{0.954}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
plea to incarceration	0.0159*	-0.0043	0.0103	-0.0130	0.0081	-0.0199
	(0.0089)	(0.0156)	(0.0082)	(0.0152)	(0.0068)	(0.0133)
	{0.355}	{0.545}	{0.358}	{0.550}	{0.357}	{0.546}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
plea to lower felony	0.0222*	0.0511*	0.0201*	0.0359	0.0193**	0.0351
	(0.0109)	(0.0249)	(0.0102)	(0.0239)	(0.0094)	(0.0220)
	{0.142}	{0.276}	{0.144}	{0.273}	{0.144}	{0.272}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
plea to misdemeanor	0.0028	-0.0002	0.0017	-0.0009	0.0038	0.0038
	(0.0051)	(0.0117)	(0.0044)	(0.0102)	(0.0044)	(0.0101)
	{0.036}	{0.054}	{0.036}	{0.053}	{0.036}	{0.053}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
case duration (in days)	-10.71	-26.61**	-15.12*	-23.57**	-9.15	-26.86***
	(7.71)	(9.85)	(7.66)	(9.44)	(6.98)	(8.52)
	{338.9}	{363.6}	{337.1}	{360.3}	{336.1}	{360.2}
	[26656]	[5880]	[35525]	[7909]	[44014]	[9782]
p-hat restriction	none	≥ 0.05	none	≥ 0.05	none	≥ 0.05
bandwidth	9	9	12	12	15	15
N	30002	6952	39955	9314	49699	11556

Notes: This table presents results of RD analysis for five different plea bargain outcome dependent variables across six different specifications. The first four variables are binary while case duration is measured in number of days to resolution. All specifications include grid fixed effects. Standard errors, clustered at the month level, are presented in parentheses. Pre-*Blakely* means are presented in curly braces and sample size in brackets. The p-hat restriction is based on the propensity score described above. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4. Effect of *Blakely* on Sentencing Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables						
incarceration	0.0190*	0.0081	0.0156*	0.0047	0.0105	-0.0079
	(0.0092)	(0.0154)	(0.0088)	(0.0146)	(0.0074)	(0.0131)
	{0.371}	{0.589}	{0.374}	{0.592}	{0.373}	{0.588}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
probation	1.138***	-0.881	0.741**	-0.781	0.444	-0.660
	(0.265)	(0.564)	(0.275)	(0.558)	(0.292)	(0.498)
	{31.41}	{35.59}	{31.52}	{35.53}	{31.49}	{35.44}
	[18824]	[2848]	[24982]	[3785]	[31143]	[4751]
mitigation	-0.0016	-0.0085	-0.0058	-0.0228*	0.0060	-0.0013
	(0.0050)	(0.0136)	(0.0053)	(0.0126)	(0.0050)	(0.0129)
	{0.096}	{0.167}	{0.096}	{0.166}	{0.095}	{0.163}
	[30002]	[6952]	[39955]	[9314]	[49699]	[11556]
p-hat restriction	none	≥ 0.05	none	≥ 0.05	none	≥ 0.05
bandwidth	9	9	12	12	15	15

Notes: This table presents results of RD analysis for three different sentencing outcome dependent variables across six different specifications. All specifications include grid fixed effects. Standard errors, clustered at the month level, are presented in parentheses. Pre-*Blakely* means are presented in curly braces. Sample sizes are presented in brackets. The p-hat restriction is based on the propensity score described above. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5. Heterogeneous effects of *Blakely* by judge leniency

	(1) enhancement	(2) enhancement	(3) min. sent	(4) min. sent
Blakely	-0.0238***	-0.0252***	2.680***	2.226***
	(0.0073)	(0.0073)	(0.635)	(0.653)
DD_50	0.0103	0.0101	-1.043*	-0.951
	(0.0083)	(0.0084)	(0.593)	(0.632)
DD_75	-0.0078	-0.0080	-2.247***	-2.221***
	(0.0096)	(0.0098)	(0.660)	(0.689)
DD_100	-0.0243***	-0.0240***	-4.222***	-4.114***
	(0.0090)	(0.0089)	(0.583)	(0.597)
Fixed effects	yes	yes	yes	yes
Additional controls	no	yes	no	yes
Pre-*Blakely* mean	0.0723	0.0723	39.20	39.20
N	27536	27536	27536	27536

Notes: This table gives the differential effects of *Blakely* on enhancement rate and sentence length across the four judge leniency groups. The sentence length leniency measure is used throughout for these effects. The *Blakely* coefficient gives the effect for the most lenient judges, while DD_50, DD_75, and DD_100 give the difference-in-difference estimates for increasingly strict judges. All specifications include grid and year fixed effects. The additional controls include age, a male dummy, race, defense type, criminal history points, the type of crime, and annual arrest and population rates. Standard errors are clustered at the judicial district level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$