

How Much do Mandatory Minimums Matter?*

Spencer Cooper

July 31, 2024

Abstract

I estimate the causal increase to sentence lengths created by mandatory minimum (MM) eligibility for federal drug crimes. I utilize a regression discontinuity design with extrapolation to decompose the impacts of MM eligibility into effects driven by the law and effects driven by legal actor decisions. I find that without manipulation, MM eligibility increases sentence length uniformly across case types by about 10 months (18%). However, cases that are manipulated to a higher MM eligibility receive an additional 7 to 11-month increase. I find that manipulation exclusively impacts minority defendants and is driven by prosecutor manipulation decisions. Using a unique dataset of US Attorneys, I show that minorities are also more likely to be manipulated to higher eligibility levels than White counterparts. Taken together, these results show that race gaps in MM sentencing are driven by prosecutor manipulation decisions.

***Affiliation:** University of Connecticut, Department of Economics. Email: spencer.cooper@uconn.edu.

Data Availability: Drug sentencing data used in this article is provided by the United States Sentencing Commission and is publicly available at www.ussc.gov/research/datafiles/commission-datafiles. United States Attorney data will be available online upon article publication.

Thanks: I thank Evan Taylor for guidance and support throughout this project. I also thank Katherine Barnes, Tiemen Wouterson, Daniel Herbst, Hidehiko Ichimura, Jason Kreag, Stephen Ross, David Simon, Kevin Schnepel, and several anonymous assistant US Attorneys and Assistant US Attorneys for helpful comments and insight.

The federal prison system is overcrowded and extremely costly. The Federal Bureau of Prisons estimates federal prisons will operate at 10% overcapacity in 2024, with the average cost of incarceration estimated to be \$43,836 per inmate, per year (United States Government Publishing Office 2023).¹ This large number of inmates is primarily comprised of drug offenders, who make up around 45% of all federal prisoners. Drug offenders bring especially high costs due to their long incarceration spells; in 2019, federal drug offenders had sentence lengths 252% higher than non-drug defendants. This was not always the case; both the volume of drug convictions and attached sentence lengths have dramatically increased since the 1980s and 90s (The Pew Charitable Trusts 2015). The cause of this increase is often attributed to legislation that increased punishment schedules for drug offenders. The most commonly cited and controversial aspect of such legislation is mandatory minimum (MM) sentencing. Yet, despite their prominence and controversy in drug reform debate, very little research has been done to estimate the causal impact MM eligibility has on sentence lengths.

Mandatory minimum sentencing restricts a judge's ability to give a sentence length below a specified amount. In practice, over 97% of all trafficking cases are resolved by plea bargain but MM sentencing is still expected to increase sentence lengths by increasing prosecutor bargaining power. MM sentencing applies to several crime types in the federal system, including weapon crimes, sex crimes, certain economic crimes, and most commonly, drug crimes. Federal drug trafficking offenses may be eligible for a 5-year or 10-year MM sentence depending on the drug type and the charged quantity of drugs. Many cases are affected by this practice; in 2019 when looking at the five most common drug types, over 86% of cases

¹The cost of incarceration estimate is for FY2021. This is just shy of the median annual wage in 2021, which was \$45,760.

were eligible for a MM sentence, with over 69% of cases meeting the threshold for a 10-year MM charge. In this paper, I provide the first causal analysis of the impact of MM eligibility on sentence length.

Mandatory minimum laws also raise equity concerns. One of the most common complaints over MM legislation is that it is believed to disproportionately impact racial minority defendants. Race disparities may emerge due to individual discrimination or through systemic forces. Specifically, because of low weight thresholds, MM sentencing raises concerns of disproportionate punishment for crack cases compared to other drug types, a drug-type in which Black defendants are overrepresented. These concerns are also supported in empirical research; Rehavi and Starr (2014) attributes significant disparities in sentence lengths to MM charges and Tuttle (2023) finds Blacks are disproportionately targeted for MM sentencing in crack cases. MM eligibility may also increase sentence lengths for non-violent or low-criminal history offenders, who may be inefficiently punished due to the constraints created by minimum sentencing. In this paper, I investigate both of these concerns.

The primary challenge in estimating the causal impact of MM laws on sentence length comes from selection issues. For most cases, MM eligibility is determined by the drug quantity listed at sentencing, measured by weight. If the weight at sentencing is greater than some threshold value, an individual is eligible to be charged with a 5 or 10-year mandatory minimum. However, this quantity does not need to equal the weight recorded at arrest. Essentially, this means that cases near the MM threshold may have their charged quantity manipulated in such a way that their eligibility is determined in part by legal actors. Tuttle (2023) notes the presence of significant bunching in the number of cases right above the 10-year MM threshold weight, evidence of this manipulation. This means that certain cases are

systematically moved from 5-year MM eligibility to 10-year MM eligibility, thereby distorting the relationship between charged weight and sentence length.

In this paper, I estimate the causal impact of MM eligibility on federal drug trafficking sentence lengths. I specifically consider the effects of 10-year MM eligibility and its impacts on race disparities in sentencing. Because cases are subject to manipulation, I decompose MM eligibility impacts into effects driven by the law and effects driven by legal actor discretion. This is done by separately fitting the distribution of sentence length over drug weight using data only in non-manipulated regions on either side of the 10-year eligibility cutoff. I then extrapolate the fitted distributions into the manipulated regions to create sentence length measures under the counterfactual scenario where no manipulation occurs. The discontinuity between these counterfactual distributions at the threshold weight is what I call the legal effect. It is the impact of MM laws for cases without manipulation. The distance between the actual sentence length and the higher counterfactual distribution (the one fit from the right-side of the distribution) is what I call the manipulation effect. It is a combination of two separate effects: (1) selection bias created by manipulation choices, and (2) the impact of being pushed from the 5-year eligibility into 10-year eligibility separate from the legal effect. That is, cases at the bunching point may have high sentence lengths because these cases are fundamentally different than non-manipulated cases, but also because the process of manipulation causes an increase in of itself. I provide evidence that manipulation effects are at least in part driven by non-selection factors at the higher eligibility level.

I find that, across the five most common drug types, 10-year MM eligibility increases sentence lengths by about 10 months (or 18% above the mean) purely through the legal effect. I find that the manipulation effect for the average case is around 17 months. Further,

the manipulation effect persists with the inclusion of controls, though the magnitude shrinks to around 7 months in the full-controls specification. This indicates that manipulation occurs most commonly for high-sentence defendants, but also that the manipulation process has a causal impact on increasing sentence lengths outside of selection bias. Taken together, these findings suggest that on average, all cases see an increase to their sentence length if eligible for the 10-year mandatory minimum, but for cases that are manipulated up to the 10-year eligibility the increase is significantly higher.

Perhaps the most interesting finding in this analysis is that racial disparities in MM sentencing are almost entirely driven by manipulation. I find that the legal effect is nearly identical across race groups and is fairly similar across drug type and criminal history level. However, the manipulation effect is entirely driven by racial minorities, with effects mostly local to cocaine and heroin cases. This means that the legal impact of MM eligibility has similar effects for most cases, but Black and Hispanic defendants with weights near the cutoff receive significantly higher sentences due to legal actor discretion. Manipulation effects are also present for low criminal history offenders, suggesting that policies such as the safety valve provision may not adequately shield low level offenders from MM effects.

Next, I utilize a unique data set of US Attorney spells to perform a series of analyses that compare estimates across treatment intensity. Doing so helps give validity to the main estimates and race results, provides evidence for prosecutor manipulation as a key mechanism, and considers race disparities on the external margin of case weight manipulation. The basic idea of this analysis is to utilize variation in case manipulation across district and time created by US Attorney leadership turnover. I find that manipulation effects increase with the timing of high-bunching US Attorney administrations taking office, and that

manipulation effects are generally driven by high-bunching attorney spells. I then compare manipulation across race and treatment intensity. I find that racial minorities are significantly more likely to be bunched compared to White counterparts. These results suggest that prosecutors disproportionately target racial minorities for case manipulation to higher MM eligibility, resulting in significantly longer sentence lengths. Taken together with the main results, this paper shows that most cases, MM eligibility has similar effects across race but when prosecutors use discretion to increase eligibility, minorities are disproportionately punished.

This paper contributes to literatures on the welfare impacts of sentencing structure, legal actor discretion, and how each of these may drive racial disparities. Many papers suggest legal actors use discretion to disproportionately target or punish racial minorities with worse court outcomes (Arnold et al. 2018; Rehavi and Starr 2014; Sloan 2022; Tuttle 2023; Yang 2016), though recent literature has suggested prosecutor discretion may actually limit racial disparities, specifically in connection with enforcement decisions (Shaffer and Harrington 2017; Shaffer 2023). Findings from this paper support previous results; that race disparities are largely driven by discretion and are likely tied to prosecutor choices. On a smaller scale, this paper also adds to the drug crime, MM sentencing literature. Findings among these studies vary, with some papers downplaying the significance of mandatory minimums (Bjerk 2005; Bjerk 2017a; Bjerk 2017b; Fischman and Schanzenbach 2012) and others finding significant impacts on sentencing generally and in contributing to racial disparities (Didwania 2020b; Didwania 2020a; Rehavi and Starr 2014; Tuttle 2023). This paper uses discontinuity techniques similar to Diamond and Persson (2017), leniency measures similar to Goncalves and Mello (2021), and exploits the same bunching point as seen in Tuttle (2023).

Bjerk (2017b) is the paper most closely related to this work. It provides a descriptive analysis of the sentence length impacts of being convicted at particular charging weights for fiscal years 2011 and 2012. The main analysis shows that MM eligibility rates are similar across drug types, the safety valve provision effectively prevents low-level offenders from receiving large sentence length increases, and that effects for crack are minimal.² I contribute to this paper in four ways: (1) by identifying the causal impact of MM laws on sentence length, (2) by disentangling causal effects into legal and manipulation effects, (3) by assessing the degree to which low criminal history defendants are protected from MM effects, and (4) by giving a more comprehensive view of MM impacts on racial and drug type disparities. To my knowledge, this is the first paper to causally estimate the impacts of MM eligibility on sentence length and the first to decompose MM race gaps by systemic and discretionary channels.

II. Background

II.A. Mandatory Minimum Sentencing of Federal Drug Cases

The main eligibility criteria for MM eligibility is drug quantity. In order for the mandatory minimum to apply, the weight at sentencing for one drug type must meet or exceed the set threshold weight.³ There are two separate thresholds, one for a 5-year and one for a 10-year mandatory minimum. A small subset of cases is eligible for MM penalty without meeting quantity thresholds. This can occur in two ways: if the crime involves death or serious injury, or if the defendant has committed a serious prior drug offense. Including,

²It's worth noting that in a brief appendix analysis, Tuttle (2023) finds that crack does indeed exhibit some sentence length effects if other years are included.

³Note that prosecutors cannot sum weights across drug types to get a MM charge.

excluding, or controlling for this subset does not substantially change results or significance of findings.⁴

Currently, MM sentencing applies to eight illegal or controlled substances: powder cocaine, crack, heroin, marijuana, methamphetamine, lysergic acid diethylamide (LSD), phenylcyclohexyl piperidine (PCP), and fentanyl. Due to a low number of cases among the last three types, this paper will focus only on cocaine, crack, heroin, marijuana, and meth offenses. Offenders still face a lower and higher threshold amount for each drug with a 1:10 weight ratio between thresholds. In this paper, all analysis considers only the higher threshold, which has stronger bunching, more severe punishment increases for eligible cases, and for which 70% of all federal trafficking cases are eligible.⁵

Cases with drug quantities at or above the MM threshold weight may not necessarily be charged with a mandatory minimum. But being charged at an eligible weight opens the possibility for the prosecutor to impose a MM charge, meaning hitting the threshold weight significantly increases prosecutor bargaining power. Some cases have initial charges that apply the MM filing, but through plea negotiations, do not carry a mandatory minimum in the final charges. In my data, I only observe final charges.⁶ For this reason, I do not focus on the MM charges themselves but only consider drug quantity and the sentence length. This ensures I do not leave out cases that do not show MM charges in the final charge data, but were still substantially impacted by MM eligibility during the plea bargaining process.

⁴While I cannot directly identify previous offense types, I can control for “serious” prior offenses in general. Fatal or serious injury crimes can be directly identified and excluded from the analysis.

⁵Many trafficking cases have a charging weight close to zero. Thus, the lower MM weight threshold has a far less prominent bunching point.

⁶Rehavi and Starr(2014) observe initial and final charges for their data, but it does not include drug cases. To my knowledge, data containing initial and final charges for federal drug crimes has not been made available to researchers.

Furthermore, MM sentences are often non-binding, with many convictions receiving sentences above or below the minimum specified sentence length. Sentences below the mandatory minimum can occur if a defendant provides “substantial assistance” in the prosecution or investigation of another offender,⁷ if the offender is eligible for the safety valve provision, or if a lower sentence is negotiated through plea bargain. Defense can apply for the safety valve provision if the defendant in question has a sufficiently low criminal history record.⁸ I directly test for MM effects for cases filed under the safety valve provision.

II.B. Drug Weight Manipulation

Discrepancies between seized, charged, and sentencing drug weights can occur for several reasons. Sources of these discrepancies may include drug weight approximations changing through the criminal justice process, law enforcement officers manipulating weights through false reporting or planting additional drugs, or convenience reporting of round numbers. However, the primary source of drug weight manipulation occurs due to changes in relevant evidence. This may come by way of additional testimony or connecting a defendant to other traffickers or cases (Lynch 2016). The impetus for this additional evidence comes from prosecutors, who can decide whether to pursue additional evidence in building a case. If prosecutors can connect a defendant to other offenders or larger organizations, they may increase the available evidence of drugs to charge a person with.

One indication of this evidentiary channel is in conspiracy charges. Qualifications for a conspiracy charge are broad, with drug conspiracy generally defined as two or more individ-

⁷See USSG § 5K1.1

⁸Before the 2018 First Step Act, this was only defendants with one or zero criminal history points. The FSA expanded this provision to include those with four or less points and only counted past crimes that were two- or three-point offenses. See USSC (2019) for details.

uals agreeing to transport, manufacture, or sell illegal substances. Thus, if prosecutors seek to connect defendants to other offenders or an organization, they are more likely to charge them with conspiracy. This appears true in my data; I find that 54% of cases that are not at the bunching point carry a conspiracy charge. However, at the MM threshold weight, 83% carry a conspiracy charge. It's also worth noting that conspiracy charge rates are similar across race, with White defendants having slightly more conspiracy charges than non-White defendants.

Tuttle (2023) and Cooper (2023) each give additional insight into drug weight manipulation. Tuttle (2023) provides a detailed description of the judicial process for recording weights, and gives evidence that drug weight manipulation in crack cases is driven by prosecutor discretion to seek evidence. Cooper (2023) provides a deeper explanation of the way in which weights can be added across defendants in a conspiracy charge. In each paper, the authors assume that manipulation only occurs upward, meaning that case weights are pushed from the 5-year MM to the 10-year MM eligibility. This assumption is supported by the findings in this paper.

III. Data

The primary drug case data is provided by the United States Sentencing Commission (USSC) and includes all federal drug trafficking cases from 2010 to 2021. Data is at the case-individual level. I restrict the data to the five most prevalent substances subject to MM sentencing: powder cocaine, crack, heroin, methamphetamine, and marijuana.⁹ I also restrict the data to the primary racial groups of study: White, Black, and Hispanic. For the intensive

⁹The other three substances subject to MM law are PCP, LSD, and fentanyl. These had too few observations for any meaningful analysis.

margin analysis, the data is further restricted to include only cases with primary drug weights at 20% to 180% of the threshold weight. This gives enough data to fit distributions on but excludes the 5-year threshold and extremely high weight cases that are less similar to cases near the 10-year threshold. This gives a total of 44,626 observations.

The USSC data provides a rich set of defendant and litigation details. Information on the defendant's sex, race, age, education, and citizenship is included. The data also contains specific statutes that are charged, the drug weight given in grams or an approximation of the weight, any factors that increased or decreased the sentence length above or below the prescribed guideline amount, whether the case was plead or not, and sentence length the defendant received.

Drug weight measurement is complicated by two factors: multiple drug types and ranges of weights. Around 27% of cases involve a defendant charged with multiple drug types. In regard to the weight threshold for a MM charge, weights are to be considered separately. Thus, I consider the primary drug type for these cases, which is the drug with the highest weight relative to the 10-year threshold. However, prosecutors may sum multiple drug type weights together when determining the base offense level, implying more drug types are likely to increase sentence length. To control for this, I include the number of charged drugs as a control in some of the main specifications. I also provide discontinuity results after controlling for the weights of other drug types to ensure results are not driven by differences in multiple drug carrying behavior. This is done for the full sample and each racial group and is reported in the robustness section. For about 20% of cases, defendants are charged with a range of weights rather than one precise measure. These cases are omitted from the main analysis but are considered across three different measurement schemes in the robustness

checks.

Table 1 gives summary statistics for several key variables in the drug data. These statistics are presented for the full sample as well as split for each racial group. The table shows that each of the five drug types have a sizable number of cases, with heroin cases making up the fewest percent of cases at 14.3% and cocaine making up the most with 26.1%. Drug type is highly correlated with race, suggesting primary specifications for race heterogeneity should control for the primary charging drug. The table also shows that Black and White defendants have very similar criminal history points, while Hispanic defendants have much lower criminal history on average. I also consider conspiracy charges as these are a primary mechanism by which drug weight manipulation occurs (Lynch 2016, Tuttle 2023, Cooper 2023). Across all three racial groups, the proportion of cases with a conspiracy charge is nearly identical. This helps alleviate concerns of unequal opportunity for manipulation across races.

I also use a unique, hand collected data set of US Attorneys from years 2013 to 2020 to consider heterogeneity by treatment intensity. The data set includes all US Attorneys from each district, including presidential appointed attorneys, Attorney General appointed attorneys, and acting and interim attorneys that took the leadership role between appointments. This data is used to construct a measure of bunching propensity by US Attorney administration, which is key for identifying racial disparities in drug weight manipulation. This data is gathered from a number of sources including direct correspondence from US attorney district offices, US attorney district office websites, Wikipedia, and news articles. The data set includes the US attorney's name, nomination date (if applicable), confirmation date, and date out of office. Dates are all recorded at the monthly level to match the drug data, with

Table 1: Drug data - summary statistics

	(1) All Cases	(2) Black	(3) Hispanic	(4) White
<i>Drug type:</i>				
cocaine	0.264	0.253	0.341	0.113
crack	0.181	0.461	0.0395	0.0453
heroin	0.142	0.144	0.164	0.0879
marijuana	0.172	0.0459	0.283	0.134
meth	0.241	0.0952	0.172	0.620
<i>Defendant characteristics:</i>				
criminal history points	4.112	5.698	2.394	5.330
female	0.124	0.0706	0.117	0.221
high school	0.569	0.661	0.423	0.739
age	35.43	35.91	34.04	37.68
<i>Outcomes:</i>				
gun involved	0.101	0.144	0.0595	0.122
trial	0.0295	0.0452	0.0233	0.0183
# of drug types charged	1.306	1.508	1.178	1.262
percent weight	0.736	0.669	0.753	0.805
conspiracy	0.479	0.501	0.466	0.474
MM imposed	0.219	0.271	0.200	0.177
safety valve	0.314	0.150	0.465	0.245
sentence length	72.59	93.85	58.84	68.82
Observations	42437	14122	19373	8942

Notes: Figures represent means. Sample includes all cases with weights between 20% and 180% of the threshold weight. All variables are binary except for criminal history points, age, percent weight, and sentence length. Sentence length is measured in months.

an attorney being counted as acting that month if the days served are greater than or equal to 16.¹⁰ I match this data with the above USSC drug data to analyze manipulation effects for cases between June 2013 and September 2020. This gives a sample of 22,606 cases.

¹⁰For some attorneys, dates of entry and exit are only available at the month level. In these rare cases, I default to the incoming attorney being the acting attorney on the month of overlap

Online Appendix Table A.1 gives a few key statistics about US Attorney administrations. For the study period of 2013 to 2020, each district had an average of 3.168 attorneys serve in the position with a total of 282 different attorneys. Each attorney served an average of 48.28 months and prosecuted 134.2 drug trafficking cases with weights between 50 percent and 150 percent of the threshold weight. The mean bunching propensity measure is 9.75 but with a standard error of 12.84, indicating high variance between administrations. To further illustrate the variance in bunching, I consider the maximum and minimum bunching propensity measures within each district. The mean maximum propensity is 16.41 and the mean minimum propensity is 2.752.

IV. Empirical Strategy

To accurately estimate the causal effects of MM sentencing, I need to know what the discontinuity looks like without any manipulation. Thus, my strategy is to create counterfactual distributions of sentence lengths over charged drug weight by estimating fits over the distribution of unmanipulated regions. These fits are then extrapolated into the regions near the eligibility cutoff where manipulation is present. This technique is a regression discontinuity design using extrapolation, often referred to as a “donut RD”. I fit two separate distributions on either side of the cutoff. The fitted distributions are determined only by cases that are assumed to be unmanipulated. I then look at the extrapolated points right at the threshold weight. The distance between these two points gives the legal effect - the causal effect of MM law in absence of manipulation activity.

I also consider how these fitted distributions compare against the actual sentence lengths. Specifically, I compare the point extrapolated from the right-hand side at the threshold

weight against the actual sentence length at the threshold weight. Essentially, I am comparing the counterfactual sentence length at the threshold weight versus the observed sentence length at this point. The distance between these two points gives the manipulation effect, which is comprised of bias and non-bias caused increases. Selection bias occurs because prosecutors do not randomly choose who they manipulate. But increases also may occur beyond selection effects - recall that the primary manipulation mechanism is additional evidence gathering by the prosecution. Evidence of more drugs and conspiracy connections are likely to increase sentence lengths beyond the effects of a higher weight. If the manipulation effect survives the inclusion of controls, it suggests the manipulation effect is not purely driven by bias, but is also impacted by these evidentiary effects.

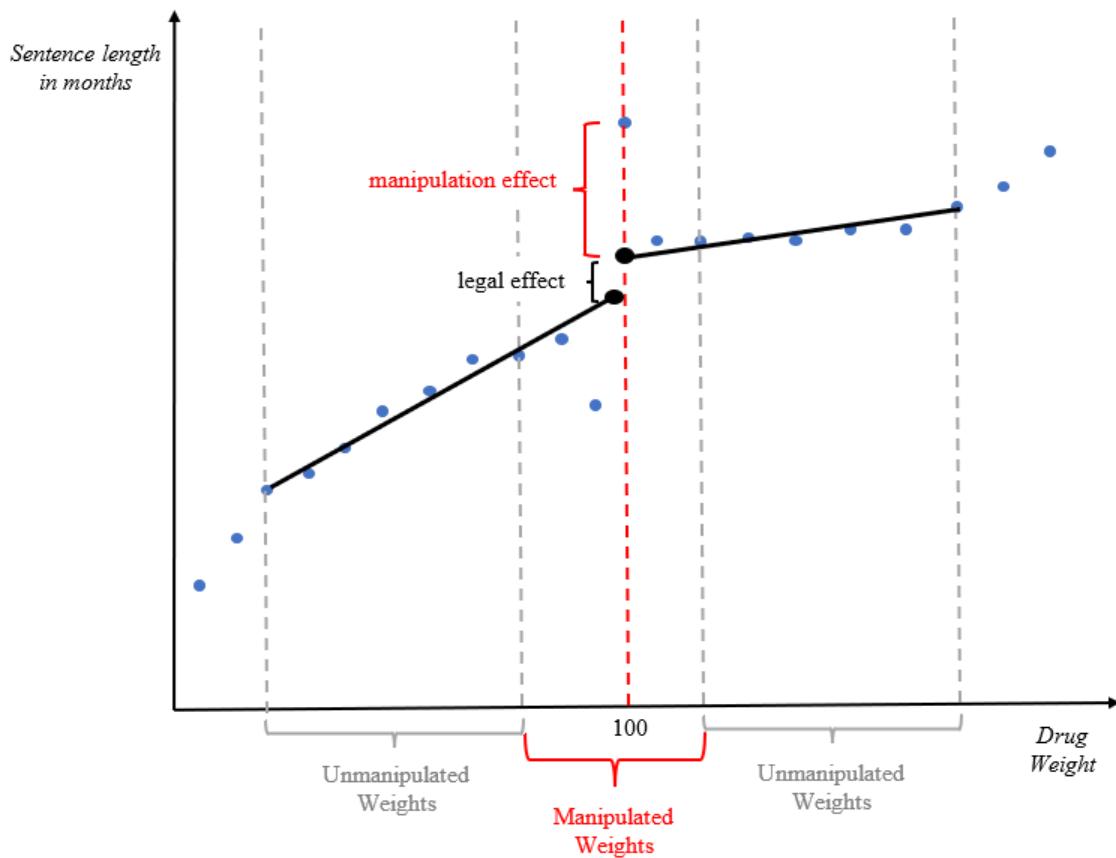
Finally, I look for a discontinuous drop in sentence length for cases just before the cutoff. If prosecutors manipulate cases that are already likely to have higher sentences, the “worst cases,” then the average sentence where they are manipulated from will be lower than the rest of the trend. Note that this is analogous to a missing mass argument. This missing sentence amount is not formally estimated since the number of cases varies drastically at the bunching point compared to at other charging weights.¹¹ However, I note its presence as evidence that prosecutors are indeed manipulating cases upward to increase sentence lengths for certain defendants.

Figure 1 illustrates the idea of this strategy. The blue points represent the actual sentence length distribution for the group being considered. Dashed vertical lines show the manipulated and unmanipulated regions. The black lines give the fitted distributions which

¹¹Ideally, I would multiply the missing sentence amounts by the missing mass in cases and compare this value against the product of the manipulation effect and the excess mass at the threshold weight. I refrain from doing this since missing masses in cases left of the threshold weight cannot be precisely estimated.

are extrapolated into manipulation regions. The large black points give the discontinuity absent manipulation. The distance between the top point and actual sentence length at the threshold gives the manipulation effect. I do not label the missing sentence amount since it is not estimated, but it can be observed as the distance between the fitted sentence just to the left of the threshold and the actual sentence lengths near this weight.

Figure 1: Empirical strategy illustration



Notes: This figure depicts the empirical strategy used to estimate the MM impacts and decompose them into legal and manipulation effects. The black lines represent the fit over the unmanipulated weights and extrapolated into the manipulated weights. The discontinuity between these fits at the threshold is the legal effect. The gap between the predicted sentence from the right-hand side of the distribution and the actual sentence length is the manipulation effect. Note that the legal effect is interpreted as causal while the manipulation effect is likely biased upward in its raw form.

To implement this empirical strategy, I must first determine which regions are manipulated and which are not. There are two primary ways in which manipulation regions have

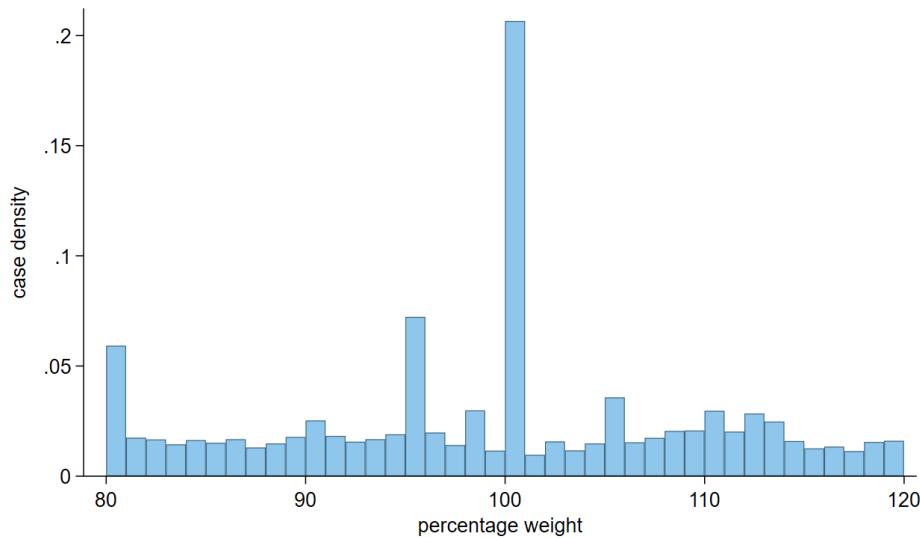
been detected and determined in past literature. The first is a formal test developed in Frandsen (2017) and practiced in Goncalves and Mello (2021) used to detect changes in the distribution attributed to manipulation. This method is unlikely to work well in this setting because small scale manipulation likely occurs as prosecutors or law enforcement round to whole numbers. The second method is simply using visual inspection, as done in Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013). This approach works when there is a clear and obvious missing mass in the distribution that is supplying the observations at the bunching point. Where the missing mass begins is assumed to be the beginning of the manipulation region. To tease out the missing mass area, I fit a fifth order polynomial with fixed effects for each 10-percentage point round figure over the main analysis weights: 20% to 180% of the threshold weight. Figure 2(b) shows this fit distribution in comparison to the actual case density. Missing mass appears following the 70% value and continues up until cases just before the MM threshold at 100%.

One common way to check whether the manipulation region is correctly specified is to compare the excess and missing mass amounts, which should be equal. The excess mass at the bunching point is just under 2,385 cases. The missing mass from 70% to 99% is about 1,129. The smaller missing mass may be generated by two sources. First, when looking at each drug individually, meth has a manipulation range that is clearly wider than the other drug types (closer to 60%). Second, the missing mass estimation does not include cases at 70%, 80%, or 90%. These round numbers are omitted as rounding bunching occurs at these points as well. Significant case reduction may occur from cases at the 80 percentage point cases.

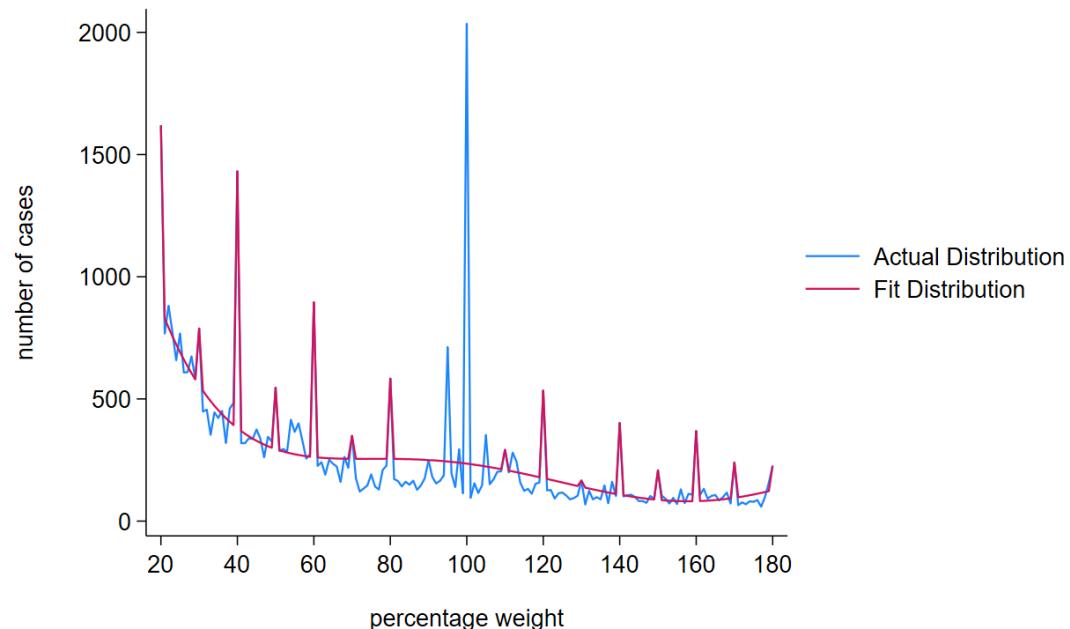
Based on the comparison between the fit and actual case distribution, I consider cases be-

Figure 2: Case distribution by sentencing weight

((a)) Distribution of all cases near the threshold



((b)) Distribution of cases with fit



Notes: This figure considers which cases are being manipulated to the bunching point. I plot the distribution of cases and then fit a 5th order polynomial controlling for internal bunching that occurs at round points (I control for every 10% value). Missing mass is identified as areas where cases are below the fit polynomial - this appears to be primarily from weights at 70% to 95%.

tween 70% and 105% to be within manipulation ranges for my main analysis. This means the left-side regression is fit using cases between 20% and 70% of the threshold weight, while the

right-side regression is fit using cases between 105% and 180%. Each of these fit predictions are then extrapolated into the manipulation region to create counterfactual distributions, or the trend of sentence length absent manipulation. In my robustness checks I consider the results under many specifications with different cutoffs for the left-side manipulation region. There is a tradeoff in setting the manipulation region cutoff; cutoffs further from the threshold are less likely to be biased since they are less likely to accidentally include cases that are manipulated, while windows closer to the threshold have more data to fit on and are thus likely to be more precise. For this reason, it is important to show that results are similar across many different set cutoffs. This is analogous to testing a regression discontinuity with different bandwidth sizes.

Following Gelman and Imbens (2019), my analysis assumes a functional form that is linear, though I include some quadratic fit predictions in my robustness checks. I use standard errors of the prediction to create confidence intervals for each fitted value to assess inference. The legal effects are considered statistically significant if the two confidence intervals do not overlap. The manipulation effect is considered statistically significant if the actual sentence length falls outside of the confidence interval from the right-side fitted regression.

The key identifying assumption of this empirical design is that without manipulation, the distribution of sentence lengths for cases in manipulation regions would have followed the counterfactual distributions fit using unmanipulated cases. Note that this assumption is standard in a traditional bunching design, only there it is typically involving the variable being bunched, which in this context would be cases. Now I assume that the relationship between sentence length and charging weight can be predicted using the polynomial coefficients among unmanipulated cases.

A second important assumption is that prosecutors are not manipulating cases from low weight amounts (around 20%-50% of the threshold weight) up to the bunching point. This assumption is supported in three ways. One, drug weight manipulation primarily occurs from additional fact finding on the part of the prosecutor (Lynch, 2016; Tuttle, 2023). In Cooper (2023) and Online Appendix D of Tuttle (2022), costs to evidence gathering are assumed to increase as the manipulation quantity increases. Thus, if the distribution of available evidence used to manipulate weights is relatively uniform across seized drug weights, the probability of manipulation should be highest for cases just below the threshold weight. This manipulation probability will decrease for lower seized weights, indicating manipulation is more likely to occur for cases near the threshold. Two, Cooper (2023) assesses decreases bunching caused by a policy and finds increased mass appear primarily for cases near the bunching point. This indicates these marginal cases were mostly manipulated near the threshold weight.¹² Three, missing sentence lengths appear at weights just below the threshold.

V. Results

Before presenting the main race disparity results, I first discuss findings for the analysis aggregated across all cases. I then compare legal and manipulation effects across racial groups. After these main results, I briefly discuss heterogeneity across drug type and criminal history group. For each of these analyses, my preferred specification is a linear functional form with the manipulation region being defined as cases between 70% and 105% of the

¹²Following Cooper (2023) I graph the weight distributions before and after the policy for the primary affected group, but on a wider distribution to check which cases are being manipulated. I find excess mass appears mostly near the threshold, with some excess mass appearing between 50% and 60% of the threshold. Figure available upon request.

threshold weight and without controls. The simple model is preferred so that any mechanisms driving disparities between groups is not controlled for. I also provide analysis with varying sets of controls, varying fit cutoff values, and with quadratic fits in the main results or in the robustness section.¹³ A discussion on which types of cases are manipulated is provided in Section A of the Online Appendix, which gives additional context for these results.

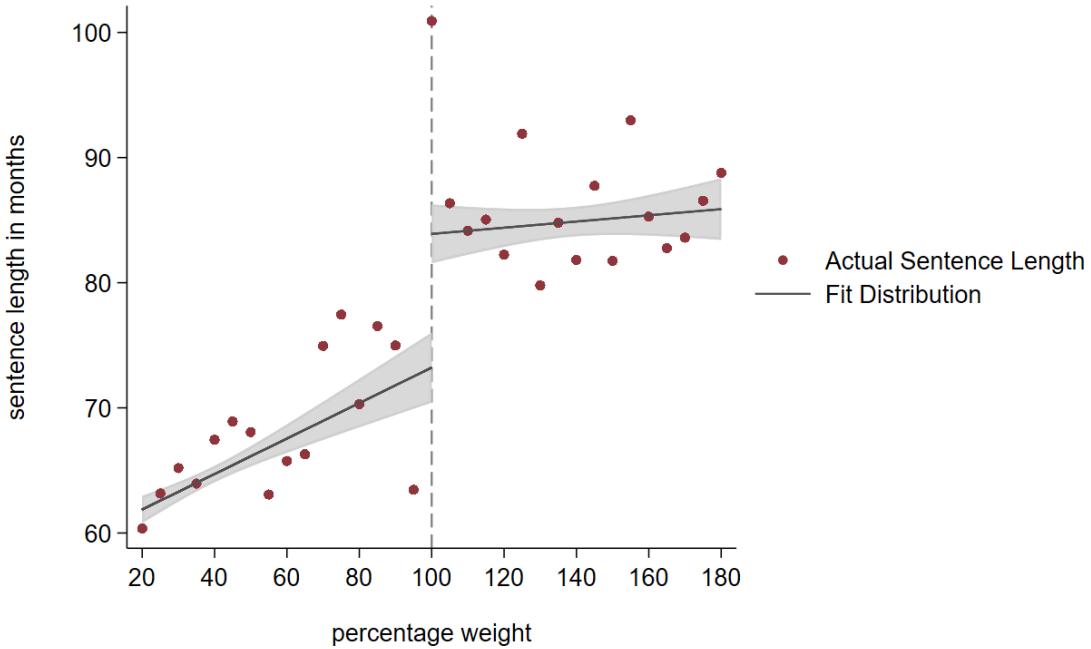
Overall Effects

Figure 3 shows the discontinuity aggregated across all cases. The figure highlights three important findings. First, the fit values approximate sentence lengths quite closely within the manipulation range except for cases closest to the threshold weight. Second, the legal effect of mandatory minimums is significant, with an increase of 10.63 months. This is an 18.04% increase over the mean at 99% threshold weight. Third, the figure shows a large manipulation effect with the actual sentence length at the bunching point far higher than the fit value. I estimate a difference of 17.01 months between these points. And fourth, the missing sentence appears to be local to cases just left of the threshold weight. The presence of missing sentence lengths gives evidence that prosecutors are manipulating cases up to the threshold weight.

Table 2 reports these results along with results for three different control schemes. The baseline controls include criminal history points, drug type, racial group, the defendant's age, their education, and whether they are female. The fixed effects are at the district and year level. The prosecutor decision controls include the a binary for gun use, the number of drug types an individual is charged with, and whether the case went to trial. It's worth

¹³Many sub group results, particularly drug type-race intersections, are not presented here for space. Results for specific subgroups are available upon request to the author

Figure 3: Discontinuity analysis for all cases



Notes: This figure gives the main discontinuity analysis for all cases with linear fits through points before 70% of the threshold weight and extrapolated past 70% to the threshold from either side. Sentence lengths are binned by 5% except at the threshold weight, which only reports average sentences for cases charged with exactly the threshold weight.

noting these prosecutor controls may be decisions made in conjunction with the manipulation decision. Thus controlling for them may reduce effect sizes lower than the true causal effect. Still, I include the specification with these prosecutor decisions to try and reduce any possible bias from the manipulation effect.

The size of the manipulation effect decreases to 11.24 when controlling for baseline and fixed effects, and to 7.05 when including prosecutor decisions. This means that even with the full set of controls, there is still a sizable sentence premium that cannot be accounted for by the included observables. If this estimate were to be interpreted as the manipulation effect free from selection, it would imply that cases that are manipulated to the 10-year mandatory minimum threshold receive a causal increase to their sentence length of 16.70 months (this

Table 2: MM effects for all cases together

	(1) sent length	(2) resid sent 1	(3) resid sent 2	(4) resid sent 3
legal effect	10.63	11.08	10.10	9.65
left side 95% CI	[70.36, 75.91]	[-0.61, 4.00]	[1.17, 5.58]	[1.62, 5.67]
right side 95% CI	[81.61, 86.23]	[10.96, 14.98]	[11.78, 15.61]	[11.77, 15.24]
manip. effect	17.01	12.40	11.24	7.05
fit value	70%	70%	70%	70%
baseline controls	no	yes	yes	yes
fixed effects	no	no	yes	yes
pros. decision controls	no	no	no	yes
N of obs fit on	33691	33691	33691	33691

Notes: This table gives the regression discontinuity results for the full sample of cases. Column 1 presents the simple linear case fitted at 70% with no controls. Columns 2, 3, and 4 then add various controls (through residualization), still fitting at 70%. Baseline controls include criminal history points, drug type, and racial group, defendant age, defendant sex, and defendant education. Fixed effects are at the district and year level. Prosecutor decision controls include whether a gun was used in the offense, the number of drug types charged in the case, and whether the case went to trial.

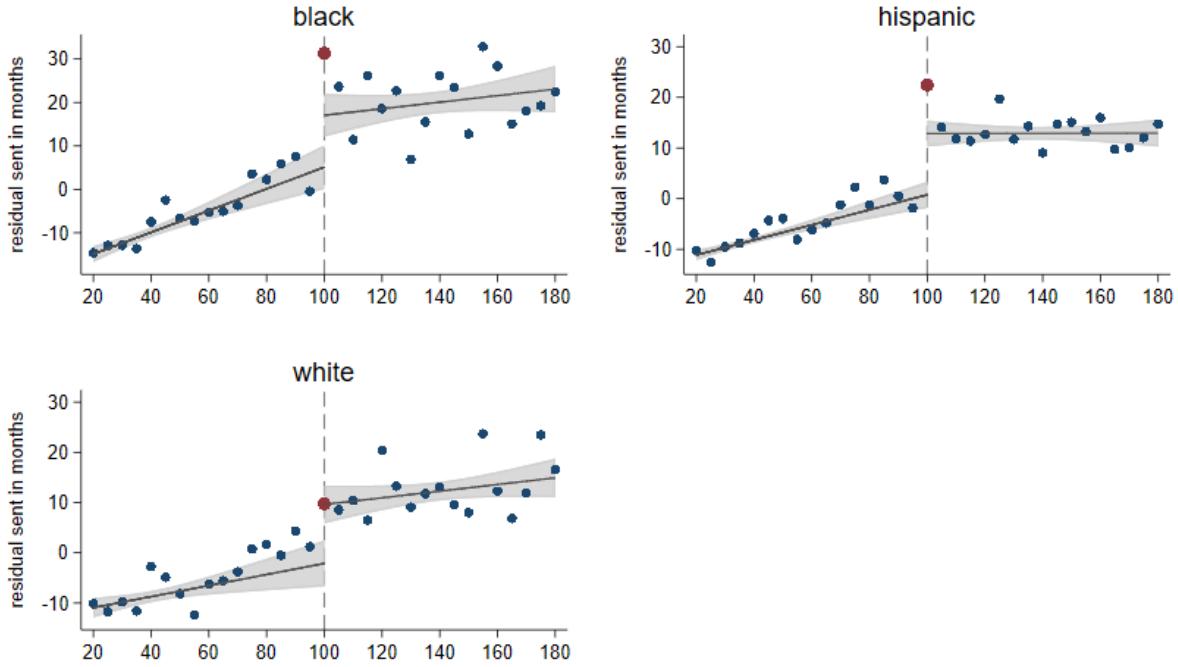
is the legal + manipulation effects after controls). Statistical significance persists through each specification.

Racial Disparities

Figure 4 illustrates the discontinuity analysis separated by race with sentence lengths residualized on the baseline controls, which in this case are criminal history points, drug type, defendant sex, education, age, and whether they are an illegal alien. These regression results are also reported in Table 3. There are two key results presented here. First, the legal effect is nearly identical across each race group. It's worth noting that these are level effects and if considered by percentage of mean sentence left of the cutoff, the effects vary more, with Hispanic and White defendants experiencing a significantly larger impact from the legal effect compared to Black defendants. However, the results still imply that, absent any manipulation, the causal impact of mandatory minimum eligibility is effectively equal

across race in terms of level sentence lengths.

Figure 4: Race comparison - discontinuity using residualized sentence lengths



Notes: This figure gives the residualized sentence discontinuity analysis by racial group with linear fits through points before 70% of the threshold weight and extrapolated past 70% to the threshold from either side. Residual sentence lengths are binned by 5% except at the threshold weight, which only reports average sentences for cases charged with exactly the threshold weight.

The second key finding is that the manipulation effects are driven exclusively by racial minority defendant cases. Black defendants at the weight threshold experience sentence lengths 14.28 months higher than cases just to the right of the threshold, even after controlling for criminal history and drug type. Similarly, Hispanic cases have a manipulation effect of 9.59 months in this specification. But White cases have a manipulation effect of essentially zero. These effects shrink under the full control scheme, with manipulation effects of 6.27 months and 4.73 months for Black and Hispanic defendants, respectively. This provides evidence that prosecutors choose to manipulate cases more frequently for minority cases than White ones, corroborating findings from Tuttle (2023). These results also indicate that minority

defendants who are manipulated receive a sentence premium, meaning the causal impact of mandatory minimum eligibility is higher for minority cases than White ones in the presence of manipulation.

Table 3: MM effects by race

	(1) black	(2) hispanic	(3) white
legal effect	11.87	12.03	11.71
left side 95% CI	[-0.04, 9.94]	[-1.91, 3.25]	[-6.74, 2.36]
right side 95% CI	[12.28, 21.96]	[10.32, 15.38]	[5.98, 13.38]
manip. effect	14.28	9.59	0.11
fit value	70%	70%	70%
N of obs fit on	11205	15296	7190

Notes: The dependant variable is sentence length residualized against criminal history points, primary drug type, defendant education, age, sex, and whether they are an illegal immigrant . All specifications are discontinuities based on local linear fits. 95% confidence intervals are presented for each fit regression on either side of the cutoff. These are calculated using the standard error of the predicted expected value, and significance is determined as no overlaps between these two intervals. Fit value represents the cutoff for where extrapolation begins. In this case, all specifications have left-hand regressions fit on cases with weights between 20% and 70% of the threshold weight.

The results can also be considered at each drug-race intersection. However, these samples become smaller and noisier, leading to less precise estimates. The overall story is the same though - legal effects are similar across groups, though estimates tend to be strongest for whichever race has the most cases for a drug type. In extreme cases, one or two race groups make up almost all of the observations within a drug type, such as Black defendants for crack cases. In these settings, legal effects are only present for the high case group. Manipulation effects are driven primarily by racial minorities in cocaine and heroin cases, though some smaller effects exist for other minority race-drug combinations. These results still highlight that the law effects defendants similarly, but manipulation decisions only penalize Black and Hispanic defendants.

Drug Type and Criminal History

Here I briefly consider heterogeneity in effects across drug type and criminal history. While these factors are controlled for in the residualized results, I consider here whether drug types are a key mechanism for manipulation decisions. Due to its importance in drug sentencing history and interaction with Black defendants, I especially consider whether results are driven primarily by crack cases. I then repeat the analysis for low level offenders.

The legal effect is estimated to be between 9 months and 14 months for each drug type besides for meth, which has a large legal effect of 20 months. The manipulation effects are driven almost entirely by only two drug types: cocaine and heroin. It's noteworthy that crack cases do not exhibit bunching, a result that corroborates findings from Bjerk (2017b). This may come from the fact that overall sentence lengths are much higher for crack cases than any other drug type. So while mandatory minimums seem to have less impact on crack compared to other drugs, crack cases are still punished at a higher level it seems.¹⁴ Estimates are displayed in the Online Appendix in Figure A.3 and Table A.2.

To measure low-level offender effects, I consider effects for two subgroups: defendants who have had no criminal history contact of any sort, and those who have had contact but have 0 or 1 criminal point, which relates to crimes with sentences less than 60 days. Because there are policies designed to protect first-time and low-level offenders from MM sentences, the expected effect of MM eligibility may be low for these two low history groups. Bjerk (2017b) reports this expected result of almost no effects for low level defendants.

I find that individuals with low criminal histories are still subject to both legal and

¹⁴This may stem from differences in criminal behavior, systemic forces, or discrimination. Explaining the sentence length disparities across drug types outside of MM eligibility is beyond the scope of this paper.

manipulation effects. Online Appendix Figure A.4 shows the regression discontinuity graphs for these two criminal history groups. Individuals with no prior criminal justice interactions experience a legal effect increase of 6.721 months and a manipulation effect of 11.49 months. Defendants with 0 or 1 points experience a legal effect of 10.18 months and a manipulation effect of 10.03 months. Both groups exhibit a relatively small missing sentence amount, suggesting that selection may not be a primary driver of the manipulation effect. These results suggest that legal protections do not fully shield low history offenders from higher sentences, nor do they protect them from effects driven by case manipulation in non-safety valve cases. See Online Appendix Table A.3.

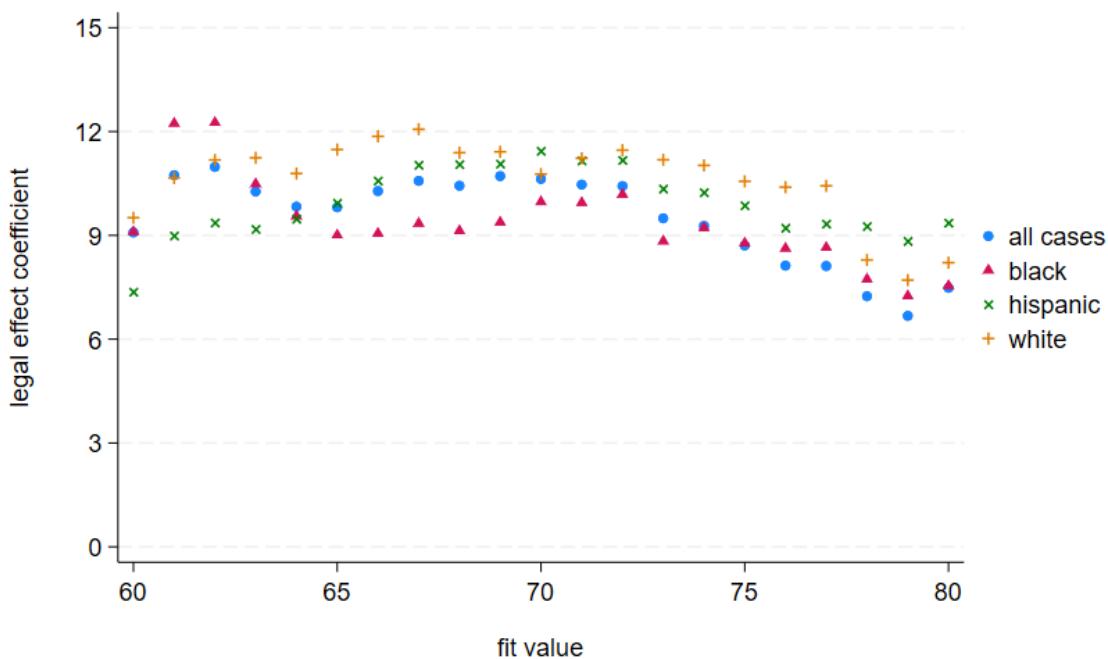
V.A. Robustness

One of the primary concerns for the main results is that using the 70% cutoff for fitting over non-manipulated cases either creates bias (if 70% is too low) or doesn't capture all the variation in the data well (if 70% is too high). To check that results and magnitudes are not driven by one specific fit, I repeat the analyses for the overall results and the heterogeneous results using a variety of different fit cutoffs. For each group, I run the main regression discontinuity again for every cutoff between weights 60% and 90% of the threshold for a total of 30 additional specifications (in addition to the 70% analysis). I present these results in four specifications - one in which all fits are aggregated and averaged into one fit, and the median, upper bound, and lower bound fits with regards to the legal effect. These are displayed for all cases together in Table A.2 in the Online Appendix. The legal effect remains largely the same, though around a month smaller for most specifications as 70% happens to be the upper bound for the full sample. The lower bound of the legal effect is 6.362 and

is still significant at the 5% level. Note that the manipulation effect doesn't change across these four specifications as the cutoff value only affects the left-side regression fit.

The heterogeneous effects presented in the main results are strongly consistent across different fit values. Groups with noisier data, like crack cases, tend to have higher variability of results by fit. Even still, these results tell a similar story. I present the legal effect coefficients across fit by race and for all cases in Figure 5. This graph shows the result for all 21 regressions for each group. The legal effect is consistent across these specifications, suggesting results are not driven by specific cutoff choices.

Figure 5: Legal effect robustness - all cases and by race



Notes: Each point gives the magnitude of the legal effect regression under a specific fit value. A fit value designates which area is considered manipulation region and which is considered unmanipulated. For example, at a fit value of 65, the regression is fit using all cases with weights between 20% and 65% of the threshold weight. The regression is extrapolated from 65% up to the 10-year MM cutoff, where it is compared against the right-hand regression to estimate the 65% legal effect coefficient.

I also check that the main results hold using a quadratic fit. For this check, I only use a cutoff value for the manipulation region of 80% of the threshold weight. I use a higher cutoff

because at 70%, the quadratic polynomial tends to overfit the data for certain groups. Thus, to assure a reasonable fit I increase the amount of data to fit on and decrease the amount of extrapolation needed. The quadratic fits do not largely change the results, only slightly increasing or decreasing effects. Following Gelman and Imbens (2019), I refrain from using higher order polynomials, especially given concerns of overfitting across the non-manipulated region of cases.

In the main analysis, I control for the number of other drug types a defendant is charged with outside of the primary drug type. Readers may be concerned that the quantity of other drugs is driving results, especially if other drug type quantities is correlated with certain racial groups. I consider the legal and manipulation effects for the full sample and across racial groups again, now controlling for the quantity of other drug types rather than just the number. These results are presented in Online Appendix Table A.5. I find effect sizes and statistical significance very close to those listed above, though with smaller manipulation effects.

Another concern may be that limiting the sample to cases with precise weights biases results if imprecise measures are strongly correlated with sentence length. I now rerun the full sample analysis including cases that are charged with a range of weights rather than one precise count. I use three different measures to do this - the minimum, median, and maximum of the range. These range cases are combined with the precise weight cases for a combined samples of 51,975 cases when using the maximum or median, and 56,189 when using the minimum measure. Regression results for this sample are presented in Online Appendix Table A.6. Across all three measures, the legal and manipulation effects have similar magnitudes to main results and remain statistically significant.

VI. Mechanism and Treatment Intensity

The main results show the presence of a large manipulation effect that is driven entirely by racial minority cases. In Section II., the primary mechanism for this manipulation effect is explained as prosecutor discretion. In this section, I explore this mechanism further by considering heterogeneity in manipulation across US Attorney spells. This analysis serves three main purposes: (1) it gives additional evidence that prosecutors drive bunching and that bunching leads to higher sentences at the threshold, (2) it considers whether higher bunching administrations exhibit stronger or weaker legal and manipulation effects, and (3) it provides evidence that higher bunching administrations bunch Black defendants more than White counterparts.

VI.A. US Attorney manipulation measure

I utilize the variation in bunching propensity by US Attorney administration as a pseudo-random measure of the probability of exposure to drug weight manipulation. The idea is that non-bunching and low-bunching administrations serve as reliable counterfactual distributions compared against high-bunching administrations. This is a similar strategy as described in Frandsen (2017) and practiced in Goncalves and Mello (2021). However, the Frandsen (2017) method relies on comparing observation counts for values neighboring and at the bunching point to identify bunching and non-bunching agents. Because of round number bias, this method does not work well in my setting - an attorney who practices little or no manipulation may still exhibit higher observation counts at the bunching point compared to neighboring values.

Instead, I construct a continuous, residualized bunching propensity measure. This ap-

proach has a flavor of judge-leniency instrumental variables as seen in a large number of law and economics papers (Kling 2006; Aizer and Doyle 2015; Mueller-Smith 2015; Bhuller et al. 2020; Di Tella and Schargrodsy 2013; Dobbie et al. 2018). However, I don't use the propensity measure as an instrument but consider it a proxy for the defendant's probability of being charged at the bunching weight. I then simply divide attorney spells into halves based on this bunching score and compare outcomes across these two groups. The bunching propensity score is calculated as the proportion of cases at the bunching point within an attorney spell, residualized on district fixed effects. Residualizing on district fixed effects accounts for geographic differences in case volumes and drug weight distributions and allows the bunching measure to capture differences in manipulation relative to other spells within a district. Figure A.2 in the Online Appendix highlights the variation in bunching by district.

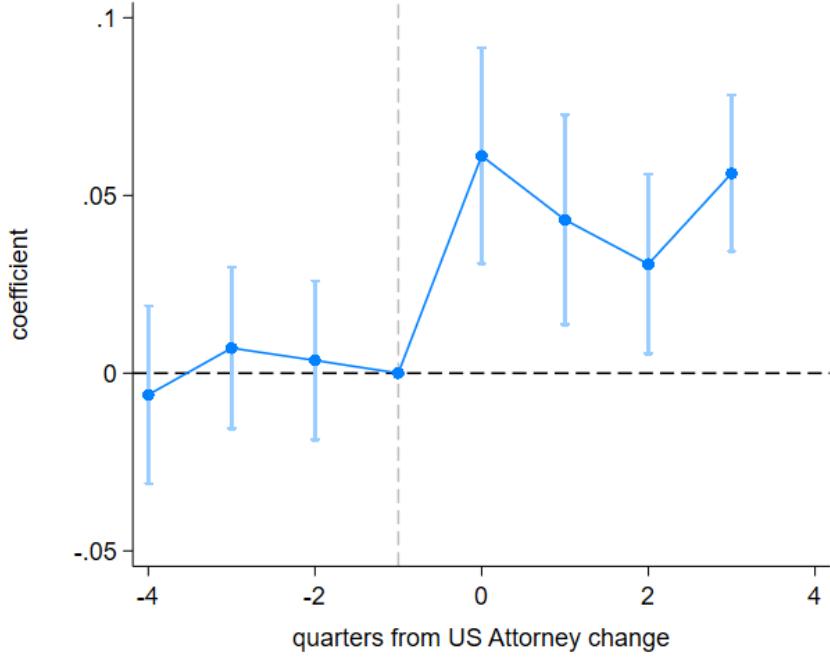
One might worry that differences between the high-bunching and low-bunching Attorney groups are not measuring differences in manipulation decisions, but are the result of changes in sample composition or criminal behavior. That is, I need that individuals are not changing their criminal behavior based on the current or recent US Attorney prosecution patterns. This seems reasonable; to have a meaningful impact on criminal behavior, criminals would need to be aware of the US Attorney's position on drug prosecution and that knowledge would need to have a strong enough incentive to change production or transportation activity. Beyond this, many federal cases pass through the state system first, meaning there is uncertainty for the defendant about which level they will be prosecuted at. A bigger threat comes by way of detection and law enforcement activity. US Attorneys work in close contact with members of the FBI and sometimes with the US Marshall's Service or the DEA. A high bunching attorney may also encourage specific types of drugs be targeted or more arrests in

general.

While I cannot test directly for changes in criminal activity or law enforcement, I can broadly test for selection on observables by regressing bunching propensity on defendant characteristics. Table A.7 in the Online Appendix reports *F*-Statistics and tests for regressions run on observables. The joint *F*-test for the simple bunching binary measure returns an *F*-value of 7.11 and a *p*-value of 0.000, indicating non-random selection of who has their case manipulated. However, when using the residualized attorney bunching propensity score, the resulting *F*-statistic decreases significantly to a value of 1.26, indicating conditional random assignment to high or low bunching US Attorney spells. Limiting the sample to attorneys who see more than 25 or 50 drug cases within the weight range yields similar results. The *p*-value also increases and is insignificant across each tested sample. To test changes in law enforcement, I consider the main specifications again but drop attorney spells with especially low or high cases per month. This is discussed in more detail in the Online Appendix, Section B.

If US Attorney spells really create variation in manipulation, bunching levels should change with US Attorney turnover. To investigate this, I create an event study for when a US Attorney changes and the new Attorney has a higher bunching propensity than the previous one within their district. I consider 12-months before and after the US Attorney change, meaning if the change resulted in a US Attorney that served for fewer than 12-months, they are not counted as treated here. Figure 6 displays the event study graph. Note that following a change to an attorney that bounces more, the proportion of cases at the threshold weight nearly doubles.

Figure 6: Bunching by US Attorney turnover



Notes: Each point gives the proportion of cases bunched within a district in quarterly time relative to the US Attorney turnover. This is specific to cases where a district changes from a US Attorney that bunches more than the previous Attorney(s).

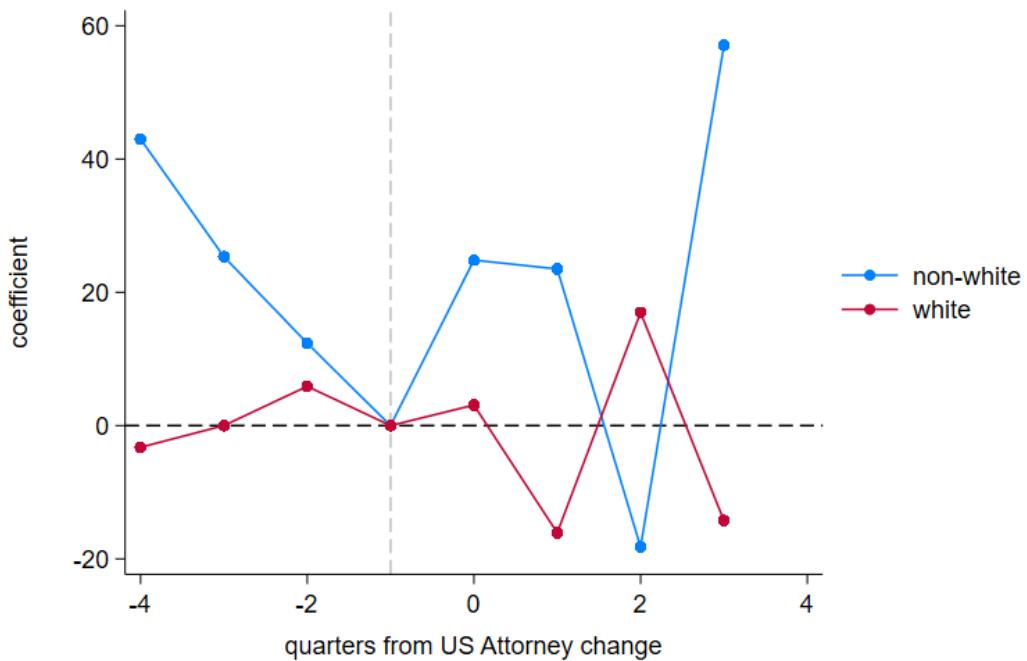
VI.B. Heterogeneity of main results

I now check how both legal and manipulation effects differ across the high and low bunching Attorney groups. I first consider the event study format again but now see whether manipulation effects change with Attorney turnover. The expected result here is not completely obvious; US Attorney spells with low bunching may still exhibit sentence lengths above the counterfactual if round number bias is low. However, if high bunching is indicative of case manipulation, the expected impact of US Attorney turnover is a stronger manipulation effect. Likewise, the main results suggest increased manipulation effects should be driven by minority cases.

Figure 7 displays the result. Note that manipulation effects are only measured for cases

at the threshold weight. Thus, by reducing the sample to cases at the threshold within a year of a qualifying US Attorney change, the number of observations becomes thin, resulting in noisy results. Even still, the graph shows a clear change in trend for the non-White groups with increases in manipulation effects following the US Attorney change and no real change for White defendants. This provides evidence that prosecutors are driving increases in sentence lengths at the threshold weight.

Figure 7: Manipulation effect change at US Attorney turnover



Notes: This figure follows the same design as the event study in Figure 6, now measuring manipulation effects. No controls or fixed effects are included due to thin data. Each point gives the average difference between sentence length and counterfactual sentence length using the fitted values of the right-hand side extrapolation.

I then run the main discontinuity analysis again split by US Attorney spell groups. Specifically, I compare effects across the split of the top half of bunching attorney spells versus the bottom half. Note that this analysis still has smaller sample sizes than the main analysis due to data constraints from the US Attorney data, which only includes years

2013-2020. In concordance with the event study, I find that the top half of spells exhibit much larger manipulation effects than the bottom half, with the high-bunch group averaging effects between 11.24 and 22.08 months and the low-bunch group averaging effects between 4.55 and 9.94 months. However, legal effects are on average slightly larger for low-bunching spells. The disparity in legal effects is considerably smaller than in manipulation effects, but is consistent across specifications. These results are displayed in Table A.8 of the Online Appendix.

VI.C. Do minorities get disproportionately bunched?

Why are manipulation effects localized to Black and Hispanic individuals? One possible explanation is that manipulation through additional evidence gathering occurs primarily for minority defendants but not for White ones. Indeed, minority defendant distributions do exhibit more bunching compared to the White distribution. However, without knowing race drug carrying behavior, it isn't clear whether this comes from additional manipulation or not. In this section, I provide a brief analysis to consider whether minority defendants actually have weights manipulated more than White counterparts.

To estimate racial disparities, I once again use the US Attorney spell bunching split and employ the following model:

$$\begin{aligned}
 bunched_{idmt} = & \alpha + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 high_bunch_{idmt} \\
 & + \beta_4 Black_i \times high_bunch_{idmt} + \beta_5 Hispanic_i \times high_bunch_{idmt} \\
 & + X_i \gamma + \lambda_d + \kappa_t + \eta_m + \epsilon_{idmt} \quad (1)
 \end{aligned}$$

where *bunched* is a binary measure of whether a case is sentenced at the threshold weight and

where $high_bunch$ is a binary variable for whether a case occurs during a high-bunch spell. β_4 and β_5 are the primary coefficients of interest, which give the differential effects of increasing bunching propensity on Black and Hispanic defendants compared to White counterparts. X_i gives observable characteristics of the case and of the defendant. These include the drug type, the total number of criminal history points applied in the case, the sex of the defendant, defendant age and age squared, a binary for whether the defendant completed some or graduated from college, and a binary for whether the defendant is an illegal alien. It also contains a measure of the percentage of cases with a White defendant per US Attorney spell to control for differences in case composition and prosecution opportunity across race. I likewise include year, month-of-year, and district fixed effects. The fixed effects control for temporal shocks, seasonality, and district differences in bunching behavior.

Columns 1-3 of Table 4 displays these results. Across these specifications, I estimate the effects for the full sample, the sample excluding attorneys with less than 25 total cases, and the sample excluding attorneys with less than 50 total cases.¹⁵ I reduce the sample this way to ensure outliers are not driving results. All columns include controls and fixed effects described above.

The estimates indicate Black defendants face significantly higher bunching odds when assigned to a high-bunching administration. The probability of bunching increases for White defendants assigned to high-bunching administrations by 3.21-3.47 percentage points (around 86% increase over low-bunch mean) compared against low-bunching spells. This increase is 1.69-1.73 percentage points larger for Black defendants compared to White ones, which is

¹⁵Note that the number of cases mentioned here is just the number of drug trafficking cases within the 20 percent - 180 percent weight window. Even US Attorneys with relatively short spells see many cases when including all weights and other types of crime.

Table 4: Racial disparity in bunching

	(1)	(2)	(3)	(4)	(5)	(6)
	bunched	bunched	bunched	sent length	sent length	sent length
Black	-0.00241 (0.00782)	-0.00260 (0.00801)	-0.00193 (0.00850)	14.97*** (1.817)	14.36*** (1.837)	14.18*** (2.033)
Hispanic	-0.00424 (0.00677)	-0.00392 (0.00698)	-0.00285 (0.00710)	7.806*** (1.531)	7.675*** (1.571)	7.774*** (1.636)
high bunch	0.0347*** (0.00665)	0.0325*** (0.00677)	0.0321*** (0.00731)	0.590 (1.695)	0.302 (1.721)	0.223 (1.906)
Black*high bunch	0.0170** (0.00757)	0.0173** (0.00768)	0.0169** (0.00829)	-0.117 (2.229)	0.333 (2.257)	0.776 (2.501)
Hispanic*high bunch	0.00367 (0.00823)	0.00322 (0.00820)	0.000791 (0.00800)	0.295 (1.974)	0.367 (2.011)	0.318 (2.115)
Attorney # of Cases	All	≥ 25	≥ 50	All	≥ 25	≥ 50
R Squared	0.130	0.127	0.131	0.312	0.314	0.316
N	22606	22202	20803	22606	22202	20803

Notes: The first three specifications consider effects for whether a defendant is sentenced at the threshold weight, while the last three consider effects on sentence length. Specifications in columns 2, 3, 5, and 6 reduce the sample to US Attorney spells that had 25 or greater or 50 or greater drug cases within the target 20%-180% weight range. Standard errors are clustered at the district level.

Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

about 50% larger than the White increase. These estimates are statistically significant at the 5 percent level across each specifications. Consistency in magnitude and significance of estimates regardless of removing low-case attorney spells suggests outliers are not driving results. Robustness checks for these results are provided in the Online Appendix.

Are higher bunching administrations simply more harsh in sentencing? To answer this, I use the same method outlined in Equation 1 with sentence length as the dependant variable. Note that this is not a causal estimate of how bunching impacts sentence. Rather, I assess whether higher bunching attorney spells correlate with higher sentence lengths throughout the distribution. These results are displayed in columns 4-6 of Table 4. I find high-bunching attorney spells have no discernible differences in sentence lengths from low-bunching ones, indicating manipulation effects are being driven by bunching behavior rather than overall

leniency or harshness of the US Attorney.

VII. Conclusion

Mandatory minimum sentencing is a controversial practice in the criminal justice system with debate as to its true impact on criminal punishment and concerns over its impact on racial equity. In this paper, I show that mandatory minimum eligibility for federal drug cases affects sentence length in two ways: through a legal effect and a manipulation effect. This decomposition is crucial for accurately estimating mandatory minimum causal effects and for understanding how policy intervention might affect disparities between race groups.

The main analysis shows that without manipulation, MM eligibility still leads to significant increases in sentence length. The average legal effect of eligibility is a 10.63 month increase in sentence length, which is an 18% increase over the counterfactual mean. The level effect is fairly consistent across drug type and racial group. This means that race gaps in sentence length are not driven by some systemic aspect of the law itself. It also means that for most cases, MM eligibility increases punishment homogeneously regardless of drug type or defendant race.

However, for a subset of specific types of cases that are close to the 10-year eligibility threshold weight, manipulation effects can lead to even larger increases in sentence length. When controlling for criminal history, drug type and defendant characteristics, I find that Black defendants receive an additional 14.28 months to their sentence due to manipulation. Similarly, Hispanic cases receive an additional 9.59 months from manipulation. Thus, racial disparities in sentence lengths caused by mandatory minimums are driven almost entirely by manipulation of legal actors. Manipulation effects shrink when including prosecutor

decision controls, but remain statistically significant. This indicates that at least part of the manipulation effect can be interpreted as increases from additional evidence against the defendant. Manipulation effects are primarily driven by powder cocaine and heroin cases. The legal and manipulation effects are also present for low-history offenders, meaning current policy does not fully shield them from MM impacts. Finally, I provide evidence that bunching is indeed driven by prosecutor discretion and that Black individuals are more likely to be manipulated to the threshold weight.

This paper suggests that reducing MM punishment or eligibility levels is likely to reduce sentence lengths for most drug trafficking cases and would do so fairly uniformly. However, it is difficult to say how policy might impact racial disparities since race gaps are driven by prosecutor discretion. Furthermore, elimination or changes to MM sentencing may be met with compensating behavior from prosecutors or judges. It is also worth noting that this paper does not address large racial disparities in sentence lengths that are not driven from MM sentencing. Future work will be needed to understand how legal institutions and legal actors create disparities for specific defendant types.

References

- Aizer, A. and J. J. Doyle, Jr. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *Quarterly Journal of Economics* 139(2), 759–803.
- Arnold, D., W. Dobbie, and C. S. Yang (2018). Racial bias in bail decisions. *Quarterly Journal of Economics* 133(4), 1885–1932.
- Bhuller, M., G. B. Dahl, K. V. Loken, and M. Magne (2020). Incarceration, recidivism and employment. *Journal of Political Economy* 128(4).
- Bjerk, D. (2005). Making the crime fit the penalty: The role of prosecutor discretion under mandatory minimum sentencing. *Journal of Law and Economics* 48(2), 591–625.
- Bjerk, D. (2017a). Mandatory minimum policy reform and the sentencing of crack cocaine defendants: An analysis of the fair sentencing act. *Journal of Empirical Legal Studies* 14(2), 370–396.
- Bjerk, D. (2017b). Mandatory minimums and the sentencing of federal drug crimes. *Journal of Legal Studies* 46(1), 93–128.
- Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *Quarterly Journal of Economics* 126(4), 1593–1660.
- Cooper, S. (2023). Prosecutor tradeoffs and race: Evidence from a circuit split.
- Di Tella, R. and E. Schargrodsy (2013). Criminal recidivism after prison and electronic monitoring. *Journal of Political Economy* 121(1), 28–73.
- Diamond, R. and P. Persson (2017). The long-term consequences of teacher discretion in grading of high-stakes tests. Nber working paper 22207, Stanford University.
- Didwania, S. (2020a). Charging leniency and federal sentences. University of wisconsin legal studies research 1746.
- Didwania, S. H. (2020b). Mandatory minimum entrenchment and the controlled substances act. *Ohio State Journal of Criminal Law* 18.
- Dobbie, W., J. Goldin, and C. S. Yang (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review* 108(2), 201–40.
- Fischman, J. B. and M. M. Schanzenbach (2012). Racial disparities under the federal sentencing guidelines: The role of judicial discretion and mandatory minimums. *Journal of Empirical Legal Studies* 9(4), 729–764.
- Frandsen, B. R. (2017). *Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete*. Bingley, UK: Emerald Publishing Limited.

- Gelman, A. and G. Imbens (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business and Economic Statistics* 37(3), 447–456.
- Goncalves, F. and S. Mello (2021). A few bad apples? racial bias in policing. *American Economic Review* 111(5), 1406–41.
- Kleven, H. and M. Waseem (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *Quarterly Journal of Economics* 128, 669–723.
- Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review* 96(3), 863–76.
- Lynch, M. (2016). *Hard Bargains: The Coercive Power of Drug Laws in Federal Court*. New York: Russell Sage Foundation.
- Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. <https://sites.lsa.umich.edu/mgms/wp-content/uploads/sites/283/2015/09/incar.pdf>.
- Rehavi, M. M. and S. B. Starr (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy* 122(6), 41320–1354.
- Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy* 2(3), 180–212.
- Shaffer, H. (2023). Prosecutors, race, and the criminal pipeline. *The University of Chicago Law Review* 90(7), 1889–1965.
- Shaffer, H. and E. Harrington (2017). Brokers of bias in the criminal justice system: Do prosecutors compound or attenuate racial disparities inherited at arrest? <https://drive.google.com/file/d/1Y4r3yuYPX6cqsFtYh0MCKMY7FyC3QB80/view>.
- Sloan, C. (2022). Do prosecutor and defendant race pairings matter? evidence from random assignment.
- The Pew Charitable Trusts (2015, August). Federal drug sentencing laws bring high cost, low return.
- Tuttle, C. (2022). Online appendix for ‘racial disparities in federal sentencing: Evidence from drug mandatory minimums’. https://codytuttle.github.io/tuttle_mandatory_minimums_online_appendix_2022.pdf.
- Tuttle, C. (2023). Racial disparities in federal sentencing: Evidence from drug mandatory minimums. https://codytuttle.github.io/tuttle_mandatory_minimums.pdf.
- United States Department of Justice (2018). Justice manual. Technical report, Washington, DC.

United States Department of Justice (2023, March 23). Bureau of prisons: Southeast region fiscal year 2024 president's budget narrative.

United States Government Publishing Office (2023, September 22). National archives and records administration: Records schedules; availability and request for comments.

United States Sentencing Commission (2010-2021). Individual offender datafiles. <https://www.ussc.gov/research/datafiles/commission-datafiles> (accessed October 1, 2021).

USSC (2019). First step act. https://www.ussc.gov/sites/default/files/pdf/training/newsletters/2019-special_FIRST-STEP-Act.pdf.

Yang, C. (2016). Resource constraints and the criminal justice system: Evidence from judicial vacancies. *American Economic Journal: Economic Policy* 8(4), 289–332.