Online Appendix for "How Much do Mandatory Minimums Matter?"

Spencer Cooper - 2024

Contents

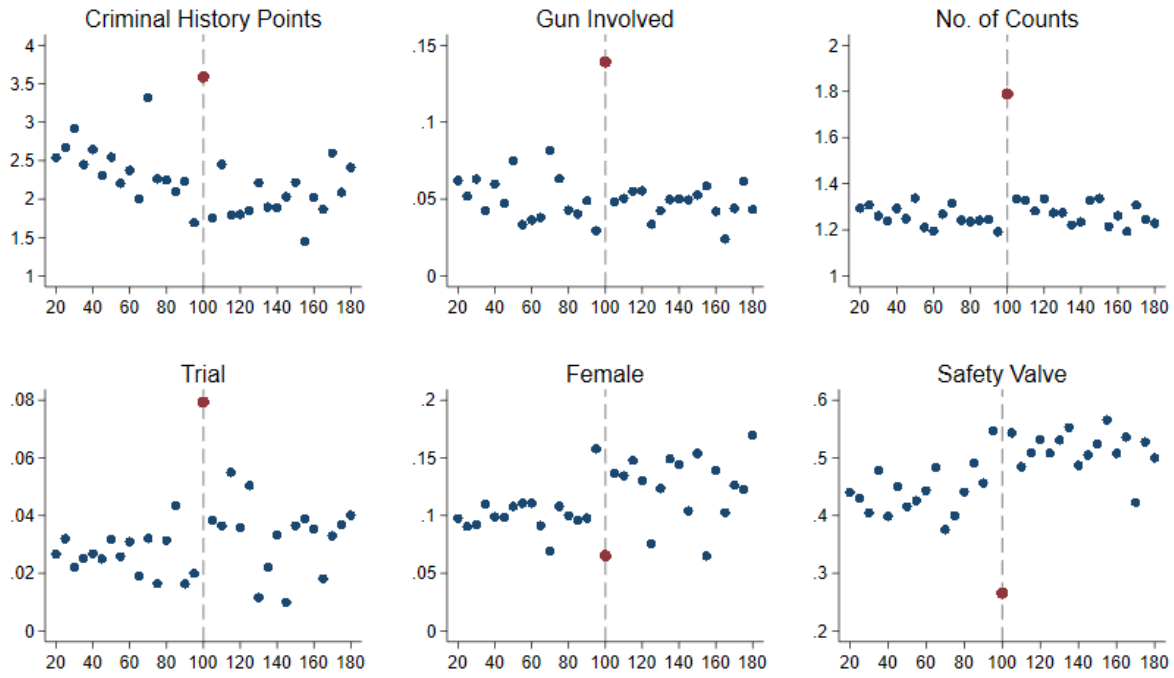
A Case type - who gets manipulated?

In this section, I consider whether cases that are manipulated to the 10-year mandatory minimum (MM) threshold share similar characteristics. I particularly look for whether manipulated cases have characteristics that correlate with higher sentence lengths outside of the MM or drug weight effects. This exercise is informative in two key ways. First, it shows that manipulation effects are likely subject to selection bias. This implies the inclusion of the richer sets of case controls is necessary for identifying the causal sentence length increase from drug weight manipulation. Second, it provides evidence that manipulation occurs from below the bunching point.

The basic motivation for this analysis is this: if prosecutors are systematically choosing defendants to get higher sentence lengths through manipulation, cases at the threshold may have similar bunching patterns across other observable factors that are highly correlated with sentence length. One way to think of this is that prosecutors might manipulate weights for the defendants with the "worst" characteristics, or those that are most likely to increase a defendant's sentence length. Thus, I consider how charging weight correlates with six factors strongly associated with higher sentences: criminal history points, whether a gun was involved in the crime, the number of counts a defendant is charged with, whether the case went to trial, whether the defendant is female, and whether the safety valve provision was applied at sentencing. I graph these factors against charging weight for the cases with the highest bunching effects - cocaine, crack, and marijuana cases for Black and Hispanic defendants.

Figure A.1 presents the results. The figure shows significant selection effects across each of these factors. The mean at the MM threshold is a significant outlier compared to other charging weights: for criminal history the bunch mean is 54.9% higher than the non-bunch mean, 158.9% higher for gun involvement, 41.3% higher for number of counts, 167.8% higher for trial, 40.3% lower for female, and 41.9% lower for safety valve. Each of these variables are positively correlated with sentence length besides female and safety valve, which are negatively correlated. Each of these bunching or dipping results suggest manipulation occurs for the cases that are already likely to receive higher sentence lengths.

Figure A.1: Factors being selected on



Notes: Each graph displays a different key variable measured across drug weight as a percentage of the threshold weight. These are shown in 5% bins, except at the threshold weight which only contains cases charged exactly at the threshold weight. Each variable has a strong, positive correlation with sentence length except for whether the defendant is female or receives the safety valve provision, which correlate negatively. The point at the threshold is a different color and size to highlight its difference compared to other charging weights.

The first takeaway from this analysis is not particularly surprising - that the manipulation effect is likely biased upward. Indeed, in the main paper I note the presence of missing sentence lengths in several of the discontinuity analyses[1]. This provides evidence that prosecutors manipulate cases with relatively high average sentences. These significant deviations in sentence trends only appear for cases in which the manipulation effect is present and large. Specifically, the regression analysis for heroin cases and for Black and Hispanic cases, given respectively in Figure A.3 and Figure 4 in the main text, show significantly lower sentence lengths for cases between 95% and 99% of the threshold weight. The above graphs corroborate this idea and suggest the raw manipulation effect is not to be interpreted as causal. It

[1]Here, missing sentence length refers to dips in sentence lengths that are obviously outside of the trend. This is analogous to a missing mass in a traditional bunching design

is worth noting that the manipulation effect shrinks, but is still significantly positive, when controlling for these six factors.

The second takeaway from Figure A.1 is the direction of manipulation. The bunching (dipping) of these high (low) sentence factors is consistent with the idea that prosecutors or other legal actors manipulate cases from below the 10-year MM threshold to the threshold weight to secure a higher sentence length. If bunching was due purely to round number convenience, these factors would exhibit bunching at other round numbers in the distribution, which they do not. Upward manipulation is further evidenced by the presence of missing sentence lengths below the threshold weight, as shown in Figure 3 in the main text. This direction of manipulation is important for identification, as it suggests that extrapolation from the right-hand side is not necessarily[2]. It also suggests that if selection bias is mitigated with controls, the right-hand side sentence lengths serve as a reasonable counterfactual for sentence without manipulation.

The mechanism driving these types of cases to the bunching point is not entirely clear. If manipulation is driven by prosecutor discretion as suggested in Lynch (2016), and prosecutors are selecting cases to be manipulated upward, the cases most likely to be manipulated are those that either have low cost to manipulation or high benefit for achieving the 10-year MM threshold (likely through higher sentence lengths).[3] Each of these channels are plausible. Cooper (2023) provides evidence that prosecutors tend to bunch more when costs to bunching are reduced. Likewise, high sentence factors may correlate with manipulation costs, such that these defendants are easier to connect to a criminal conspiracy. Indeed, a simple OLS regression finds that each of these factors correlate strongly with conspiracy except for gun use, which moves in the opposite direction. However, the presence of the manipulation effect, which survives controls, suggests there is an expected sentence length premium to prosecutors for manipulating a marginal case. Without more granular data on evidence and case connection, disentangling these two mechanisms is difficult.

[2]This is why in the main intensive margin analysis, extrapolation from the right-hand side is only across 5 percentage points compared to 30 percentage points from the left

[3]Note that this assumes that prosecutors generally prefer higher sentence lengths to lower ones. This may not be true in every case, but it is not inconsistent with proposed models, whether prosecutors seek to minimize errors, maximize convictions, maximize sentence lengths, or some combination of these. See Landes (1971), Grossman and Katz (1983), and Silveira (2017).

B US Attorney background and robustness

Background

Because I use US Attorney spells as a source of variation in bunching, I provide a brief background for their role in the charging process. United States Attorneys serve as the chief federal law enforcement agent within their district. There are 93 US Attorneys serving at all times, one for each district. US Attorneys are typically appointed by the president of the United States and serve until they choose to step down or are asked to resign. Resignation requests often occur after a new president is sworn in but may also occur within presidential administrations. In times of vacancy, an assistant attorney already serving in that district fills the leadership role and is considered the Acting US Attorney. For extended vacancies, the US Attorney General may also appoint an interim attorney to fill the leadership role until a new presidential appointment.

Attorneys are given immense discretion to dictate the focus and procedure of prosecution within their district. In the Principles of Prosecution section of the United States Justice Manual, it states that "...individual United States Attorneys are required to establish their own priorities (in consultation with law enforcement authorities), within the national priorities, in order to concentrate their resources on problems of particular local or regional significance."[4] Other sections of the Justice Manual describe the US Attorney as having "the broadest discretion in the exercise of such authority" in relation to prosecuting criminal matters.[5] This implies US Attorneys have significant impacts on the types of cases that are prosecuted and the manner in which prosecution should occur. This appears to hold true in practice; in my discussion with a number of federal prosecutors, they described the US Attorney in office as having a significant impact in the day to day operations of the Assistant US Attorneys.[6]

[4] See section 9-27.230.

[5] See section 9-2.001. This section also gives specifics about which aspects of prosecution US Attorneys are allowed to exercise their discretion. This includes authorizing prosecution and determining the manner of prosecuting and deciding trial related questions.

[6] I spoke with a handful of US Attorneys and Assistant US Attorneys about the impacts of who is in the role of US Attorney, who are kept anonymous by request. While responses varied, the general consensus is that US Attorneys have a lot of flexibility in what types of cases should be prosecuted and how prosecution should be carried out. Conversations with legal scholars have told a similar story.

This broad discretion in prosecution leads to variation in bunching across US Attorney spell. In my data, I find bunching variation both across and within districts. As shown in Table A.1, the US Attorney bunching propensity measure has a high standard deviation of 12.84. I also calculate the standard deviation within districts to ensure variation is not only driven by geographic differences. The within district standard deviation is 7.45, a number still high relative to the overall mean. Likewise, the table demonstrates variance in within district bunching by reporting mean maximum and minimum bunching scores per district, with the maximum mean at 16.41 cases and the minimum at 2.752 cases. I also show that variation is not driven only by temporal differences; Figure A.2 displays bunching for each district averaged across each US Attorney in the district. The figure shows high variation in bunching across districts.

Robustness

Here I provide a couple of alternative specifications for extensive margin analysis presented in Table 4 of the main text. To check that these results are not driven by law enforcement changes or substantial changes in volume of cases prosecuted, I run the above analysis again omitting attorney spells with especially high or low cases per month. Note that this is different than omitting attorneys with less than 25 or 50 total cases; that restriction removes attorneys with very few total cases, mostly removing very short US Attorney spells. Case-per-month restrictions control for times in which there happens to be an especially high volume of cases entering the system, a measure that will strongly correlate with law enforcement behavior relating to drug trafficking. Table A.9 in the Online Appendix displays these results. Columns 1 and 2 omit US Attorney spell cases with the bottom 10 percent of cases per month while columns 3 and 4 omit the top 10 percent. These estimates have similar magnitudes to the main specifications and most estimates keep statistical significance.

I also consider this analysis again by simply interacting the continuous bunching propensity measure with race rather than the attorney half-split binary variable. Now rather than drawing some cutoff between high and low bunching attorneys, I measure the average effect of the intensity of bunching. The results are then standardized such that coefficients represent the effect of a 1 standard deviation increase in bunching propensity. Table A.10 displays

the effects from this analysis. While the effects are no longer statistically significant, the signs and magnitudes of the coefficients are consistent with the main results.

C Additional tables

Table A.1: US attorney data - summary statistics

	mean	sd
number of US attorneys per district	3.125	(0.894)
months served	48.62	(25.94)
number of cases	343.2	(390.4)
percent of cases bunched	3.130	(3.594)
max bunch pct within district	5.681	(7.559)
min bunch pct within district	1.299	(2.001)
# of Attorneys		282

Notes: This table considers characteristics of the US Attorneys data set. Note that number of cases only includes drug trafficking cases with weights between 20% and 180% of the threshold weight; many more cases are prosecuted during an Attorney's tenure.

Table A.2: MM effects with varying fits

	(1) sent length	(2) sent length	(3) sent length	(4) sent length
legal effect	9.502	10.98	9.832	6.676
left side 95% CI	[71.36, 77.15]	[70.63, 77.22]	[69.34, 76.24]	[74.68, 79.39]
right side 95% CI	[81.61, 86.23]	[81.61, 86.23]	[81.61, 86.23]	[81.61, 86.23]
manip. effect	17.01	17.01	17.01	17.01
fit value	60%-80%	64%	62%	79%
N of obs fit on	-	32177	31706	35067

Notes: This table repeats the simple linear discontinuity displayed in Table 2, but now uses various fitting schemes. Column 1 averages all fits between 60% and 80%, while columns 2-4 give the median, upper bound, and lower bound fits, respectively. Confidence intervals are determined using the standard error of the expected prediction.

Table A.3: Main effects by drug type comparison

	(1) cocaine	(2) crack	(3) heroin	(4) marijuana	(5) meth
legal effect	15.00	9.873	15.65	12.51	15.51
left side 95% CI	[-6.26, 1.90]	[3.86, 18.28]	[-8.02, 4.08]	[7.87, 14.99]	[-8.05, 1.89]
right side 95% CI	[8.98, 17.02]	[13.42, 29.02]	[8.26, 19.30]	[18.19, 30.16]	[7.74, 17.33]
manip. effect	14.43	2.181	15.19	5.430	-1.474
fitted mean at 99%	-0.157	12.56	-1.276	14.57	-1.716
fit value	70%	70%	70%	70%	70%
N of obs fit on	5865	5233	3035	5256	4272

Notes: The dependant variable is sentence length residualized against criminal history points and race. 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. The legal effect is the regression discontinuity between the two extrapolated fits at the 10-year MM cutoff. The manipulation effect is the difference between the right-hand regression fit and the actual sentence length at the threshold weight. 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, with extrapolation occurring from 71% up to the cutoff.

Table A.4: Main effects by criminal history and safety valve

	(1) no prior history	(2) low history	(3) no safety valve	(4) yes safety valve
legal effect	6.708	10.46	16.11	5.397
left side 95% CI	[1.41, 7.12]	[-2.16, 3.53]	[-3.79, 2.88]	[0.84, 3.43]
right side 95% CI	[8.11, 14.16]	[8.38, 14.25]	[13.01, 18.69]	[6.30, 9.04]
bunch effect	9.830	11.21	11.93	-5.984
fitted mean at 99%	4.26	0.68	2.14	2.61
fit value	70%	70%	70%	70%
N of obs fit on	3395	6828	16314	7197

Notes: The dependant variable is sentence length residualized against criminal history points and race. 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. The legal effect is the regression discontinuity between the two extrapolated fits at the 10-year MM cutoff. The manipulation effect is the difference between the right-hand regression fit and the actual sentence length at the threshold weight. 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, with extrapolation occurring from 71% up to the cutoff.

Table A.5: Discontinuity analysis with multiple drug weight controls

	(1) All cases	(2) Black	(3) Hispanic	(4) White
legal effect	9.520	11.86	10.18	11.02
left side 95% CI	[1.63, 5.64]	[2.60, 11.07]	[0.54, 4.97]	[-6.01, 2.01]
right side 95% CI	[11.65, 15.11]	[14.72, 23.23]	[10.89, 15.16]	[6.09, 12.42]
manip. effect	5.573	3.665	3.856	-2.249
fit value	70%	70%	70%	70%
N of obs on fit	33691	11205	15296	7190

Notes: The dependant variable is residual sentence length. All specifications include controls for defendant characteristics, time and district fixed effects, and the new other drug type weight controls. These are weight as a percent of the 10-year mandatory minimum threshold for up to 4 other drug types other than the primary type. Each specification is fit at 70%.

Table A.6: Discontinuity analysis with weight range observations included

	(1) sentence length	(2) resid sent 1	(3) resid sent 2	(4) resid sent 3
Panel A: Minimum of weight range				
legal effect	9.467	13.00	9.192	11.02
left side 95% CI	[71.70, 76.97]	[-4.43, 0.11]	[-5.05, -0.84]	[-1.84, 2.19]
right side 95% CI	[81.55, 86.20]	[8.86, 12.96]	[4.37, 8.09]	[9.52, 13.03]
	{33911}	{33911}	{33911}	{33911}
manip. effect	15.20	7.156	12.38	5.319
Panel B: Median of weight range				
legal effect	13.69	11.68	10.91	11.09
left side 95% CI	[68.19, 73.37]	[-1.40, 3.077]	[-5.90, -1.76]	[-0.67, 3.26]
right side 95% CI	[82.19, 86.81]	[10.56, 14.64]	[5.15, 8.86]	[10.71, 14.21]
	{33546}	{33546}	{33546}	{33546}
manip. effect	16.02	14.79	5.372	8.309
Panel C: Maximum of weight range				
legal effect	13.22	16.93	9.799	12.06
left side 95% CI	[67.44, 72.39]	[-6.69, -2.42]	[-6.32, -2.40]	[-1.78, 1.94]
right side 95% CI	[80.86, 85.51]	[10.37, 14.47]	[3.52, 7.25]	[10.44, 13.97]
	{33546}	{33546}	{33546}	{33546}
manip. effect	11.37	6.788	6.706	3.471
fit value	70%	70%	70%	70%
baseline controls	no	yes	yes	yes
additional controls	no	no	yes	yes
fixed effects	no	no	no	yes

Notes: This table gives the regression discontinuity results for the full sample of cases including cases with imprecise weight measures. For each measure, I use the precise weight for cases where it is available and then vary how the range measure is considered. In Panel A, I use the minimum value of the range as the drug weight measure. In Panel B, I use the median of the drug weight range. And in Panel C, I use the maximum weight in the drug range. The four specifications use the same control schemes as used in the main analysis. All regressions are fit at 70% of the threshold weight and use a linear fit. Sample size is given in curly braces.

Table A.7: Bunching propensity randomization check

	(1) bunch	(2) bunch_score	(3) bunch_score	(4) bunch_score
F-Value:	7.11	1.26	0.90	0.94
F-Test:	0.000	0.263	0.544	0.507
Attorney # of Cases	All	All	≥ 25	≥ 50
N	22606	22606	20803	16986

Notes: Here I regress the residualized bunching propensity measure on defendant characteristics omitting district fixed effects. Covariates included in the regression are drug type, sex, criminal history points, age and age squared, a binary measure for college, a binary measure for illegal alien, and the proportion of cases with a White defendant for each US Attorney.

Table A.8: High vs low bunching US Attorney spell - legal and manipulation effects

	(1) sent length	(2) resid sent 1	(3) resid sent 2	(4) resid sent 3
Panel A: Low-bunch				
legal effect	15.71	15.11	13.49	12.08
left side 95% CI	[62.57, 74.11]	[-8.50, 2.09]	[-7.01, 3.24]	[-4.48, 4.91]
right side 95% CI	[79.70, 88.62]	[7.85, 16.21]	[7.80, 15.70]	[8.84, 16.06]
manip. effect	7.490	8.267	9.938	4.550
N of obs fit on	4812	4812	4812	4812
Panel B: High-bunch				
legal effect	10.23	9.875	9.088	8.918
left side 95% CI	[63.78, 72.41]	[-4.65, 3.28]	[-3.37, 4.36]	[-2.47, 4.65]
right side 95% CI	[74.72, 82.28]	[5.88, 12.92]	[6.39, 13.20]	[7.11, 13.34]
manip. effect	22.08	17.68	16.16	11.24
N of obs fit on	7393	7393	7393	7393
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

Notes: This table presents the main discontinuity effects again, now split by US Attorney bunching groups. Panel A gives the discontinuity estimates for the low-bunching Attorney spells while Panel B reports effects for high-bunching spells. All extrapolation is fitting between 20% and 70% for the left side of the threshold and 105% and 180% on the right side.

Table A.9: Prosecutor analysis - omitting high and low case-per-month attorneys

	(1)	(2)	(3)	(4)
	bunch	sent length	bunch	sent length
Black	-0.00278	-0.00291	-0.00376	-0.00448
	(0.00818)	(0.00837)	(0.00802)	(0.00807)
Hispanic	-0.00416	-0.00385	-0.00335	-0.00368
	(0.00703)	(0.00734)	(0.00714)	(0.00718)
high bunch	0.0355***	0.0341***	0.0362***	0.0365***
	(0.00731)	(0.00798)	(0.00662)	(0.00674)
Black*high bunch	0.0155*	0.0132	0.0179**	0.0179**
	(0.00808)	(0.00831)	(0.00764)	(0.00771)
Hispanic*high bunch	0.00212	0.000618	0.00374	0.00374
	(0.00854)	(0.00871)	(0.00899)	(0.00926)
	drop	drop	drop	drop
Sample restriction	lowest 5%	lowest 10%	highest 5%	highest 10%
R Squared	0.131	0.132	0.132	0.131
N	21555	20370	21705	20196

Notes: All specifications include the full set of controls; defendant characteristics and time and district fixed effects. The first two specifications omit US Attorney spell cases with very low number of cases-per-month. Columns 3 and 4 omit the spells with the very highest cases-per-month volume. Note that this is different than the sample restrictions in the main specifications - there I omit attorneys who see very few cases total which removes several attorneys who only served for a short amount of time. In this analysis I omit attorneys on a cases-per-month basis which may keep short-term attorneys if they fall within the assigned ranges.

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

Table A.10: Racial disparity in bunching using a continuous measure

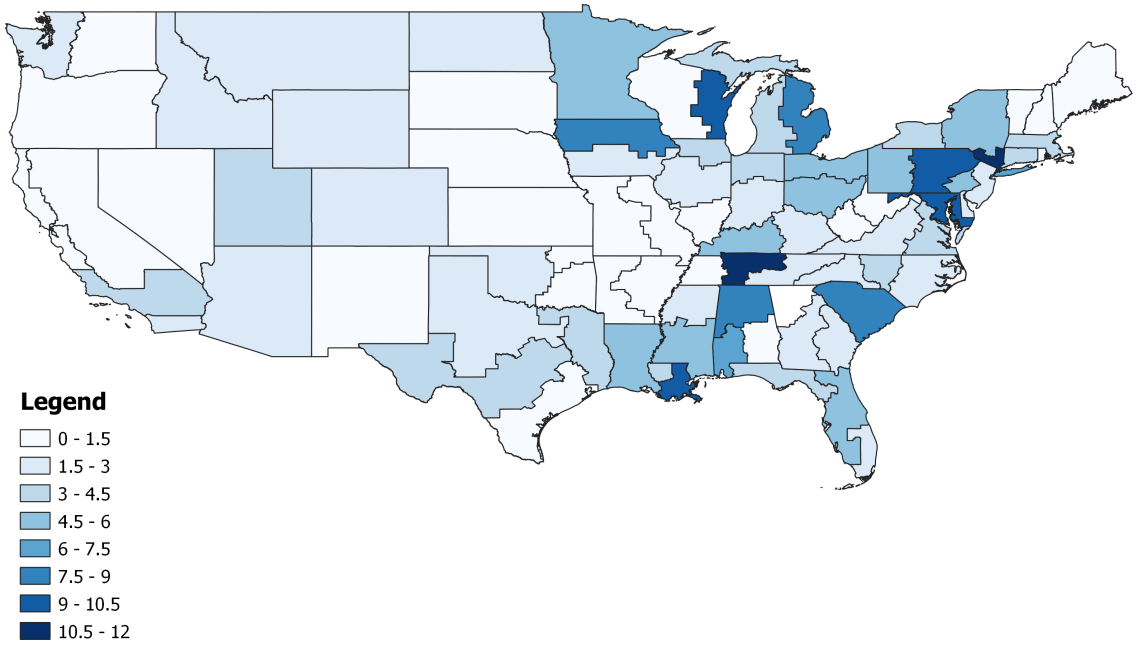
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.0359***	0.00527	0.00966	0.00723	0.00717	0.00779
	(0.0105)	(0.0111)	(0.00586)	(0.00704)	(0.00703)	(0.00776)
Hispanic	0.0104	-0.0161**	0.0107**	-0.00215	-0.00208	-0.00237
	(0.0115)	(0.00694)	(0.00525)	(0.00502)	(0.00505)	(0.00545)
bunch_score	0.00658	0.00639	0.00567	0.00537	0.00569	0.00670
	(0.00461)	(0.00455)	(0.00446)	(0.00448)	(0.00525)	(0.00576)
Black*bunch_score	0.00514	0.00439	0.00442	0.00417	0.00363	0.00417
	(0.00801)	(0.00820)	(0.00630)	(0.00639)	(0.00725)	(0.00866)
Hispanic*bunch_score	0.00472	0.00613	0.000342	0.00132	-0.000752	-0.00295
	(0.00483)	(0.00467)	(0.00604)	(0.00587)	(0.00661)	(0.00531)
Attorney # of Cases	All	≥ 25	≥ 50	All	≥ 25	≥ 50
R Squared	0.007	0.029	0.113	0.124	0.122	0.127
N	22606	22606	22606	22606	22202	20803

Notes: The dependent variable is a binary measure for whether the case sentenced is at the threshold weight. The key difference from Table **??** is that instead of separating high vs low US Attorney spells, I simply interact the continuous bunching propensity measure. All specifications contain the full set of controls.

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

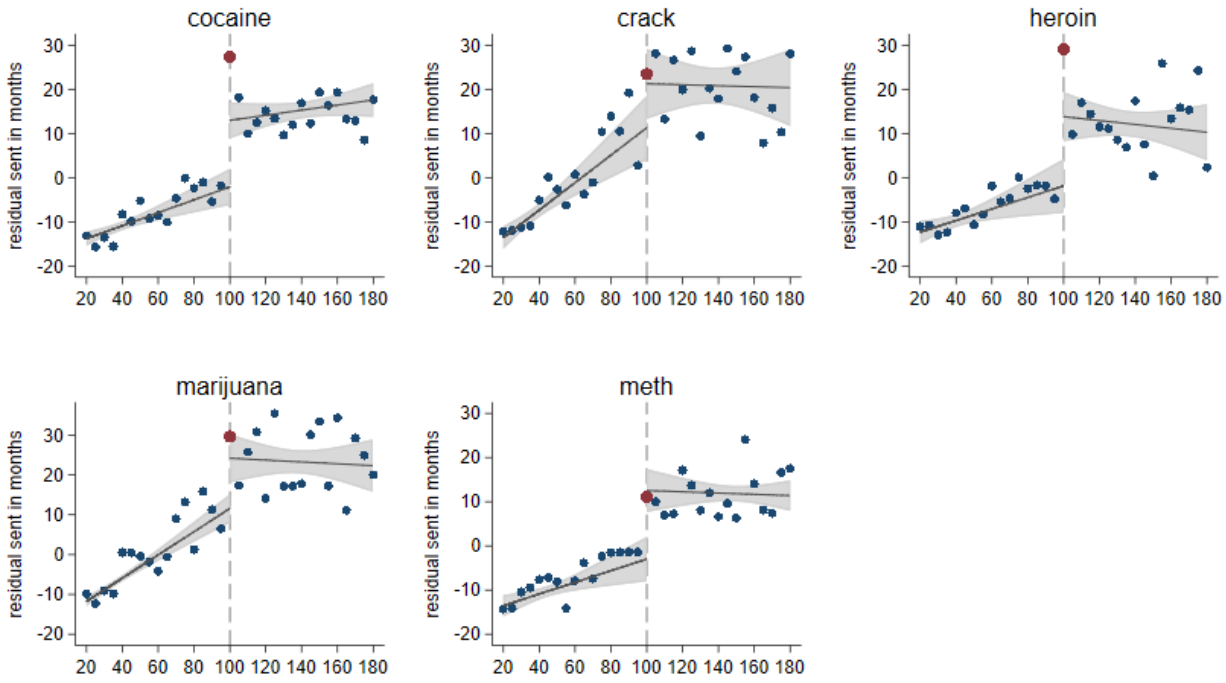
D Additional figures

Figure A.2: Percent of cases bunched by district



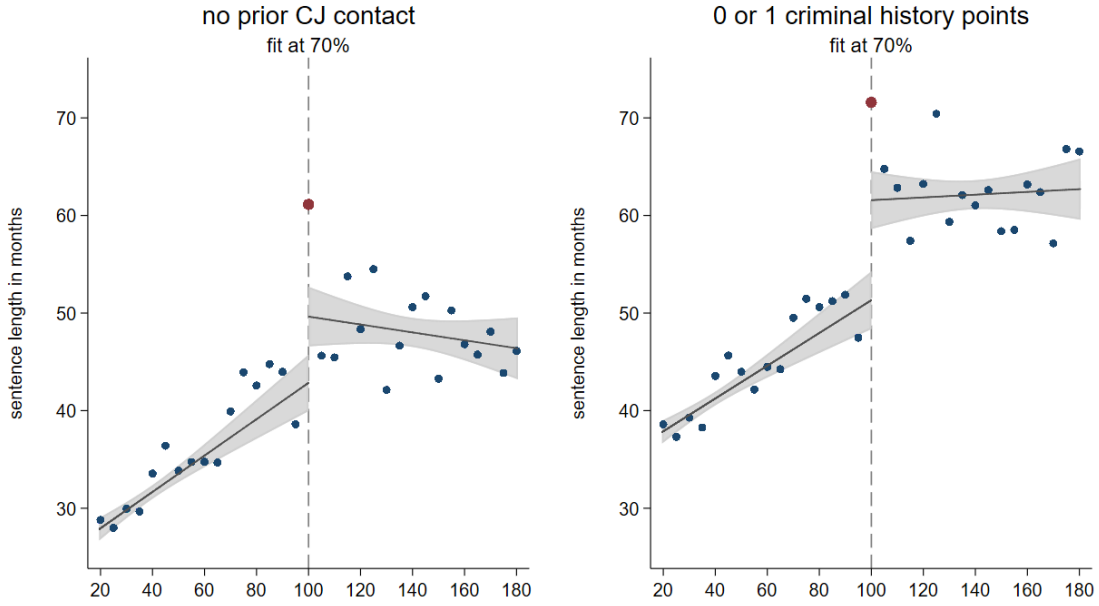
Notes: This figure shows variation in bunching propensity across district. It presents the percent of cases bunched for all cases before December of 2018 across each district.

Figure A.3: Main results by drug type



Notes: This figure illustrates the regression discontinuity design controlling for race and criminal history points, fit at 70% for each drug type. Data at the bunching point is larger and with a different color simply to emphasize differences in manipulation effects between drug types.

Figure A.4: Main results by criminal history group



Notes: These are the main regression discontinuity results for cases in the first two criminal history categories, generally considered low-history defendants. Group 1 includes only defendants who have had no previous encounters with the criminal justice system, including events that would lead to zero criminal history points, such as arrest. Group 2 includes individuals who have no points but have had some encounters with the justice system, and individuals with one point. Both discontinuities are fit using the 70% cutoff for extrapolation and use linear fits.

Online Appendix References

Cooper, S. (2023). Prosecutor tradeoffs and race: Evidence from a circuit split.

Grossman, G. M. and M. L. Katz (1983). Plea bargaining and social welfare. *The American Economic Review 73*(4), 749–757.

Landes, W. (1971). An economic analysis of the courts. *Journal of Law and Economics 14*(1), 61–107.

Lynch, M. (2016). *Hard Bargains: The Coercive Power of Drug Laws in Federal Court*. New York: Russell Sage Foundation.

Silveira, B. S. (2017). Bargaining with asymmetric information: An empirical study of plea negotiations. *Econometrica 85*(2), 419–452.