Online Appendix for "Prosecutor Tradeoffs and Race: Evidence from a Circuit Split"

Spencer Cooper - 2024

Contents

A Notes on Enhancement and Safety Valve Provision

Sentence enhancement refers to an increase in a punishment given when the crime involves a specified undesirable quality or behavior. In context of the First Step Act (FSA), the enhancement refers to an increased mandatory minimum length given for offenders with past drug or violent crimes. Prior to FSA, offenders with a prior drug or violent felony conviction were eligible for an enhancement that would double the mandatory minimum sentence length. For enhancement to be applied, the prosecution must explicitly file for a notice of enhancement under 21 U.S.C. §851 and these must be confirmed by a jury.[1] This means that if an offender is convicted of trafficking drugs at a weight above the high threshold and they receive the enhancement, they would be subject to a mandatory minimum of 20 years rather than the typical statutory 10 years. FSA changed enhancement procedure in two ways. First, it decreased the high threshold penalty from 20 years to 15 years. Second, the qualification for enhancement changed from a prior drug or violent felony conviction to a "serious" prior drug or violent felony conviction. Here, "serious" is defined as a drug or violent crime for which the offender served a term of imprisonment for longer than 12 months.[2] Taken together, these two effects imply there are less individuals who qualify for enhancement and the penalty for enhancement at the higher threshold is decreased.

The second major sentencing revision introduced by the FSA is an expansion of the safety valve provision. Safety valve provides a means whereby offenders can receive a sentence below the mandatory minimum even if the quantity of drugs they are charged with exceeds the threshold. The intent of the provision was to provide relief for first-time, non-violent offenders. To qualify for the safety valve provision, the defense must provide proof at a preponderance of evidence level that the defendant did not use violence or possess a dangerous weapon in connection with the crime, the offense did not result in death or serious injury of any persons, the defendant did not act as a leader or organizer, the defendant cooperates with law enforcement and gives all known information of the crime (if applicable), and the

[1] Prior to *Alleyne v. United States* enhancement factors were to be determined by the judge. However, the observation period only considers cases post *Alleyne*, so all enhancements in this study come by approval of a jury.

[2] The details of which crimes qualify as "serious" are more technical than explained here, but include most drug or violent crimes for which the offenders was imprisoned for over 12 months. See Public Law 115-391 for more details of seriousness as well as which statutes qualify as a violent crime.

defendant does not have more than one criminal history point. One criminal history point translates to having one or less prior offense, and this offense must have resulted in a sentence of less than 60 days incarceration. This criminal history stipulation is especially binding as many non-violent offenders have more than one criminal history point.[3] FSA expanded the safety valve provision to include individuals with higher criminal history points (up to four, excluding one point offenses). However, offenders with a prior conviction that resulted in more than 12 months incarceration or a violent prior conviction that resulted in more than 60 days incarceration are still exempt.

B Conceptual Framework

To conceptualize the prosecutor incentives and formalize predictions for the three policies, I present a simple model of the prosecutor's manipulation choice. At the moment, I only give predictions for the FSA, though predictions for the other two policies are similar. I only model upward manipulation, and thus only consider cases with weights below the mandatory minimum threshold weight. Manipulation increases the sentence length for a case but is costly to the prosecutor. Prosecutors will only choose manipulation in cases where a higher sentence will increase utility and outweigh the manipulation costs.

Prosecutors choose a binary manipulation choice m that maximizes their utility function:

$$\max_{m} U(s^*, s, w, \nu) = \pi([s^* - s]^2) - m \cdot C(w; \nu) \tag{1}$$

π gives the utility gained from benefits of sentence length, and is a decreasing function. s^* is the optimal sentence length to the prosecutor for the specific defendant, and reflects a combination of career benefits and what the prosecutor considers fair. Thus, I assume π is maximized when $s^* = s$. C is a cost function that is decreasing in seized weight w. I only model the costs that are added from manipulation and assume all other costs constant regardless of manipulation behavior. ν reflects all other institutional factors that may affect the cost of bunching. s represents the sentence length at the time of disposition. For

[3] In 2017, around 66% of all federal trafficking cases involved offenders with criminal history points over one point.

simplicity I assume s only takes on three values as described here:

$$s = \begin{cases} s_l & \text{if} \quad m = 0 \\ s_m & \text{if} \quad m = 1 \text{ and } FSA = 1 \\ s_h & \text{if} \quad m = 1 \text{ and } FSA = 0 \end{cases}$$

where $s_l < s_m < s_h$ and $FSA = 0$ if the case happens before the policy change and equals 1 if after.

If the prosecutor chooses not to manipulate weights to the threshold ($m = 0$), they receive the lower sentence length of s_l with no additional costs. If they choose manipulation ($m = 1$), they receive a higher sentence length, s_h or s_m, which may increase utility if s_m or s_h is closer to s^*. They also receive the additional costs of manipulation. Manipulation costs are most likely are made up of costs to search for additional evidence, but may also include increased bargaining costs if the higher sentence increases the time needed to reach a plea deal. C may also include any psychic costs to the prosecutor associated with causing the sentence length to increase.

In my first analysis, I consider variation in bunching behavior by US Attorney spell to identify whether bunching is driven by prosecutor choice. In context of the model, US Attorneys change the utility function π for prosecutors working under their direction. US Attorneys dictate which cases to focus on and how prosecution should be carried out in their district. Thus, for a US Attorney that places high emphasis on drug prosecution the career benefits for the individual prosecutor are likely to be higher than when working for a US Attorney that puts low priority on drug cases.

I assume there are no systematic differences in π by circuit type, meaning prosecutors receive the same benefit of manipulation in each circuit, all else constant. However, conspiracy method circuits are expected to have significantly lower costs to manipulation since evidence gathering costs are explicitly lower. The model suggests that overall bunching should then be higher in conspiracy circuits compared to individual circuits.

Racial disparities in manipulation decisions may come from differences in benefits or costs. Prosecutors may get more benefit from manipulating racial minority charging weights

through some form of taste based discrimination or if bargaining conditions cause sentence lengths to be higher for racial minorities at the threshold weight. Differences in costs could be driven by ease of gathering evidence, unobserved defendant behavior that change bargaining power, different types of conspiracies, or drug carrying behavior being correlated with race.

The First Step Act Predictions

The First Step Act effectively lowered the expected sentence length of mandatory minimum cases. This is reflected in the model as a change from s_h to s_m as the received sentence length for a prosecutor who chooses manipulation. If s^* is sufficiently high, the expected effect of the FSA would be to lower manipulation. As the expected benefits of manipulation decrease, there will be some subset of cases for which the cost will now be too high to justify manipulation. These marginal defendants satisfy two conditions:

1. $U(s^*, s_h, w, \nu) > U(s^*, s_l, w, \nu)$, and
2. $U(s^*, s_m, w, \nu) < U(s^*, s_l, w, \nu)$

These conditions represent a defendant who would have been bunched before the FSA, but no longer are bunched due to the lower benefit. Similarly, the FSA may have no effect on a defendant if they wouldn't have been bunched pre-FSA anyway, or if they would still be bunched after the FSA. That is, there are some defendants who will always be bunched regardless of the FSA, likely because the cost to manipulate their drug weight is extremely low. There are also some defendants who won't be bunched whether their case is before or after the FSA, likely because there is little evidence of a conspiracy or of additional drugs within a conspiracy. Essentially, no change in bunching occurs if manipulation costs are lower than benefits both pre and post-FSA or if manipulation costs are higher than benefits both pre and post-FSA. But bunching may also increase after the FSA. For a given defendant, if costs to manipulation are sufficiently low and s^* is closer to s_l than to s_h but closer to s_m than s_l, the defendant would actually be bunched after the FSA but not before it.

To illustrate how bunching may increase post FSA, I provide a simple example: suppose

for a given defendant and prosecutor, the model parameters are as follows:

$s^* = 4, \quad s_h = 10, \quad s_m = 5$

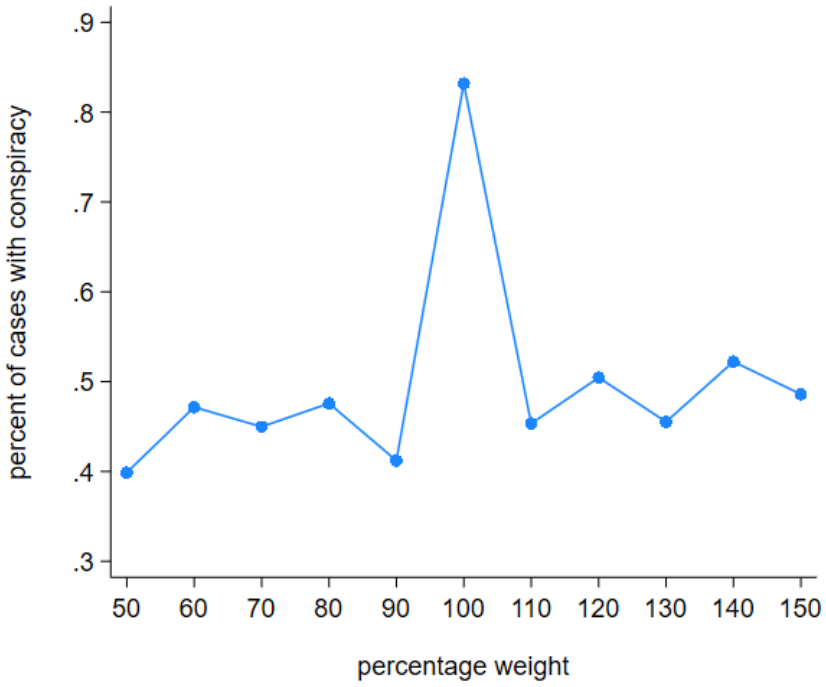
$s_l = 2, \quad C(w;\nu) = 0$

Then before the FSA, the distance between optimal and received sentences is lower without manipulation. This means the prosecutor will choose to not manipulate and the defendant will receive a sentence length of 2 years. But after the FSA, the optimal sentence s^* is closer to s_m than to s_l, so manipulation becomes optimal. Then the prosecutor will choose to manipulate after the FSA, but not before. Thus, the effects of the FSA on bunching direction is ambiguous and may vary by circuit and racial group, depending on differences in s^* and $C(w;\nu)$.

The model gives several key insights as to how the results of this analysis should be interpreted. First, decreases in bunching will only happen for groups with sufficiently high costs. Second, groups with high costs but low s^* are likely to have little to no bunching both pre and post-FSA. Third, groups with low cost and high s^* are likely to have high bunching both pre and post-FSA. And finally, increases in bunching will only occur for groups with low costs to bunching and a specific window of s^* that isn't too high or low. These key facts highlight how this analysis uncovers the way prosecutors respond to incentives and how prosecutors may consider manipulation costs between racial groups differentially.

C Additional Figures

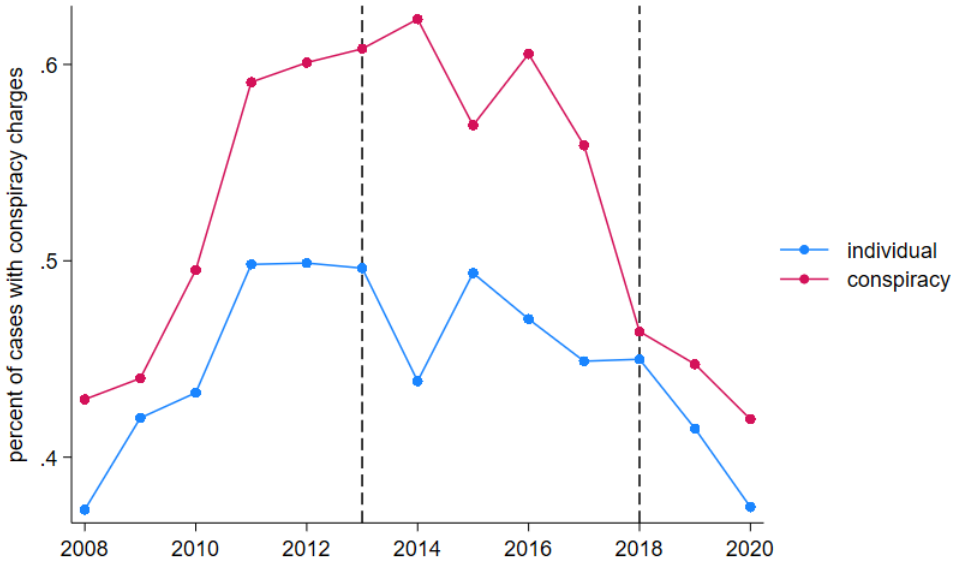
Figure A.1: Percent of cases with a conspiracy charge by weight



Each point gives the average percent of cases with a conspiracy charge for the ten percent bin, besides the bunching point. For example, the point at 50 includes all cases with a drug weight between 50 and 60. The point at 110 contains all cases with weights between 100 and 110, not including the bunching point. The bunching point is all cases with weights exactly at 100.

Figure A.2: Percent of cases with a conspiracy charge by year

((a)) All cases



((b)) By race

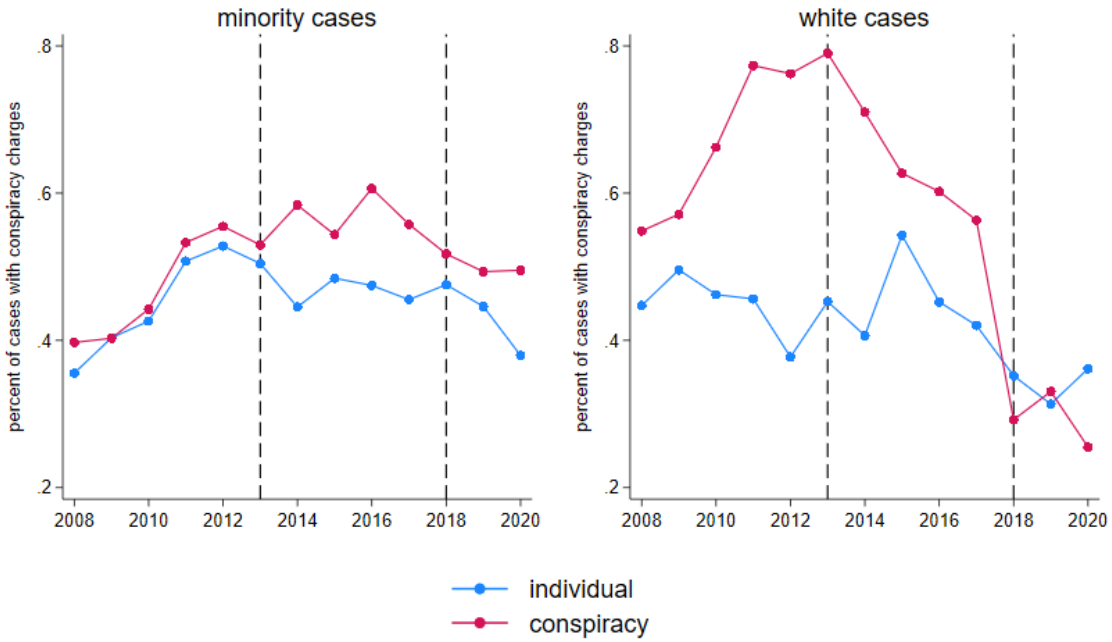
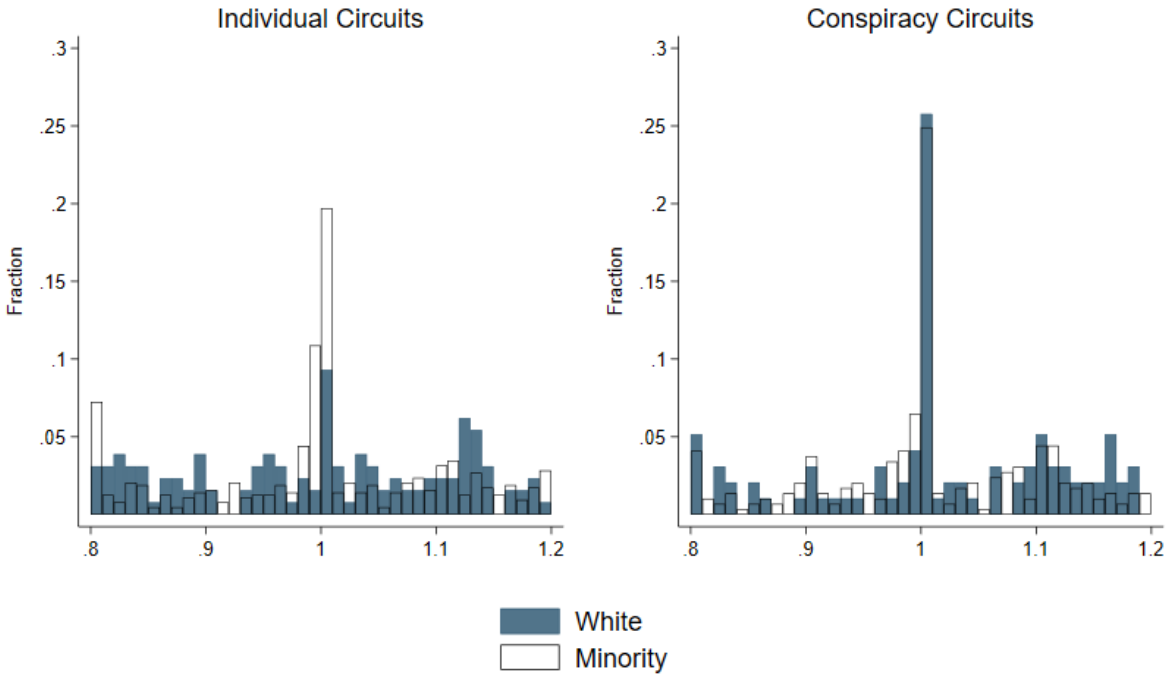
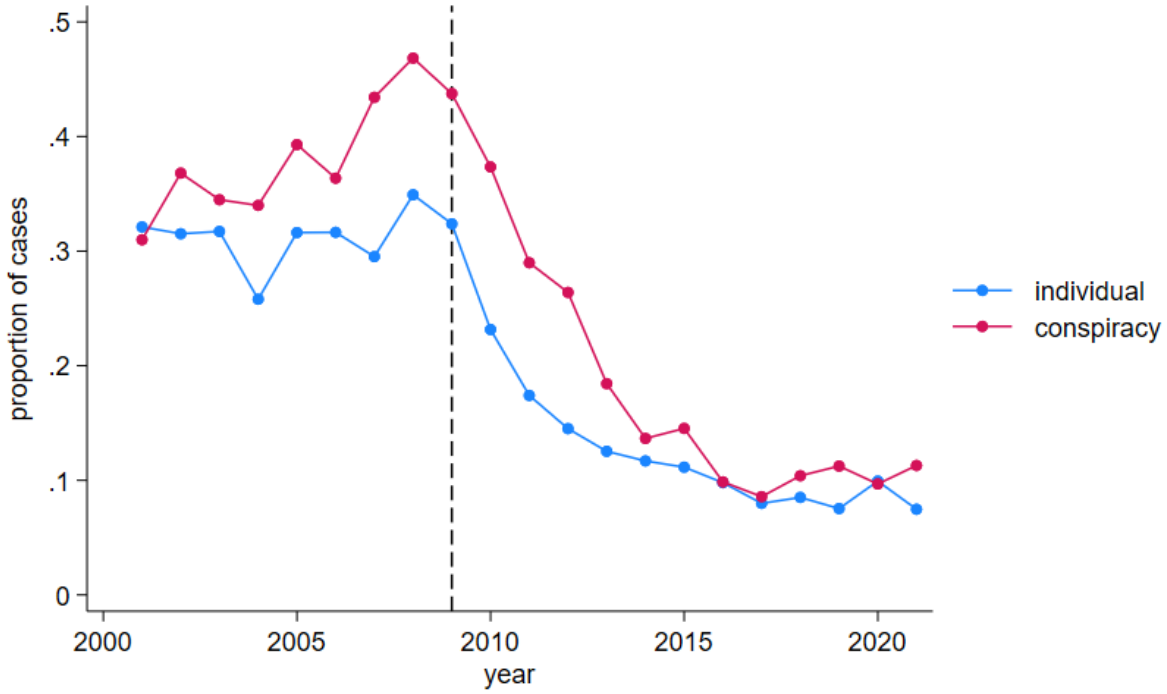


Figure A.3: Percent of cases bunched by race



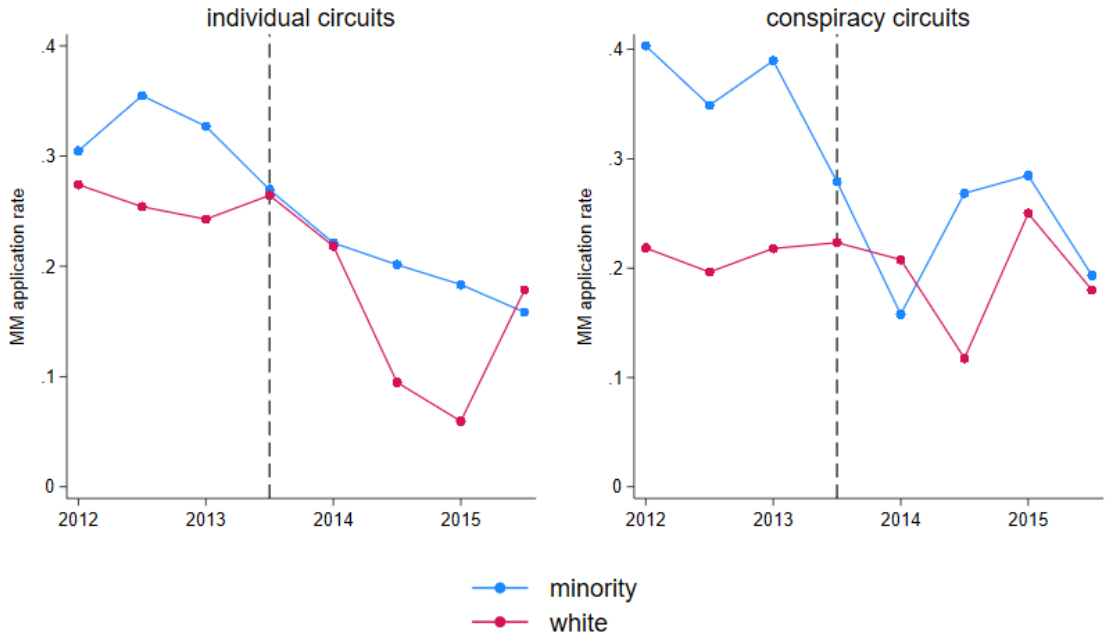
The following histogram present distributions by race-circuit-type intersection for two years preceding the timing of the First Step Act.

Figure A.4: Proportion of cases crack



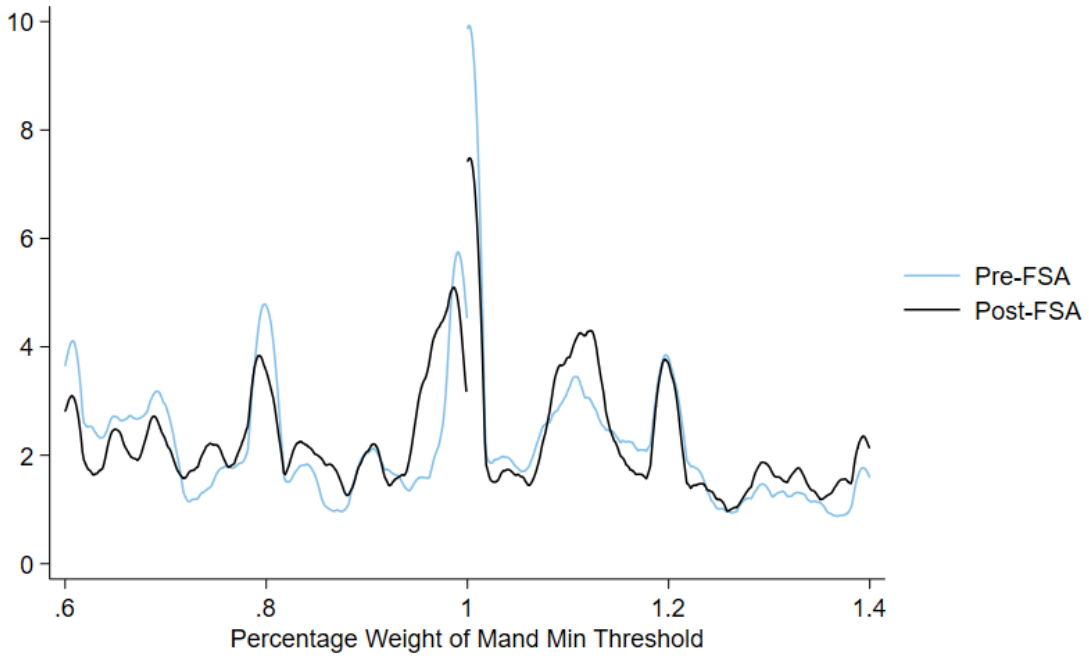
This graph shows what proportion of drug cases had crack cocaine as the primary drug type by year and circuit type.

Figure A.5: Minus-2 effects on race gaps



Notes: This figure gives the race gap changes in residual formal MM charges for all drug cases across the two evidence regimes. The y-axis measures the percentage point difference in MM charges between minority and White cases. Charges are residualized on drug type, sex, criminal history, age, age squared, education level, citizenship status, and district and monthly fixed effects.

Figure A.6: FSA kernel density



Notes: This figure combines White-conspiracy cases with minority-individual cases to see where excess mass arises following the FSA. These two groups are combined as they are the cases that experience the largest decrease in bunching throughout the analysis.

D Additional Tables

Table A.1: MM weight thresholds

Drug Type	5-Year Threshold Amount (in grams)	10-Year Threshold Amount (in grams)
cocaine	500	5000
crack	28	280
heroin	100	1000
marijuana	100000	1000000
meth	5	5000

Table A.2: USSC Drug Case Data - Summary Statistics

	All Cases	Individual Circuits	Conspiracy Circuits
Panel A. Drug Types			
cocaine	0.265	0.269	0.245
crack	0.206	0.171	0.250
heroin	0.144	0.122	0.169
marijuana	0.136	0.201	0.0856
meth	0.250	0.237	0.250
Panel B. Defendant Characteristics			
minority	0.780	0.810	0.748
female	0.132	0.140	0.119
age	35.17	34.81	35.22
criminal history points	3.943	3.654	4.422
college	0.183	0.169	0.181
illegal alien	0.161	0.188	0.141
Panel C. Outcomes			
percent weight of threshold	0.920	0.919	0.920
bunched (binary)	0.0914	0.0713	0.0923
10-year MM applied	0.282	0.266	0.298
conspiracy charge	0.490	0.449	0.526
safety valve applied	0.320	0.348	0.265
assistance to government	0.283	0.297	0.297
sentence length (in months)	80.36	78.43	84.19
Share	1	0.500	0.247
Observations	27832	13924	6872

Table A.3: Alternative timing results for Minus-2

	Non-Conspiracy Circuits		**Conspiracy Circuits**	
	(1) bunch	(2) MM app.	(3) bunch	(4) MM app.
Panel A: Alleyne Results				
minority	-0.0121	0.0150	0.00133	0.0846*
	(0.0150)	(0.0313)	(0.0224)	(0.0492)
post	-0.0206	-0.0639	0.0410*	-0.0619
	(0.0173)	(0.0419)	(0.0210)	(0.0418)
post × minority	0.0197	-0.0366	-0.0365	-0.0999*
	(0.0173)	(0.0384)	(0.0273)	(0.0536)
N	6774	6774	2288	2288
Panel B: Holder Results				
minority	-0.0465*	-0.0126	0.0251	0.0459
	(0.0250)	(0.0276)	(0.0254)	(0.0526)
post	-0.0262	-0.0693	0.0793**	-0.0502
	(0.0320)	(0.0510)	(0.0382)	(0.0687)
post × minority	0.0266	-0.0418	-0.0955*	-0.138*
	(0.0301)	(0.0399)	(0.0514)	(0.0767)
N	3685	3685	1090	1090

Notes: This table is analogous to Table **??**, but uses the timing of *Alleyne* and Holder instead, with results for *Alleyne* presented in Panel A and Holder in Panel B. Note that the Holder analysis only contains cases that are eligible under the Holder decision. Each regression controls for primary drug type, defendant sex, criminal history points, age, age squared, education, citizenship status, and district and month of year fixed effects. Standard errors are clustered at the district level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Robustness - triple differences

	Fair-2010		**Minus-2**		**FSA**	
	(1) bunch	(2) MM app.	(3) bunch	(4) MM app.	(5) bunch	(6) MM app.
Panel A: Wildbootstrap Clusters						
ddd	-0.0796	-0.238*	-0.0618	-0.119**	0.148***	0.00205
	(0.0842)	(0.140)	(0.0459)	(0.0536)	(0.0429)	(0.0588)
Panel B: Circuit-Specific Time Trends						
ddd	-0.0794	-0.236*	-0.0612	-0.120**	0.152***	0.0123
	(0.0829)	(0.139)	(0.0485)	(0.0600)	(0.0413)	(0.0596)
Panel C: No Cartel-Heavy Districts						
ddd	-0.0414	-0.207	-0.0980**	-0.130**	0.135***	0.00623
	(0.0887)	(0.145)	(0.0469)	(0.0579)	(0.0432)	(0.0601)
N	3072	3072	8413	8413	6372	6372

Notes: This table considers three different robustness checks of the the triple-differences analysis, shown across three panels. Panel A contains only crack cases, while Panels B and C contain all drug types. The ddd coefficient is the triple differences effect, representing the interaction of minority, conspiracy circuit, and post. All specifications include defendant controls, population controls, and district and monthly fixed effects. All standard errors are clustered at the district level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Wider drug weight windows

	(1) bunch	(2) MM app.	(3) bunch	(4) MM app.
Panel A: Fair Sentencing Act				
ddd	-0.0276	-0.307***	-0.0150	-0.185***
	(0.0327)	(0.0978)	(0.0167)	(0.0672)
N	6486	6486	11749	11749
Panel B: Minus-2 Amendment				
ddd	-0.0402*	-0.0895***	-0.0193*	-0.0374*
	(0.0228)	(0.0336)	(0.0111)	(0.0191)
N	17413	17413	40401	40401
Panel C: First Step Act				
ddd	0.0844***	0.00488	0.0405***	-0.0203
	(0.0232)	(0.0375)	(0.0129)	(0.0235)
N	12266	12266	26937	26937
Drug-weight in %	20-200	20-200	0-300	0-300

Notes: This table replicates the main MM application analysis allowing for a larger window of weights. Each panel reports separate analyses for each policy. Panel A contains only crack cases, while Panels B and C contain all drug types. Columns 1 and 2 give all cases between 20 % and 200 % of the 10-year MM threshold weight. Columns 3 and 4 contain all cases with weights at or under 300% of the threshold weight. All specifications contain defendant characteristics, population controls, and district and monthly fixed effects. Standard errors are clustered at the district level.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$