Disclosure Statement for “The Unintended Consequences of Raising the Standard of Proof in Criminal Sentencing,” 2026

Coauthor Spencer Cooper declares that he has no relevant or material financial interests that relate to the research described in this paper.

Disclosure Statement for “The Unintended Consequences of Raising the Standard of Proof in Criminal Sentencing,” 2026

Coauthor Andy Yuan declares that he has no relevant or material financial interests that relate to the research described in this paper.

The Unintended Consequences of Raising the Standard of Proof in Criminal Sentencing

Spencer Cooper* & Andy Ye Yuan†

Abstract

We study how raising the standard of proof affects charging and sentencing outcomes. Exploiting a Supreme Court decision that increased the evidentiary burden for sentencing enhancements, we document a sharp decline in enhancement rates. While this change should mechanically reduce average sentence lengths, average sentences instead increase. We find no evidence of prosecutorial adjustment across 13 charging and bargaining outcomes. By contrast, judges respond heterogeneously: more lenient judges increase sentence lengths, while stricter judges reduce them, compressing the distribution of sentences. These results show that tightening procedural constraints can unintentionally lead to harsher punishment through judicial responses.

* Assistant Professor of Economics, University of Connecticut
† Assistant Professor of Law, University of Florida Levin College of Law.

1. Introduction

To what degree do legal actors adjust their behavior in response to procedural changes? We analyze how behavioral responses by prosecutors and judges to a policy change can lead to unintended consequences in criminal court outcomes. Understanding and anticipating such responses is essential for designing effective policy and case law. In this paper we focus on detecting responses from prosecutors and judges, two actors who exercise broad discretion with little oversight. Specifically, we consider how each adjusts behavior in response to an increase in the standard of proof required for certain charges. We find no evidence of a behavioral change by prosecutors, but significant changes in sentencing behavior from judges in both trial and plea cases.

In the United States criminal justice system, prosecutors and judges exercise a high degree of discretion across multiple margins. Prosecutors decide which cases to charge and which charges to file against a defendant. This initial charging choice, together with subsequent prosecutorial choices, impacts many of the potential outcomes a defendant faces, including bail determinations, plea offers, and final sentence lengths. Judges preside over trials, instruct juries, rule on evidence, decide whether to accept or reject plea agreements, and determine sentences upon conviction. Because of this wide-ranging impact, prosecutors and judges may be able to respond to changes in incentives, costs, or discretion by substituting toward other decision margins. We refer to this process as compensatory behavior, whereby a reduction in discretion along one dimension is offset by more punitive decisions along another.

Measuring compensatory behavior 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, isolating the causal source of outcome changes becomes difficult. Moreover, highly salient or consequential 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. In this paper, we exploit Supreme Court case *Blakely v. Washington* (hereafter referred to as *Blakely)* to mitigate these concerns and estimate prosecutorial and judicial behavioral responses.

Blakely works as an optimal reform for our study for two main reasons. First, the decision impacted charging and sentencing through a single dimension – the standard of proof. *Blakely*

held that any facts that increase the statutory maximum sentence length of a charge must be found by jury. This effectively shifts the burden of proof required for sentence enhancements from a preponderance-of-the-evidence standard to proof beyond a reasonable doubt. Because the decision left other features of charging and sentencing unchanged and applied only to cases eligible for enhanced sentences, the direct, mechanical effects of *Blakely* absent compensation are confined to a small set of outcomes.

The second benefit of the *Blakely* decision lies in the predictability of its effects. By effectively raising the cost of pursuing enhanced sentences, the decision should, absent compensatory adjustments by legal actors, lead to a reduction in the frequency of sentence enhancements. If such reductions occur, the implications for sentence length are straightforward: because enhanced sentences are always at least as long as their non-enhanced counterparts for a given conviction, any decline in enhancements necessarily results in shorter average sentences, provided prosecutorial and judicial charging and sentencing practices remain unchanged.

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-compensation 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-compensation prediction. If the difference between these two effects is statistically 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 observed increase in sentence lengths. To do so, we first present conceptual models which offer predictions of how prosecutors and judges may compensate. We then 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 each of the 13 dependent variables, effect sizes are either small and insignificant or move in the opposite direction of the sentence length effects. These null effects carry small standard errors, giving confidence that our insignificant results are not due to noise but truly indicate a lack of response from prosecutors. Even for variables that show small increases, such as charge severity or the number of plea deals, our 95% confidence intervals rule out any increases over 3.3% - far too low to explain the increases we see in sentence length. Furthermore, small changes in these variables are not consistent across various bandwidths or control schemes. The only notable prosecutorial response is a 6.5 percent decline in case duration, which likewise cannot explain the increase in sentence length. Taken together, these results provide no evidence that prosecutorial behavior drives the post-*Blakely* increase in sentences.

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 next 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. From our conceptual models, we propose two potential mechanisms for judicial response – a change in information as signals from prosecutors become noisier, and a

discretionary response to judicial authority tightening. While both mechanisms may occur, we argue for the latter. We support this with a brief correlative analysis that shows that strict judges are more frequently constrained by binding sentencing guideline ceilings, limiting their ability to compensate for reduced enhancement use. Taken together, these findings indicate a compensatory judicial response, whereby judges impose higher sentences for otherwise comparable offenses following *Blakely*.

This paper contributes to the growing literature on prosecutorial and judicial discretion. Despite the central role these actors play in shaping case outcomes, relatively little empirical work examines how prosecutors respond to a change in their incentives. Boylan (2005) analyzes whether pursuing longer sentences or higher conviction rates yields greater long-run career benefits and finds that career advancement correlates more strongly with securing longer sentences than with accumulating convictions. Silveira (2017), using a formal model of prosecutorial choice, similarly shows that prosecutors may substitute between conviction rates and sentence lengths when the relative costs or benefits of each shift. A related literature examines whether prosecutorial discretion mitigates or amplifies punishment. Recent work shows substantial effects of prosecutorial choices on case resolutions and downstream outcomes (Agan, Doleac, and Harvey 2023; Sloan 2020), and several studies suggest that discretion can exacerbate racial or gender disparities (Cooper 2025; Didwania 2022; Rehavi and Starr 2014; Sloan 2024; Tuttle 2019; Yuan and Cooper 2025). Other research, however, finds discretion being used to reduce punishment: Bjerk (2005) shows that in jurisdictions with three-strikes laws, prosecutors more frequently downgrade charges for eligible defendants, and Shaffer (2023) and Shaffer and Harrington (2025) provide evidence of discretion being exercised in ways that lessen racial disparities. Taken together, this literature offers no clear prior prediction about whether prosecutors will respond to higher standards of proof for enhancements by compensating with more severe baseline charges.

Literature on judicial discretion has largely shown strong heterogeneity in case outcomes dependent on the ruling judge. Judge assignment is so impactful it is often used as an instrument due to its strong correlation with punishment received (Arnold, Dobbie, and Yang 2018; Dobbie

et al. 2018; Kling 2006; Leslie and Pope 2017; Mueller-Smith 2015).[1] Reasons for heterogeneity across judges have not been deeply explored, but prior literature has shown variation created by resource constraints (Yang 2016), political ideology (Cohen and Yang 2019; Gordon 2007; Lim 2013), and competency compared to algorithmic assessment (Angelova, Dobbie, and Yang 2025). Bhuller and Sigstad (2024) provides evidence of judicial updating, where judges change behavior in response to a decision of theirs being reversed. A handful of papers provide models of judicial choice under constrained optimization, but these are typically within the context of reelection or reappointment (Cohen et al. 2015; Huber and Gordon 2007). As with the prosecutor literature, previous studies on judicial discretion suggest judges have considerable impacts on case outcomes but these studies provide no clear prediction on how judges will respond to *Blakely.*

This paper contributes to both of these literatures, examining the impacts of *Blakely* on prosecutorial and judicial discretion. Whereas prior studies have often focused on how either prosecutorial or judicial discretion affects particular categories of defendants, we consider a broader set of discretionary responses that apply across all case and defendant types. Our results suggest that prosecutors respond far less to the heightened standard of proof than judges. More broadly, we show that reforms intended to restrict discretion can generate unintended consequences when legal actors shift to alternative mechanisms of punishment. In this case, *Blakely* led to the unexpected outcome of higher average sentences, driven by compensatory judicial behavior.

The remainder of the paper proceeds as follows. Section 2 describes North Carolina's structured sentencing system and the role of sentence enhancements. In Section 3 we provide conceptual models for both prosecutors and judges. These offer predictions for sentence length changes following *Blakely.* Section 4 outlines our empirical strategy and identification approach. In Section 5, we document the direct effects of *Blakely* on sentence enhancements and sentence length, showing that although enhancement use declines sharply, average sentence lengths increase relative to the pre-*Blakely* period. We then examine a wide range of prosecutorial outcomes and find little evidence of systematic prosecutorial adjustment. In Section 6, we show

[1] These are only a small proportion of papers that utilize the "judge fixed effect" instrument design. Frandsen, Lefgren and Leslie (2023) provides a wider, though still incomplete, description of the papers that use this methodology.

that the increase in sentence length is instead driven by judicial behavior. We document significant heterogeneity across judges and discuss how these findings relate to the proposed mechanisms discussed in Section 3.2. Section 7 concludes.

2. Background

2.1. Sentence Enhancement

Sentence enhancement is a legal process whereby the severity or length of punishment is increased above some base level. In North Carolina, these base levels are strictly prescribed for all state-level felonies and misdemeanors by structured sentencing guidelines. That is, based on the crime severity and criminal history level of the defendant, the sentence length for a conviction must be within a legally specified range. However, if a defendant is found guilty of certain aggravating factors, this range is increased. Sentence enhancements therefore expand the upper bound of legally permissible punishment, increasing the discretion available to sentencing judges. Judges differ substantially in how they use this discretion, particularly in their propensity to sentence at or near guideline maxima. As a result, changes that restrict access to aggravated ranges may differentially constrain judges depending on their baseline sentencing preferences.

The decision to seek an aggravated sentence originates with the prosecutor, who must allege the relevant aggravating factors alongside the charged offense. To trigger the higher guideline window, the prosecutor must prove the existence of aggravating factors in conjuncture with the charged offenses. Some of these factors are factually based, such as the defendant being armed with a deadly weapon during the offense. Others rely on the opinions of the court, including the offense being committed with cruelty. There are 20 listed aggravating factors in NC state statute § 15A-1340.16.(d) & (e) for which a defendant may be convicted. This list is provided in the online appendix.

The default sentence length range, called the presumptive range, is weakly lower than this higher, enhanced range. Thus, for a given crime severity-criminal history intersection, an aggravated sentence will always be equal or greater than a presumptive one. The size of the increase between presumptive and aggravated ranges is dependent on the severity of the offense. For example, for individuals with no prior history convicted of the lowest offense level, the presumptive range of minimum sentence lengths is 4-6 months, and the aggravated range is 6-8

months. For a more severe offense, like a class C conviction, the presumptive range is 58-73 months while the aggravated is 73-92.[2] These examples illustrate that the absolute and proportional gap between presumptive and aggravated ranges increases with offense severity. As a result, sentence enhancements are most consequential in more serious cases, where they can shift expected punishment by several years. This feature of the guidelines makes enhancements a particularly important margin of discretion for both prosecutors and judges.

2.2. *Blakely v. Washington*

In Supreme Court case *Blakely v. Washington* 542 U.S. 296, the defendant pled guilty to second-degree kidnapping involving domestic violence, an offense with a presumptive sentence length range between 49 and 53 months under Washington state's sentencing guidelines. However, the trial judge gave a minimum sentence length of 90 months, a sentence nearly 70% higher than the upper limit of the prescribed guideline range. In June of 2004, the Supreme Court held that in states with mandatory sentencing guidelines, the Sixth Amendment right to a jury trial forbids a judge from increasing sentence lengths above the guideline upper limit based on facts that were not decided by the jury or admitted to by the defendant. As a result, *Blakely* effectively raised the evidentiary standard required to impose a sentence enhancement from a preponderance-of-the-evidence standard to proof beyond a reasonable doubt.

Because North Carolina employs a mandatory structured sentencing system, the *Blakely* decision generated a sharp and immediate change in the evidentiary burden required to secure an aggravated sentence. After *Blakely*, prosecutors seeking sentence enhancements were required to either obtain a jury finding on aggravating factors or secure an explicit admission from the defendant, increasing the cost and difficulty of pursuing enhancements. Consequently, for a subset of cases in which aggravating evidence was sufficient to justify enhancement prior to *Blakely* but insufficient to meet the heightened standard afterward, the expected direct effect of the decision is a decline in the use of sentence enhancements.

Because sentence enhancements apply only to a well-defined subset of cases and operate through a single institutional channel, changes in enhancement use following *Blakely* generate sharp, testable predictions for both mechanical and behavioral responses in sentencing outcomes. Because aggravated sentences are always weakly higher than sentences in the presumptive range,

[2] Ranges for each crime severity-criminal history combination are provided in the sentencing guidelines, which are also presented in the online appendix.

this mechanical effect implies that, holding behavior fixed, *Blakely* should lead to a reduction in average sentence lengths. Any deviation from this mechanical prediction therefore reflects behavioral responses by legal actors to the increased burden of proof.

3. Conceptual Models

In this section, we present conceptual models for each of our legal actors, prosecutors and judges, and consider how *Blakely* may impact their charging or sentencing decisions.

3.1. Prosecutors

Landes (1971) presents the seminal model of prosecutor utility in the context of plea bargaining. While subsequent work has extended this framework, we rely on the Landes model here for simplicity and transparency. In this model, the prosecutor chooses charging and bargaining strategies to maximize expected punishment across cases, subject to a binding resource constraint. Expected punishment is increasing in both the probability of conviction and the severity of the sentence.

Formally, the prosecutor's objective can be written as:

Equation 1: $$E(C) = \sum_{i=1}^{n} P_i S_i + \lambda(B - \sum_{i=1}^{n} R_i)$$

where the prosecutor is choosing whether or not to charge some number of n defendants, the probability of defendant i receiving conviction upon trial is P_i, the sentence imposed upon conviction at trial is S_i, R_i denotes prosecutorial resources devoted to case i, B is the total resource budget, and λ is the shadow value of prosecutorial resources.

Landes models plea bargaining by comparing this objective to the defendant's utility function and derives conditions under which cases are resolved by plea rather than trial. We abstract from the defendant's full optimization problem but retain one key implication: because trials are costly, prosecutors are willing to offer plea bargains that reduce expected sentences relative to trial outcomes. Let $\Delta S_i > 0$ denote the sentence concession offered in a plea bargain, which is increasing in the difference between trial and settlement costs. The minimum sentence a prosecutor is willing to offer in a plea agreement can then be written as $S_{0i} = P_i S_i - \Delta S_i$.

The *Blakely* decision increases the standard of proof required to secure enhanced sentences, thereby raising the cost of achieving a given sentence length through the enhancement channel. In the context of the model, this can be represented as an increase in the resource cost required to produce higher sentences, particularly for cases near the enhancement margin. Holding other factors fixed, this leads to a mechanical decline in enhanced sentences and, for affected cases, lower values of both $S_i$and S_{0i}.

In principle, prosecutors could attempt to offset this decline by adjusting other margins of behavior - such as pursuing harsher charges, filing additional charges, or bargaining more aggressively - to raise sentence lengths from $S_i^0$to S_i^1. However, such adjustments are also costly. A prosecutor will pursue these strategies only if the resulting increase in expected punishment exceeds the additional resource cost, that is, if

$$P_i S_i^1 - \lambda R_i^1 > P_i S_i^0 - \lambda R_i^0.$$

If the sentencing gains from these alternative strategies are small or the associated resource costs are high, the model predicts little or no compensatory response from prosecutors following *Blakely*.

In addition to affecting expected punishment directly, prosecutorial decisions, particularly enhancement filings, also shape the information available to judges at sentencing, thereby influencing judicial inference about case severity. Thus, the impact of *Blakely* on prosecutor choice will have downstream impacts on judicial decisions as well. This is discussed in detail below in Section 3.2.1.

This framework motivates our empirical analysis of prosecutorial behavior. While the model allows for compensation along multiple margins, it also clarifies why prosecutorial responses are not inevitable. If *Blakely* raises the marginal cost of enhancement while alternative sentencing margins yield limited returns net of resource costs, the model predicts little prosecutorial adjustment. In this case, the primary downstream response to *Blakely* will operate through judicial decision-making rather than prosecutorial substitution.

3.2. Judges

In the context of judicial decision making, enhancement can be considered in two distinct ways: an information channel and a discretionary tool to increase sentences. We discuss these two mechanisms separately, but it is important to note that they may occur simultaneously.

3.2.1. Information and pooling under reduced enhancement signals

Enhancement filing from the prosecutor can be considered as a channel of information for crime severity. When a prosecutor files for an enhanced sentence, they must provide evidence of statutory aggravating factors. This additional evidence tells the judge more about the case and may impact their decision to apply a lenient or harsh sentence. Because *Blakely* raised the cost of enhancement, prosecutors have less incentive to gather and present evidence of aggravating factors. Thus, post-*Blakely,* judges now have less information about which cases carry these aggravating factors and thus less information about which types of cases may merit a longer sentence length.

We model a judge who has a preferred sentence length for a given defendant i based on their severity-type. The defendant may be one of two unobserved types: $\theta = \{L, H\}$, where L denotes low severity and H high severity. Judges do not directly observe a defendant's severity type θ, but instead infer it from observable case characteristics and prosecutorial actions. Judges have a prior belief about the probability of type: $\Pr(\theta = H) = \pi$. Based on this belief, judges choose a sentence length $s \in [\,\underline{s}, \bar{s}\,]$ to maximize the following loss function:

Equation 2: $$U_j(s, \theta) = -\big(s - s^*(\theta)\big)^2.$$

Note that $s^*(L) < s^*(H)$ for all judges. However, judges may differ in preferences for overall sentence length levels, such that for judges m and n even if θ was known to be H, $s_m^*(H)$ does not necessarily equal $s_n^*(H)$.

We model the prosecutor's decision to file for enhancement as a noisy but informative signal of defendant severity. Let $e = \{0,1\}$ be the binary enhancement filing choice of the prosecutor. This noisy signal introduces two more probability measures: the probability of enhancement occurring given the defendant type:

$$\Pr(e = 1|\theta = H) = q_H,$$
$$\Pr(e = 1|\theta = L) = q_L$$
$$\text{where } \; q_H > q_L.$$

Utilizing the enhancement signal, the judge arrives at the following posterior probabilities of defendant type:

$$\Pr(H|e = 1) = \frac{q_H \pi}{q_H \pi + q_L(1 - \pi)},$$

$$\Pr(H|e=0) = \frac{(1-q_H)\pi}{(1-q_H)\pi + (1-q_L)(1-\pi)}$$

Based on the observed signal e, the judge chooses a sentence equal to the posterior optimal sentence length:

$$s(e) = E[s^*(\theta)|e].$$

When q_H and q_L are sufficiently large, this implies enhancement filings are sufficiently informative. When this occurs, posterior beliefs will diverge and judges will select sentences at or near $s^*(H)$ or $s^*(L)$, depending on the signal type. When enhancement filings become weakly informative, posterior beliefs converge toward the prior π, and judges increasingly impose pooled sentences near $s^{pool} = \pi s^*(H) + (1-\pi)s^*(L)$.

Blakely does exactly this - it makes filing for enhancement more costly to the prosecutor, decreasing the probability of enhancement filing such that $q_L^{post} < q_L^{pre}$ and $q_H^{post} < q_H^{pre}$. As these probabilities decrease, the signal weakens and the posterior probabilities of defendant type collapse towards π. This will induce judges to use the pooling equilibrium more frequently post-*Blakely.* This increase in pooling equilibrium leads to a compressing of sentence lengths toward the center of the guideline window. Importantly, this mechanism does not predict an overall increase in average sentence lengths.

3.2.2. Loss of discretion and compensatory sentencing

In addition to the informational channel, Blakely also removes a direct margin of judicial discretion by shifting aggravating fact-finding from judges to juries. Under pre-*Blakely* sentencing rules, judges could directly influence sentence severity by finding aggravating factors and imposing enhanced sentences. By eliminating this margin, Blakely reduces judges' ability to implement their preferred sentencing outcomes through enhancements.

While we do not formally model this institutional channel, its qualitative implication is straightforward. Judges who prefer higher sentences may respond to the loss of enhancement authority by increasing punishment along remaining discretionary margins—most notably, by selecting higher sentences within the applicable guideline range. In this sense, judges compensate for the reduction in enhancement-based discretion by reallocating sentencing discretion within the presumptive range.

This mechanism alone predicts upward pressure on sentence lengths across judges. However, because sentencing guideline ceilings bind more frequently for stricter judges, the scope for compensation differs across the judicial distribution. As a result, lenient judges, who retain greater room to adjust sentences upward, drive increases in average sentence length, while stricter judges are constrained. When combined with the information channel described above, these forces generate both increased average sentence lengths and compression of sentencing outcomes across judges.

4. 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.

4.1. 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, $\widehat{\Delta}_{blakely}$.

4.2. 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 3: $\Pr(enhancement_i = 1 | \, crime\ type_i, history\ points_i) =$ $F(crime\ type_i + history\ points_i)$

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 4, which gives our three main models for our econometric analysis:

Equation 4:

(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 4 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 Figure 1 in the main text and Figure A.1 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*). 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.

4.3. 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. Because North Carolina employs a structured sentencing system similar to those used in many other states, our findings are likely to generalize to other jurisdictions that employ structured or guideline-based sentencing regimes. The data contains detailed defendant characteristics and rich case variables throughout the legal process, including at the charging, conviction, and disposition stages. Defendant characteristics include race, sex, criminal history points, age, and the defendant's type of counsel. We observe the specific statutes a defendant is charged with, the statutes at conviction, whether the case was resolved by plea bargain, and the minimum sentence length at disposition. We leverage this detailed charging information to identify whether *Blakely* sentence length effects are driven by changes in prosecutor charging or bargaining behavior. We also observe the sentencing judge, allowing us to distinguish prosecutorial from judicial responses and to examine heterogeneity across judge types.

It is important to note that we do not explicitly measure whether a case receives enhancement or what factors may be driving enhancement petitions from the prosecutor. However, we can measure whether a case receives enhancement by comparing the minimum sentence length against the prescribed guideline sentence windows. 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 noted above in Section 4.2, we reduce the full sample to a set of cases that are most likely to be impacted by the *Blakely* decision. This is done using an enhancement propensity score. The primary difference between the reduced and full sample is offense type. 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%). Most of these cases are excluded in the reduced sample, reflecting the institutional reality that sentence enhancements are rarely applied in lower-severity drug and property offenses and are therefore unlikely to be directly affected by *Blakely*. 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. By restricting attention to these cases with meaningful enhancement risk, we focus on the population for which the *Blakely* decision plausibly altered incentives.

Table 1 provides means and standard deviations for the key variables in our data. We provide statistics for both the full sample and the main, propensity score reduced sample. Note that defendant characteristics such as race, sex, age, and defense type are similar across the two samples. Unsurprisingly, cases in the reduced sample have significantly higher sentence lengths, charge severities, criminal histories, incarceration rate, and enhancement rate. The similarity in observable defendant characteristics across samples suggests that the restriction primarily alters offense composition rather than defendant selection. For each sample, over 95% of cases are resolved through a plea deal. Cases in the propensity score-reduced sample have especially long sentence lengths, with an average of 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.

5. The Sentencing Puzzle and Prosecutorial Non-Response

This section documents the effect of *Blakely* on sentence enhancements and sentence length and examines whether prosecutorial behavior can account for the observed increase in sentences. The following section then describes and presents the judicial response analysis.

5.1. Mechanical Effects and the Sentencing Puzzle

The direct effects of *Blakely* are presented in Table 2 across three separate panels, one for each component of our main models outlined in Equation 4. 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.1 and Figure A.3.

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 gives the correlative effects 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 estimates of Panel A and B gives the benchmark value $\Delta_{blakely}$. 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. Online appendix Figure A.2 gives the event study corresponding figure to Figure 2, once again showing evidence of no decrease following the *Blakely* decision. We also show that both enhancement and sentence length discontinuity estimates are robust to various kernel functions and a quadratic polynomial order in Table B.4 and to various types of standard errors in Table B.5. This increase in sentence

length is surprising and motivates our analysis into other legal outcomes that could be driven by prosecutor compensating activity.

5.2. Prosecutorial Response: Evidence of No Compensation

In this section, we consider whether prosecutor decisions can explain the sentence length effects. When *Blakely* increased costs for pursuing enhancement, prosecutors may have responded by changing other charging or bargaining decisions. Thus, the expected decrease in sentence length following *Blakely* may not occur because prosecutors compensate for less enhancement with higher charges. We consider this possibility by measuring *Blakely* effects across three categories of variables for which prosecutors may have direct or indirect impact: charging, bargaining, and sentencing.

We perform similar analyses as above for a host of dependent variables, estimating effects across different bandwidths for both the 0.05 propensity score restricted sample and the full sample. Specifically, we use 9-, 12-, and 15-month bandwidths across each sample for a total of 6 specifications. These are reported in Tables 3-5. We also show findings are consistent under the event study framework. These results are displayed in the online appendix in Figures A.5 and A.6.

5.2.1. Charging decisions

If prosecutors respond to the higher burden of proof by substituting toward harsher charges, we would expect to observe an increase in charge severity following Blakely. Across 5 different charge-related measures, we find no evidence of such a change. These measures are the overall charge severity,[5] a binary measure of whether the charged offense was an index crime,[6] the number of charges before and after consolidation,[7] and a binary measure for whether the case was dismissed. Table 3 contains the discontinuity results for each of these charging dependent variables. Note that the dismissal analysis is only considered for the full sample as the dismissed cases lack the data needed to create the propensity score. Across each specification for each dependent variable, the regression results report small and statistically insignificant effect sizes with small standard errors, indicating no changes in charging behavior following *Blakely*. Even

[5] Charge severity is measured simply on a categorical scale from 1-11, with higher numbers representing more severe crimes.
[6] Index crimes include assault, burglary, robbery, sexual assault, larceny, and homicide.
[7] We created two measures for the number of charges, with or without including the charges that are consolidated. See Appendix D for more details on consolidated charges.

when considering the upper limits of 95% confidence intervals for these estimates, effect sizes are close to zero. The largest of these effects would be charge severity, which has a 95% confidence interval upper bound of around 0.2. This is still a small and insignificant impact given mean charge severity is around 5.9. Furthermore, many of the estimated coefficients have signs that correlate with lower sentence length, indicating they would more likely decrease sentence lengths following *Blakely.* The corresponding event study graphs in Panel A of Figures A4 and A5 likewise show no discernable effects. Thus, increased sentence lengths are not occurring due to prosecutors pursuing more counts or more severe charges.

5.2.2. Plea bargaining outcomes

Even though prosecutors are not changing initial charging decisions, they may still adjust to *Blakely* through more stringent bargaining. If prosecutors are less likely to lower charges in a plea deal compared to pre-*Blakely,* it may explain the null sentence length effects. Thus, we perform the event study and discontinuity analysis again to consider whether prosecutors compensate for fewer enhancements with more stringent plea outcomes. Specifically, we check for changes in the probability of a case having a guilty plea, whether a case's guilty plea leads to incarceration, whether a defendant decreases initial felony charges to a lower felony charge through a plea deal, or whether a defendant decreases initial felony charges to a misdemeanor charge through a plea deal. Each of these are measured using a binary variable. We also include a measure of case duration in this analysis. While case duration is likely driven by many factors, changes in bargaining stringency is one of the main ways a prosecutor can impact case durations.

Table 4 presents the discontinuity results. For the four binary measures of bargaining, we find no discernable changes at the timing of *Blakely,* with small effect sizes tightly centered at zero. The only significant effect is a decrease in plea bargains when using the 12-month bandwidth. However, this effect disappears entirely in other bandwidth specifications. Even when considering this effect, the 95% CI rules out impact sizes larger than -0.032, a relatively small effect compared to the pre-*Blakely* mean of 0.95. The event study analyses largely corroborate this finding. However, case duration appears to decrease following the *Blakely* decision. In the propensity score reduced sample, we estimate a decrease of 23.57 to 26.86 days in case duration, a 6.5 to 7.5 percent decrease from the pre-*Blakely* mean. However, case duration is strongly correlated with higher sentence lengths, even when controlling for many case, area, and timing factors. Thus, a decrease in case duration would predict a decrease in

sentence length following *Blakely*; instead, we observe the opposite. Thus, plea decision changes also do not explain the increase we see in sentence length post-*Blakely.*

It's worth noting that *Blakely*'s effect on case duration is an important finding in and of itself. By decreasing prosecutors' incentive to pursue enhancement, the average time taken to case resolution falls. While we cannot directly measure at which point in the justice process time is reduced, the most intuitive mechanisms would be in evidence gathering or plea negotiations. If prosecutors wish pursue enhancement, they may work with law enforcement to gather additional evidence of aggravating factors that occurred in connection with the offense. With lower incentive to pursue enhancement, prosecutors may move to negotiations faster. Likewise, bargaining based on enhancement during plea negotiations are likely to decrease with lower enhancement rate, leading to shorter case durations. In either case, *Blakely* effectively reduces the time needed for case resolution.

5.2.3. Other sentencing and case outcomes

We next consider whether other sentencing practices can explain the sentence length effects of *Blakely.* Changes in other punishment or leniency measures of sentencing may drive our sentence length results and help indicate whether prosecutors compensate for *Blakely* by increasing other types of punishment. We specifically check whether incarceration rates or probation lengths are affected by *Blakely.* We also consider whether *Blakely* reduced mitigation which would also lead to a sentence length increase. Table 5 presents the results. Once again, the effect sizes are small and insignificant, apart from two specifications with probation as the dependent variable. These two exceptions are both for the full sample of cases and their effects do not hold in the event study analysis, which shows tight null effects for probation as seen in Figure A5. For the propensity score reduced sample, the 95% CI rules out incarceration effects larger than 0.0383. Our preferred specification also shows a noisy decrease in mitigation. Like the effects on plea bargain, this effect disappears entirely when looking at other bandwidth specifications. Across each specification, standard errors once again remain relatively small, indicating null results are most likely truly zero rather than simply noisy. Taken together, the evidence in Section 5 rules out observable prosecutorial channels as an explanation for the post-Blakely increase in sentence lengths, leaving judicial behavior as the remaining plausible source of compensation.

6. Judicial Response and Mechanism

The main results present a surprising finding – the *Blakely v. Washington* decision drastically reduced enhancement rates, but sentence lengths increased rather than decreasing. This is evidence of some sort of compensating behavior by a legal actor. The preceding section provides numerous checks by which a prosecutor may increase sentence length through charges, plea negotiations, or other sentencing practices; none of these can explain the sentence length increase.

Having ruled out prosecutorial responses, we now show that the post-*Blakely* increase in sentence length is driven by judicial behavior. Judges may impact sentence lengths in several different ways. Most directly, if a case goes to trial, upon conviction the judge determines the sentence length from within the guideline window. The judge has full discretion in choosing the sentence length within this 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. That is, prosecutors and defense attorneys bargain in the shadow of the judge, such that they adjust their expectations for a plea outcome based on how harsh or lenient an assigned judge is. 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.

How can we detect whether effects are driven by judicial behavioral changes? First, because judges direct court proceedings, judicial effects may be stronger for cases resolved by jury trial. Thus, we begin our judicial analysis by simply repeating the main sentence length discontinuity analysis for jury-trial cases only. Judicial effects may also be determined by considering heterogeneity in our results across judge type. If our main effects are localized to specific judges, this is evidence that judicial discretion is driving compensation effects, either directly or indirectly. We specifically consider how the discontinuity effects may differ by judge leniency. While much of the law and economics literature uses judge leniency as an instrument, here we simply construct the leniency measure as a tool to consider whether judges drive results

rather than prosecutors. Below we describe how we construct the judge leniency measure, then present the heterogeneous results.

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 interpretation of the outcome changes depending on which leniency measure is used. Using the enhancement measure gives the heterogeneous impact of *Blakely* on enhancement and sentence length by judges' different propensities to adjudicate cases with enhancement. But the sentence length measure considers heterogeneity by judges' propensity to give a higher sentence length at conviction. These two measures may not necessarily give the same impact – judges who are harsh in sentencing may not be any more likely to allow for enhancements. To check the similarity between these measures, we plot them against each other. The resulting graph, displayed as Figure A.6 in the online appendix, suggests that most judges who are strict in enhancement are also strict in sentencing. Thus, the results are likely to be similar regardless of leniency measure. We present results for analyses using the sentence length measure, and show results are consistent when using the enhancement measure in the online appendix. We favor the sentence length measure because it is more likely to pick up indirect effects. This is because enhancement is rare, meaning prosecution and defense have less observations to create a

prediction of enhancement behavior for a judge. On the other hand, sentence lengths are observed in most cases, meaning sentence leniency is likely a better signal for judge leniency when legal actors engage in 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 5:
$$Y_{itg} = \alpha + \beta_1 Blakely_t + \beta_2 leniency_50_{it} + \beta_3 leniency_75_{it} + \beta_4 leniency_100_{it} + \beta_5(Blakely \cdot leniency_50_{it}) + \beta_6(Blakely \cdot leniency_75_{it}) + \beta_7(Blakely \cdot leniency_100_{it}) + \eta_g + \rho_t + u_{it}$$

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 4 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.6 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 indicate a probable link between the compensating behavior found in Section 5 and judicial behavior and motivates our heterogeneous analysis.

Table 6 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 3. Note that *Blakely* completely eliminates the pre-period discrepancy in enhancement rate.

Columns 3 and 4 of Table 6 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 4 visualizes these effects for the most lenient and most strict judicial groups.

6.3. Interpretation and Mechanisms

These results uncover three important findings: (1) the sentence length increase shown in Section 5.1 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 take the scaling of effect size with judicial leniency as evidence that compensation is driven by judge behavior. While we cannot fully rule out prosecutor behavior that is unobserved, such as more intensively seeking evidence in each case, it is not clear why

such behavior would change across judicial leniency. Furthermore, we repeat the prosecutor analysis in 4.2 again split by judicial leniency quartile and once again find consistent null effects.

Our second main result is somewhat surprising. One might expect that cases assigned to the strictest judges would exhibit the largest compensating increase in sentence length, since these judges should be most directly affected by *Blakely*. Instead, we find significant *decreases* in sentence length, which align with the observed decline in enhancement rates. However, recall that the two proposed mechanisms in Section 3.2 both predict higher increases in sentences among the lenient judges. For the discretionary pushback mechanism described in Section 3.2.2, less response from strict judges only occurs if strict judges face tighter constraints on their ability to increase punishment following *Blakely*.

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

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. These are reported in online appendix Table B.8. These facts suggest that binding guideline ceilings limit the ability of strict judges to offset reduced enhancement use, while leaving greater scope for lenient judges to increase sentences within the presumptive 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 sentence 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 respond by increasing sentences within the remaining discretionary range. As a result, sentencing outcomes converge toward the middle of the distribution. As illustrated in Figure 4, both lenient and strict judges move toward a common median sentence length following *Blakely*, generating a significant compression of sentence lengths across judges.

6.3.1. Information versus discretion

While both the information and discretion channels predict compression in sentencing outcomes, they differ in their implications for average sentence lengths. In the information channel, reduced enhancement filing weakens the signal judges receive about defendant severity, leading to more frequent pooling toward an interior sentence. Absent strong assumptions about judicial priors or asymmetric loss functions, this mechanism primarily predicts reduced dispersion rather than an increase in average sentence length. [8]

By contrast, the discretion channel, through which Blakely removes judges' ability to impose enhanced sentences while leaving within-guideline discretion intact, naturally generates upward pressure on sentences. Judges may compensate for lost enhancement authority by selecting higher sentences within guideline ranges, with this response concentrated among more lenient judges for whom guideline ceilings are less binding. Consistent with this distinction, we find direct evidence that guideline constraints bind more tightly for harsher judges and that increases in sentence length are driven by upward shifts within guideline ranges rather than by changes in case composition. Taken together, these patterns suggest that while the information channel may contribute to compression, the observed increase in average sentence lengths is most consistent with a discretionary pushback response by judges.

6.3.2. Other explanations and details

How do judges generate higher sentence lengths following *Blakely*? The primary channel operates through judges' choice of sentence length within the applicable guideline range. As shown in Section 5.2, prosecutorial charging behavior remains unchanged after *Blakely*, implying that judges are sentencing observationally similar cases before and after the decision. Conditional on a given conviction and criminal history, however, judges retain discretion to

[8] In principle, the information channel could generate increases in average sentence lengths if judicial priors or loss functions differ systematically across judges. We do not observe such objects directly, and we therefore refrain from making claims about their role. Our interpretation instead relies on institutional features of guideline constraints and directly observable sentencing behavior.

select a specific sentence from within the prescribed guideline window. Following *Blakely*, judges, particularly those with more discretionary room under the guidelines, respond by selecting higher sentences within these ranges, even as access to enhanced ranges becomes more limited.

We also consider alternative judicial channels, such as changes in pretrial detention or trial-related behavior. Prior work shows that pretrial detention can affect sentencing outcomes (Didwania 2020), raising the possibility that judges might indirectly increase sentences through stricter bail decisions. While we cannot directly observe bail in our data, several patterns suggest this channel is unlikely to drive our results. In particular, sentence length increases are present in both trial and plea cases, indicating that changes in trial procedure or trial selection are not the primary mechanism. Taken together, the evidence points to within-guideline sentencing decisions as the dominant source of judicial compensation following *Blakely*.

Finally, one might wonder whether results vary across other judicial characteristics outside of leniency. The data provided to us by the North Carolina AOC does not contain any demographic information about these judges but does give first and last name and the year they started as a judge. Using this information, we found pictures for 53 of the 146 judges in our dataset. With these pictures, we inferred judge race (White versus non-White) and sex. While this is a substantial reduction in our sample, the observation count is slightly higher at around 47% of the original as many of the judges we could not find information for are those with relatively low case counts. We find no evidence of heterogeneity across judge race, sex, or tenure for enhancement results and only marginally significant results for tenure, which is positively correlated with leniency. We likewise find that results for leniency hold when including these demographic factors as controls, though significance levels shrink slightly. We also check for political ideology of the judge, but this results in an even smaller sample – only 22 judges. We likewise find no evidence of heterogeneity here. These results are presented in appendix Table B.11. Taken together, the results in Section 6 illustrate how constraining discretion along one institutional margin can reallocate discretion along another, with distributional and efficiency consequences for sentencing outcomes.

7. Conclusion

This paper studies how legal actors respond to an increase in the standard of proof using the Supreme Court's decision in *Blakely v. Washington*. 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 predictions of sentencing guidelines 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 compensation. 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. The compression of sentence lengths along with the overall increase in sentences supports the idea that judges push back against loss of enhancement authority with harsher sentencing.

These findings illustrate how constraining discretion along one institutional margin can reallocate discretion along another. In settings with structured sentencing, procedural reforms that limit access to enhanced penalties may unintentionally intensify judicial discretion within guideline ranges, altering both the level and distribution of punishment. More broadly, our results caution against evaluating legal reforms solely through their direct mechanical effects, without accounting for adaptive behavior by legal actors.

Our analysis focuses on sentencing outcomes and does not directly observe all potential judicial channels, such as pretrial detention decisions. Future work could explore how procedural constraints affect earlier stages of the criminal process or interact with defendant behavior. More generally, examining how different institutional designs shape compensatory responses remains an important direction for understanding the full consequences of legal reform.

References

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

Angelova, Victoria, Will Dobbie, and Crystal Yang. 2025. "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).

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).

Cohen, Alma, Alon Klement, and Zvika Neeman. 2014. "Judicial Decision Making: A Dynamic Reputation Approach." *Journal of Legal Studies.* 44: 133-159.

Cohen, Alma, and Crystal S. Yang. 2019. "Judicial Politics and Sentencing Decisions." *American Economic Journal: Economic Policy* 11 (1).

Cooper, Spencer. 2025. "How Much Do Mandatory Minimums Matter?" Working paper.

Didwania, Stephanie. 2020. "The Immediate Consequences of Federal Pretrial Detention." *American Law & Economics Review* 22 (1): 24-74.

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

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.

Gordon, Sanford C. 2007. "The Effect of Electoral Competitiveness on Incumbent Behavior." *Quarterly Journal of Political Science* 2 (2).

Huber, Gregory A., and Sanford C. Gordon. 2007. Directing Retribution: On the Political Control of Lower Court Judges. *Journal of Law, Economics and Organization.* 23: 386-420.

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

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.

Lim, Claire S.H. 2013. "Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges." *American Economic Review*.

Marit Rehavi, M., 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.

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.

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

Sloan, CarlyWill. 2020. "How Much Does Your Prosecutor Matter? An Estimate of Prosecutorial Discretion." Working paper.

Sloan, CarlyWill. 2024. “Do Prosecutor and Defendant Race Pairings Matter? Evidence from Random Assignment.” Working paper.

Tuttle, Cody. 2025. “Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums.” Working paper.

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).

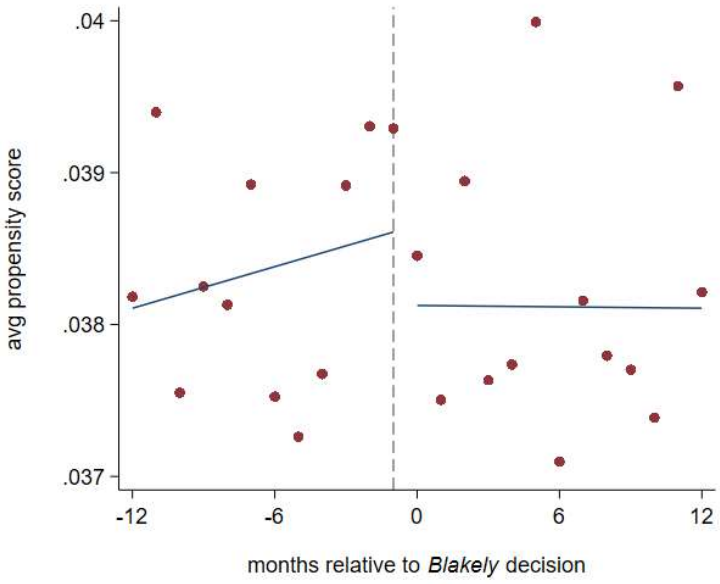
Yang, Crystal S. 2016. “Resource Constraints and the Criminal Justice System: Evidence from Judicial Vacancies.” *American Economic Journal: Economic Policy* 8 (4).

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

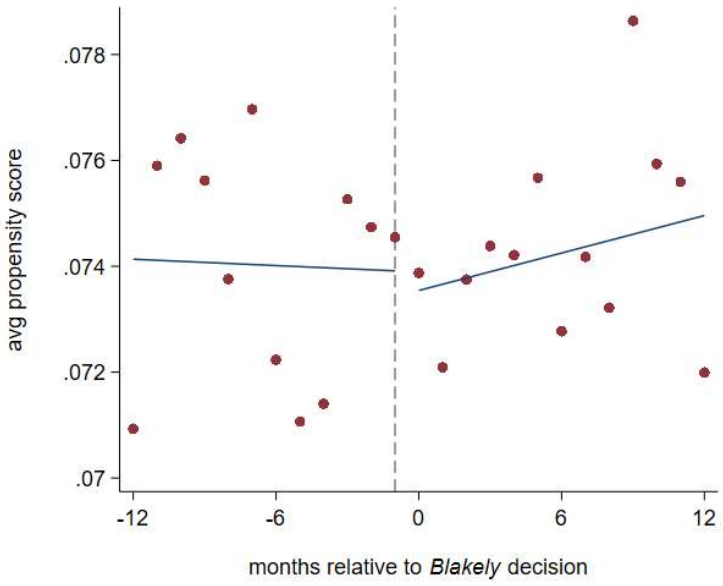
Figures

Figure 1. Smoothness of propensity score

Panel A: *Full sample*

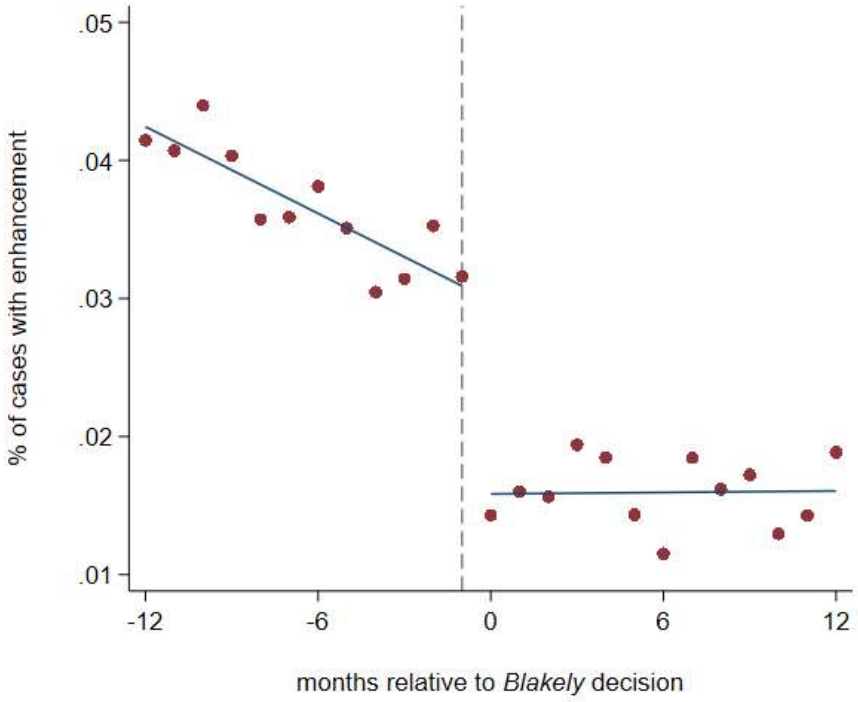


Panel B: *Phat ≥ 0.5 sample*



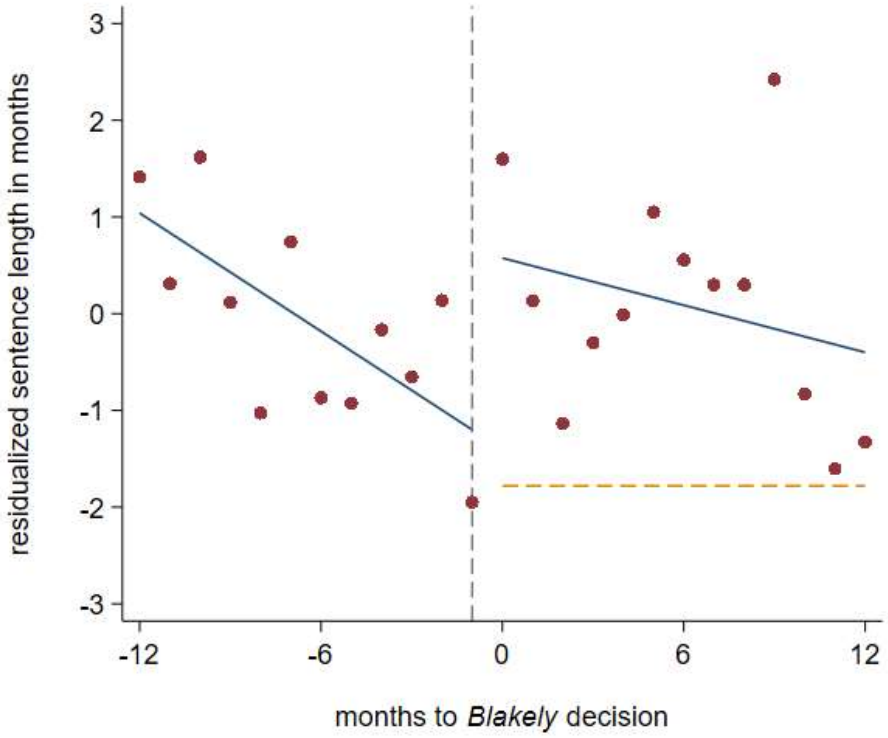
Notes: These graphs illustrate how the aggravated propensity score changes across the timing of *Blakely.* Panel A gives mean propensity scores across time for the full sample while Panel B illustrates the propensity scores for the sample of cases with scores above 0.05. In both cases, *Blakely* has no discernible change in bunching propensities.

Figure 2. Effect of *Blakely* on enhancement



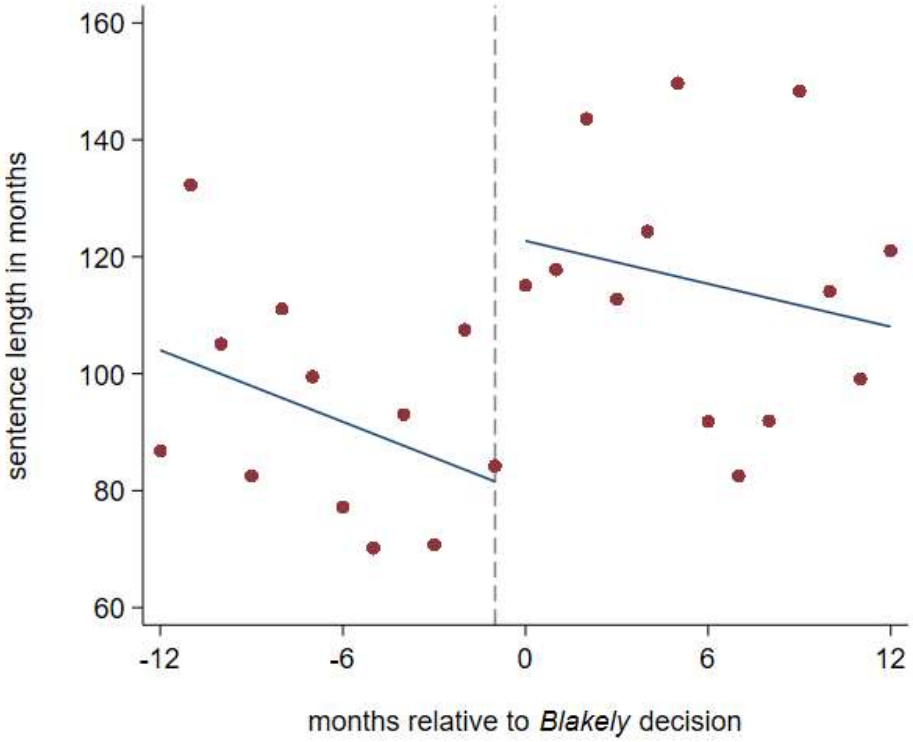
Notes: This graph shows the discontinuity for the percentage of cases with enhancement under a 12-month bandwidth with linear fits for the full sample of cases. The dashed vertical line indicates the timing of the *Blakely* decision.

Figure 3. Effect of *Blakely* on sentence length



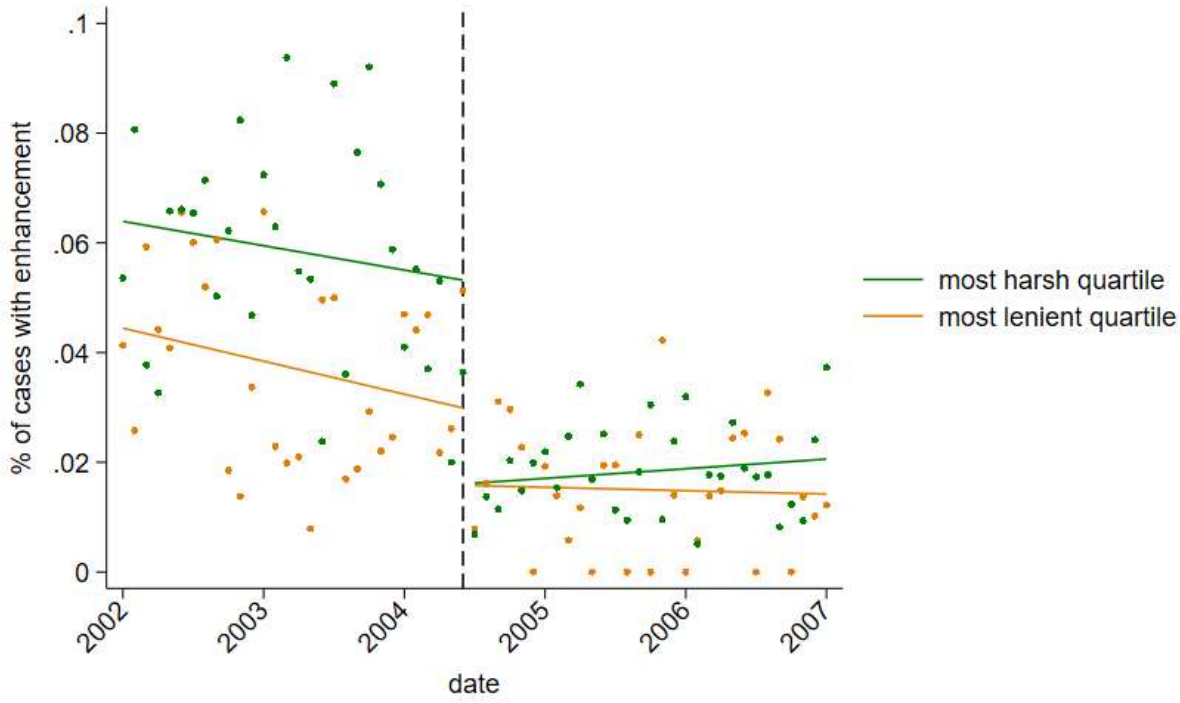
Notes: This graph gives the discontinuity for residualized sentence length using a 12-month bandwidth and linear fit. Sentence length is residualized on grid fixed effects. The vertical dashed line indicates the timing of the *Blakely* decision. The orange, horizontal dashed line gives the predicted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

Figure 4. Effect of *Blakely* on sentence length – jury trial cases



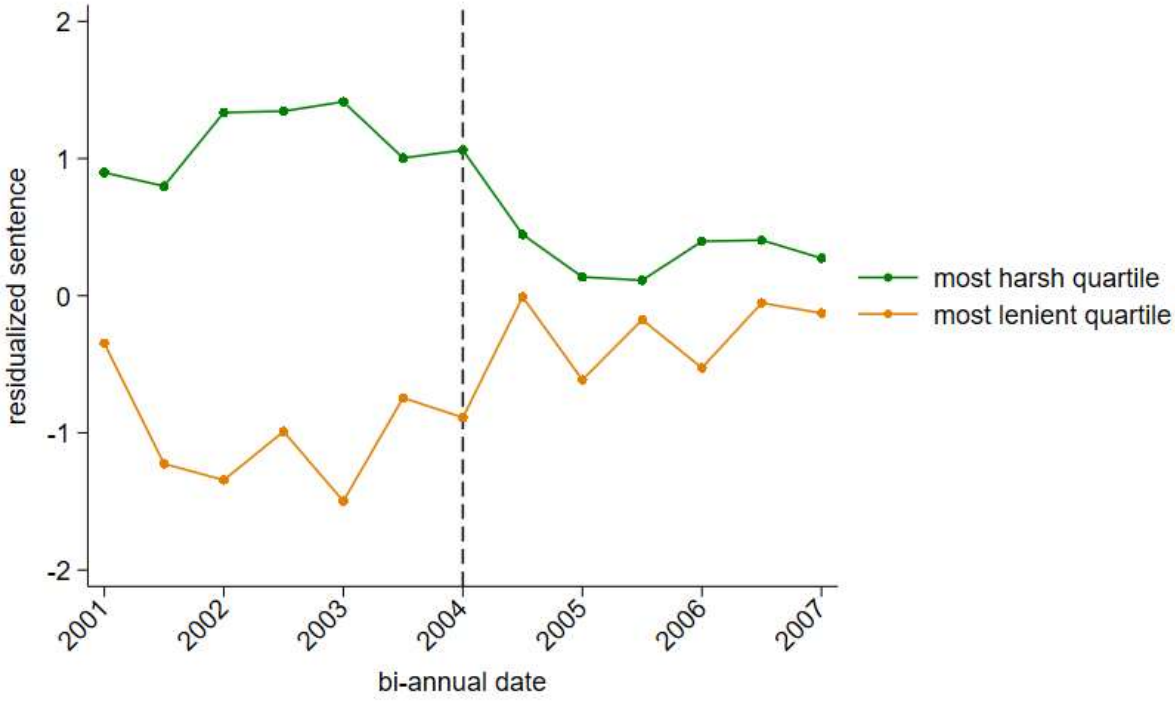
Notes: This graph gives the discontinuity for sentence lengths for cases resolved by jury trial and with a propensity scores greater than 0.05. The fit is linear and bandwidth is 12 months.

Figure 5. 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 6. 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. Summary statistics

	full sample		phat ≥0.05 sample	
	mean	sd	mean	sd
Panel A: Defendant Characteristics				
age at charge	30.23	10.19	30.38	11.09
male	0.857	0.350	0.922	0.268
minority race	0.590	0.492	0.611	0.488
private attorney	0.211	0.408	0.168	0.374
public defender	0.259	0.438	0.262	0.440
criminal history points	4.986	5.490	6.025	7.793
Panel B: Case Outcomes				
enhancement	0.027	0.162	0.051	0.220
min sentence length in months	18.50	31.37	38.85	51.78
charge severity	3.609	1.996	5.874	2.282
index crime	0.410	0.492	0.832	0.374
probation length	31.33	10.47	35.13	11.29
incarcerated	0.374	0.484	0.592	0.492
mitigated	0.096	0.295	0.165	0.371
plea bargain	0.979	0.144	0.955	0.208
plead to lower felony	0.144	0.352	0.271	0.445
plead to misdemeanor	0.036	0.186	0.055	0.228
days of case duration	335.6	297.6	359.5	335.5
N	118379		27536	

Notes: Figures are presented for the main analysis sample, which consists of cases with a propensity score over 0.05, and the full sample of cases. Each of these samples contain cases 3 years before or after the timing of *Blakely.* Variables enhancement, male, minority race, private attorney, public defender, index crime, incarcerated, mitigated, plea bargain, plead to lower felony, and plead to misdemeanor are binary. Charge severity is measured on a scale from 1-11, with higher numbers indicating more severe charges.

Table 2. 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 4, with Panel A giving estimates of β_1 in equation 4.1, Panel B giving estimates of δ_1 in Equation 4.2, and Panel C giving estimates of γ_1 in Equation 4.3. $\widehat{\Delta}_{blakely}$ 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 3. 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 4. 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 5. 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 6. 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Online Appendix for

The Unintended Consequences of Raising the Standard of Proof in Criminal Sentencing

Spencer Cooper* & Andy Ye Yuan†

* Assistant Professor of Economics, University of Connecticut
† Assistant Professor of Law, University of Florida Levin College of Law.

Table of Contents

Appendix A: Additional Figures *3*

Figure A.1 Smoothness of other characteristics 3

Figure A.2 Propensity score distribution 4

Figure A.3 Enhancement event study 5

Figure A.4 Sentence length event study 6

Figure A.5 Changes in prosecutor decisions 7

Figure A.6 Changes in prosecutor decisions – full sample 9

Figure A.7 Leniency measure correlation 11

Figure A.8. Jury trial cases sentence discontinuity – full sample and residuals 12

Figure A.9. Sample restriction at the *Blakely* cutoff 13

Appendix B: Additional Tables *15*

Table B1. The effect of *Blakely* on enhancement rate – event study 15

Table B.2 Effect of enhancement on sentence length – event study 16

Table B.3 Effect of *Blakely* on Sentence Length – event study 17

Table B.4 *Blakely* effects – kernel and quadratic robustness 18

Table B.6 *Blakely* effects on sentence length – jury trial cases only 20

Table B.7 Heterogeneous effects of *Blakely* by judge leniency 21

Table B.8 Binding Sentencing Guidelines Mechanism Analysis 22

Table B.9 Enhancement Effects by Judge Leniency – RD Design 23

Table B.10 Sentence Effects by Judge Leniency – RD Design 24

Table B.11 Heterogeneous effects of *Blakely* by other judge characteristics 25

Table B.12 Sample restrictions and the *Blakely* discontinuity 26

Appendix C: More Institutional Details *27*

C.1: Sentencing Guidelines 27

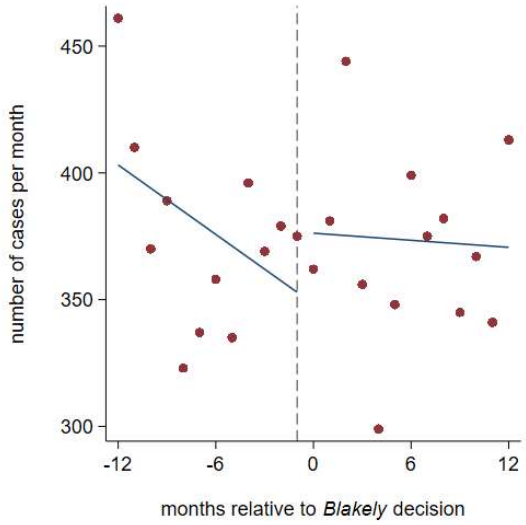
C.2 Statutory Aggravating Factors 28

Appendix D: Data cleaning *30*

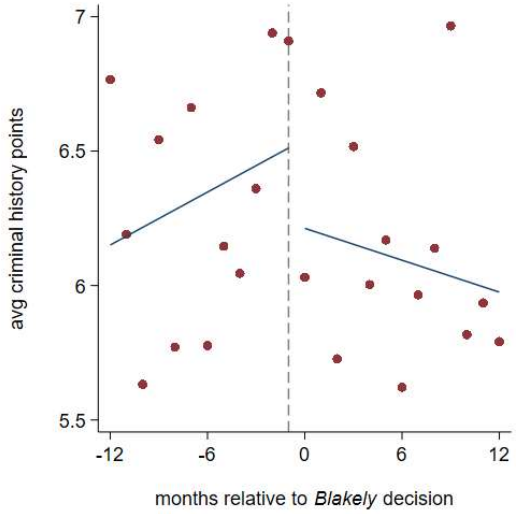
Appendix A: Additional Figures

Figure A.1 Smoothness of Other Characteristics

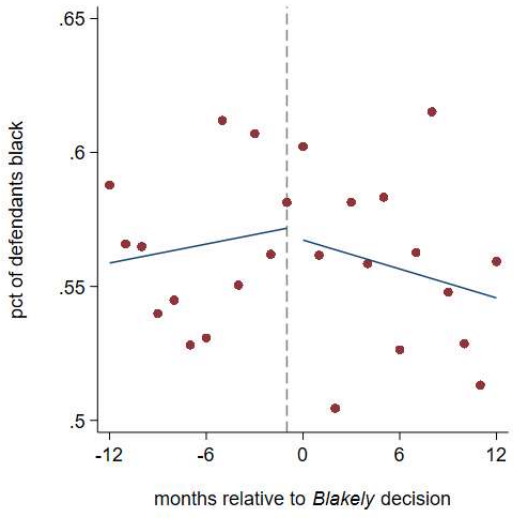
Panel (A): Case Density



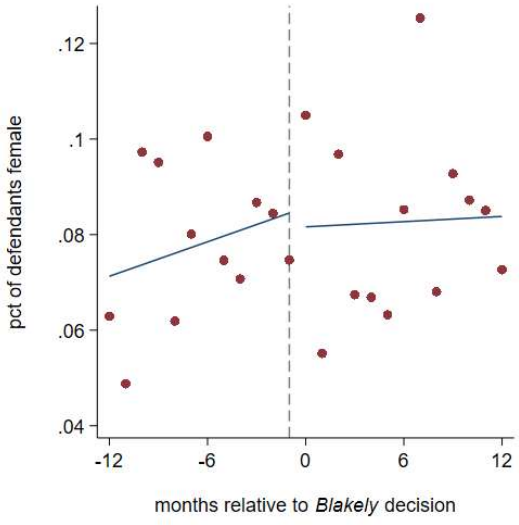
Panel (B): Criminal history points



Panel (C): Black

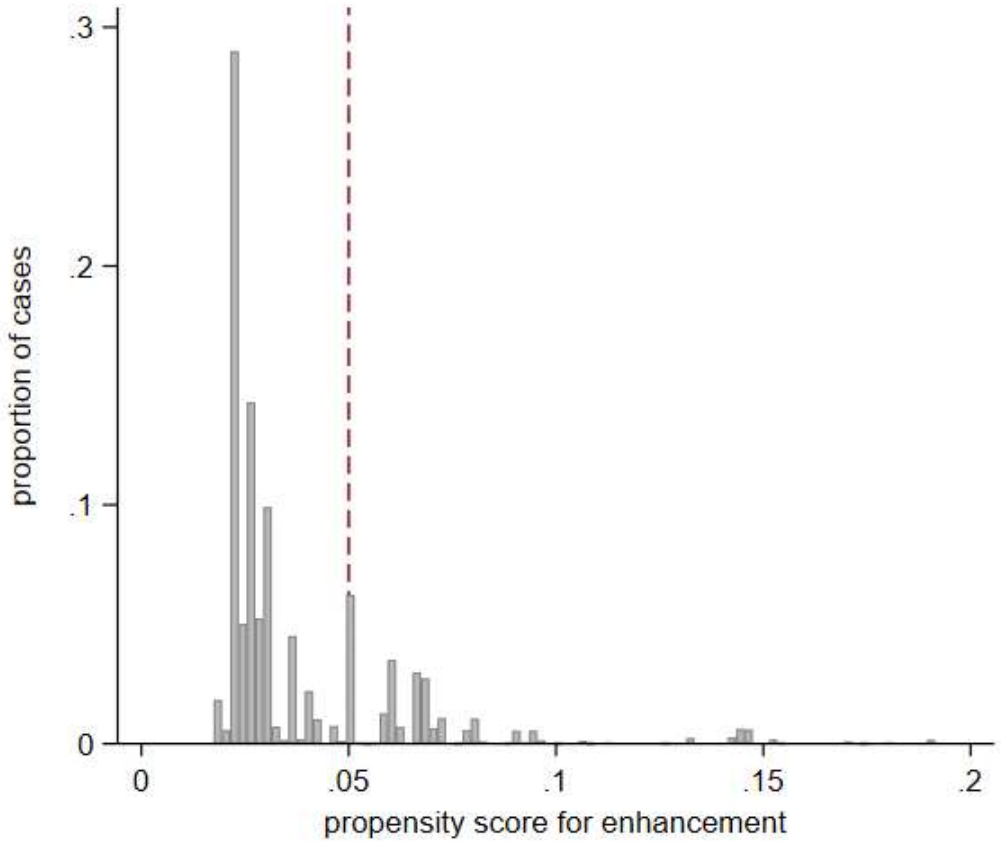


Panel (D): Female



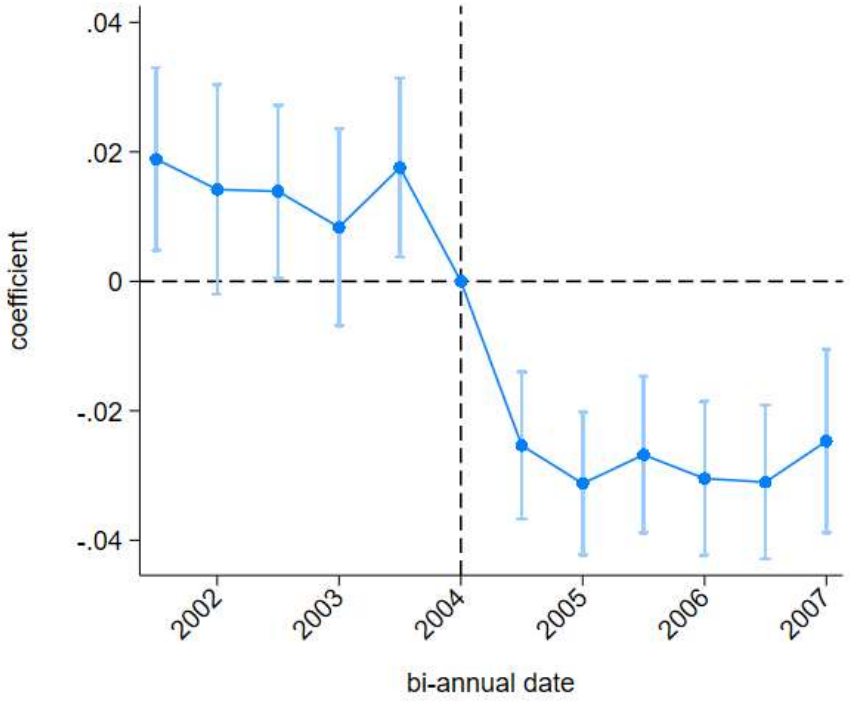
Notes: This figure presents additional smoothness checks to ensure the sample is not changing across the timing of *Blakely.* Panel A gives the number of cases per month, while the other panels give averages of defendant characteristics. All figures represent tests for the phat ≥0.05 sample.

Figure A.2 Propensity Score Distribution



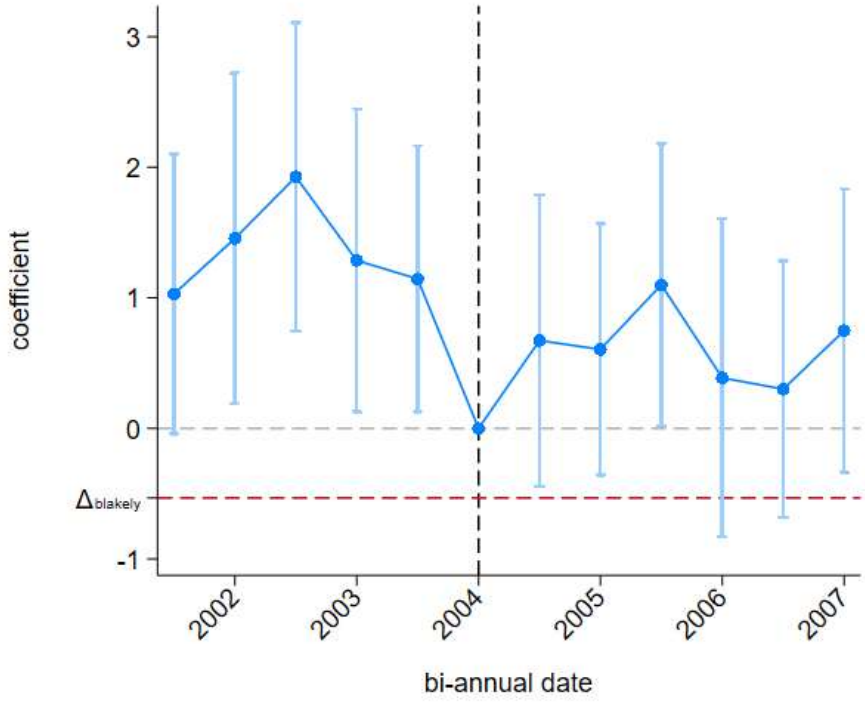
Notes: This histogram shows the distribution of cases by the estimated propensity of enhancement. The propensity score was estimated using a logit on cases decided before the timing of *Blakely.* See Section 4.2 for more details on the propensity score estimation. The dashed line at 0.05 indicates where the sample is separated for much of the analysis. In many cases we drop cases with a propensity score under 0.05. This decreases the sample size considerably, allowing for a higher proportion of cases likely to receive enhancement.

Figure A.3 Enhancement Event Study



Notes: This graph shows the event study analysis for the percentage of cases with enhancement. This specification controls for defendant characteristics, defense type, grid fixed effects, and county fixed effects. The dashed vertical line indicates the timing of *Blakely*. The sample includes cases with propensity scores greater than 0.05. Bars represent 95% confidence intervals.

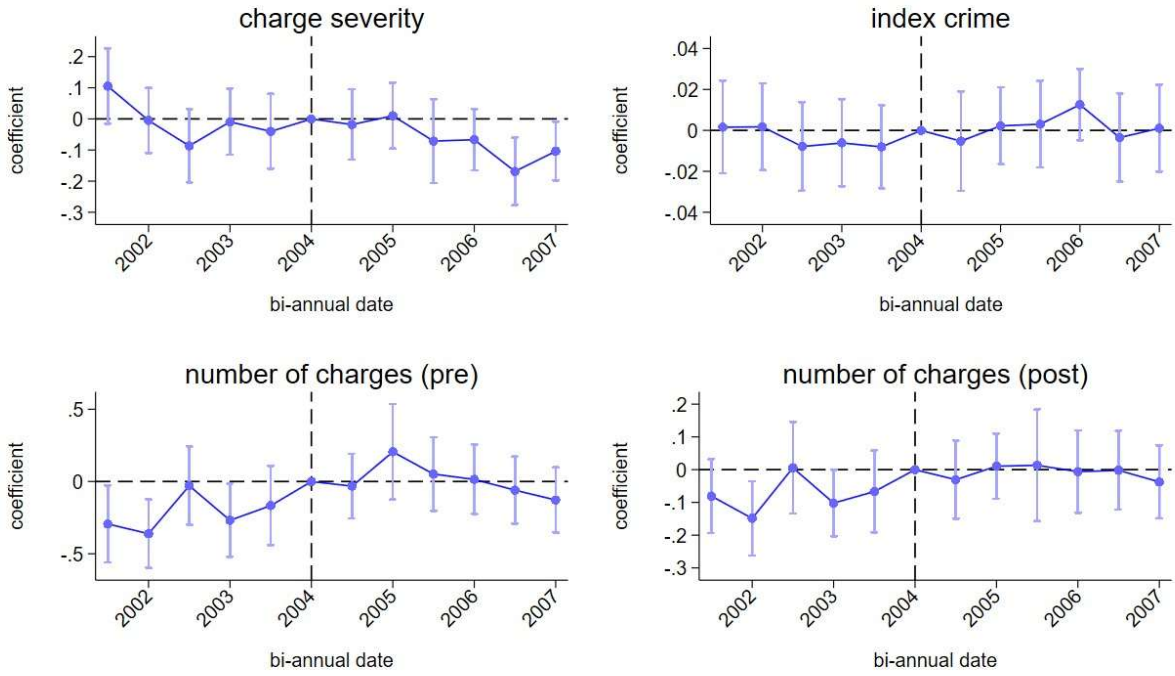
Figure A.4 Sentence Length Event Study



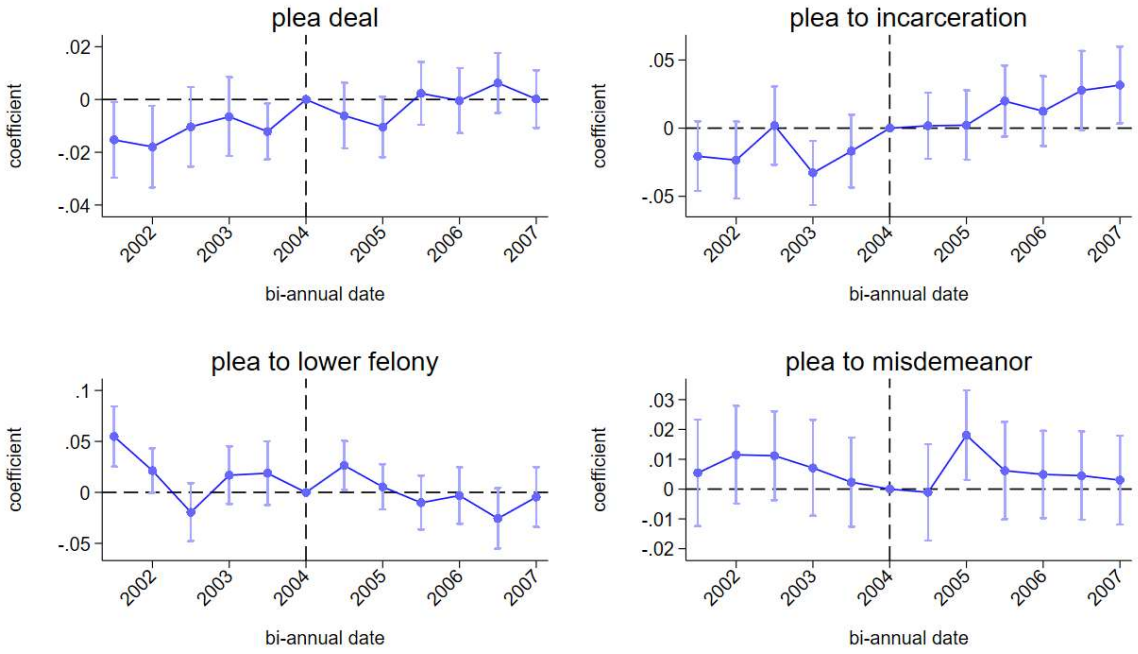
Notes: This graph gives the event study for sentence length that occurs from *Blakely.* This specification controls for defendant characteristics, defense type, grid fixed effects, and county fixed effects. The vertical dotted line indicates the timing of the *Blakely* decision. The red dashed line gives the predicted decrease in sentence length following *Blakely,* termed the benchmark value. The sample includes cases with propensity scores greater than 0.05. Bars represent 95% confidence intervals.

Figure A.5 Changes in Prosecutor Decisions

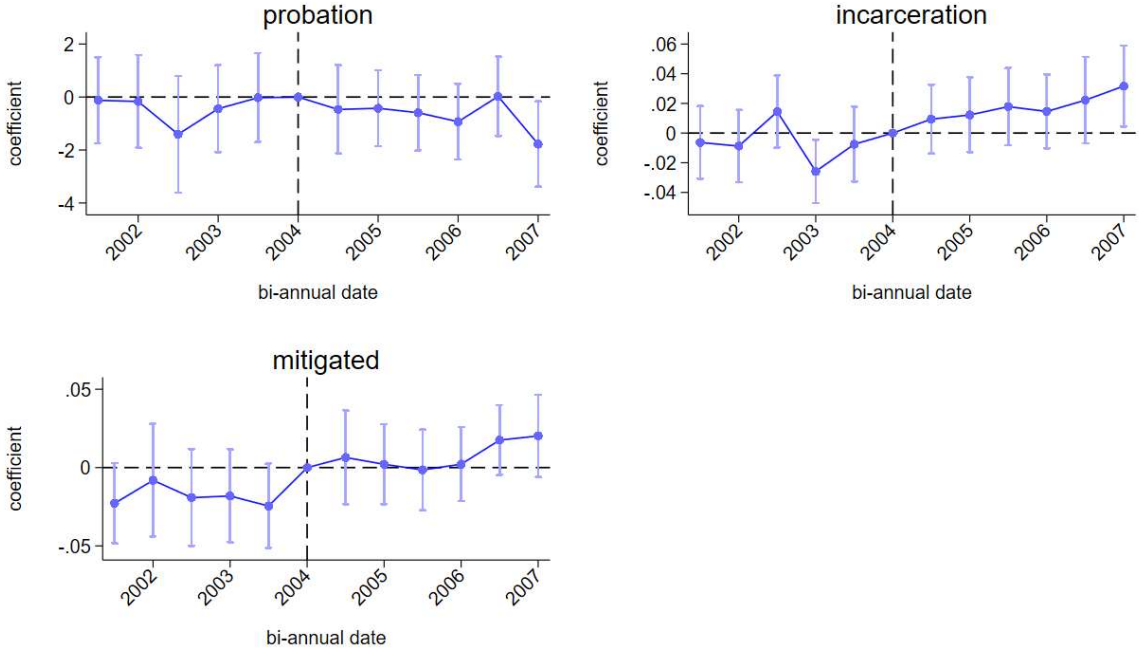
Panel (A): Charging outcomes



Panel (B): Plea bargain outcomes



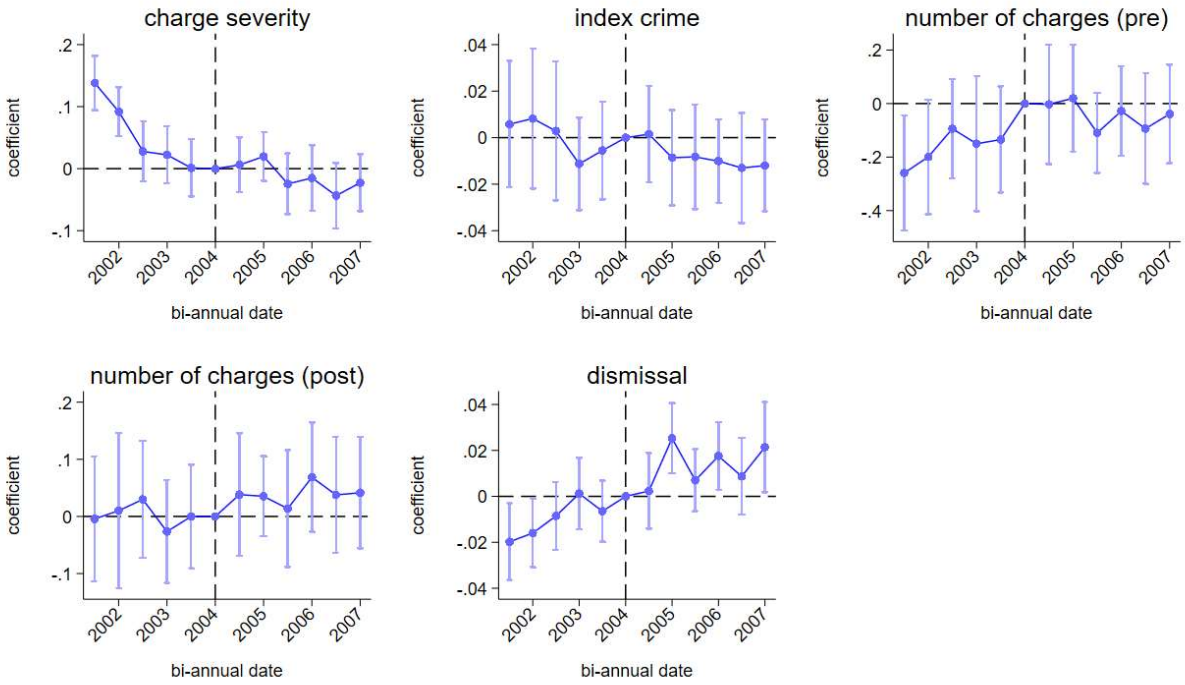
Panel (C): Sentencing outcomes



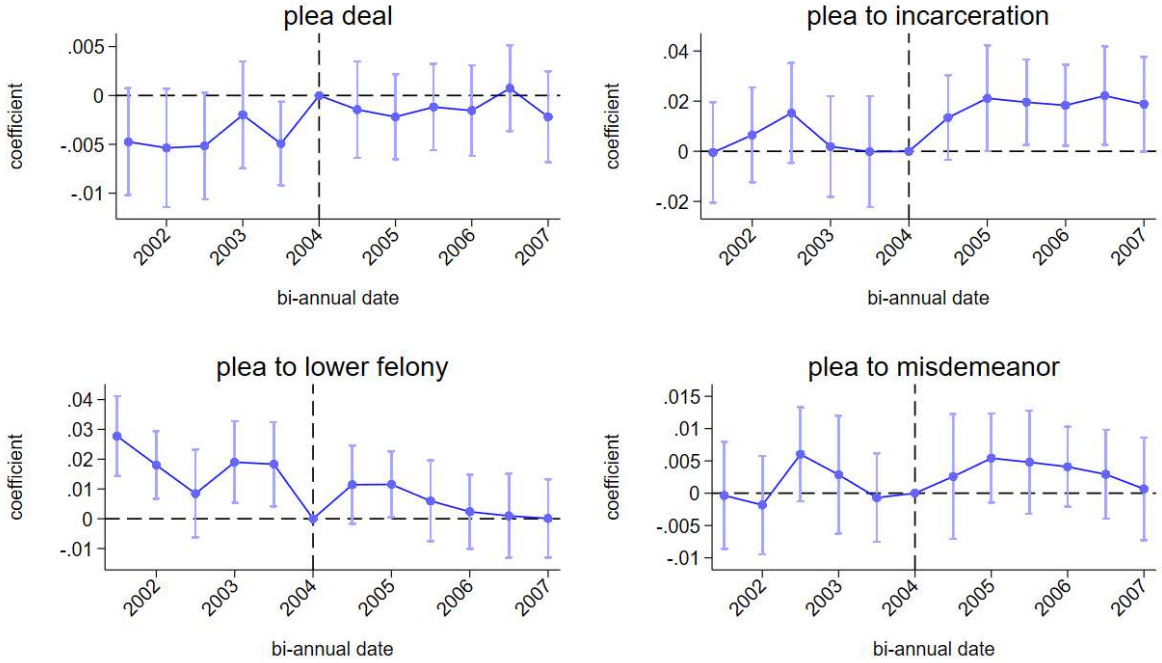
Notes: This graph gives the event study plots for outcomes related to prosecutor choice. Each regression includes grid fixed effects. The sample restricted to cases with propensity scores above 0.05 is used throughout.

Figure A.6 Changes in Prosecutor Decisions – Full Sample

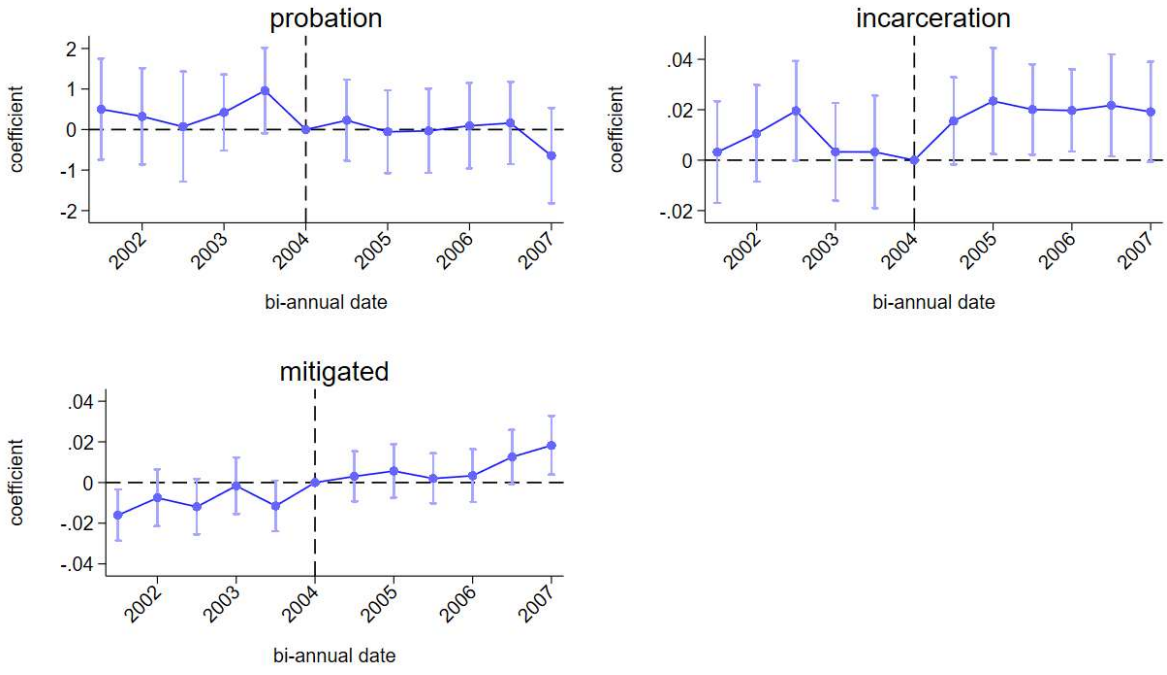
Panel (A): Charging outcomes



Panel (B): Plea bargain outcomes

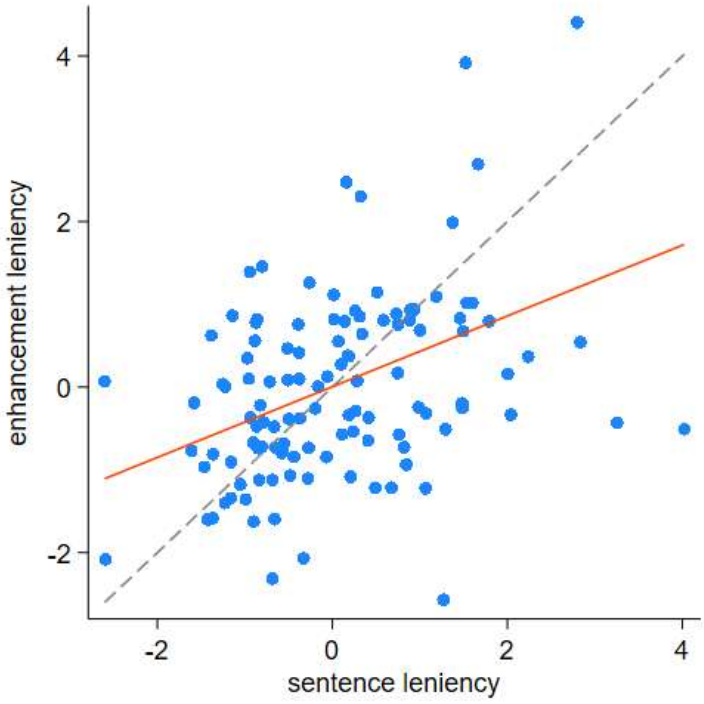


Panel (C): Sentencing outcomes



Notes: This figure reports event study analysis for each of the prosecutor outcomes as in Figure A4, but now includes the full sample of all cases.

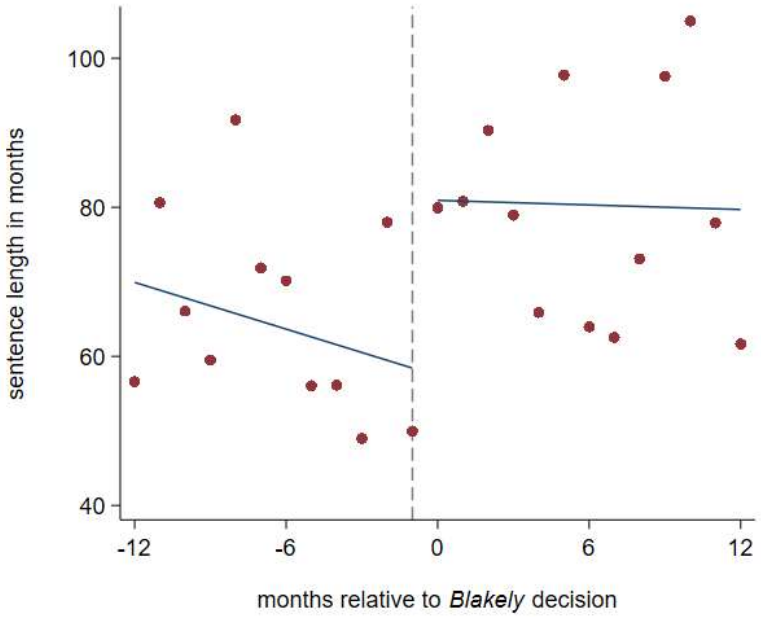
Figure A.7 Leniency Measure Correlation



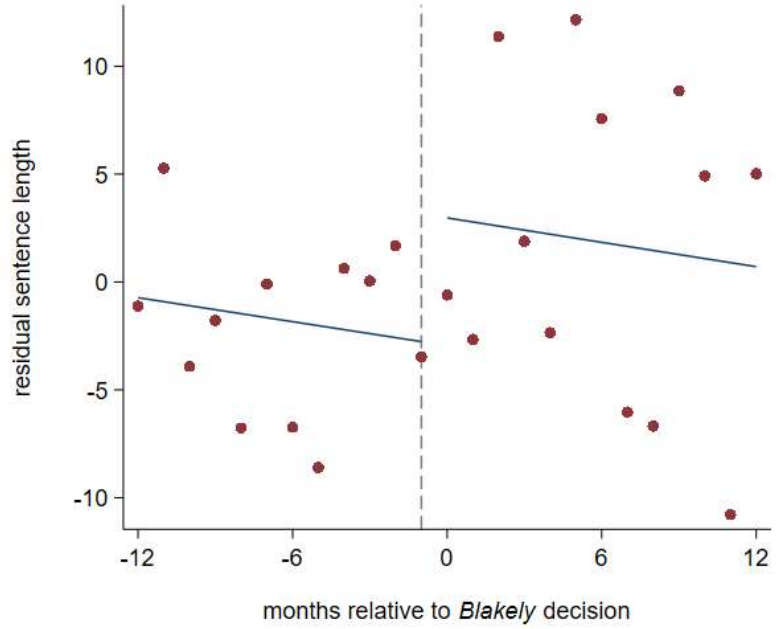
Notes: We plot the two leniency measures against each other to see to what degree judges are lenient across sentence length and enhancement. The dashed 45-dgree line gives reference for a judge that is equally lenient/harsh by sentence and enhancement measures. The orange line gives the actual linear correlation between these two measures. Each leniency measure is standardized to have a mean of 0 and a standard deviation of 1. The figure is truncated to exclude one outlier judge for readability.

Figure A.8. Jury Trial Cases Sentence Discontinuity – Full Sample and Residuals

Panel (A): Full Sample



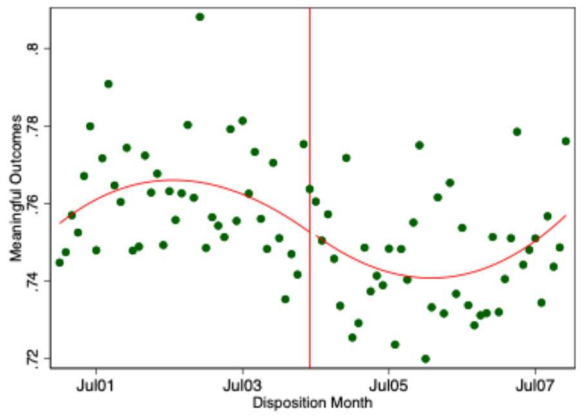
Panel (B): Residuals



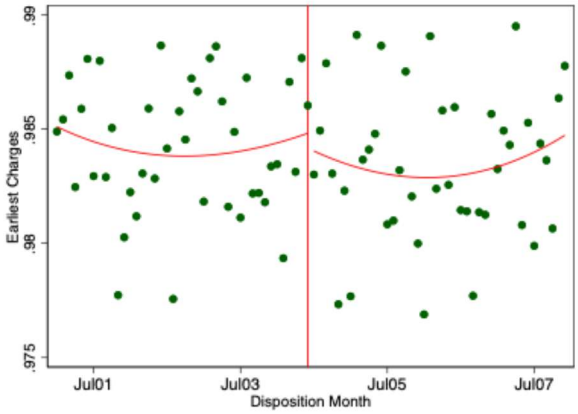
Notes: These figures show the discontinuity in sentence length at the time of *Blakely* for cases resolved at jury trial. Panel A contains all trial cases within the 12-month bandwidth of *Blakely* while Panel B gives results for the phat restricted sample and plots residuals rather than raw sentence lengths.

Figure A.9. Sample Restriction at the *Blakely* Cutoff

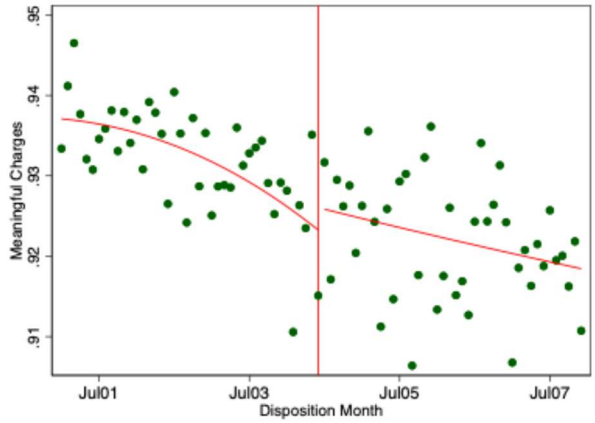
Panel (A): Meaningful Outcomes



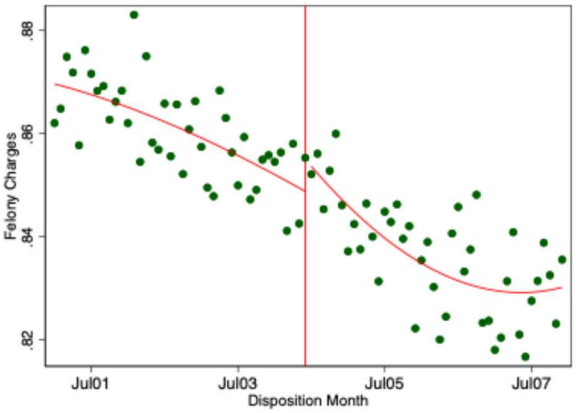
Panel (B): Earliest Charges



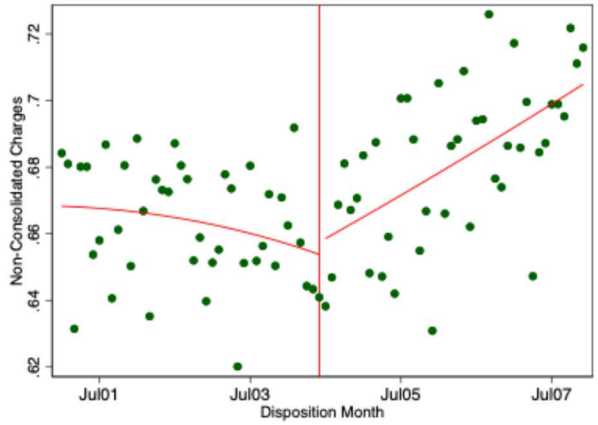
Panel (C): Meaningful charges



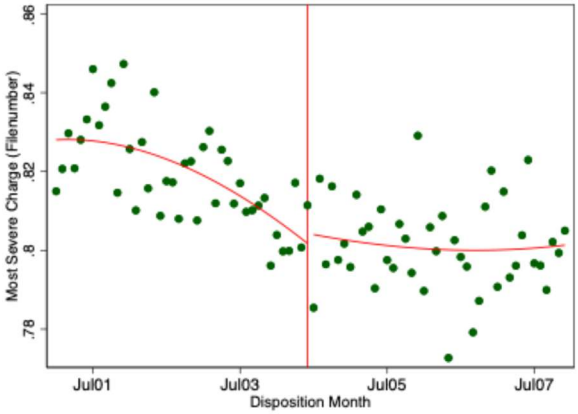
Panel (D): Felony Charges



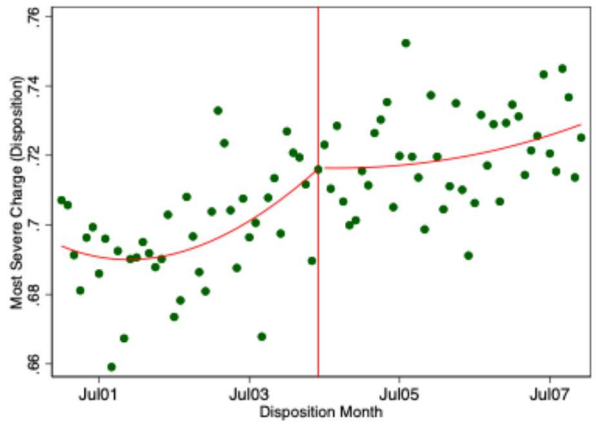
Panel (E): Non-consolidated Charges



Panel (F): Most Severe Charge 1



Panel (G): Most Severe Charge 2



Notes: The figures show the fraction of observations that survived each sample restriction and the changes in the fractions at the *Blakely* cutoff. The line in the middle is the month when *Blakely* was decided. Appendix D has detailed explanation on the meaning of each sample restriction. The curves are fitted on each side of *Blakely* with a second-degree polynomial function of the time. The dataset comes from the North Carolina Administrative Office of Courts (AOC). The sample period is from 2001 to 2007.

Appendix B: Additional Tables

Table B1. The Effect of *Blakely* on Enhancement Rate – Event Study

	(1)	(2)	(3)	(4)
Blakely	-0.0282***	-0.0282***	-0.0288***	-0.0292***
	(0.00625)	(0.00627)	(0.00630)	(0.00643)
Pre-Blakely mean	0.0721	0. 0721	0. 0721	0. 0721
fixed effects	yes	yes	yes	yes
defendant controls	no	yes	yes	yes
counsel/crime controls	no	no	yes	yes
arrest/unemp/population	no	no	no	yes
N	27636	27636	27636	27636

Notes: The dependent variable is a binary measure for whether a case is sentenced with enhancement. Standard errors clustered at the judicial district level are presented in parentheses. Grid, county, and year fixed effects are included in each specification. Defendant controls include age, gender, race, and criminal history points. Counsel/crime controls include binaries for whether the defense type is a private attorney or a public defender and a set of specific crime type fixed effects. The specification in column 4 also controls for the county-year level arrest rate, population level, and unemployment rate. The p-hat restriction is based on the propensity score described above. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.2 Effect of Enhancement on Sentence Length – Event Study

	(1)	(2)	(3)	(4)
Aggravated	18.18***	18.16***	17.86***	17.85***
	(1.056)	(1.056)	(1.057)	(1.055)
Pre-Blakely mean	39.2	39.2	39.2	39.2
fixed effects	yes	yes	yes	yes
defendant controls	no	yes	yes	yes
counsel/crime controls	no	no	yes	yes
arrest/unemp/population	no	no	no	yes
N	13177	13177	13177	13177

Notes: The dependent variable is the minimum sentence length given in months at the time of disposition. Sample only considers pre-*Blakely* observations 3 years before the timing of *Blakely.* Standard errors clustered at the judicial district level are presented in parentheses. Grid, county, and year fixed effects are included in each specification. Defendant controls include age, gender, race, and criminal history points. Counsel/crime controls include binaries for whether the defense type is a private attorney or a public defender and a set of specific crime type fixed effects. The specification in column 4 also controls for the county-year level arrest rate, population level, and unemployment rate. The p-hat restriction is based on the propensity score described above. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.3 Effect of *Blakely* on Sentence Length – Event Study

	(1)	(2)	(3)	(4)
Blakely	0.932	0.912	0.825	0.912
	(0.550)	(0.549)	(0.566)	(0.578)
pre-*Blakely* sentence length	39.2	39.2	39.2	39.2
$\widehat{\Delta}_{blakely}$ (benchmark)	-0.532	-0.531	-0.533	-0.540
t-score (benchmark)	2.66	2.63	2.40	2.51
fixed effects	yes	yes	yes	yes
defendant controls	no	yes	yes	yes
counsel/crime controls	no	no	yes	yes
arrest/unemp/population	no	no	no	yes
N	27536	27536	27536	27536

Notes: The dependent variable is the minimum sentence length given in months at the time of disposition. All specifications use the sample that includes only cases with a propensity score above 0.05 3 years before and after the timing of *Blakely*. $\widehat{\Delta}_{blakely}$ gives the null hypothesis value predicted effect of *Blakely.* t-score (benchmark) gives the t-score value calculated using this benchmark value as the null hypothesis. Standard errors are clustered at the district level, presented in parentheses. Each specification includes grid fixed, county, and year fixed effects.

Table B.4 *Blakely* Effects – Kernel and Quadratic Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
	enhanced	enhanced	enhanced	sentence length	sentence length	sentence length
Blakely	-0.0225**	-0.0205*	-0.0257*	1.522*	1.457*	1.074
	(0.0112)	(0.0109)	(0.0150)	(0.842)	(0.825)	(1.357)
kernel	triangle	epanech.	uniform	triangle	epanech.	uniform
polynomial order	1	1	2	1	1	2
N	8440	8440	9314	8440	8440	9314

Notes: This table presents results for the main discontinuity analysis again using several different kernels and a second order polynomial. All specifications consider Blakely effects using the 0.05 restricted propensity score sample and a 12-month bandwidth. Standard errors presented in parentheses are clustered at the month level.
* p < 0.1, ** p < 0.05, *** p < 0.01

Table B.5 *Blakely* Effects –Robustness Across Various Standard Errors

	(1)	(2)	(3)	(4)	(5)	(6)
	enhanced	enhanced	enhanced	sentence length	sentence length	sentence length
Blakely	-0.0208**	-0.0208*	-0.0208**	1.996***	1.996**	1.996**
	(0.0089)	-	(0.0097)	(0.6795)	-	(0.7807)
p-value	0.021	0.056	0.032	0.003	0.022	0.011
SE type	EHW	wild bootstrap	cluster on district	EHW	wild bootstrap	cluster on district
N	9314	9314	9314	9314	9314	9314

Notes: This table presents results for the main discontinuity analysis again using different standard errors for inference. Following motivation given in Kolesar & Rothe (2018), columns 1 and 4 use conventional Eicker-Huber-White standard errors. Columns 2 and 5 use wild clustered bootstrap standard errors clustered at the month level. We also check that clustering on geographic level does not matter in columns 3 and 6, which cluster at the district level. All specifications use a 12-month bandwidth, a uniform kernel, and a local linear design.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.6 *Blakely* Effects on Sentence Length – Jury Trial Cases Only

	(1)	(2)	(3)	(4)	(5)
Blakely	3.791	8.670**	3.796	4.799	13.10***
	(4.435)	(4.368)	(4.363)	(3.839)	(4.686)
Pre-*Blakely* mean	106.1	106.3	106.0	72.8	95.3
p-hat restriction	≥ 0.05	≥ 0.05	≥ 0.05	none	≥ 0.03
bandwidth	9	12	15	12	12
N	341	446	546	853	566

Notes: This table presents results for the main sentence length discontinuity analysis again using several different kernels and a second order polynomial. All specifications consider *Blakely* effects using the 0.05 restricted propensity score sample and a 12-month bandwidth. Standard errors presented in parentheses are clustered at the month level.

* p < 0.1, ** p < 0.05, *** p < 0.01

Table B.7 Heterogeneous Effects of *Blakely* by Judge Leniency

	(1) enhancement	(2) enhancement	(3) min. sent	(4) min. sent
Blakely	0.0044	0.0027	1.569**	1.160*
	(0.0067)	(0.0066)	(0.664)	(0.694)
DD_50	-0.0210***	-0.0219***	-0.687	-0.784
	(0.0059)	(0.0058)	(0.881)	(0.781)
DD_75	-0.0403***	-0.0391***	-0.246	-0.135
	(0.0056)	(0.0057)	(0.595)	(0.608)
DD_100	-0.0707***	-0.0709***	-1.947***	-1.919***
	(0.0080)	(0.0080)	(0.549)	(0.540)
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 is analogous to Table 6 in the main text, but instead uses the enhancement leniency measure rather than the sentence length one. It gives the differential effects of *Blakely* on enhancement rate and sentence length across the four judge leniency groups. 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 month level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.8 Binding Sentencing Guidelines Mechanism Analysis

	(1) top guideline	(2) top guideline	(3) top guideline	(4) top guideline
sent_leniency_50	0.0080 (0.0300)	0.0004 (0.0288)	0.0052 (0.0398)	-0.0016 (0.0400)
sent_leniency_75	0.0913*** (0.0319)	0.0832*** (0.0279)	0.1280*** (0.0357)	0.1190*** (0.0346)
sent_leniency_100	0.0936*** (0.0340)	0.0850*** (0.0289)	0.0857** (0.0369)	0.0769** (0.0348)
Fixed effects	Yes	yes	yes	yes
Additional controls	no	yes	no	yes
p-hat restriction	≥ 0.05	≥ 0.05	none	none
Pre-*Blakely* mean	0.3099	0.3099	0.4575	0.4575
N	25508	25508	110599	110599

Notes: This table presents results for the regression specified in Equation 6 in the main text. The dependent variable is a binary measure that is equal to one if the minimum sentence imposed equals the maximum amount permitted under the guidelines. Each reported variable represents a quartile of judge leniency, with the most lenient quartile being the omitted group. The sample is limited to 36 months before and after the timing of *Blakely* for consistency with other analyses. Columns 1 and 2 perform the analysis for the p-hat ≥ 0.05 restricted sample while columns 3 and 4 contain all observations. 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 district level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.9 Enhancement Effects by Judge Leniency – RD Design

	(1)	(2)	(3)	(4)	(5)	(6)
leniency_50	0.0134	0.0142	0.0174**	0.0190	0.0143	0.0153*
	(0.0106)	(0.00898)	(0.00781)	(0.0119)	(0.00989)	(0.00855)
leniency_75	0.0331***	0.0335***	0.0350***	0.0344***	0.0337***	0.0308***
	(0.0109)	(0.00951)	(0.00834)	(0.0121)	(0.0104)	(0.00908)
leniency_100	0.0589***	0.0722***	0.0706***	0.0655***	0.0706***	0.0636***
	(0.0122)	(0.0109)	(0.00930)	(0.0133)	(0.0115)	(0.00972)
Blakely	0.000511	0.00418	-0.00395	0.00322	0.00533	-0.00510
	(0.0.122)	(0.0107)	(0.00937)	(0.0128)	(0.0110)	(0.00957)
leniency_50**Blakely*	-0.0105	-0.0127	-0.0167*	-0.0105	-0.0107	-0.0110
	(0.0133)	(0.0114)	(0.0100)	(0.0144)	(0.0119)	(0.0104)
leniency_75**Blakely*	-0.0202	-0.0224*	-0.0290***	-0.0209	-0.0242*	-0.0279**
	(0.0138)	(0.0120)	(0.0105)	(0.0147)	(0.0125)	(0.0110)
leniency_100**Blakely*	-0.0585***	-0.0698***	-0.0668***	-0.0674***	-0.0741***	-0.0659***
	(0.0137)	(0.0124)	(0.0109)	(0.0150)	(0.0130)	(0.0113)
p-hat restriction	≥ 0.05	≥ 0.05	≥ 0.05	≥ 0.05	≥ 0.05	≥ 0.05
bandwidth	9	12	15	9	12	15
demographic controls	no	no	no	yes	yes	yes
county fixed effects	no	no	no	yes	yes	yes
grid fixed effects	yes	yes	yes	yes	yes	yes
N	6987	9369	11627	6978	9356	11611

Notes: The dependent variable is a binary variable for whether the case received enhancement. The leniency variable is determined according to judge propensity for enhancement, as described in Section REF. Variables leniency_50, leniency_75, and leniency_100 are binary variables signifying whether the assigned judge for each case is in the 2nd, 3rd, or top quartile of harshness, respectively. Here, leniency is based on the percent of a judge's cases that receive enhancement. The bottom quartile (most lenient judges) variable is the omitted group. Demographic controls include age, sex, race, criminal history, and representation controls. Standard errors are presented in parentheses. Note that we use conventional Eicker-Huber-White following Kolesar & Rothe (2018). The p-hat restriction is based on the propensity score described above. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.10 Sentence Effects by Judge Leniency – RD Design

	(1)	(2)	(3)	(4)	(5)	(6)
leniency_50	1.076	2.029***	2.057***	1.189	2.039***	2.158***
	(0.765)	(0.651)	(0.607)	(0.905)	(0.740)	(0.662)
leniency_75	1.889**	3.081***	3.092***	2.360***	3.307***	3.042***
	(0.773)	(0.680)	(0.640)	(0.952)	(0.781)	(0.709)
leniency_100	4.215***	5.844***	5.463***	4.948***	6.030***	5.562***
	(0.884)	(0.777)	(0.694)	(0.992)	(0.847)	(0.769)
Blakely	1.652	3.748***	2.718***	2.085**	3.839***	2.852***
	(1.036)	(0.905)	(0.813)	(1.042)	(0.915)	(0.821)
leniency_50**Blakely*	-0.244	-1.490	-1.658**	-0.288	-1.048	-1.449*
	(1.083)	(0.918)	(0.835)	(1.137)	(0.953)	(0.853)
leniency_75**Blakely*	-1.139	-2.506***	-2.273***	-1.834**	-2.646***	-2.531***
	(1.137)	(0.970)	(0.903)	(1.208)	(1.008)	(0.928)
leniency_100**Blakely*	-1.015	-3.403***	-3.150***	-1.911	-3.694***	-3.328***
	(1.221)	(1.056)	(0.940)	(1.259)	(1.082)	(0.960)
p-hat restriction	≥ 0.05	≥ 0.05	≥ 0.05	≥ 0.05	≥ 0.05	≥ 0.05
bandwidth	9	12	15	9	12	15
demographic controls	no	no	no	yes	yes	yes
county fixed effects	no	no	no	yes	yes	yes
grid fixed effects	yes	yes	yes	yes	yes	yes
N	6966	9339	11589	6957	9326	11573

Notes: The dependent variable is the minimum sentence length given in months at the time of disposition. The leniency variable is a mean of other sentence lengths assigned by the judge, as described in Section REF. Variables leniency_50, leniency_75, and leniency_100 are binary variables signifying whether the assigned judge for each case is in the 2nd, 3rd, or top quartile of harshness, respectively. The bottom quartile (most lenient judges) variable is the omitted group. Demographic controls include age, sex, race, criminal history, and representation controls. Standard errors are presented in parentheses. Note that we use conventional Eicker-Huber-White following Kolesar & Rothe (2018). The p-hat restriction is based on the propensity score described above. c

Table B.11 Heterogeneous Effects of *Blakely* by Other Judge Characteristics

	(1) male	(2) white	(3) tenure	(4) democrat
Panel A: Enhancement Effects				
DD	-0.0139	-0.0095	-0.0063	0.0179
	(0.0161)	(0.0137)	(0.0074)	(0.0424)
Panel B: Sentence Length Effects				
DD	0.2597	-0.7187	-1.028*	1.530
	(1.313)	(1.096)	(0.6027)	(1.971)
N	13227	13077	26692	4803

Notes: For all regressions in Panel A, the dependent variable is a binary for enhanced sentence while Panel B gives results for minimum sentence length. The column headers represent the judge characteristic being tested. DD is an interaction between this characteristic and the timing of *Blakely*. Note that tenure is a binary measure separating the top half tenured vs bottom half tenure judges. All regressions are based on the 0.05 propensity score reduction and include grid and year fixed effects. Standard errors are presented in parentheses and are clustered at the district level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.12 Sample Restrictions and the *Blakely* Discontinuity

	(1) Meaningful Outcomes	(2) Earliest Charges	(3) Meaningful Charges	(4) Felony Charges	(5) Non-Consolidated Charges	(6) Most Severe Charge 1	(7) Most Severe Charge 2
Blakely	0.00130	-0.00229	0.00657	-0.00156	0.00898	-0.00117	-0.00593
	(0.00360)	(0.00193)	(0.00466)	(0.00351)	(0.00936)	(0.00434)	(0.00933)
N	332435	249999	245940	227714	193190	127942	96498

Notes: The table shows the sample balance under each sample restriction criterion outlined in Appendix D. From Column (1) to (7), the outcome variables are dummies indicating meaningful outcome, earliest charge, meaningful charge, felony, non-consolidated charge, case aggregation 1, and case aggregation 2. The regression specification in each column is local linear regression with county fixed effects, charging offense fixed effect, and county level controls (population, arrest rates, and employment rates). We restrict the sample to 12 months before and after *Blakely*. Significance level * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix C: More Institutional Details

C.1: Sentencing Guidelines

**** Effective for Offenses Committed on or after 12/1/95 Through 11/30/09 ****

FELONY PUNISHMENT CHART

PRIOR RECORD LEVEL

OFFENSE CLASS	I 0 Points	II 1-4 Points	III 5-8 Points	IV 9-14 Points	V 15-18 Points	VI 19+ Points
A	Death or Life Without Parole					
B1	A *240 - 300* **192 - 240** *144 – 192*	A *288 - 360* **230 - 288** *173 – 230*	A *336 - 420* **269 - 336** *202 – 269*	A *384 - 480* **307 - 384** *230 – 307*	A *Life Without Parole* **346 - 433** *260 – 346*	A *Life Without Parole* **384 - 480** *288 - 384*
B2	A *157 - 196* **125 - 157** *94 - 125*	A *189 - 237* **151 - 189** *114 - 151*	A *220 - 276* **176 - 220** *132 - 176*	A *251 - 313* **201 - 251** *151 - 201*	A *282 - 353* **225 - 282** *169 - 225*	A *313 - 392* **251 - 313** *188 - 251*
C	A *73 – 92* **58 - 73** *44 - 58*	A *100 – 125* **80 - 100** *60 - 80*	A *116 – 145* **93 - 116** *70 - 93*	A *133 - 167* **107 - 133** *80 - 107*	A *151 - 188* **121 - 151** *90 - 121*	A *168 - 210* **135 - 168** *101 - 135*
D	A *64 - 80* **51 - 64** *38 - 51*	A *77 - 95* **61 - 77** *46 - 61*	A *103 - 129* **82 - 103** *61 - 82*	A *117 - 146* **94 - 117** *71 - 94*	A *133 - 167* **107 - 133** *80 - 107*	A *146 - 183* **117 - 146** *88 - 117*
E	I/A *25 - 31* **20 - 25** *15 - 20*	I/A *29 - 36* **23 - 29** *17 - 23*	A *34 - 42* **27 - 34** *20 - 27*	A *46 - 58* **37 - 46** *28 - 37*	A *53 - 66* **42 - 53** *32 - 42*	A *59 - 74* **47 - 59** *35 - 47*
F	I/A *16 - 20* **13 - 16** *10 - 13*	I/A *19 - 24* **15 - 19** *11 - 15*	I/A *21 - 26* **17 - 21** *13 - 17*	A *25 - 31* **20 - 25** *15 - 20*	A *34 - 42* **27 - 34** *20 - 27*	A *39 - 49* **31 - 39** *23 - 31*
G	I/A *13 - 16* **10 - 13** *8 - 10*	I/A *15 - 19* **12 - 15** *9 - 12*	I/A *16 - 20* **13 - 16** *10 - 13*	I/A *20 - 25* **16 - 20** *12 - 16*	A *21 - 26* **17 - 21** *13 - 17*	A *29 - 36* **23 - 29** *17 - 23*
H	C/I/A *6 - 8* **5 - 6** *4 - 5*	I/A *8 - 10* **6 - 8** *4 - 6*	I/A *10 - 12* **8 - 10** *6 - 8*	I/A *11 - 14* **9 - 11** *7 - 9*	I/A *15 - 19* **12 - 15** *9 - 12*	A *20 - 25* **16 - 20** *12 - 16*
I	C *6 - 8* **4 - 6** *3 - 4*	C/I *6 - 8* **4 - 6** *3 - 4*	I *6 - 8* **5 - 6** *4 - 5*	I/A *8 - 10* **6 - 8** *4 - 6*	I/A *9 - 11* **7 - 9** *5 - 7*	I/A *10 - 12* **8 - 10** *6 - 8*

DISPOSITION
Aggravated Range
PRESUMPTIVE RANGE
Mitigated Range

A – Active Punishment **I – Intermediate Punishment** **C – Community Punishment**

Numbers shown are in months and represent the range of minimum sentences

C.2 Statutory Aggravating Factors

Aggravating factors listed in § 15A-1340.16.(d):

(1) The defendant induced others to participate in the commission of the offense or occupied a position of leadership or dominance of other participants.

(2) The defendant joined with more than one other person in committing the offense and was not charged with committing a conspiracy.
(2a) The offense was committed for the benefit of, or at the direction of, any criminal gang as defined by G.S. 14-50.16A(1), with the specific intent to promote, further, or assist in any criminal conduct by gang members, and the defendant was not charged with committing a conspiracy.

(3) The offense was committed for the purpose of avoiding or preventing a lawful arrest or effecting an escape from custody.

(4) The defendant was hired or paid to commit the offense.

(5) The offense was committed to disrupt or hinder the lawful exercise of any governmental function or the enforcement of laws.

(6) The offense was committed against or proximately caused serious injury to a present or former law enforcement officer, employee of the Division of Adult Correction and Juvenile Justice of the Department of Public Safety, jailer, fireman, emergency medical technician, ambulance attendant, social worker, justice or judge, clerk or assistant or deputy clerk of court, magistrate, prosecutor, juror, or witness against the defendant, while engaged in the performance of that person's official duties or because of the exercise of that person's official duties.
(6a) The offense was committed against or proximately caused serious harm as defined in G.S. 14-163.1 or death to a law enforcement agency animal, an assistance animal, or a search and rescue animal as defined in G.S. 14-163.1, while engaged in the performance of the animal's official duties.

(7) The offense was especially heinous, atrocious, or cruel.

(8) The defendant knowingly created a great risk of death to more than one person by means of a weapon or device which would normally be hazardous to the lives of more than one person.

(9) The defendant held public elected or appointed office or public employment at the time of the offense and the offense directly related to the conduct of the office or employment.
(9a) The defendant is a firefighter or rescue squad worker, and the offense is directly related to service as a firefighter or rescue squad worker.

(10) The defendant was armed with or used a deadly weapon at the time of the crime.

(11) The victim was very young, or very old, or mentally or physically infirm, or handicapped.

(12) The defendant committed the offense while on pretrial release on another charge.
(12a) The defendant has, during the 10-year period prior to the commission of the offense for which the defendant is being sentenced, been found by a court of this State to be in willful violation of the conditions of probation imposed pursuant to a suspended sentence or been found by the Post-Release Supervision and Parole Commission to be in willful violation of a condition of parole or post-release supervision imposed pursuant to release from incarceration.

(13) The defendant involved a person under the age of 16 in the commission of the crime.
(13a) The defendant committed an offense and knew or reasonably should have known that a person under the age of 18 who was not involved in the commission of the offense was in a position to see or hear the offense.

(14) The offense involved an attempted or actual taking of property of great monetary value or damage causing great monetary loss, or the offense involved an unusually large quantity of contraband.

(15) The defendant took advantage of a position of trust or confidence, including a domestic relationship, to commit the offense.

(16) The offense involved the sale or delivery of a controlled substance to a minor.
(16a) The offense is the manufacture of methamphetamine and was committed where a person under the age of 18 lives, was present, or was otherwise endangered by exposure to the drug, its ingredients, its by-products, or its waste.
(16b) The offense is the manufacture of methamphetamine and was committed in a dwelling that is one of four or more contiguous dwellings.

(17) The offense for which the defendant stands convicted was committed against a victim because of the victim's race, color, religion, nationality, or country of origin.

(18) The defendant does not support the defendant's family.
(18a) The defendant has previously been adjudicated delinquent for an offense that would be a Class A, B1, B2, C, D or E felony if committed by an adult.

(19) The serious injury inflicted upon the victim is permanent and debilitating.

(20) Any other aggravating factor reasonably related to the purposes of sentencing.

Appendix D: Data cleaning

Our main dataset comes from the North Carolina Administrative Office of Courts. The data cleaning is largely based on Silveria (2017) which used the same dataset.

The initial database includes all of criminal charges filed in North Carolina. The state court system typically defines "case" with a common file number, which is frequently associated with multiple counts of charges. However, such definition of "case" does not accurately reflect the reality of criminal cases are resolved and could cause many issues in empirical analysis. First, the same file number reappears in the data if defendants violate probation conditions, which should not be counted as new criminal offenses. Second, the data suggests that defendants are often disposed of various charges associated with different file numbers on the same date. What's more, anecdotal evidence suggests that plea bargain applies to all outstanding charges. As a result, we decided to adopt an alternative definition of "case" by using a pair of individual identifier and disposition date, which was used in Abrams & Fackler (2020).

The data cleaning begins with a total number of 2,364,028 charges which were filed in the North Carolina superior courts between 1995 and 2009. This paper focuses on the sentence enhancement outcomes among defendants charged with felony offenses. In North Carolina, superior courts have the exclusive jurisdiction over all felony cases. We dropped all the cases and charges resolved at the NC district courts because the district courts only hear misdemeanor and infraction cases which are not relevant to our analysis of sentence enhancements.

Firstly, we keep charges with meaningful outcomes, meaning either dismissed or determined by judges or juries. This procedure eliminated irrelevant cases where there are changes of venues, decisions were made state district courts, etc. The observations dropped to 1,774,336 (25% decrease). Secondly, among charges associated with the same file numbers, we keep the counts which were disposed at the earliest date, which eliminated all the charges filed due to probation violations. The number of observations dropped to 1,745,225 (1.6% decrease). Thirdly, we dropped all charges which do not consist of meaningful information on charged offense and case outcomes. The observations dropped to 1,621,345 (7.1% decrease). Fourthly, we drop the charges where defendants are charged of misdemeanor offenses as the sentencing guidelines in Figure D.1 does not apply to misdemeanor offenses. This step reduces the observation to 1,379,454 (14.9% decrease). Lastly, we dropped charges which are consolidated with other ones, which further reduced the number of observations to 937,817 (32% decrease).[1] North Carolina state laws allow courts to consolidate less serious charges and to only issue one judgment on the basis of the most serious offense. Consolidated charges do not have influence the final sentence outcomes.

Then, we further aggregate charges at two levels in order to define criminal "cases." First, for each file number, we keep the count associated with the most severe charge.[2] This step

[1] The data includes a variable called "crdcfjf" and "crdcfjo," which identifies the charge being consolidated to.
[2] For charges associated with the same file number, we first order them on the basis of punishment severity, which we measure by using a pair of punishment type and minimum sentence length. We keep the count(s) associated the most severe punishment. However, such procedure could not help us determine the most severe charge in some

reduces the number of observations to 713,404 (24% decrease). Furthermore, we implemented the same aggregation procedure for each pair of individual identifiers and disposition date. The number of observations dropped to 504,498 (27.5% decrease). In the end, we are able to proxy criminal "case" by the most severe charge that defendants are disposed on the same date. In developing the empirical analysis, we further restrict time frame from 2001 to 2007.

One might be concerned that our extensive data cleaning process might introduce sample selection bias that might be correlated with the *Blakely* discontinuity. To mitigate such concern, we performed both graphical and regression analysis showing that the fraction of the sample that survives each sample restriction criteria is smooth at the discontinuity. Figure A.9 shows that the fraction of sample, calculated at the monthly level, that survived each of the seven data cleaning criteria. We implemented density test for each sample restriction criterion around the timing of the discontinuity. The graphs that none of the sample restriction criterion creates artificial discontinuity at the timing of *Blakely*. In addition, we performed regression analysis examining whether *Blakely* affects the likelihood of our data surviving each sample restriction criterion. Table B.10 shows the results. All the regression coefficients are small and statistically indistinguishable from 0. The regression results reassure that none of our sample restrictions criteria introduced artificial discontinuities at the timing of *Blakely*.

situations. For example, dismissed cases or charges with identical sentences. For those cases, we keep the charge associated with the most severe filed offense.