

Judicial Scarring

By Karthik Srinivasan ¹
January 05, 2023

I document that experienced decision makers can be influenced by irrelevant events in a high stakes setting, felony sentencing in Cook County. Using a difference-in-differences design, I estimate that judges hand down sentences that are 12% longer around the date of a first degree murder sentencing. The effect is concentrated among defendants who resemble the murderer along the dimensions of race and charge severity. The bias affects 9% of defendants and temporarily increases the Black sentencing penalty by 60%.

Sequential judgments are a feature of many important decision making settings. Evaluating job applicants, searching for housing, and dating all involve repeated assessments. Yet despite their importance and ubiquity, sequential judgments may be subject to systematic psychological biases such as anchoring and contrast effects.

Predicting and correcting these biases is challenging in practice because behavioral theories can make opposite predictions in the same setting. This is the case in the context I study, judicial sentencing decisions, where anchoring and contrast effects are both plausible and suggest opposite biases. Consider a judge who has recently handed down a long sentence for a particularly gruesome crime and now must sentence a typical defendant. If the long sentence serves as an anchor, the judge may adjust down insufficiently resulting in over-sentencing. Alternatively, after handling the gruesome crime, the judge may find a run-of-the-mill crime to appear mild in contrast resulting in under-sentencing. These opposite predictions are not purely theoretical: experimental evidence documents anchoring and contrast effects in the context of judicial sentencing (Pepitone and DiNubile, 1976; Guthrie et al., 2000; Enough and Mussweiler, 2001). Field evidence can help adjudicate whether one bias dominates in practice.

I study felony sentencing decisions in Cook County. This is an attractive setting to test behavioral theories in the field because institutional features push against finding bias. The stakes are high: the average defendant is sentenced to over 2 years in prison, and judges are given substantial discretion. The decision makers are educated and experienced: judges are required to have a post-graduate degree and typically serve on the bench for multiple 6-year terms. Moreover, judges have a legal obligation to hand down fair sentences.

Despite these factors, I find that Cook County judges exhibit a persistent behavioral bias. Using administrative data on the universe of Cook County felony sentencing decisions from March 2011 to September 2022, I estimate a difference-in-differences design around the date of a first degree murder sentencing. Judges hand down sentences that are 88 days longer in the 10 days before and 20 days after a first degree murder sentencing, an effect which I refer to as scarring. This is a sizable effect: it represents 24% of the median sentence length. Because judges handle first degree murders regularly, scarring distorts sentencing

¹University of Chicago, Booth School of Business (email: ks@chicagobooth.edu). I thank Alex Frankel, Eric Zwick, Devin Pope, Walter Zhang, Michael Galperin, Benedict Guttman-Kenney, Claire Bergey and participants in Booth's Student Research in Economics Seminar and Behavioral Economics Lab for valuable comments. I thank Era Lauder milk, Hayley Hopkins, Alexi Stocker, and Reuben Bauer for providing important legal context. I thank James Kiselik for helpful edits. Thanks to Arjun Srinivasan, Anand Srinivasan, Shanthi Srinivasan, Muthayyah Srinivasan, Gabi Hirsch, Alex Duner, and Jaclyn Zhou, Sam Osburn, Anna Cormack, Janani Nathan and Mochi.

for 9% of defendants on an ongoing basis. The scarring effect is consistent with anchoring, but I cannot rule out some alternative behavioral explanations (See Section 3.4).

I decompose the main effect into an extensive margin (moving defendants from probation into prison) and an intensive margin (making prison sentences longer). I find that effects are primarily driven by the intensive margin: defendants who would have been sentenced to prison face sentences that are 140 days longer, but I can reject increases in the probability of going to prison larger than 2%.

To get at mechanisms, I estimate triple-difference designs comparing treatment effects between defendants who are similar to and different from the murderer along the dimensions of race and charge severity. Treatment effects are 101 days longer for same-race defendants, and 283 days longer for high class felonies. These heterogeneous treatment effects are suggestive evidence that cases which call to mind the murderer are most affected by scarring. These findings are consistent with recent theoretical models which posit that salience underlies many behavioral biases (Bordalo et al., 2022).

While the mechanism I propose is race neutral in principle, the fact that 76% of defendants convicted of first degree murders in Cook County are Black means that scarring compounds existing inequalities in practice. I estimate a triple difference design comparing treatment effects for Black and non-Black defendants and find that differential sentencing persists for 60 days after a first degree murder (results are similar in magnitude in shorter windows, but are estimated imprecisely). This disparity is 60% of the size of the observed sentencing penalty faced by Black defendants in Cook County after conditioning on observables.

Identifying each of these effects presents a non-standard causal inference challenge because each treatment unit (a judge) can be treated multiple times (sentence multiple first degree murders) and time between treatment varies. My primary causal inference design departs from the standard difference-in-differences setup in order to leverage these multiple treatments for identification. I construct an event-by-event dataset where each event is a first degree murder. Treated sentencing decisions are those made by the judge who sentenced the murder in a window of time around the sentencing date. Control sentencing decisions are taken from judges in the same courtroom at the same point in calendar time who do not sentence a murder within the event time window. Assignment to treatment and control is random because cases in Cook County are randomly assigned conditional on courtroom and calendar time. This event-by-event construction means that judges serve as both treatment and control at different points in calendar time resolving a common critique of difference-in-differences designs that control units are not comparable to treatment units due to unobserved selection.

The primary concern with this identification strategy is that the timing of a first degree murder sentencing may be correlated with a broader rise in the severity of crime biasing my estimates upwards. All specifications I estimate control for felony class which prevents changes in the statutorily relevant feature of a charge from driving my results since felony class determines sentencing guidelines. However, it may still be the case that the timing of a first degree murder is correlated with cases becoming unobservably more severe within each felony class. Since the control group is composed of sentences that are randomly assigned and taken from the same point in calendar time, trends in control sentencing decisions should capture unobservable changes in the severity of cases. Finally, it may be the case

that treated judges postpone more severe cases until after a first degree murder. I test for evidence of sorting on observables, and I can rule out large changes around event time.

This paper sits at the intersection of two large interdisciplinary literatures. A literature that spans economics, political science, and empirical legal studies considers the factors that influence judicial decision making. Much of this literature documents the importance of judge characteristics, defendant characteristics, and institutional features.² This paper relates most closely to a newer strand of the literature which tests whether normatively irrelevant factors influence judicial decisions. Hunger (Danziger et al., 2011), sleep deprivation (Cho et al., 2016), newspaper coverage (Lim et al., 2015; Philippe and Ouss, 2018), football results (Eren and Mocan, 2018) temperature (Heyes and Saberian, 2019) and terrorist attacks (Shayo and Zussman, 2011; Brodeur and Wright, 2019; McConnell and Rasul, 2021; Asadi, 2022) have all been linked to judicial decisions, though some of these findings are contested (Weinshall-Margel and Shapard, 2011; Spamann, 2018, 2022).

A nearly century old literature³ in psychology and economics theorizes and estimates biases that arise in the context of sequential decision making (Beebe-Center, 1929; Sherif et al., 1958; Tversky and Kahneman, 1974). This paper joins a recent strand documenting bias in the field. Evidence of anchoring in the field has been found in art auctions, house sales, sports betting, game shows and procurement auctions (Beggs and Graddy, 2009; Bucchianeri and Minson, 2013; McAlvanah and Moul, 2013; Jetter and Walker, 2017; Ferraro et al., 2022). Evidence of contrast effects in the field has been found in commuting patterns, speed dating and financial markets (Simonsohn, 2006; Bhargava and Fisman, 2014; Hartzmark and Shue, 2018).

I am most closely related to a small literature investigating the role of behavioral biases in judicial decision making in the field. Chen et al. (2016) provides evidence of the gambler’s fallacy in sequential decision making for baseball umpires, loan officers, and asylum judges. Using data from the Pennsylvania courts, Leibovitch (2016) argues that judges make decisions relative to the historical severity of their caseload.

I make two contributions. I am the first to document an anchoring-like effect in judicial decision making in the field. The scarring effect I find is large, persistent and causes ongoing distortions in sentencing. This effect is surprising given previous findings in the literature. I find a directionally opposite effect to the lab experiment that most closely matches my setting (Pepitone and DiNubile, 1976). Additionally, the sequential contrast effects mechanism documented in Chen et al. (2016), the most closely related study to this paper, would predict that sentencing a severe case would be followed by more lenient sentencing.

My second contribution is to provide empirical evidence of the role that salience plays as a mediating factor in behavioral biases. While salience has been proposed as a theoretical mechanism underlying behavioral biases such as anchoring and contrast effects (Bordalo et al., 2015), my heterogeneity results provide evidence from the field showing that the size

²Harris and Sen (2019) provides a recent review emphasizing the role of a judge’s characteristics such as race, gender, and political ideology. Defendant characteristics such as race, gender and education have been shown to influence incarceration rates and sentence lengths (Mustard, 2001; Abrams et al., 2012; Rehavi and Starr, 2014). A few papers emphasize the relationship between judge and defendant characteristics (Shayo and Zussman, 2011; Depew et al., 2017). A smaller literature in law and economics highlights the role of institutional features with a particular emphasis on caseload (Engel and Weinshall, 2020; Shumway and Wilson, 2022).

³See Furnham and Boo (2011) for a literature review.

of a behavioral bias varies systematically with a measure of salience.

The remainder of the paper proceeds as follows. Section 1 provides institutional details of the Cook County judicial system. Section 2 outlines the econometric framework and identification strategy of the paper. Section 3 presents the main results and discusses limitations.

1 Institutional Background and Data

1.1 Advantages of the Cook County Setting

Cook County, Illinois is the second largest county in America by population, and houses the second largest prosecutor’s office. I choose to study Cook County for two reasons.

First, Cook County has particularly fine grained sentencing data. In March 2018, State’s Attorney Kim Foxx began releasing data on the county’s handling of felony cases with historical records going back to 2011.⁴ Crucially, the felony sentencing records include judge names, defendant identifiers, charge class, sentencing dates and sentence lengths.

Second, Cook County has two helpful institutional features from an identification perspective. First, according to a conversation I had with an attorney at the Cook County Public Defender’s office, felony sentencing hearings are typically scheduled at least a few weeks in advance and exact dates depend on mutual agreement in the schedules of the judge, the prosecutor and the defense attorney which makes precise sorting of cases around the date of a first degree murder sentencing hearing unlikely. Second, felony cases in Cook County are randomly assigned conditional on courthouse and calendar time.⁵ Random assignment allows me to construct a group of control sentencing decisions which account for unobservable changes in case severity over calendar time.

1.2 Felony Classes in Cook County

There are six felony classes in Cook County. In descending order of severity, they are Class M, X, 1, 2, 3, and 4. The nature of crimes across felony classes varies widely. Examples of Class 4 felonies include domestic battery, obstruction of justice, and theft of less than \$300. Examples of Class X felonies include armed robbery, sexual assault, and home invasion.

Sentencing guidelines are pegged to felony class. For example, Class 4 felonies carry sentences between 1 and 3 years while Class 1 felonies carry sentences between 5 and 15 years. Judges have additional discretion in applying aggravating factors such as the defendant’s criminal history or motivations which may double the maximum sentence length.

The Class M felony charge is reserved for first degree murders which are murders that are intentional and premeditated. Class M felonies carry a sentence of 20-60 years, though defendants can face a life sentence if there are aggravating circumstances.

⁴To download the data, visit <https://datacatalog.cookcountyil.gov/Courts/Sentencing/tg8v-tm6u>.

⁵According to Loeffler (2013), case assignment is handled by the presiding justice who uses a program called “the randomizer” to assign new cases to available judges in view of the Cook County State’s Attorneys Office and Cook County Public Defender’s Office. However, if a defendant has another matter in front of a judge then the case goes to that judge. Additionally, some cases are siphoned off to specialized drug and mental health dockets. Any remaining cases are then randomly assigned to one of 30 general assignment judges.

1.3 Defining Prison Sentence Length

My primary outcome variable is prison sentence length. I standardize the length of all prison sentences to days. Out of the set of defendants initially charged with a Class 1-X felony who are ultimately sentenced, 44% of sentences are not prison sentences (nearly all non-prison sentences are probation sentences).⁶ I assign a length of 0 to non-prison sentences.

I cap sentence length at 75 years and assign life sentences this length. This cap affects a small number of observations (20 sentences), and is intended to reflect the fact that years beyond this point are very unlikely to be served.

One drawback of this measure is that I observe days sentenced, not days served. These outcomes may be different due to credit for time served or good behavior, but I cannot account for this difference.

1.4 Sample Restrictions

Since treatment is defined using Class M felony sentences, I restrict attention to defendants initially charged with Class 1-X felonies.

Following Abrams et al. (2012), I define an observation as a judge-date-defendant tuple, keeping the longest sentence in cases where there are two sentences handed down by the same judge to the same defendant on a given day. The reason to do this is that such sentences are unlikely to be independent. I do not observe whether multiple sentences are concurrent or consecutive so I cannot combine sentences. This decision results in dropping about 10% of sentences.

Judges are divided into circuit judges and associate judges. Circuit judges are elected by the public to 6 year terms (first in a competitive election, and subsequently in retention elections). Circuit judges collectively appoint associate judges to 4 year terms. Circuit judges may judge all cases but associate judges must be authorized by the Supreme Court to hear felonies. I hand collect information on whether judges are circuit judges, and restrict attention to circuit judges for my primary results to preserve the guarantee of randomization of cases to judges.

After imposing these sample restrictions, I'm left with a sample of 95,821 sentencing decisions made by 45 judges from March 2011 to September 2022.

2 Econometric Framework

2.1 Identification Challenges

The goal of my identification strategy is to estimate how many additional days the typical defendant faces in prison if they are sentenced by a judge who has recently presided over a first degree murder. Following potential outcomes notation, denote the sentence length judge j would hand down to defendant d as Y_{jd}^0 if the judge had not recently sentenced a

⁶I define sentences at the Illinois Department of Corrections, Cook County Department of Corrections, Cook County Boot Camp, Cook County Impact Incarceration Program, and Juvenile IDOC as prison sentences. The vast majority of prison sentences are at Illinois Department of Corrections and Cook County Department of Corrections.

first degree murder and as Y_{jd}^1 if the judge had recently sentenced a first degree murder. The parameter of interest is the average treatment effect $\mathbb{E}[Y_{jd}^1 - Y_{jd}^0]$.

Estimating this effect presents a non-standard causal inference challenge because each treatment unit can be treated multiple times and time between treatments can vary. Observing multiple treatments is potentially helpful because it increases the sample of events, but it simultaneously complicates inference because treatment from an earlier event could pollute the control period of a later event, biasing the estimated coefficient downwards. This bias is particularly concerning if the murder sentencing hearings happen in quick succession: for example, if a judge sentences two first degree murders in a month, we might be worried that the sentences handed down between the murders are not a good control for the sentences handed down after the second murder. This problem is further complicated by the fact that the duration of the treatment effect is not known *ex ante*. In the extreme case, if treatment were to last forever (i.e. sentencing a murder changes a judge’s judicial philosophy permanently), the event study estimate of the treatment effect for 2nd and later murders would be zero.

Additionally, multiple treatments complicate the application of the rapidly growing literature on using two-way fixed effects models to estimate staggered difference-in-differences designs (Callaway and Sant’Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021). This is because the theoretical work in this literature assumes that treatment is an absorbing state (i.e. once treated, a units stays treated). I borrow some insights from this literature: I estimate effects in fixed length event time windows instead of as a panel in order to avoid differential weighting of events based on the length of the available pre and post-treatment periods. However, one qualitative finding in this literature is that using already-treated units as controls for future treated-units (a mechanical result of estimating a standard two-way fixed effects model) can be problematic. I do not estimate a two-way fixed effect regression, but my design does use previously treated units as control units. I think this is a reasonable decision in this setting given the fade-out nature of the treatment effect that I find.

2.2 Identification Strategy

The main effect is estimated using a difference-in-differences design. However, I depart from the standard difference-in-differences framework in order to leverage multiple treatments and random assignment of cases for identification. Rather than consider treatment and control judges, I consider treatment and control judging spells. I define a spell as a set of sentencing decisions handed down by a judge in a window around the date of a Class M felony sentencing decision.

I construct the primary dataset using the following procedure. First, I define an event as a Class M felony sentence with a 100 day clean pre-period (i.e. the sentencing judge has not sentenced another Class M felony in the previous 100 days). Then, for each event, I look at sentences in a 200 day window around the event by the sentencing judge. These sentences make up the treatment spells. Finally, for each event, I identify a set of control judges who work in the same courtroom as the treated judge and who do not sentence a Class M felony in the 100 day pre-period or the 200 day post-period. The control spells are the set of sentences handed down in a 200 day window around the event for each control judge. Because this dataset is constructed event-by-event, the same judge’s decisions can appear as both treatment spells and control spells, though these will be different sentences

handed down at different points in calendar time.

Choosing the length of the clean pre-period introduces a bias-variance trade-off because longer clean pre-periods throw out additional events reducing the precision of estimates, while shorter clean pre-periods make contamination of the pre-period by previous treatment more likely. Similarly, the length of the clean post-period for control spells introduces a trade-off because a longer clean post-period throws out control judges reducing the precision of estimates but allows results to be traced out over a longer time horizon. The choice of a 100 day clean pre-period and a 200 day clean post-period is arbitrary, and robustness to these assumptions is checked in the online appendix.

The treated spells form an event study dataset. Mean pre-treatment sentencing length is an estimate of Y_{jd}^0 , and mean post-treatment sentence length is an estimate of Y_{jd}^1 . Differencing these means gives an event study estimate of the average treatment effect.

A concern with the event study estimate is that the timing of treatment may be correlated with changes in underlying case severity. The control judge spells allow me to estimate changes in unobservable within-class variation in case severity. Because cases are randomly assigned to judges conditional on courtroom and calendar time, control cases are of the same average severity as treatment cases in expectation.

The average sentence length of control judge spells also provides an estimate of Y_{jd}^0 which could be differenced from the mean sentence length of treatment judge spells in the post-treatment period to estimate the average treatment effect. Given that pre-period treatment spells and control spells can both be used to estimate Y_{jd}^0 , it's reassuring that average sentence length for these two groups is similar not only in trends but also in levels, particularly after controlling for observables.⁷

Using “Pre” and “Post” superscripts to denote the control and treatment time periods and using “Treated” and “Control” subscripts to denote the treated and control judge spells, I report difference-in-differences estimates of the form

$$(\hat{Y}_{\text{Treated}}^{\text{Post}} - \hat{Y}_{\text{Treated}}^{\text{Pre}}) - (\hat{Y}_{\text{Control}}^{\text{Post}} - \hat{Y}_{\text{Control}}^{\text{Pre}}) \quad (1)$$

where \hat{Y} is the mean residual sentence length after controlling for observables in the given sample in the given period.

I compute standard errors assuming that sentence length is independent and identically distributed after conditioning on observables and treatment status. This assumption is justified by the fact that cases are randomly assigned to judges conditional on courthouse (Loeffler, 2013). Abadie et al. (2017) argue that when treatment status is randomly assigned clustering standard errors is excessively conservative.⁸

⁷The difference between mean sentence lengths of treated and control judge spells in the pre-period is 32 days without controls and 7 days with controls which are approximately 4% and 1% of mean sentence length. See Figure 1.

⁸Quoting Abadie et al. (2017), “If one has a random sample of units from a large population with randomized treatment assignment at the unit level, there is no reason to cluster the standard errors of the least squares estimator. ... Similarly, in a judge-leniency design—where defendants are randomly assigned to judges—standard errors should not be clustered at the level of the judge.”

2.3 Estimating Residual Sentence Length

In principle, the difference-in-differences strategy does not require controls. Random assignment as well as the relatively short event time window mean that control cases should be comparable to treatment cases. Nevertheless, controlling for observables increases the precision of my estimates. Moreover, controlling for observables prevents any changes in the composition of cases that are correlated with observables from driving my results.

I account for observables by estimating the following poisson regression.

$$y_{dje} = \alpha + \beta X_d + FE_j + FE_e + \epsilon \quad (2)$$

In this equation, d indexes the defendant, j indexes the judge, and e indexes the event. Y_{dje} is sentence length measured in days. X_d is a set of defendant-level characteristics. The characteristics are race (black, hispanic, white, other), gender (male, female), defendant age and initial felony charge class. Defendant age is included as a set of 5 year age bin fixed effects to allow for nonlinearities in how age affects sentencing. A small number of observations are missing age; these observations are assigned a missing age fixed effect. I choose to control for initial felony charge class because previous work argues that final charge class is an endogenous outcome of the interaction between the judge and the prosecutor and so could be affected by treatment. FE_j is a fixed effect for the sentencing judge j . FE_e is a fixed effect for event e . Since each event identifies a treatment judge spell and a set of control judge spells where all judges come from the same courtroom, FE_e can be thought of as a courtroom-calendar time fixed effect.

I choose a poisson regression because sentence length is strictly positive and has a right skewed distribution.

I define residual sentence length as the difference between observed sentence length and predicted sentence length.

The choice to estimate these effects in a two-step procedure (controlling for observables and then taking a difference in means) is essentially equivalent to estimating a single poisson regression specification which includes *Event Time* \times *Treated* fixed effects and reporting the coefficients on the treatment dummies. One advantage of the two-step estimation procedure is that I can report estimates of the standard error for every event time period (a single-step procedure would required dropping one time period fixed effect to avoid perfect collinearity).

3 Results

3.1 Difference-in-Differences Estimates and Heterogeneous Treatment Effects

In order to provide a top line statistic summarizing each treatment effect, I compute a difference-in-differences estimate comparing a pre-period (defined as event time < -10) to a treatment period (defined as event time between -10 and 20) where event time is measured in days relative to the date of a first degree murder sentencing. The choice of this treatment period is arbitrary and driven by inspection of the figures. The figures report coefficients

for every 10 day event time bin which provides a fine-grained view of how treatment effects evolve dynamically.

Figure 1 visualizes the primary difference-in-differences design. Panel A plots mean sentence length (without controls) for treated and control judge spells. In the pre-period, treated and control averages are similar in level and trend. At event time -10 , sentence length in the treated group spikes and this treatment effect persists until at least event time 40. This graph provides some reassurance that my results are not an artifact of specification choices as the main treatment effect is visible by taking means. The remaining treatment effect graphs plot residual sentence length after controlling for race, gender, age, felony class, judge and event fixed effects (see Equation 2).

Panel B of Figure 1 visualizes the main treatment effect by plotting the difference in mean residual sentence length between treatment and control judge spells. After controlling for observables, treatment and control judge spells are less than 7 days apart in the pre-period. In the treatment period, residual sentence length spikes by an average of 88 days ($p < 0.001$). The effect size is large relative to the length of typical sentences: 88 days is 24% of the median (365 days) and 12% of the mean length (756 days) of Class 1-X felony sentences. To test whether this effect persists over a longer time horizon, I compute a difference-in-differences estimate comparing sentencing after event time 20 to sentencing in the pre-period: sentence length remains slightly elevated with the average sentence being 28 days longer ($p = 0.0015$) in the 180 day post-period.

The main effect could be driven by sentencing defendants to prison who counterfactually would have gotten probation (the extensive margin), or it could be driven by making the sentences of defendants who would have gone to prison longer (the intensive margin). Both margins are plausible ex ante as 44% of defendants are sentenced to probation rather than prison.

Panel A of Figure 2 shows that the extensive margin is small. Residual prison probability is computed analogously to residual sentence length: I predict an indicator for whether a defendant was sentenced to prison using race, age, gender, felony class, judge and event fixed effects (See Equation 2). Residual prison probability is the difference between actual and predicted prison probability. Panel A estimates the main difference-in-differences strategy replacing residual sentence length with residual prison probability. A positive treatment effect would indicate that judging a first degree murder cause a judge to sentence some defendants to positive prison time who would have been sentenced to probation. I find a null effect: the 95% confidence interval of the estimate of the effect of a first degree murder on the probability of being sentenced to prison time is $[-1.58\%, 1.95\%]$.

Panel B of Figure 2 shows that the intensive margin effect is large. This figure estimates a difference-in-differences design restricting attention to prison sentences. This design identifies the intensive margin effect under the assumption that sentences in the post-period with observed positive sentence length would have had positive sentence length in the absence of treatment (i.e. the extensive margin effect is zero). If some sentences are moved from 0 to positive sentence length by treatment, then this difference-in-differences estimate is likely to understate the true effect size as defendants on the margin of positive sentence length are likely to be given low sentences. The difference-in-differences estimate of the treatment effect is 140 days ($p < 0.001$). This effect size is qualitatively similar to the main effect in in percentage terms: it is 10% of the mean (1350 days) sentence length in the sample of

prison sentences. As an accounting exercise, increasing the sentence length of the 56% of defendants who face prison sentences by an average of 140 days implies a treatment effect of 79 days over the whole sample, so the intensive margin effect accounts for 90% of the observed treatment effect.

In order to get at mechanisms, I compute heterogeneous treatment effects by estimating triple difference designs comparing defendants who are similar to and different from the murderer along the dimensions of race and charge severity. Panel A of Figure 3 compares same-race defendants to different-race defendants. The triple difference estimate is that the scarring effect is 101 days longer for same-race defendants than different-race defendants ($p = 0.047$). This result is not a mechanical consequence of differential sentencing by race for two reasons. First, residual sentence length controls for race. Second, the difference-in-differences design compares same-race defendants in the pre-period to the treatment period.

Panel B of Figure 3 splits defendants into high and low class felonies. I define high class felonies as Class X and 1 felonies and low class felonies as Class 2, 3 and 4 felonies. This triple-difference has clean pre-trends and a very sharp effect: the scarring effect is 283 days longer for high class felonies compared to low class felonies ($p = 0.021$). A note of caution here is that high class felonies differ from low class felonies in two ways: high class felonies are more similar to the murder event because they're more severe, but they also offer judges more discretion as the sentencing guidelines for these crimes are much looser. Nevertheless, the clean pre-period trends show that result is not a mechanical consequence of additional discretion.

Finally, I show that the scarring effect exacerbates existing inequalities because it differentially affects Black defendants. Figure 4 plots a triple difference between Black and non-Black defendants. Panel A plots treatment effects for Black and non-Black defendants separately. Non-Black pools White, Hispanic, and Asian defendants who together make up 29% of defendants. The Black coefficients in Panel A show a sharp jump in sentencing at event time -10 which fades over the next 60 days consistent with my other results. However, while a difference-in-differences estimate of the coefficient on Black defendants is significant, the triple-difference estimate is not as treatment effects for non-Black defendants are estimated imprecisely. The triple difference point estimate is 50 days ($p = 0.36$). Panel B shows that differential sentencing persists for about 60 days and a triple-difference estimate over this longer period is similar in magnitude (82 days) and significant ($p = 0.01$). This effect represents 60% of the 131 day sentencing penalty that Black defendants face for the same felony class conditional on observables.⁹ This estimate should be interpreted with caution as it's significance is sensitive to how treatment length is defined.

One reason these results are concerning is that judges sentence first degree murders regularly, so scarring distorts sentencing on an ongoing basis. As a back-of-the-envelope calculation, 16% of defendants are sentenced in the 10 days before and 20 days after a first degree murder. However, some of these defendants are much more affected than others. Scarring is entirely explained by defendants who face prison sentences, and 9% of defendants

⁹I estimate a sentencing penalty of 131 days between Black and Non-Black defendants in Cook County. This comes from regressing sentence length on a dummy for Black as well fixed effects for gender, age (5 year bins), initial charge class and sentencing judge. This specification does not distinguish between two potential explanations for a sentencing gap. The average Black defendants within each class may have committed an unobservably more severe crime resulting in a longer sentence. Alternatively, judges may be prejudiced against Black defendants, handing down longer for crimes of the same severity. Decomposing these two explanations is beyond the scope of this paper.

are sentenced to prison sentences in the 10 days before and 20 days after a first degree murder in my 11 year sample. This statistic should be thought of as a rough approximation: it could be made smaller by restricting to same-race, high-class, or Black defendants, or larger by defining treatment on a longer time interval (the main difference-in-difference estimate is significant when estimated out to event time 40, for example).

3.2 Threats to Identification

The primary concern with the identification strategy is the severity of cases around the time of a first degree murder sentencing may not be random. For example, crime spikes in the summer in Cook County. If judges are more likely to be treated at the start of summer and summer crimes are more severe, this would bias the treatment effect estimate upwards.¹⁰

I argue that this concern is unlikely to drive my results for a few reasons. First, I control for initial felony class. All sentencing guidelines in Cook County are pegged to felony class, so results cannot be driven by changes in the composition of the statutorily relevant feature of a charge.

Second, control cases are randomly assigned and taken from the same point in calendar time as treatment cases. Then, to the extent that there are unobservable changes in the severity of cases within each felony class, these changes should be picked up by trends in control group sentencing. The fact that sentence length in the control group is relatively flat through event time in Panel A of Figure 1 is reassuring.

In addition to trends in crime severity, it may also be the case that treated judges are sorting cases around event time. If the burden of judging a difficult first degree murder case cause a judge to push off sentencing of other severe cases, it would bias my estimates upwards. Moreover, this sorting would not be picked up by the control cases since judges without a first degree murder would have no incentive to sort.

Institutional features push against this kind of sorting: felony sentencing hearings in Cook County are typically scheduled at least a few weeks in advance by mutual agreement of the defense attorney, the prosecutor, and the judge.

Figure 5 directly tests for patterns in the sorting of cases. Each panel in Figure 5 is estimated analogously to the primary difference-in-differences specification replacing *Residual Sentence Length* for a defendant characteristic. Since *Black*, *Female*, and *High Class* are all binary variables, I compute residuals by estimating logit models rather than a poisson model. Each coefficient gives the mean difference in the predicted probability of a defendant characteristic between treated and control judges. If the 95% confidence interval includes 0 then I cannot reject that the samples are balanced along that demographic characteristic after controlling for observables. This is the largely the case across all 4 panels.

However, I note two points of caution: first, there appears to be an increase in variance in the age of defendants around the time of treatment though defendant age does not move sharply on average. More worryingly, the share of high class cases does appear to rise somewhat in the treatment period though I cannot reject a difference of 0 in most 10 day bins. Charge class is controlled for in residual sentence length so this does not directly drive

¹⁰This story is a bit too simple because there is a long lag between when a crime occurs and when a crime is sentenced, but we could imagine other reasons for secular trends in the intensity of crime.

of my results, but it could be suggestive of other unobserved trends in case severity that I cannot control for.

3.3 Interpretation of Treatment Variation and Limitations

A limitation of the Cook County data is that I do not observe detailed information on the timing of events within each court case. This complicates the interpretation of treatment because the process of judging a court case consists of a series of tasks (e.g. hearing arguments, viewing evidence, holding a trial, sentencing) which could each be influential. My treatment variation, the timing of the sentencing hearing, is a noisy measurement for when many tasks related to a defendant will occur in a short span of time.

Given this limitation, treatment is likely a combination of the events that occur in close proximity to the sentencing hearing. This is my explanation for why sentencing spikes in the 10 days before the sentencing date: other case-related events such as a pre-sentencing investigation or sentencing hearings are likely to take place shortly before the date of sentencing.

The murder cases I study often sit on a judge's docket for a year or more during which time a judge holds occasional hearings and both sides gather evidence. I observe the time when each Class M case enters the prosecutor's office, but the extended lag between intake and sentencing makes producing credible estimates using this variation difficult because a judge may not meaningfully interact with a case on their docket while some legal processes (e.g. discovery) occur. Given how long cases stay on the docket, the treatment effect should be thought of as capturing something over and above any effects of having a Class M felony on the docket which is likely to be the case for all of the pre-treatment period.

A related limitation of my strategy is that it can only be used to identify the effects of crimes that occur rarely. Lower class felonies all occur frequently so I cannot construct clean pre-periods. However, the question of whether or not 1-X felony cases have scarring effects on nearby cases is obviously of interest to the broader question of behavioral biases in judicial sentencing.

One specification that could have avoided some of the arbitrary definition of clean pre-periods would have been to estimate effects separately for a judge's 1st murder. This is not possible in my setting due to a data limitation. The vast majority of judges are already judging before the start of my sample: it's typical for judges to serve multiple 6 year terms, and I only observe 11 years of data. Because of this, I cannot be sure how many murders each judge has seen for the majority of judges who start judging before the start of my sample.

3.4 Anchoring and Other Potential Mechanisms

Scarring is an anchoring-like effect in that a severe case results in over-sentencing. However, given that scarring picks up before the sentencing date, it's likely that the mechanism encompasses more than just the large numeric value of the sentence. It could be that the severe case temporarily influences a judge's perceptions of the prevalence of violent crime or the benefits to taking potential offenders off the streets.

However, note that scarring is inconsistent with learning or changes to preferences because the effects I find sharply fade after 20 days past the sentencing date. This means the

effect cannot be driven by updates to priors unless these updates also sharply fade which is not typically what is meant by learning.

An intuitive alternative explanation that I cannot rule out is an emotionally-driven story: judging a violent murder could leave a judge angry or scared resulting in over-sentencing. While this is plausible, it's worth noting that felony sentencing judges regularly sentence other violent crimes and the average judge is quite experienced, so these emotional effects would have to exist over and above a baseline emotional state. Additionally, the 30 day length of the treatment effect is inconsistent with typical descriptions of emotional states which are usually thought to last for shorter periods of time.

4 Conclusion

I document a sizable and persistent behavioral bias in felony sentencing in Cook County. Identifying and correcting biases in this setting is important given the stakes: unjustly harsh sentences rob defendants of freedom, and equal justice under the law is a foundational legal principle.

Judges sentence defendants to an average of 88 additional days in prison around the date of a first degree murder sentencing. Since judges repeatedly sentence first degree murders, scarring affects 9% of defendants on an ongoing basis. Effects are largest for defendants who are the same race as the murderer and defendants who face high class felony charges. Coupled with the spike-and-fade pattern of the main effect, these results suggest that defendants who call to mind a recent murderer face harsher sentences. These differential treatment effects exacerbate existing racial disparities temporarily increasing the Black sentencing penalty by 60% for around two months.

As a piece of law and economics, these results are important because they shed light on a novel and sizable violation of the Equal Justice principle: defendants are sentenced differently due to the luck of timing. From a behavioral perspective, this is the first paper to demonstrate an anchoring-like effect in the judicial context in the field.

The scarring effect emphasizes how psychological biases interact with the institutional design of the court system to produce injustice. This is important because the way that sentencing decisions are scheduled is under the purview of judges and policymakers. Institutional reforms could plausibly mitigate scarring: it would be feasible to institute tighter sentencing guidelines, to have judges specialize in sentencing high class felonies, or to mandate cooling off periods after judge handles a particularly challenging case, though a cost-benefit analysis of such policies is beyond the scope of this paper.

References

- Alberto Abadie, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research, 2017.
- David S Abrams, Marianne Bertrand, and Sendhil Mullainathan. Do judges vary in their treatment of race? *The Journal of Legal Studies*, 41(2):347–383, 2012.
- Masoud Asadi. Essays on intersection of economics and judicial decision-making. *Working Paper*, 2022.
- John G Beebe-Center. The law of affective equilibrium. *The American Journal of Psychology*, pages 54–69, 1929.
- Alan Beggs and Kathryn Graddy. Anchoring effects: Evidence from art auctions. *American Economic Review*, 99(3):1027–39, 2009.
- Saurabh Bhargava and Ray Fisman. Contrast effects in sequential decisions: Evidence from speed dating. *Review of Economics and Statistics*, 96(3):444–457, 2014.
- Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience theory of judicial decisions. *The Journal of Legal Studies*, 44(S1):S7–S33, 2015.
- Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience. *Annual Review of Economics*, 14:521–544, 2022.
- Abel Brodeur and Taylor Wright. Terrorism, immigration and asylum approval. *Journal of Economic Behavior & Organization*, 168:119–131, 2019.
- Grace W Buccianeri and Julia A Minson. A homeowner’s dilemma: Anchoring in residential real estate transactions. *Journal of Economic Behavior & Organization*, 89:76–92, 2013.
- Brantly Callaway and Pedro HC Sant’Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.
- Daniel L Chen, Tobias J Moskowitz, and Kelly Shue. Decision making under the gambler’s fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *The Quarterly Journal of Economics*, 131(3):1181–1242, 2016.
- Kyoungmin Cho, Christopher M Barnes, and Cristiano L Guanara. Sleepy punishers are harsh punishers: Daylight saving time and legal sentences. *Psychological science*, 2016.
- Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17):6889–6892, 2011.
- Briggs Depew, Ozkan Eren, and Naci Mocan. Judges, juveniles, and in-group bias. *The Journal of Law and Economics*, 60(2):209–239, 2017.
- Christoph Engel and Keren Weinshall. Manna from heaven for judges: Judges’ reaction to a quasi-random reduction in caseload. *Journal of Empirical Legal Studies*, 17(4):722–751, 2020.
- Birte Engh and Thomas Mussweiler. Sentencing under uncertainty: Anchoring effects in the courtroom 1. *Journal of applied social psychology*, 31(7):1535–1551, 2001.

- Ozkan Eren and Naci Mocan. Emotional judges and unlucky juveniles. *American Economic Journal: Applied Economics*, 10(3):171–205, 2018.
- Paul J Ferraro, Kent D Messer, Pallavi Shukla, and Collin Weigel. Behavioral biases among producers: Experimental evidence of anchoring in procurement auctions. *The Review of Economics and Statistics*, pages 1–40, 2022.
- Adrian Furnham and Hua Chu Boo. A literature review of the anchoring effect. *The journal of socio-economics*, 40(1):35–42, 2011.
- Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.
- Chris Guthrie, Jeffrey J Rachlinski, and Andrew J Wistrich. Inside the judicial mind. *Cornell L. Rev.*, 86:777, 2000.
- Allison P Harris and Maya Sen. Bias and judging. *Annual Review of Political Science*, 22: 241–259, 2019.
- Samuel M Hartzmark and Kelly Shue. A tough act to follow: Contrast effects in financial markets. *The Journal of Finance*, 73(4):1567–1613, 2018.
- Anthony Heyes and Soodeh Saberian. Temperature and decisions: evidence from 207,000 court cases. *American Economic Journal: Applied Economics*, 11(2):238–65, 2019.
- Michael Jetter and Jay K Walker. Anchoring in financial decision-making: Evidence from jeopardy! *Journal of Economic Behavior & Organization*, 141:164–176, 2017.
- Adi Leibovitch. Relative judgments. *The Journal of Legal Studies*, 45(2):281–330, 2016.
- Claire SH Lim, James M Snyder Jr, and David Strömberg. The judge, the politician, and the press: newspaper coverage and criminal sentencing across electoral systems. *American Economic Journal: Applied Economics*, 7(4):103–35, 2015.
- Charles E Loeffler. Does imprisonment alter the life course? evidence on crime and employment from a natural experiment. *Criminology*, 51(1):137–166, 2013.
- Patrick McAlvanah and Charles C Moul. The house doesn’t always win: Evidence of anchoring among australian bookies. *Journal of Economic Behavior & Organization*, 90:87–99, 2013.
- Brendon McConnell and Imran Rasul. Contagious animosity in the field: Evidence from the federal criminal justice system. *Journal of Labor Economics*, 39(3):739–785, 2021.
- David B Mustard. Racial, ethnic, and gender disparities in sentencing: Evidence from the us federal courts. *The Journal of Law and Economics*, 44(1):285–314, 2001.
- Albert Pepitone and Mark DiNubile. Contrast effects in judgments of crime severity and the punishment of criminal violators. *Journal of Personality and Social Psychology*, 33(4):448, 1976.
- Arnaud Philippe and Aurélie Ouss. “no hatred or malice, fear or affection”: Media and sentencing. *Journal of Political Economy*, 126(5):2134–2178, 2018.
- M Marit Rehavi and Sonja B Starr. Racial disparity in federal criminal sentences. *Journal of Political Economy*, 122(6):1320–1354, 2014.

- Moses Shayo and Asaf Zussman. Judicial ingroup bias in the shadow of terrorism. *The Quarterly journal of economics*, 126(3):1447–1484, 2011.
- Muzafer Sherif, Daniel Taub, and Carl I Hovland. Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of experimental psychology*, 55(2):150, 1958.
- Clayson Shumway and Riley Wilson. Workplace disruptions, judge caseloads, and judge decisions: Evidence from ssa judicial corps retirements. *Journal of Public Economics*, 205:104573, 2022.
- Uri Simonsohn. New yorkers commute more everywhere: contrast effects in the field. *Review of Economics and Statistics*, 88(1):1–9, 2006.
- Holger Spamann. Are sleepy punishers really harsh punishers? comment on cho, barnes, and guanara (2017). *Psychological science*, 29(6):1006–1009, 2018.
- Holger Spamann. Comment on” temperature and decisions: Evidence from 207,000 court cases”. *American Economic Journal: Applied Economics*, 14(4):519–28, 2022.
- Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.
- Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131, 1974.
- Keren Weinshall-Margel and John Shapard. Overlooked factors in the analysis of parole decisions. *Proceedings of the National Academy of Sciences*, 108(42):E833–E833, 2011.

Figures

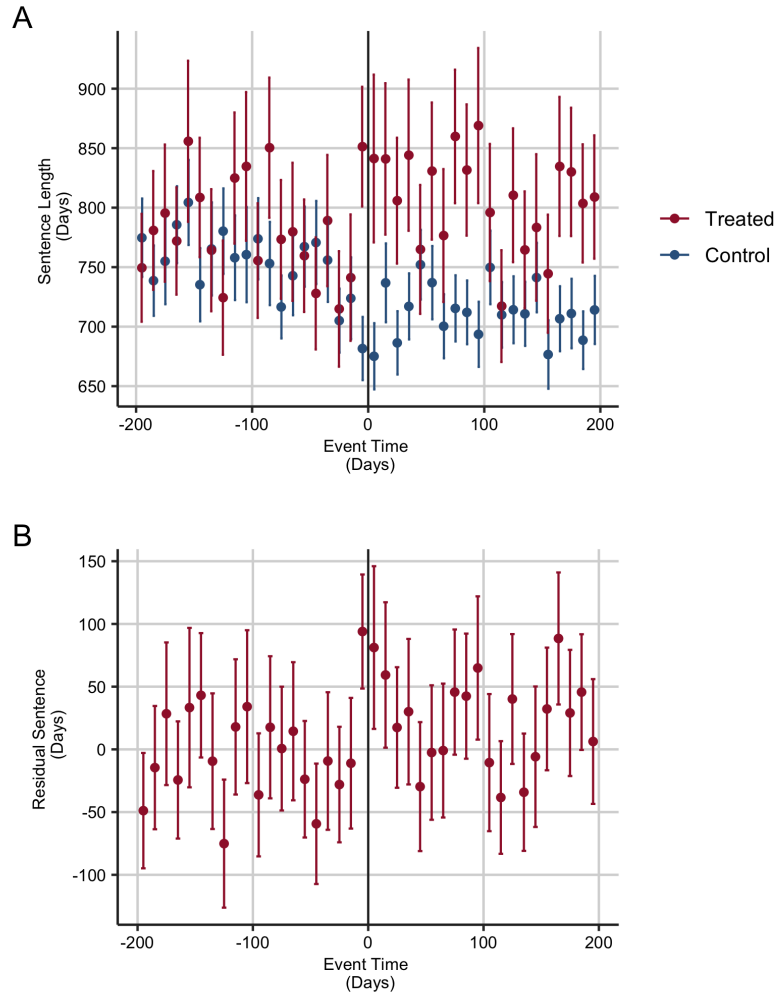


Figure 1: This figure visualizes a difference-in-differences design showing that prison sentence length spikes around the time of a first degree murder. I define an event as the sentencing date of a first degree murder by a judge who has not sentenced a first degree murder in the prior 100 days. A treated judge spell is the set of sentences in a 200 day window around the event handed down by the judge who sentenced the first degree murder. A control judge spell is the set of sentences in a 200 day window around the event handed down by a judge in the same courtroom who did not sentence a first degree murder in the 100 days prior and the 200 days post the event. Panel A plots mean sentence length for treated and control judge spells. Panel B plots the difference in mean residual sentence length between treated and control judge spells. Residual sentence length is the difference between actual sentence length and predicted sentence length after controlling for race, gender, age, felony class, judge and event fixed effects (see Equation 2). Each point represents 10 days of event time, and bars represent 95% confidence intervals.

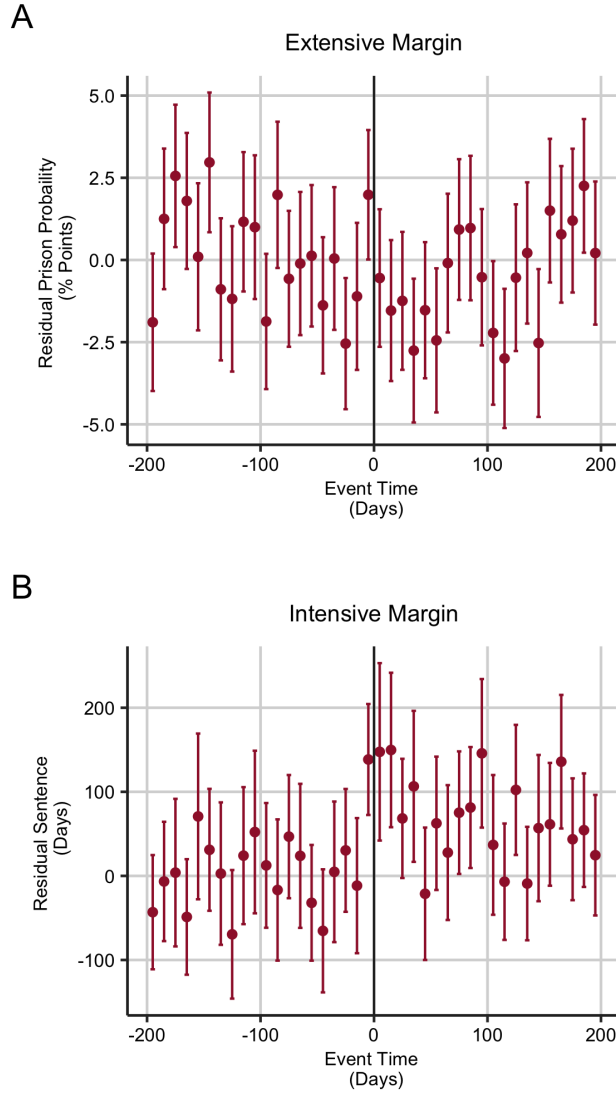


Figure 2: This figure decomposes the main treatment effect into intensive and extensive margins. The intensive margin quantifies how much treatment increases sentence length for defendants who would have gone to prison in the pre-period. The extensive margin quantifies how much treatment causes judges to sentence defendants to prison who otherwise would not have gone. Panel A visualizes the extensive margin effect by estimating the treatment effect on probability of a prison sentence. Prison is a variable that takes on 1 if sentence length is positive, and residual prison probability is difference between actual and predicted prison probability after controlling for race, gender, age, felony class, judge and event fixed effects (see Equation 2). Panel B visualizes the intensive margin effect by plotting the mean difference in residual sentence length between treated and control judge spells conditional on positive sentence length. Residual sentence length is the difference between actual sentence length and predicted sentence length after controlling for race, gender, age, felony class, judge and event fixed effects (see Equation 2). Each point represents 10 days of event time, and bars represent 95% confidence intervals.

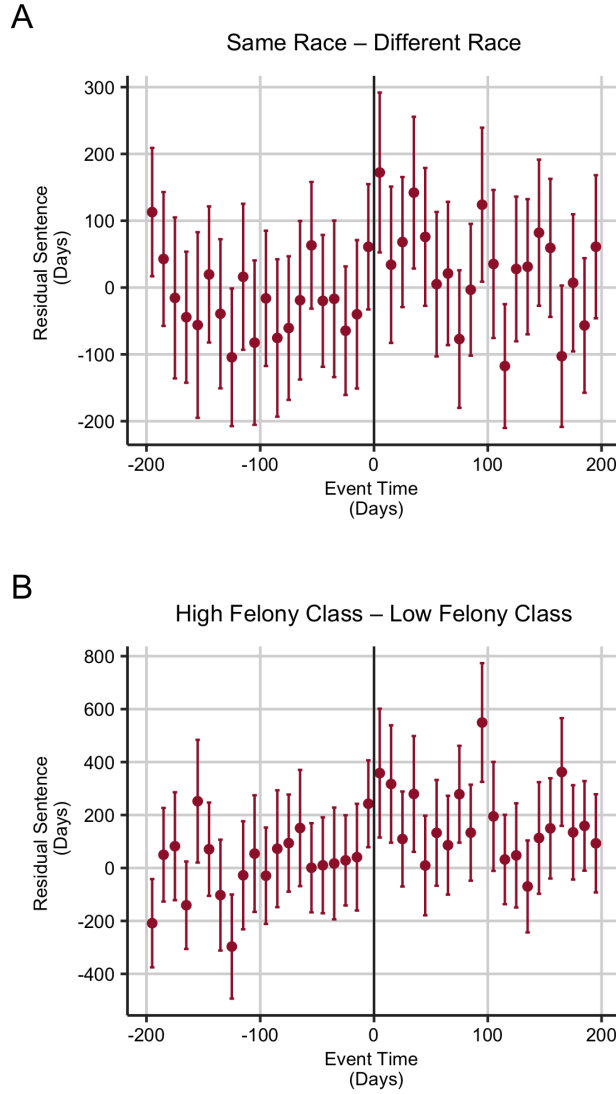


Figure 3: This figure visualizes a triple-difference design comparing treatment effects by similarity between the defendant and the murderer along the dimensions of race and felony class. The three differences are: (1) sentences that are handed down (before/after) a first degree murder sentence to (2) defendants who are (similar to/different from) the murderer by (3) judges who (did/did not) hand down the first degree murder sentence. Panel A differences treatment effects for defendants who are the same race as the murderer from treatment effects for defendants who are a different race than the murderer. Panel B differences treatment effects for defendants who face a high class felony charge (Class X or 1) from treatment effects for defendants who face a low class felony charge (Class 2, 3, or 4). Residual sentence length is the difference between actual sentence length and predicted sentence length after controlling for race, gender, age, felony class, judge and event fixed effects (see Equation 2). Each point represents 10 days of event time, and bars represent 95% confidence intervals.

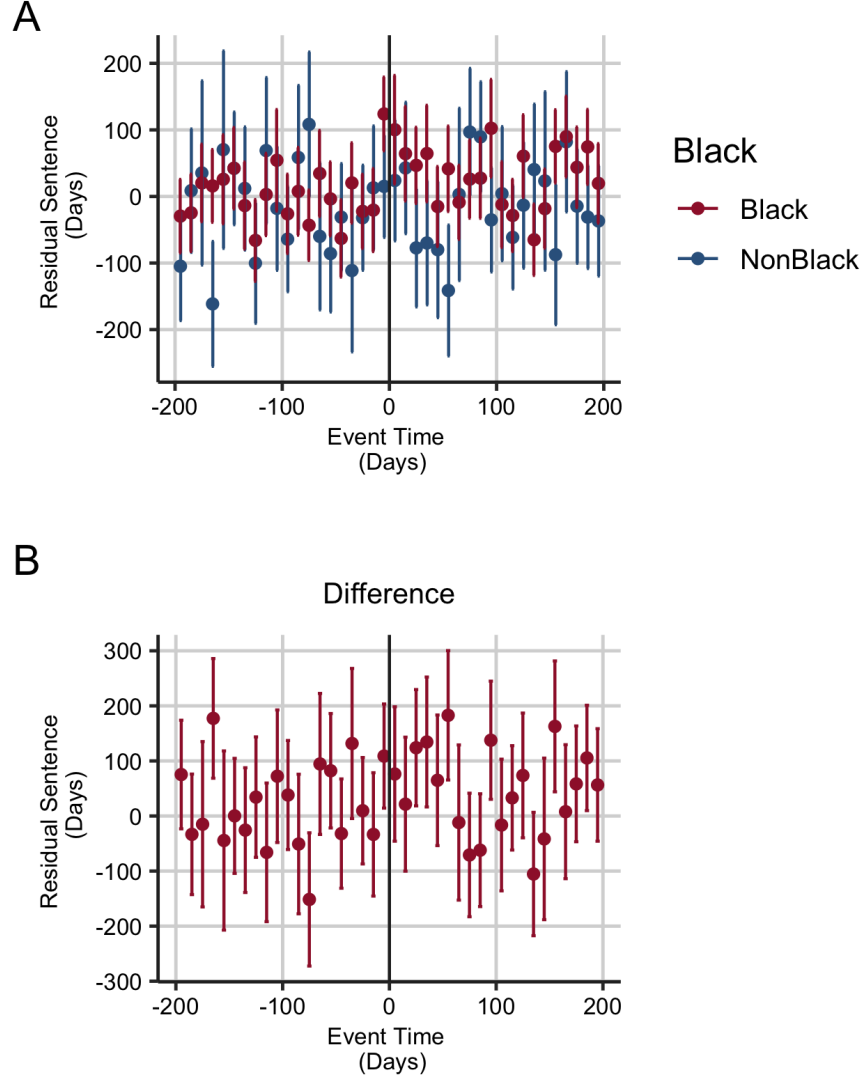


Figure 4: This figure visualizes a triple-difference design comparing treatment effects between Black and non-Black defendants. The three differences are: (1) sentences that are handed down (before/after) a first degree murder sentence to (2) defendants who are (Black/non-Black) by (3) judges who (did/did not) hand down the first degree murder sentence. Panel A visualizes treatment effects for each of the difference-in-difference designs: it plots the mean difference in residual sentence length between treated and control judge spells for Black defendants and non-Black defendants. Panel B plots the difference between these treatment effects. Residual sentence length is the difference between actual sentence length and predicted sentence length after controlling for race, gender, age, felony class, judge and event fixed effects (see Equation 2). Each point represents 10 days in event time, and bars represent 95% confidence intervals.

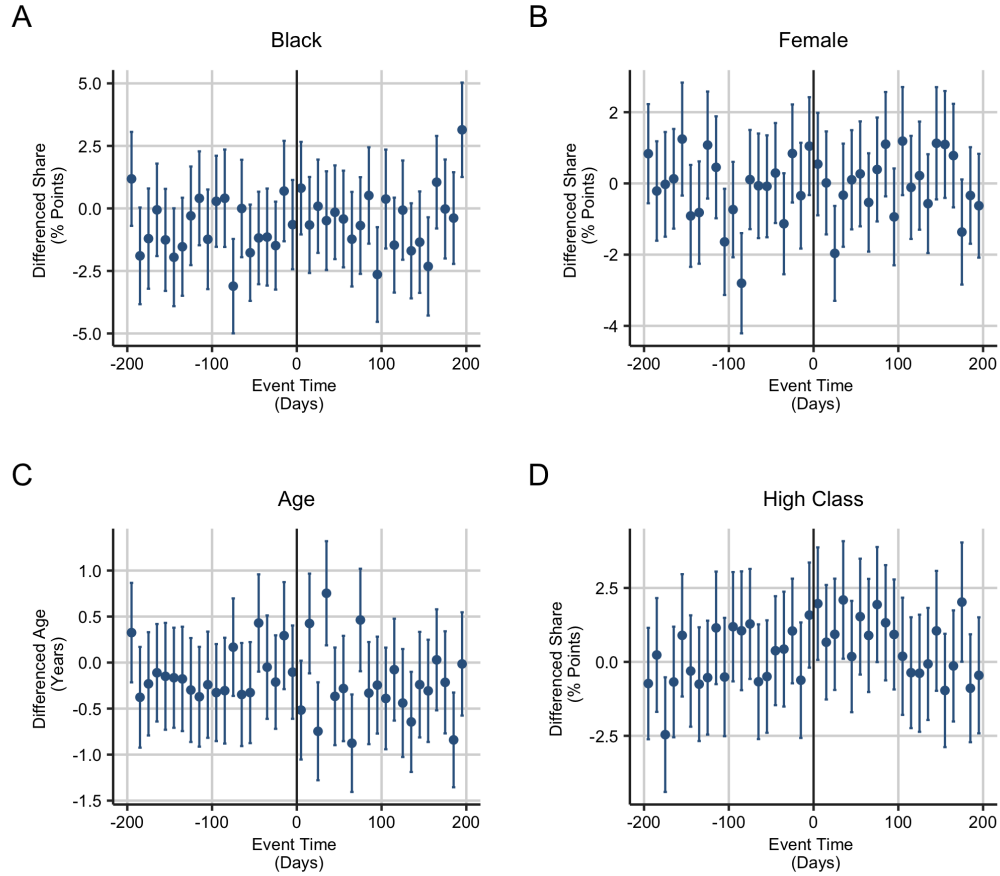


Figure 5: This figure plots tests for violations of random assignment of cases across event time. Each panel visualizes a difference-in-differences design equivalent to Panel B of Figure 1 replacing the primary outcome variable (residual sentence length) for another defendant characteristic. Residual defendant characteristics are computed by estimating Equation 2 replacing sentence length for the characteristic as the left hand side variable, and taking the difference between actual and predicted defendant characteristic. I estimate a logit regression rather than a poisson regression when the characteristic is binary. Each point represents 10 days in event time, and bars represent 95% confidence intervals.